

IBM SPSS Neural Networks 19



Note: Before using this information and the product it supports, read the general information under Notices el p. 96.

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright SPSS Inc. 1989, 2010.

Prefacio

IBM® SPSS® Statistics es un sistema global para el análisis de datos. El módulo adicional opcional Redes neuronales proporciona las técnicas de análisis adicionales que se describen en este manual. El módulo adicional Redes neuronales se debe utilizar con el sistema básico de SPSS Statistics y está completamente integrado en dicho sistema.

Acerca de SPSS Inc., an IBM Company

SPSS Inc., an IBM Company, es uno de los principales proveedores globales de software y soluciones de análisis predictivo. La gama completa de productos de la empresa (recopilación de datos, análisis estadístico, modelado y distribución) capta las actitudes y opiniones de las personas, predice los resultados de las interacciones futuras con los clientes y, a continuación, actúa basándose en esta información incorporando el análisis en los procesos comerciales. Las soluciones de SPSS Inc. tratan los objetivos comerciales interconectados en toda una organización centrándose en la convergencia del análisis, la arquitectura de TI y los procesos comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de SPSS Inc. como ventaja ante la competencia para atraer, retener y hacer crecer los clientes, reduciendo al mismo tiempo el fraude y mitigando los riesgos. SPSS Inc. fue adquirida por IBM en octubre de 2009. Para obtener más información, visite <http://www.spss.com>.

Asistencia técnica

El servicio de asistencia técnica está a disposición de todos los clientes de mantenimiento. Los clientes podrán ponerse en contacto con este servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de SPSS Inc. o sobre la instalación en alguno de los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia técnica, consulte el sitio web de SPSS Inc. en <http://support.spss.com> o encuentre a su representante local a través del sitio web <http://support.spss.com/default.asp?refpage=contactus.asp>. Tenga a mano su identificación, la de su organización y su contrato de asistencia cuando solicite ayuda.

Servicio de atención al cliente

Si tiene cualquier duda referente a la forma de envío o pago, póngase en contacto con su oficina local, que encontrará en el sitio Web en <http://www.spss.com/worldwide>. Recuerde tener preparado su número de serie para identificarse.

Cursos de preparación

SPSS Inc. ofrece cursos de preparación, tanto públicos como in situ. Todos los cursos incluyen talleres prácticos. Los cursos tendrán lugar periódicamente en las principales ciudades. Si desea obtener más información sobre estos cursos, póngase en contacto con su oficina local que encontrará en el sitio Web en <http://www.spss.com/worldwide>.

Publicaciones adicionales

Los documentos *SPSS Statistics: Guide to Data Analysis*, *SPSS Statistics: Statistical Procedures Companion* y *SPSS Statistics: Advanced Statistical Procedures Companion*, escritos por Marija Norušis y publicados por Prentice Hall, están disponibles y se recomiendan como material adicional. Estas publicaciones cubren los procedimientos estadísticos del módulo SPSS Statistics Base, el módulo Advanced Statistics y el módulo Regression. Tanto si da sus primeros pasos en el análisis de datos como si ya está preparado para las aplicaciones más avanzadas, estos libros le ayudarán a aprovechar al máximo las funciones ofrecidas por IBM® SPSS® Statistics. Si desea información adicional sobre el contenido de la publicación o muestras de capítulos, consulte el sitio web de la autora: <http://www.norusis.com>

Contenido

Parte I: Manual del usuario

1	Introducción a las redes neuronales	1
	¿Qué es una red neuronal?	1
	Estructura de red neuronal	2
2	Perceptrón multicapa	4
	Particiones	9
	Arquitectura	10
	Entrenamiento	13
	Salida	15
	Guardar	18
	Exportar	19
	Opciones	20
3	Función de base radial	22
	Particiones	26
	Arquitectura	27
	Resultado	29
	Guardar	31
	Exportar	33
	Opciones	34

Parte II: Ejemplos

4 Perceptrón multicapa 36

Uso del perceptrón multicapa para evaluar riesgos de crédito	36
Preparación de los datos para su análisis	36
Ejecución del análisis	39
Resumen de procesamiento de casos	41
Información sobre la red	42
Resumen del modelo	42
Classification.	43
Corrección del sobreentrenamiento.	44
Resumen	54
Uso del perceptrón multicapa para estimar los costes de asistencia sanitaria y las duraciones de las estancias	54
Preparación de los datos para su análisis	54
Ejecución del análisis	55
Advertencias.	62
Resumen de procesamiento de casos	63
Información sobre la red	64
Resumen del modelo	65
Gráficos de pronosticados por observados	66
Gráficos de residuos por pronosticados.	68
Importancia de las variables independientes	70
Resumen	70
Lecturas recomendadas	71

5 Función de base radial 72

Uso de la función de base radial para la clasificación de clientes de telecomunicaciones	72
Preparación de los datos para su análisis	72
Ejecución del análisis	73
Resumen de procesamiento de casos	77
Información sobre la red	78
Resumen del modelo	79
Classification.	79
Gráfico de pronosticados por observados	80
Curva COR.	82
Gráficos de elevación y ganancias acumuladas.	83
Lecturas recomendadas	84

Apéndices

A Archivos muestrales **86**

B Notices **96**

Bibliografía **98**

Índice **100**

Parte I:
Manual del usuario

Introducción a las redes neuronales

Las redes neuronales son la herramienta preferida para muchas aplicaciones de minería de datos predictiva por su potencia, flexibilidad y facilidad de uso. Las redes neuronales predictivas son especialmente útiles en las aplicaciones cuyo proceso subyacente sea complejo; por ejemplo:

- Prever la demanda de los consumidores para racionalizar los costes de producción y entrega.
- Predecir la probabilidad de respuesta al marketing mediante correo directo para determinar a qué unidades familiares de una lista de correo debe enviarse una oferta.
- Puntuar a un solicitante para determinar el riesgo que supone concederle un crédito.
- Detectar transacciones fraudulentas en una base de datos de reclamaciones de seguros.

Las redes neuronales utilizadas en las aplicaciones predictivas, como las redes de **perceptrones multicapa (MLP)** y las de **función de base radial (RBF)**, se supervisan en el sentido de que los resultados pronosticados por el modelo se pueden comparar con los valores conocidos de las variables de destino. La opción de redes neuronales de le permite ajustar las redes MLP y RBF y guardar los modelos resultantes para la puntuación.

¿Qué es una red neuronal?

El término **red neuronal** se aplica a una familia de modelos relacionada de manera aproximada que se caracteriza por un gran espacio de parámetro y una estructura flexible y que proviene de los estudios sobre el funcionamiento del cerebro. Conforme fue creciendo la familia, se diseñó la mayoría de los nuevos modelos para aplicaciones no biológicas, aunque gran parte de la terminología asociada refleja su origen.

Las definiciones específicas de redes neuronales son tan variadas como los campos en que se utilizan. Aunque no hay una sola definición que abarque a toda la familia de modelos, tenga en cuenta la siguiente descripción por ahora (Haykin, 1998):

Una red neuronal es un procesador distribuido en paralelo de forma masiva con una propensión natural a almacenar conocimiento experimental y convertirlo en disponible para su uso. A semeja al cerebro en dos aspectos:

- El conocimiento se adquiere por la red mediante un proceso de aprendizaje.
- Las fuerzas de conexión interneuronal, conocidas como ponderaciones sinápticas, se utilizan para almacenar el conocimiento.

Si desea ver información detallada acerca de por qué esta definición puede ser demasiado restrictiva, consulte (Ripley, 1996).

Para poder diferenciar las redes neuronales de los métodos estadísticos tradicionales que usan esta definición, lo que *no* se dice es tan significativo como el texto real de la definición. Por ejemplo, el modelo de regresión lineal tradicional puede adquirir conocimiento a través del método de mínimos cuadrados y almacenar ese conocimiento en los coeficientes de regresión. En este sentido, es una red neuronal. De hecho, se puede argumentar que la regresión lineal es un caso especial de determinadas redes neuronales. Sin embargo, la regresión lineal tiene una estructura de modelo rígida y un conjunto de supuestos que se imponen antes de aprender a partir de los datos.

Por lo contrario, la definición anterior plantean exigencias mínimas sobre la estructura y los supuestos del modelo. Por tanto, una red neuronal puede aproximar una amplia gama de modelos estadísticos sin que tenga que hipotetizar de antemano determinadas relaciones entre las variables dependientes e independientes. En lugar de eso, la forma de las relaciones se determina durante el proceso de aprendizaje. Si una relación lineal entre las variables dependientes e independientes es adecuada, los resultados de la red neuronal deben aproximarse lo máximo posible a los del modelo de regresión lineal. Si una relación no lineal es más adecuada, la red neuronal se aproximará automáticamente a la estructura del modelo “correcta”.

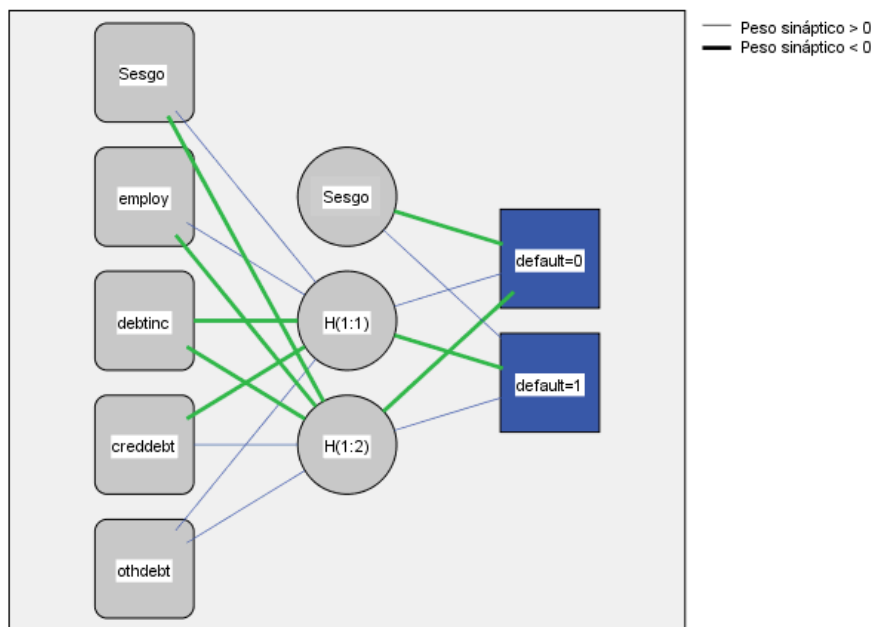
El inconveniente que supone esta flexibilidad radica en que las ponderaciones sinápticas de una red neuronal no pueden interpretarse fácilmente. Por tanto, si intenta explicar un proceso subyacente que produce las relaciones entre las variables dependientes e independientes, sería mejor usar un modelo estadístico más tradicional. Sin embargo, si la interpretación del modelo no es importante, puede obtener rápidamente modelos con buenos resultados mediante

Estructura de red neuronal

Aunque las redes neuronales plantean exigencias mínimas sobre los supuestos y la estructura del modelo, resulta útil comprender la **arquitectura** general de la red. La red de perceptrones multicapa (MLP) o de función de base radial (RBF) es una función de predictores (denominados también entradas o variables independientes) que minimiza el error de predicción de las variables de destino (también denominadas salidas).

Tomemos como ejemplo el conjunto de datos *bankloan.sav* incluido con el producto, en el que desea poder identificar a los posibles morosos entre un grupo de solicitantes de préstamos. Una red MLP o RBF aplicada a este problema es una función de las medidas que minimiza el error al pronosticar la posibilidad de mora. La figura siguiente resulta muy útil para ilustrar la forma de esta función.

Figura 1-1
Arquitectura Feedforward con una capa oculta



Función de activación de capa oculta: Tangente hiperbólica

Función de activación de capa de salida: Softmax

Esta estructura se denomina **arquitectura feedforward** porque las conexiones de la red fluyen unidireccionalmente desde la capa de entrada hasta la capa de salida sin ciclos de retroalimentación. En esta figura:

- La **capa de entrada** contiene los predictores.
- La **capa oculta** contiene nodos (o unidades) no observables. El valor de cada unidad oculta es una función de los predictores; la forma exacta de la función depende, por un lado, del tipo de red y, por otro lado, de especificaciones controlables por el usuario.
- La **capa de salida** contiene las respuestas. Como el historial de moras es una variable categórica con dos categorías, se recodifica como dos variables de indicador. Cada unidad de salida es una función de las entradas ocultas. Nuevamente, la forma exacta de la función depende, por un lado, del tipo de red y, por otro lado, de especificaciones controlables por el usuario.

La red MLP permite una segunda capa oculta; en ese caso, cada unidad de la segunda capa oculta es una función de las unidades de la primera capa oculta, y cada respuesta es una función de las unidades de la segunda capa oculta.

Perceptrón multicapa

El procedimiento Perceptrón multicapa (MLP) genera un modelo predictivo para una o más variables dependientes (de destino) basada en los valores de las variables predictoras.

Ejemplos. A continuación, se describen dos situaciones en las que se utiliza el procedimiento MLP:

El encargado de préstamos de un banco desea poder identificar características que sean indicativas de personas que puedan causar mora en los créditos y utilizar dichas características para identificar riesgos de crédito positivos y negativos. Utilizando una muestra de clientes anteriores, puede entrenar un perceptrón multicapa, validar el análisis utilizando una muestra reservada de clientes anteriores y posteriormente utilizar la red para clasificar a los posibles clientes según los riesgos de crédito que presenten.

Un sistema hospitalario está interesado en realizar un seguimiento de los costes y las duraciones de las estancias de los pacientes ingresados para tratamientos de infartos de miocardio (IM o “ataque al corazón”). La obtención de estimaciones precisas de estas medidas permite a la administración gestionar adecuadamente las camas disponibles a medida que se trata a los pacientes. Utilizando los registros de tratamiento de una muestra de pacientes que recibieron tratamiento a consecuencia de un infarto de miocardio, el administrador puede entrenar una red que pronostique tanto el coste como la duración de la estancia.

Variables dependientes. Las variables dependientes pueden ser:

- **Nominal.** Una variable se puede tratar como nominal si sus valores representan categorías que no obedecen a una ordenación intrínseca (por ejemplo, el departamento de la empresa en el que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- **Ordinal.** Una variable puede tratarse como ordinal cuando sus valores representan categorías con alguna ordenación intrínseca (por ejemplo, los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho). Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- **Escala.** Una variable puede tratarse como escala (continua) cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

El procedimiento supone que se ha asignado el nivel de medida adecuado a todas las variables dependientes. No obstante, puede cambiar temporalmente el nivel de medida para una variable pulsando con el botón derecho en la variable en la lista de variables de origen y seleccionar un nivel de medida en el menú contextual.

Un icono situado junto a cada variable de la lista de variables identifica el nivel de medida y el tipo de datos.

Nivel de medida	Tipo de datos			
	Numérico	Cadena	Fecha	Hora
Escala (Continuo)		n/a		
Ordinal				
Nominal				

Variables predictoras. Los predictores se pueden especificar como factores (categóricos) o covariables (de escala).

Codificación de la variable categórica. El procedimiento recodifica temporalmente predictores categóricos y variables dependientes utilizando la codificación “una de c ” para todo el procedimiento. Si hay c categorías de una variable, la variable se almacena como vectores c , con la primera categoría denotada $(1,0,\dots,0)$, la siguiente categoría $(0,1,0,\dots,0)$, ..., y la última categoría $(0,0,\dots,0,1)$.

Este esquema de codificación incrementa el número de ponderaciones sinápticas y puede dar como resultado un entrenamiento más lento. No obstante, los métodos de codificación más “compactos” normalmente conducen a redes neuronales con un peor ajuste. Si el entrenamiento de la red avanza muy lentamente, pruebe a reducir el número de categorías en los predictores categóricos combinando categorías similares o eliminando los casos que tengan categorías extremadamente raras.

Toda codificación “una de c ” se basa en los datos de entrenamiento, incluso si se define una muestra reservada o de comprobación (consulte [Particiones](#) el p. 9). De este modo, si las muestras reservadas o de comprobación contienen casos con categorías de predictores que no están presentes en los datos de entrenamiento, esos casos no son utilizados por el procedimiento o en la puntuación. De este modo, si las muestras reservadas o de comprobación contienen casos con categorías de variables dependientes que no están presentes en los datos de entrenamiento, esos casos no son utilizados por el procedimiento o en la puntuación.

Cambio de escala. Por defecto, se cambia la escala de las covariables y las variables dependientes de escala para mejorar el entrenamiento de la red. Todo cambio de escala se realiza basándose en los datos de entrenamiento, incluso si se define una muestra reservada o de comprobación (consulte [Particiones](#) el p. 9). Es decir, dependiendo del tipo de cambio de escala, la media, la desviación típica, el valor mínimo o el valor máximo de una covariable o variable dependiente se calculan utilizando sólo los datos de entrenamiento. Si especifica una variable para definir particiones, es importante que estas covariables o variables dependientes tengan distribuciones similares en todas las muestras reservadas, de entrenamiento o comprobación.

Ponderaciones de frecuencia. Este procedimiento ignora las ponderaciones de frecuencia.

Replicación de los resultados. Si desea replicar exactamente los resultados, utilice el mismo valor de inicialización para el generador de números aleatorios, el mismo orden de datos y el mismo orden de variables, además de utilizar la misma configuración del procedimiento. A continuación se incluyen más detalles sobre este tema:

- **Generación de números aleatorios.** El procedimiento utiliza la generación de números aleatorios durante la asignación aleatoria de particiones, el submuestreo aleatorio para la inicialización de las ponderaciones sinápticas, el submuestreo aleatorio para la selección automática de arquitectura y el algoritmo de recocido simulado que se utiliza en la inicialización de las ponderaciones y en la selección automática de arquitectura. Para reproducir los mismos resultados aleatorios en el futuro, utilice el mismo valor de inicialización para el generador de números aleatorios antes de cada ejecución del procedimiento Perceptrón multicapa. Consulte [Preparación de los datos para su análisis](#) el p. 36 para obtener instrucciones específicas.

- **Orden de casos.** Los métodos de entrenamiento en línea y por mini lotes (consulte [Entrenamiento](#) el p. 13) dependen explícitamente del orden de casos; sin embargo, incluso el entrenamiento por lotes depende del orden de casos ya que la inicialización de las ponderaciones sinápticas implica el submuestreo del conjunto de datos.

Para minimizar los efectos del orden, ordene los casos aleatoriamente. Para comprobar la estabilidad de una solución dada, puede obtener varias soluciones distintas con los casos ordenados en distintos órdenes aleatorios. En situaciones con tamaños de archivo extremadamente grandes, se pueden llevar a cabo varias ejecuciones con una muestra de casos ordenados con distintos órdenes aleatorios.

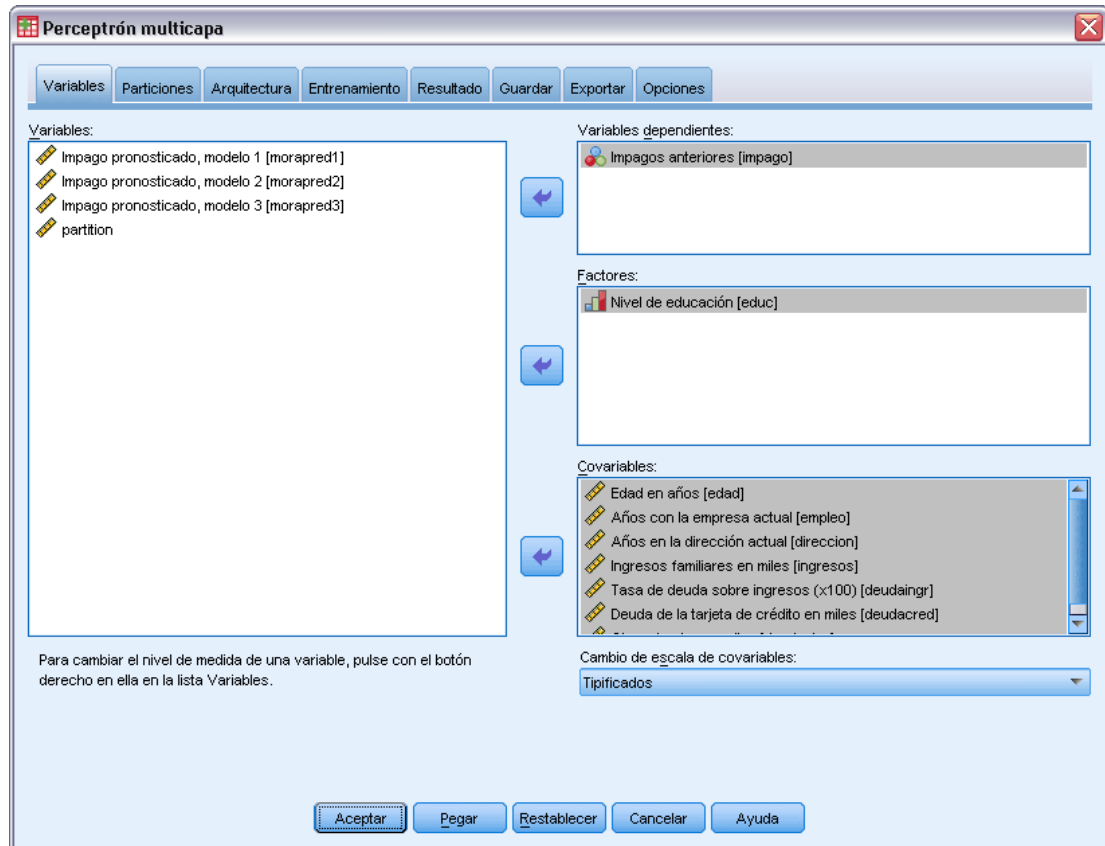
- **Orden de las variables.** Los resultados pueden verse afectados por el orden de las variables en las listas de factores y covariables, debido al diferente patrón de los valores iniciales asignados al cambiar el orden de las variables. Como con los efectos del orden de casos, puede probar diferentes órdenes de variables (sencillamente arrastrándolas y soltándolas en las listas de factores y covariables) para evaluar la estabilidad de una solución determinada.

Creación de una red de perceptrones multicapa

Seleccione en los menús:

Analizar > Redes neuronales > Perceptrón multicapa...

Figura 2-1
Perceptrón multicapa: pestaña Variables



- ▶ Seleccione al menos una variable dependiente.
- ▶ Seleccione al menos un factor o covariable.

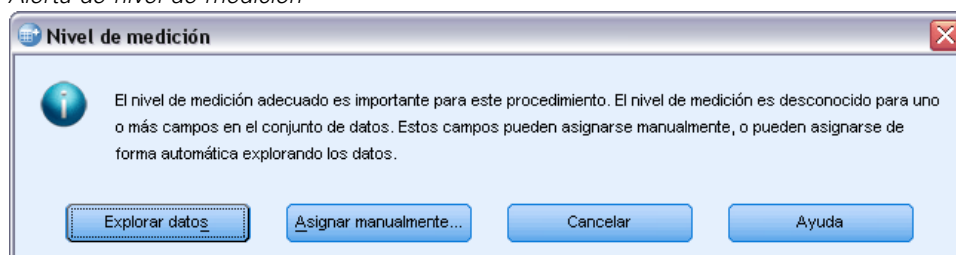
Si lo desea, en la pestaña Variables puede cambiar el método para cambiar la escala de las covariables. Las diferentes opciones son:

- **Tipificados.** Se resta la media y se divide por la desviación típica, $(x - \text{media})/s$.
- **Normalizados.** Se resta el mínimo y se divide por el rango, $(x - \text{mín.})/(\text{máx.} - \text{mín.})$. Los valores normalizados quedan comprendidos entre 0 y 1.
- **Normalizados corregidos.** Versión corregida de restar el mínimo y dividir por el rango, $[2 * (x - \text{mín.}) / (\text{máx.} - \text{mín.})] - 1$. Los valores normalizados corregidos quedan comprendidos entre -1 y 1.
- **Ninguna.** Sin cambio de escala de covariables.

Campos con un nivel de medición desconocido

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 2-2
Alerta de nivel de medición

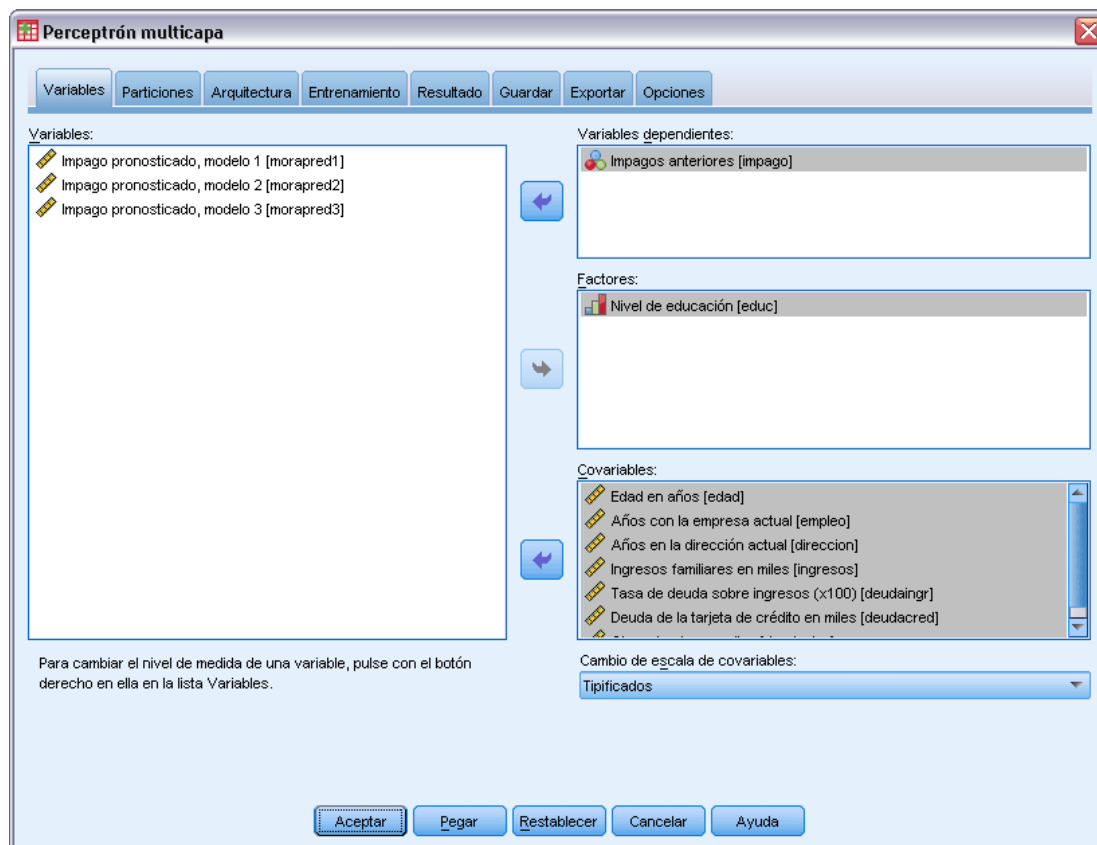


- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

Particiones

Figura 2-3
Perceptrón multicapa: Pestaña Particiones



Conjunto de datos de partición. Este grupo especifica el método de crear particiones en el conjunto de datos activo correspondientes a las muestras de entrenamiento, prueba y reserva. La **muestra de entrenamiento** comprende los registros de datos utilizados para entrenar la red neuronal; cierto porcentaje de casos del conjunto de datos debe asignarse a la muestra de entrenamiento para poder obtener un modelo. La **muestra de prueba** es un conjunto independiente de registros de datos que se utiliza para realizar un seguimiento de los errores durante el entrenamiento, con el fin de evitar un exceso de entrenamiento. Es muy recomendable crear una muestra de entrenamiento y el entrenamiento de la red generalmente será más eficiente si la muestra de prueba es más pequeña que la muestra de entrenamiento. La **muestra reservada** es otro conjunto independiente de registros de datos que se utiliza para evaluar la red neuronal final; el error de la muestra reservada ofrece una estimación “sincera” de la capacidad predictora del modelo, ya que los casos reservados no se utilizan para crear el modelo.

- **Asigne aleatoriamente los casos en función del número relativo de casos.** Especifique el número relativo (proporción) de casos asignados aleatoriamente a cada muestra (entrenamiento, prueba y reserva). Las columnas de % indican el porcentaje de casos que se asignarán a cada muestra según los números relativos que se han especificados.

Por ejemplo, especificar 7, 3, 0 como números relativos de las muestras de entrenamiento, prueba y reserva, equivale a especificar 70%, 30% y 0%. Especificar 2, 1, 1 como números relativos equivale a 50%, 25% y 25%; 1, 1, 1 equivale a dividir el conjunto de datos en tres partes iguales para entrenamiento, prueba y reserva.

- **Utilizar variable de partición para asignar los casos.** Especifique una variable numérica que asigne cada caso del conjunto de datos activo a la muestra de entrenamiento, prueba o reserva. Los casos con un valor positivo de la variable se asignarán a la muestra de entrenamiento, los casos con un valor 0 a la muestra de prueba y los casos con un valor negativo se asignarán a la muestra reservada. Los casos con un valor perdido del sistema se excluirán del análisis. Todos los valores perdidos de usuario de la variable de partición se tratarán siempre como válidos.

Nota: utilizar una variable de partición no garantiza resultados idénticos en posteriores ejecuciones de este procedimiento. Consulte “Replicado de resultados” en el tema principal [Perceptrón multicapa](#).

Arquitectura

Figura 2-4
Perceptrón multicapa: Pestaña Arquitectura

La pestaña Arquitectura se utiliza para especificar la estructura de la red. El procedimiento puede seleccionar automáticamente la “mejor” arquitectura o el usuario puede especificar una arquitectura personalizada.

La selección automática de arquitectura genera una red con una capa oculta. Especifique el número máximo y el mínimo de unidades permitido en la capa oculta, y la selección automática de arquitectura calculará cuál es el “mejor” número de unidades de la capa oculta. La selección automática de arquitectura utiliza las funciones de activación por defecto para las capas ocultas y de salida.

La selección de arquitectura personalizada le da un control de experto sobre las capas ocultas y de salida, y puede ser muy útil cuando se sabe de antemano qué arquitectura se desea o cuando necesita refinar los resultados de la selección automática de arquitectura.

Capas ocultas

La capa oculta contiene nodos de red no observables (unidades). Cada unidad oculta es una función de la suma ponderada de las entradas. La función es la función de activación y los valores de las ponderaciones se determinan mediante el algoritmo de estimación. Si la red contiene una segunda capa oculta, cada unidad oculta de la segunda capa es una función de la suma ponderada de las unidades de la primera capa oculta. La misma función de activación se utiliza en ambas capas.

Número de capas ocultas. Una percepción multicapa puede tener una o dos capas ocultas.

Función activación. La función de activación "relaciona" la suma ponderada de unidades de una capa, con los valores de unidades en la capa correcta.

- **Tangente hiperbólica.** Esta función tiene la siguiente forma: $\gamma(c) = \tanh(c) = (e^c - e^{-c}) / (e^c + e^{-c})$. Toma argumentos de valor real y los transforma al rango $(-1, 1)$. Cuando se utiliza la selección automática de arquitectura, ésta es la función de activación para todas las unidades de las capas ocultas.
- **Sigmoide.** Esta función tiene la siguiente forma: $\gamma(c) = 1 / (1 + e^{-c})$. Toma argumentos de valor real y los transforma al rango $(0, 1)$.

Número de unidades. El número de unidades en cada capa oculta se puede especificar explícitamente o el algoritmo de estimación puede determinarlo automáticamente.

Capa de salida

La capa de salida contiene las variables (dependientes) de destino.

Función activación. La función de activación "relaciona" la suma ponderada de unidades de una capa, con los valores de unidades en la capa correcta.

- **Identidad.** Esta función tiene la siguiente forma: $\gamma(c) = c$. Toma argumentos de valor real y los devuelve sin modificar. Cuando se utiliza la selección automática de arquitectura, ésta es la función de activación para las unidades de la capa de salida si hay variables dependientes de escala.
- **Softmax.** Esta función tiene la siguiente forma: $\gamma(c_k) = \exp(c_k) / \sum_j \exp(c_j)$. Toma un vector de argumentos de valor real y lo transforma en un vector cuyos elementos quedan comprendidos en el rango $(0, 1)$ y suman 1. Softmax está sólo disponible si todas las variables dependientes son categóricas. Cuando se utiliza la selección automática de arquitectura, ésta es la función de activación para las unidades de la capa de salida si todas las variables dependientes son categóricas.

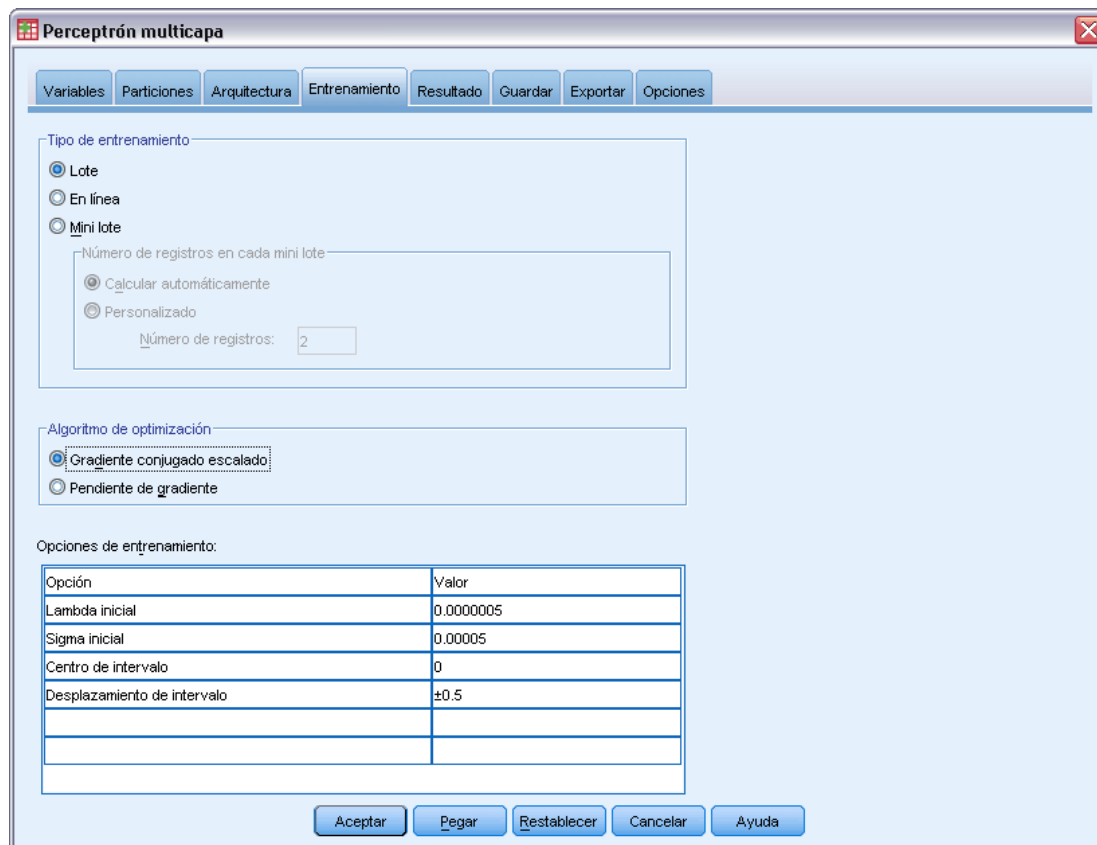
- **Tangente hiperbólica.** Esta función tiene la siguiente forma: $\gamma(c) = \tanh(c) = (e^c - e^{-c}) / (e^c + e^{-c})$. Toma argumentos de valor real y los transforma al rango $(-1, 1)$.
- **Sigmoide.** Esta función tiene la siguiente forma: $\gamma(c) = 1 / (1 + e^{-c})$. Toma argumentos de valor real y los transforma al rango $(0, 1)$.

Cambio de escala de variables dependientes de escala. Estos controles sólo se encuentran disponibles si se ha seleccionado al menos una variable dependiente de escala.

- **Tipificados.** Se resta la media y se divide por la desviación típica, $(x - \text{media}) / s$.
- **Normalizados.** Se resta el mínimo y se divide por el rango, $(x - \text{mín.}) / (\text{máx.} - \text{mín.})$. Los valores normalizados quedan comprendidos entre 0 y 1. Éste es el método de cambio de escala necesario para las variables dependientes de escala si la capa de salida utiliza la función de activación sigmoide. La opción de corrección especifica un pequeño número ϵ que se aplica como corrección a la fórmula de cambio de escala; esta corrección garantiza que los valores de todas las variables dependientes reescaladas quedan comprendidos dentro del rango de la función de activación. En concreto, los valores 0 y 1, que se producen en la fórmula no corregida cuando x toma los valores mínimo o máximo, definen los límites del rango de la función sigmoide, pero no se encuentran dentro de dicho rango. La fórmula corregida es $[x - (\text{mín} - \epsilon)] / [(\text{máx} + \epsilon) - (\text{mín} - \epsilon)]$. Especifique un número mayor o igual que 0.
- **Normalizados corregidos.** Versión corregida de restar el mínimo y dividir por el rango, $[2 * (x - \text{mín}) / (\text{máx} - \text{mín})] - 1$. Los valores normalizados corregidos quedan comprendidos entre -1 y 1 . Éste es el método de cambio de escala necesario para las variables dependientes de escala si la capa de salida utiliza la función de activación de tangente hiperbólica. La opción de corrección especifica un pequeño número ϵ que se aplica como corrección a la fórmula de cambio de escala; esta corrección garantiza que los valores de todas las variables dependientes reescaladas quedan comprendidos dentro del rango de la función de activación. En concreto, los valores -1 y 1 , que se producen en la fórmula no corregida cuando x toma los valores mínimo o máximo, definen los límites del rango de la función tangente hiperbólica, pero no se encuentran dentro de dicho rango. La fórmula corregida es $\{2 * [(x - (\text{mín} - \epsilon)) / ((\text{máx} + \epsilon) - (\text{mín} - \epsilon))]\} - 1$. Especifique un número mayor o igual que 0.
- **Ninguna.** Sin cambio de escala de las variables dependientes de escala.

Entrenamiento

Figura 2-5
Perceptrón multicapa: Pestaña Entrenamiento



La pestaña Entrenamiento se utiliza para especificar el modo en que debe entrenarse la red. El tipo de entrenamiento y el algoritmo de optimización determinan qué opciones de entrenamiento se encuentran disponibles.

Tipo de entrenamiento. El tipo de entrenamiento determina el modo en que la red procesa los registros. Seleccione uno de los siguientes tipos de entrenamiento:

- **Lote.** Actualiza las ponderaciones sinápticas sólo tras pasar todos los registros de datos de entrenamiento; es decir, el entrenamiento por lotes utiliza la información de todos los registros del conjunto de datos de entrenamiento. El entrenamiento por lotes se suele preferir porque minimiza directamente el error total; sin embargo, el entrenamiento por lotes puede obligar a actualizar muchas veces las ponderaciones hasta que se cumpla alguna de las reglas de parada y por tanto pueden ser necesarias muchas lecturas de datos. Resulta muy útil para conjuntos de datos “más pequeños”.
- **En línea.** Actualiza las ponderaciones sinápticas después de cada registro de datos de entrenamiento; es decir, el entrenamiento en línea utiliza la información de un solo registro por vez. El entrenamiento en línea obtiene continuamente un registro y actualiza las ponderaciones hasta que se cumple una regla de parada. Si se han utilizado todos los registros una vez y no se cumple ninguna regla de parada, el proceso continúa reciclando los registros

de datos. El entrenamiento en línea es superior al entrenamiento por lotes para los grupos de datos “más grandes” con predictores asociados; es decir, si hay muchos registros y muchas entradas, y sus valores no son independientes entre sí, el entrenamiento en línea puede obtener una respuesta razonable con mayor rapidez que el entrenamiento por lotes.

- **Por mini lotes.** Divide los registros de datos de entrenamiento en grupos de tamaño parecido y actualiza las ponderaciones sinápticas tras pasar un grupo; es decir, el entrenamiento por mini lotes utiliza la información de un grupo de registros. A continuación, el proceso recicla el grupo de datos si es necesario. El entrenamiento por mini lotes ofrece una solución intermedia entre el entrenamiento por lotes y en línea, y puede ser el idóneo para conjuntos de datos de “tamaño medio”. El procedimiento puede determinar automáticamente el número de registros de entrenamiento por mini lote, o bien puede especificar un entero mayor que 1 y menor o igual que el número máximo de casos para almacenar en memoria. Puede establecer el número máximo de casos que se almacenarán en memoria en la pestaña [Opciones](#).

Algoritmo de optimización. Éste es el método utilizado para estimar las ponderaciones sinápticas.

- **Gradiente conjugado escalado.** Los supuestos que justifican el uso de métodos de gradiente conjugado se aplican únicamente a los tipos de entrenamiento por lotes, de modo que este método no se encuentra disponible para el entrenamiento en línea o por mini lotes.
- **Pendiente de gradiente.** Este método debe utilizarse con el entrenamiento en línea o por mini lotes; también puede utilizarse con el entrenamiento por lotes.

Opciones de entrenamiento. Las opciones de entrenamiento le permiten ajustar el algoritmo de optimización. Generalmente no tendrá que cambiar estos ajustes a menos que la red experimente problemas con la estimación.

Entre las opciones de entrenamiento del algoritmo de gradiente conjugado escalado se incluyen:

- **Lambda inicial.** El valor inicial del parámetro lambda para el algoritmo de gradiente conjugado escalado. Especifique un número mayor que 0 y menor que 0,000001.
- **Sigma inicial.** El valor inicial del parámetro sigma para el algoritmo de gradiente conjugado escalado. Especifique un número mayor que 0 y menor que 0,0001.
- **Centro de intervalo y desplazamiento de intervalo.** El centro del intervalo (a_0) y el desplazamiento de intervalo (a) definen el intervalo $[a_0 - a, a_0 + a]$, en el que se generan aleatoriamente vectores de ponderación cuando se utiliza el recocido simulado. El recocido simulado se utiliza para salir de un mínimo local, con el objetivo de buscar el mínimo global, durante la aplicación del algoritmo de optimización. Este enfoque se utiliza en la inicialización de ponderaciones y la selección automática de arquitectura. Especifique un número para el centro de intervalo y un número mayor que 0 para el desplazamiento de intervalo.

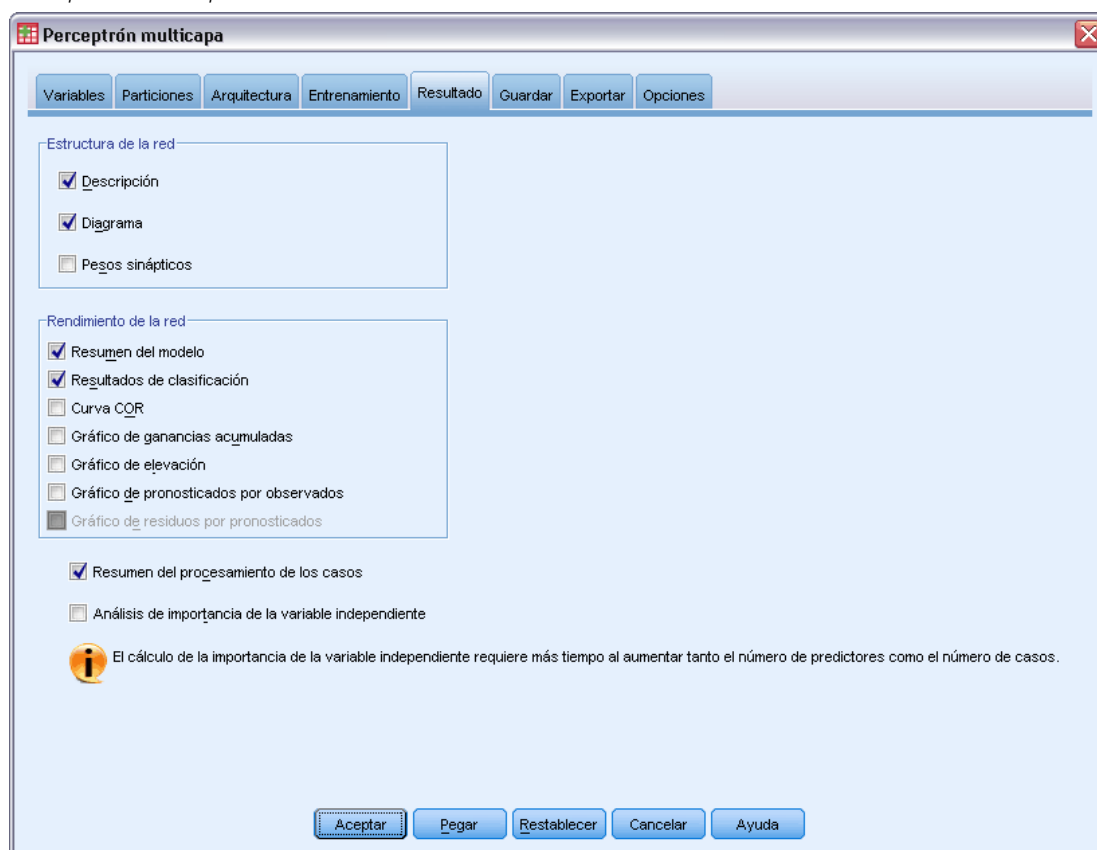
Entre las opciones de entrenamiento del algoritmo de pendiente de gradiente se incluyen:

- **Tasa de aprendizaje inicial.** El valor inicial de la tasa de aprendizaje para el algoritmo de pendiente de gradiente. Una tasa de aprendizaje superior significa que la red se entrenará más rápidamente, con el inconveniente de que puede volverse inestable. Especifique un número mayor que 0.
- **Límite inferior de la tasa de aprendizaje.** El límite inferior de la tasa de aprendizaje para el algoritmo de pendiente de gradiente. Este ajuste se aplica sólo al entrenamiento en línea y por mini lotes. Especifique un número mayor que 0 y menor que la tasa de aprendizaje inicial.

- **Impulso.** El parámetro de impulso inicial para el algoritmo de pendiente de gradiente. El término de impulso ayuda a evitar las inestabilidades provocadas por una tasa de aprendizaje demasiado elevada. Especifique un número mayor que 0.
- **Reducción de la tasa de aprendizaje, en épocas.** El número de épocas (p), o lecturas de datos de la muestra de entrenamiento, para reducir la tasa de aprendizaje inicial al límite inferior de la tasa de aprendizaje cuando la pendiente de gradiente se utiliza con el entrenamiento en línea o por mini lotes. Esto le da control sobre el factor de decaimiento de la tasa de aprendizaje $\beta = (1/pK) * \ln(\eta_0/\eta_{low})$, donde η_0 es la tasa de aprendizaje inicial, η_{low} es el límite inferior de la tasa de aprendizaje y K es el número total de mini lotes (o el número de registros de entrenamiento, para el entrenamiento en línea) en el conjunto de datos de entrenamiento. Introduzca un entero mayor que 0.

Salida

Figura 2-6
Perceptrón multicapa: Pestaña Resultados



Estructura de red. Muestra información resumida sobre la red neuronal.

- **Descripción.** Muestra información sobre la red neuronal, incluyendo las variables dependientes, el número de unidades de entrada y de salida, el número de unidades y capas ocultas, y las funciones de activación.

- **Diagrama.** Muestra el diagrama de red como un gráfico que se puede editar. Tenga en cuenta que a medida que aumenta el número de covariables y niveles del factor, el diagrama se hace más difícil de interpretar.
- **Ponderaciones sinápticas.** Muestran las estimaciones de coeficiente que muestran la relación entre las unidades de una capa determinada con las unidades de la capa siguiente. Las ponderaciones sinápticas se basan en la muestra de entrenamiento incluso si el conjunto de datos activo se divide en datos de entrenamiento, comprobación y reservados. Tenga en cuenta que el número de ponderaciones sinápticas puede llegar a ser bastante elevado, y estas ponderaciones generalmente no se utilizan para interpretar los resultados de red.

Rendimiento de red. Muestra los resultados utilizados para determinar si el modelo es “bueno”.

Nota: los gráficos de este grupo se basan en las muestras de entrenamiento y comprobación combinadas, o sólo la muestra de entrenamiento si no existe muestra de comprobación.

- **Resumen de modelo.** Muestra un resumen de los resultados de la red neuronal por partición y global, incluyendo el error, el error relativo o el porcentaje de pronósticos incorrectos, la regla de parada utilizada para detener el entrenamiento y el tiempo de entrenamiento.

El error es el error de sumas cuadráticas cuando se aplica la función de activación de tangente hiperbólica, identidad o sigmoide a la capa de salida. Éste es el error de entropía cruzada cuando se aplica la función de activación softmax a la capa de salida.

Los porcentajes o errores relativos de pronósticos incorrectos se muestran dependiendo de los niveles de medida de la variable dependiente. Si alguna variable dependiente tiene un nivel de medida de escala, se muestra el error relativo global promedio (relativo al modelo de la media). Si todas las variables dependientes son categóricas, se muestra el porcentaje promedio de pronósticos incorrectos. También se muestran los errores relativos o porcentajes de pronósticos incorrectos para las variables dependientes individuales.

- **Resultados de clasificación.** Muestra una tabla de clasificación para cada variable dependiente categórica por partición y global. Cada tabla proporciona el número de casos clasificados correcta e incorrectamente para cada categoría de variable dependiente. También se informa del porcentaje de casos totales que se clasificaron correctamente.
- **Curva COR.** Muestra una curva COR (del inglés Receiver Operating Characteristic, características operativas del receptor) para cada variable dependiente categórica. También muestra una tabla que proporciona el área bajo cada curva. Para una variable dependiente dada, el gráfico COR muestra una curva para cada categoría. Si la variable dependiente tiene dos categorías, cada curva trata la categoría en cuestión como el estado positivo frente a la otra categoría. Si la variable dependiente tiene más de dos categorías, cada curva trata la categoría en cuestión como el estado positivo frente a la agregación de las demás categorías.
- **Gráfico de ganancias acumuladas.** Muestra un gráfico de ganancias acumuladas para cada variable dependiente categórica. La presentación de una curva para cada categoría de variable dependiente es la misma que para las curvas COR.
- **Gráfico de elevación.** Muestra un gráfico de elevación para cada variable dependiente categórica. La presentación de una curva para cada categoría de variable dependiente es la misma que para las curvas COR.
- **Gráfico de pronosticados por observados.** Muestra un gráfico de valores pronosticados por observados para cada variable dependiente. Para las variables dependientes categóricas, se muestran diagramas de caja agrupados de pseudoprobabilidades pronosticadas para

cada categoría de respuesta, con la categoría de respuesta observada como la variable de conglomerado. Para las variables dependientes de escala se muestra un diagrama de dispersión.

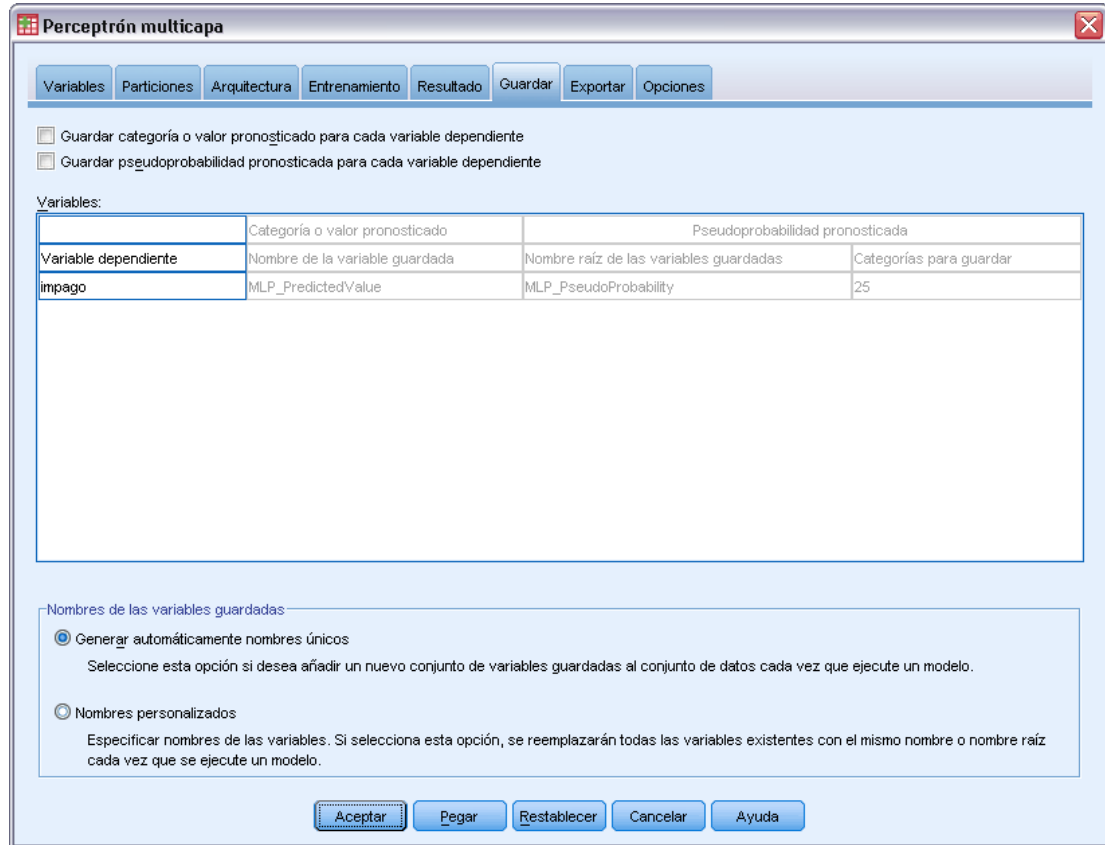
- **Gráfico de residuos por pronosticados.** Muestra un gráfico de residuos por valores pronosticados para cada variable dependiente de escala. No debería haber patrones visibles entre los residuos y los valores pronosticados. Este gráfico sólo se genera para las variables dependientes de escala.

Resumen de procesamiento de casos. Muestra la tabla de resumen de procesamiento de casos, que resume el número de casos incluidos y excluidos en el análisis, en total y por muestras de entrenamiento, comprobación y reservadas.

Análisis de importancia de la variable independiente. Realiza un análisis de susceptibilidad, que calcula la importancia de cada predictor en la determinación de la red neuronal. El análisis se basa en las muestras de entrenamiento y comprobación combinadas, o sólo la muestra de entrenamiento si no existe muestra de comprobación. Así se crea una tabla y un gráfico que muestra la importancia y la importancia normalizada de cada predictor. Tenga en cuenta que el análisis de susceptibilidad supone un proceso de cálculo muy extenso que puede tardar bastante si el número de predictores o de casos es elevado.

Guardar

Figura 2-7
Perceptrón multicapa: Pestaña Guardar



La pestaña Guardar se utiliza para guardar pronósticos como variables en el conjunto de datos.

- **Guardar categoría o valor pronosticado para cada variable dependiente.** Esta opción guarda el valor pronosticado para variables dependientes de escala y la categoría pronosticada para variables dependientes categóricas.
- **Guardar pseudoprobabilidad pronosticada o categoría para cada variable dependiente.** Esta opción guarda las pseudoprobabilidades pronosticadas para variables dependientes categóricas. Para cada una de las primeras n categorías se guarda una variable diferente, donde n se especifica en la columna Categorías para guardar.

Nombres de las variables guardadas. La generación automática de nombres garantiza que conserva todo su trabajo. Los nombres personalizados le permiten descartar/reemplazar los resultados de las ejecuciones anteriores sin eliminar antes las variables guardadas en el Editor de datos.

Probabilidades y pseudoprobabilidades

Las variables dependientes categóricas con activación softmax y error de entropía cruzada tendrán un valor pronosticado para cada categoría, donde cada valor pronosticado es la probabilidad de que el caso pertenezca a la categoría.

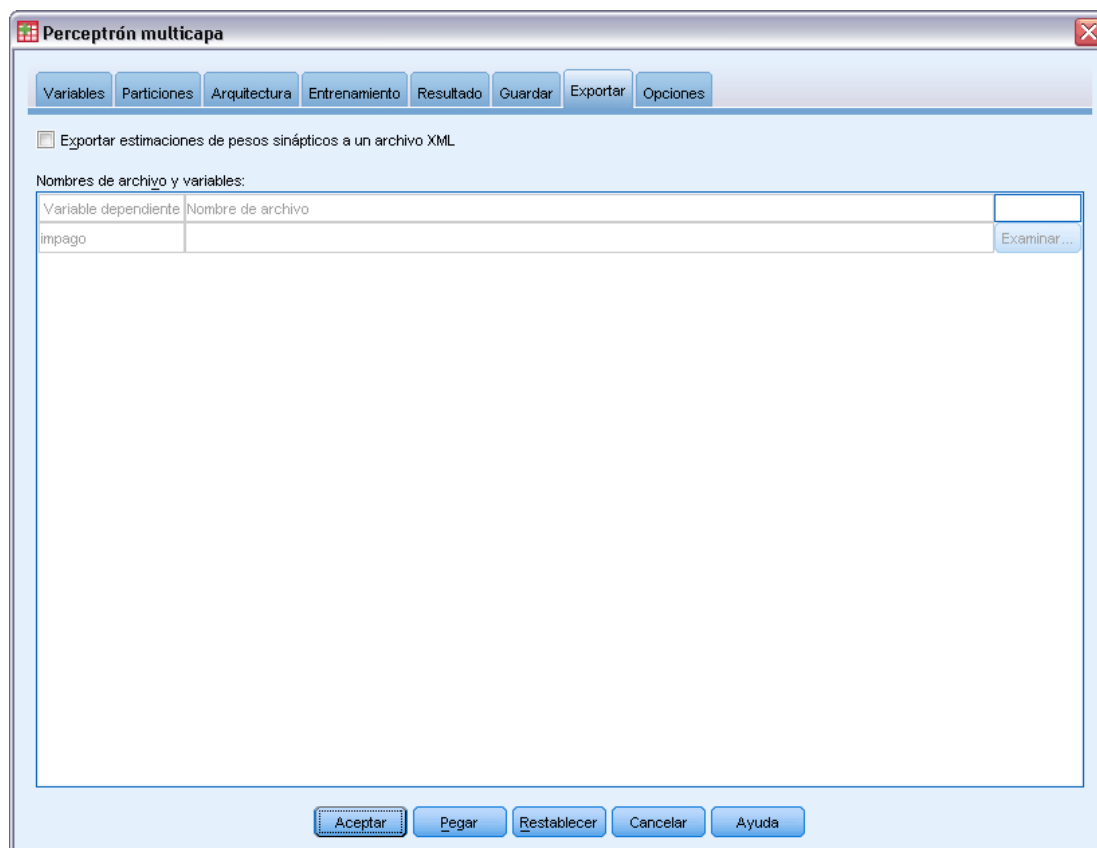
Las variables dependientes categóricas con error de sumas cuadráticas tendrá un valor pronosticado para cada categoría, pero los valores pronosticados no se pueden interpretar como probabilidades. El procedimiento guarda estas pseudoprobabilidades pronosticadas incluso si son menores que 0 o mayores que 1, o la suma para una variable dependiente dada no es 1.

Los gráficos de curva COR, ganancias acumuladas y elevación (consulte [Salida](#) el p. 15) se crean basándose en pseudoprobabilidades. En el caso de que alguna de las pseudoprobabilidades sea menor que 0 o mayor que 1, o la suma para una variable dada no sea 1, primero se reescalan para que estén entre 0 y 1 y para que sumen 1. Las pseudoprobabilidades se reescalan dividiéndolas por su suma. Por ejemplo, si un caso tiene pseudoprobabilidades pronosticadas de 0,50, 0,60 y 0,40 para una variable dependiente de tres categorías, cada pseudoprobabilidad se divide por la suma 1,50 para obtener 0,33, 0,40 y 0,27.

Si alguna de las pseudoprobabilidades es negativa, se suma el valor absoluto de la menor a todas las pseudoprobabilidades antes del cambio de escala anterior. Por ejemplo, si las pseudoprobabilidades son -0,30, 0,50 y 1,30, primero se añade 0,30 a cada valor para obtener 0,00, 0,80 y 1,60. A continuación, se divide cada nuevo valor por la suma 2,40 para obtener 0,00, 0,33 y 0,67.

Exportar

Figura 2-8
Perceptrón multicapa: pestaña Exportar



La pestaña Exportar se utiliza para guardar las estimaciones de ponderaciones sinápticas para cada variable dependiente en un archivo XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo. Esta opción no se encuentra disponible si se han definido archivos segmentados.

Opciones

Figura 2-9
Perceptrón multicapa: pestaña Opciones

Perceptrón multicapa

Variables Particiones Arquitectura Entrenamiento Resultado Guardar Exportar Opciones

Valores definidos como perdidos por el usuario

Especifique cómo desea tratar los casos con valores definidos como perdidos por el usuario en los factores y las variables dependientes categóricas.

Excluir Incluir

Los casos con valores definidos como perdidos por el usuario en las covariables o en las variables dependientes de escala siempre se excluyen.

Reglas de parada

Las reglas de parada se prueban en el orden que se indica a continuación.

Número máximo de pasos sin una disminución del error: 1

Datos para usar para calcular el error de predicción:

Elegir automáticamente Datos de entrenamiento y de prueba

Tiempo máximo de entrenamiento Minutos: 15

Número máximo de épocas de entrenamiento

Calcular automáticamente Especificar valor personalizado Número máximo de épocas:

Cambio mínimo relativo del error de entrenamiento: 0.0001

Cambio mínimo relativo de la tasa de errores de entrenamiento: 0.001

Número máximo de casos que se almacenarán en memoria: 1000

Aceptar Pegar Restablecer Cancelar Ayuda

Valores definidos como perdidos por el usuario. Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores definidos como perdidos por el usuario se deben tratar como válidos entre los factores y las variables dependientes categóricas.

Reglas de parada. Son las reglas que determinan cuando detener el entrenamiento de la red neuronal. El entrenamiento pasa al menos por una lectura de los datos. El entrenamiento puede entonces detenerse de acuerdo con los siguientes criterios, que se comprueban en el orden indicado

en la lista. En las siguientes definiciones de regla de parada, un paso corresponde a una lectura de datos para los métodos en línea y por mini lotes, una iteración para el método por lotes.

- **Número máximo de pasos sin una disminución del error.** El número de pasos que se permitirán antes de comprobar si existe una disminución del error. Si no existe una disminución del error tras el número especificado de pasos, se detiene el entrenamiento. Especifique un número entero mayor que 0. También puede especificar la muestra de datos que se va a utilizar para calcular el error. Elegir automáticamente utiliza la muestra de comprobación si la hay y, si no, la muestra de entrenamiento. Tenga en cuenta que el entrenamiento por lotes garantiza una disminución de los errores de la muestra de entrenamiento tras cada lectura de los datos, por lo que esta opción se aplica sólo al entrenamiento por lotes si existe una muestra de comprobación. Datos de entrenamiento y de prueba comprueba el error para cada una de estas muestras; esta opción se aplica únicamente si existe una muestra de comprobación.

Nota: tras cada lectura completa de los datos, el entrenamiento en línea y por mini lotes requiere una lectura adicional de los datos para calcular el error de entrenamiento. Esta lectura adicional de los datos puede ralentizar considerablemente el entrenamiento, por lo que generalmente se recomienda suministrar una muestra de comprobación y seleccionar Elegir automáticamente en cualquier caso.

- **Tiempo máximo de entrenamiento.** Elija si desea especificar un número máximo de minutos para que se ejecute el algoritmo. Especifique un número mayor que 0.
- **Número máximo de épocas de entrenamiento.** El número máximo de épocas (lecturas de datos) permitido. Si se supera el número máximo de épocas, se detiene el entrenamiento. Introduzca un entero mayor que 0.
- **Cambio mínimo relativo del error de entrenamiento.** El entrenamiento se detiene si el cambio relativo del error de entrenamiento comparado con el del paso anterior es menor que el criterio. Especifique un número mayor que 0. Para el entrenamiento en línea y por mini lotes, este criterio se ignora si sólo se utilizan datos de comprobación para calcular el error.
- **Cambio mínimo relativo de la tasa de errores de entrenamiento.** El entrenamiento se detiene si el cociente del error de entrenamiento respecto al error del modelo nulo es menor que el valor del criterio. El modelo nulo pronostica el valor promedio para todas las variables dependientes. Especifique un número mayor que 0. Para el entrenamiento en línea y por mini lotes, este criterio se ignora si sólo se utilizan datos de comprobación para calcular el error.

Número máximo de casos que se almacenarán en memoria. Esta opción controla los siguientes ajustes de los algoritmos del perceptrón multicapa. Introduzca un entero mayor que 1.

- En la selección automática de la arquitectura, el tamaño de la muestra utilizada para determinar la arquitectura de la red es $\min(1000, \text{tamañoem})$, donde *tamañoem* es el número máximo de casos que se pueden almacenar en la memoria.
- En el entrenamiento por mini lotes con cálculo automático del número de mini lotes, el número de mini lotes es $\min(\max(M/10, 2), \text{tamañoem})$, donde *M* es el número de casos de la muestra de entrenamiento.

Función de base radial

El procedimiento Función de base radial (RBF) genera un modelo predictivo para una o más variables dependientes (de destino) basado en los valores de las variables predictoras.




Ejemplo. Un proveedor de telecomunicaciones ha segmentado su base de clientes por patrones de uso del servicio, clasificando a los clientes en cuatro grupos. Una red RBF que utiliza los datos demográficos para pronosticar la pertenencia a los grupos permite a la compañía personalizar las ofertas para cada uno de los posibles clientes.









VARIABLES DEPENDIENTES. Las variables dependientes pueden ser:

- **Nominal.** Una variable se puede tratar como nominal si sus valores representan categorías que no obedecen a una ordenación intrínseca (por ejemplo, el departamento de la empresa en el que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- **Ordinal.** Una variable puede tratarse como ordinal cuando sus valores representan categorías con alguna ordenación intrínseca (por ejemplo, los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho). Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- **Escala.** Una variable puede tratarse como escala (continua) cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

El procedimiento supone que el nivel de medida adecuado se ha asignado a todas las variables dependientes, aunque puede cambiar temporalmente el nivel de medida para una variable si pulsa el botón derecho la variable en la lista de variables de origen y selecciona un nivel de medida en el menú contextual.

Un icono situado junto a cada variable de la lista de variables identifica el nivel de medida y el tipo de datos.

Nivel de medida	Tipo de datos			
	Numérico	Cadena	Fecha	Hora
Escala (Continuo)		n/a		

Ordinal				
Nominal				

Variables predictoras. Los predictores se pueden especificar como factores (categóricos) o covariables (de escala).

Codificación de la variable categórica. El procedimiento recodifica temporalmente predictores categóricos y variables dependientes utilizando la codificación “una de c ” para todo el procedimiento. Si hay c categorías de una variable, la variable se almacena como vectores c , con la primera categoría denotada $(1,0,\dots,0)$, la siguiente categoría $(0,1,0,\dots,0)$, ..., y la última categoría $(0,0,\dots,0,1)$.

Este esquema de codificación incrementa el número de ponderaciones sinápticas y puede dar como resultado un entrenamiento más lento, pero los métodos de codificación más “compactos” normalmente conducen a redes neuronales con un peor ajuste. Si el entrenamiento de la red avanza muy lentamente, pruebe a reducir el número de categorías en los predictores categóricos combinando categorías similares o eliminando los casos que tengan categorías extremadamente raras.

Toda codificación “una de c ” se basa en los datos de entrenamiento, incluso si se define una muestra reservada o de comprobación (consulte [Particiones](#) el p. 26). De este modo, si las muestras reservadas o de comprobación contienen casos con categorías de predictores que no están presentes en los datos de entrenamiento, esos casos no son utilizados por el procedimiento o en la puntuación. De este modo, si las muestras reservadas o de comprobación contienen casos con categorías de variables dependientes que no están presentes en los datos de entrenamiento, esos casos no son utilizados por el procedimiento o en la puntuación.

Cambio de escala. Por defecto, se cambia la escala de las covariables y las variables dependientes de escala para mejorar el entrenamiento de la red. Todo cambio de escala se realiza basándose en los datos de entrenamiento, incluso si se define una muestra reservada o de comprobación (consulte [Particiones](#) el p. 26). Es decir, dependiendo del tipo de cambio de escala, la media, la desviación típica, el valor mínimo o el valor máximo de una covariable o variable dependiente se calculan utilizando sólo los datos de entrenamiento. Si especifica una variable para definir particiones, es importante que estas covariables o variables dependientes tengan distribuciones similares en todas las muestras reservadas, de entrenamiento o comprobación.

Ponderaciones de frecuencia. Este procedimiento ignora las ponderaciones de frecuencia.

Replicación de los resultados. Si desea replicar exactamente sus resultados, utilice el mismo valor de inicialización para el generador de números aleatorios y el mismo orden de datos, además de utilizar la misma configuración del procedimiento. A continuación se incluyen más detalles sobre este tema:

- **Generación de números aleatorios.** El procedimiento utiliza la generación de números aleatorios durante la asignación de particiones. Para reproducir los mismos resultados aleatorios en el futuro, utilice el mismo valor de inicialización para el generador de números

aleatorios antes de cada ejecución del procedimiento Función de base radial. Consulte [Preparación de los datos para su análisis](#) el p. 72 para obtener instrucciones específicas.

- **Orden de casos.** Los resultados dependen también del orden de los datos porque se utiliza el algoritmo de conglomeración de dos fases para determinar las funciones de base radial.

Para minimizar los efectos del orden, ordene los casos aleatoriamente. Para comprobar la estabilidad de una solución dada, puede obtener varias soluciones distintas con los casos ordenados en distintos órdenes aleatorios. En situaciones con tamaños de archivo extremadamente grandes, se pueden llevar a cabo varias ejecuciones con una muestra de casos ordenados con distintos órdenes aleatorios.

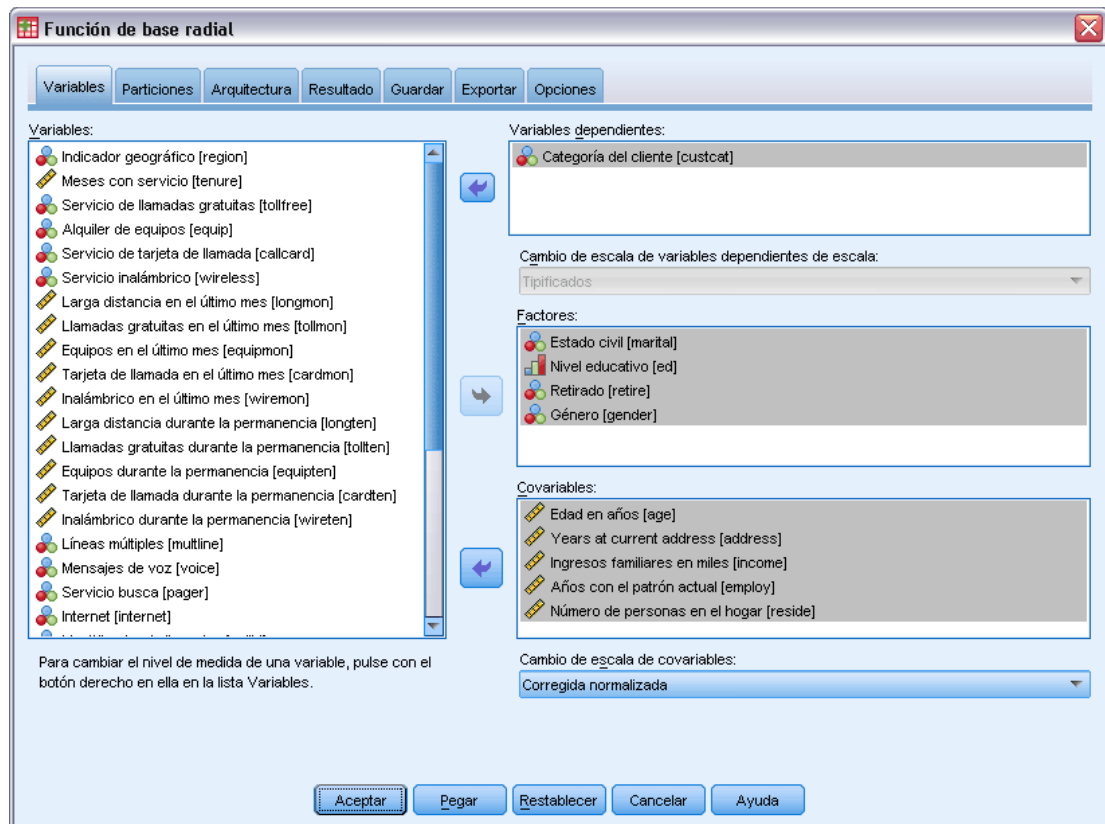
Creación de una red de función de base radial

Seleccione en los menús:

Analizar > Redes neuronales > Función de base radial...

Figura 3-1

Función de base radial: pestaña Variables



- ▶ Seleccione al menos una variable dependiente.
- ▶ Seleccione al menos un factor o covariable.

Si lo desea, en la pestaña Variables puede cambiar el método para cambiar la escala de las covariables. Las diferentes opciones son:

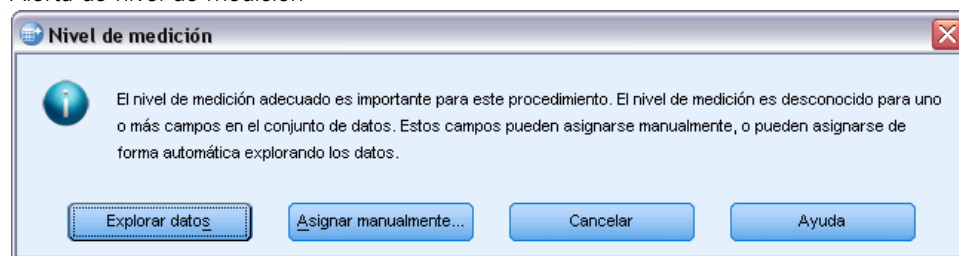
- **Tipificados.** Se resta la media y se divide por la desviación típica, $(x-\text{media})/s$.
- **Normalizados.** Se resta el mínimo y se divide por el rango, $(x-\text{mín.})/(\text{máx.}-\text{mín.})$. Los valores normalizados quedan comprendidos entre 0 y 1.
- **Normalizados corregidos.** Versión corregida de restar el mínimo y dividir por el rango, $[2*(x-\text{mín.})/(\text{máx.}-\text{mín.})]-1$. Los valores normalizados corregidos quedan comprendidos entre -1 y 1.
- **Ninguna.** Sin cambio de escala de covariables.

Campos con un nivel de medición desconocido

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 3-2

Alerta de nivel de medición

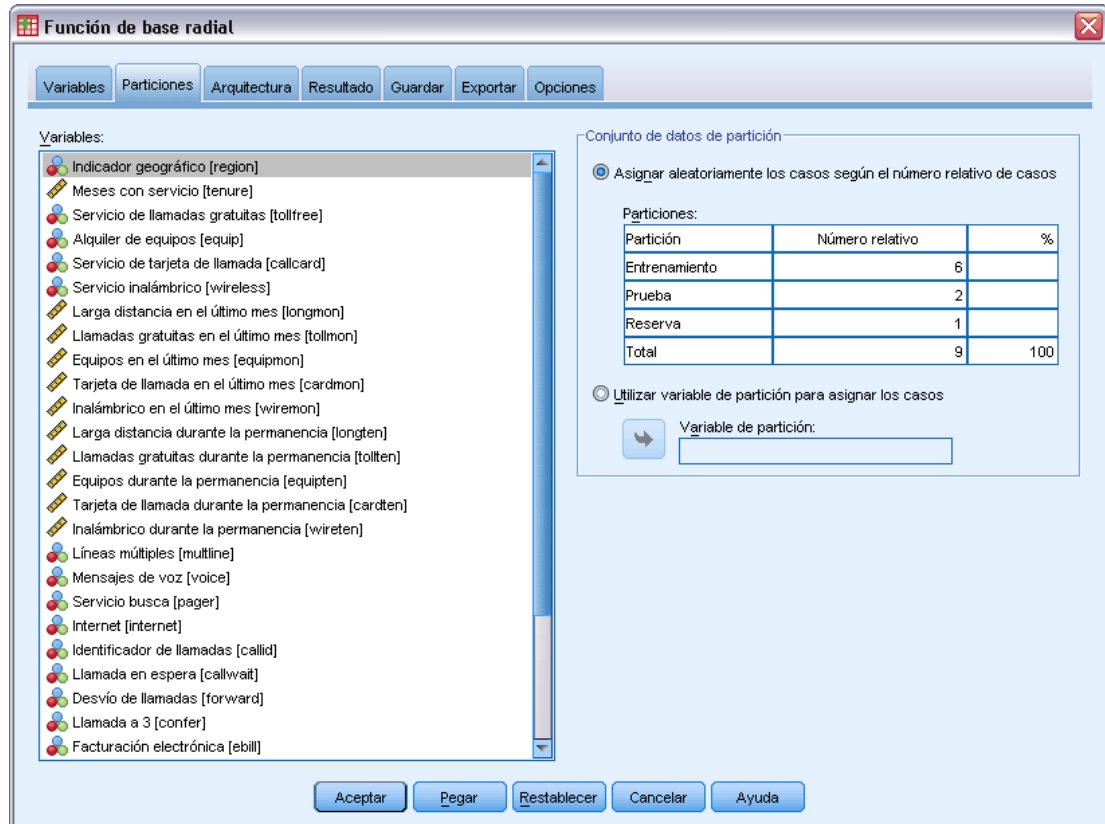


- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

Particiones

Figura 3-3
Función de base radial: Pestaña Particiones



Conjunto de datos de partición. Este grupo especifica el método de crear particiones en el conjunto de datos activo correspondientes a las muestras de entrenamiento, prueba y reserva. La **muestra de entrenamiento** comprende los registros de datos utilizados para entrenar la red neuronal; cierto porcentaje de casos del conjunto de datos debe asignarse a la muestra de entrenamiento para poder obtener un modelo. La **muestra de prueba** es un conjunto independiente de registros de datos que se utiliza para realizar un seguimiento de los errores durante el entrenamiento, con el fin de evitar un exceso de entrenamiento. Es muy recomendable crear una muestra de entrenamiento y el entrenamiento de la red generalmente será más eficiente si la muestra de prueba es más pequeña que la muestra de entrenamiento. La **muestra reservada** es otro conjunto independiente de registros de datos que se utiliza para evaluar la red neuronal final; el error de la muestra reservada ofrece una estimación “sincera” de la capacidad predictora del modelo, ya que los casos reservados no se utilizan para crear el modelo.

- **Asigne aleatoriamente los casos en función del número relativo de casos.** Especifique el número relativo (proporción) de casos asignados aleatoriamente a cada muestra (entrenamiento, prueba y reserva). Las columnas de % indican el porcentaje de casos que se asignarán a cada muestra según los números relativos que se han especificados.

Por ejemplo, especificar 7, 3, 0 como números relativos de las muestras de entrenamiento, prueba y reserva, equivale a especificar 70%, 30% y 0%. Especificar 2, 1, 1 como números relativos equivale a 50%, 25% y 25%; 1, 1, 1 equivale a dividir el conjunto de datos en tres partes iguales para entrenamiento, prueba y reserva.

- **Utilizar variable de partición para asignar los casos.** Especifique una variable numérica que asigne cada caso del conjunto de datos activo a la muestra de entrenamiento, prueba o reserva. Los casos con un valor positivo de la variable se asignarán a la muestra de entrenamiento, los casos con un valor 0 a la muestra de prueba y los casos con un valor negativo se asignarán a la muestra reservada. Los casos con un valor perdido del sistema se excluirán del análisis. Todos los valores perdidos de usuario de la variable de partición se tratarán siempre como válidos.

Arquitectura

Figura 3-4
Función de base radial: Pestaña Arquitectura

The screenshot shows the 'Función de base radial' dialog box with the 'Arquitectura' tab selected. The dialog has a title bar with a close button and a menu bar with options: Variables, Particiones, Arquitectura, Resultado, Guardar, Exportar, and Opciones. The main content area is divided into three sections:

- Número de unidades de la capa oculta:** Contains three radio button options:
 - Buscar el mejor número de unidades dentro de un rango. Below this is a 'Rango' section with two radio buttons:
 - Calcular rango automáticamente
 - Usar un rango especificado. Below this are two input fields: 'Mínimo:' and 'Máximo:'.
 - Utilizar un número especificado de unidades. Below this is an input field: 'Número:'.
- Función de activación para la capa oculta:** Contains two radio button options:
 - Función de base radial normalizada
 - Función de base radial ordinaria
- Superponer entre las unidades ocultas:** Contains two radio button options:
 - Calcular automáticamente la cantidad de superposición permitida
 - Permitir la cantidad especificada de superposición. Below this is an input field: 'Factor de superposición:'.

At the bottom of the dialog are five buttons: Aceptar, Pegar, Restablecer, Cancelar, and Ayuda.

La pestaña Arquitectura se utiliza para especificar la estructura de la red. El procedimiento crea una red neuronal con una capa “función de base radial”; en general, no será necesario cambiar esta configuración.

Número de unidades en la capa oculta. Hay tres formas de elegir el número de unidades ocultas.

1. **Busque el mejor número de unidades en un rango calculado automáticamente.** El procedimiento calcula automáticamente los valores mínimo y máximo del rango y localiza el mejor número de unidades ocultas en el rango.

Si se define una muestra de comprobación, el procedimiento utiliza el criterio de los datos de comprobación: El mejor número de unidades ocultas es el que produce el error menor en los datos de comprobación. Si no se define una muestra de comprobación, el procedimiento utiliza el criterio de información bayesiano (BIC): El mejor número de unidades ocultas es el que produce el menor BIC según los datos de entrenamiento.

2. **Busque el mejor número de unidades en un rango especificado.** Puede proporcionar su propio rango y el procedimiento encontrará el “mejor” número de unidades ocultas en ese rango. Como antes, el mejor número de unidades ocultas del rango se determina utilizando el criterio de datos de comprobación o el BIC.
3. **Utilizar un número especificado de unidades.** Puede anular el uso de un rango y especificar un número determinado de unidades directamente.

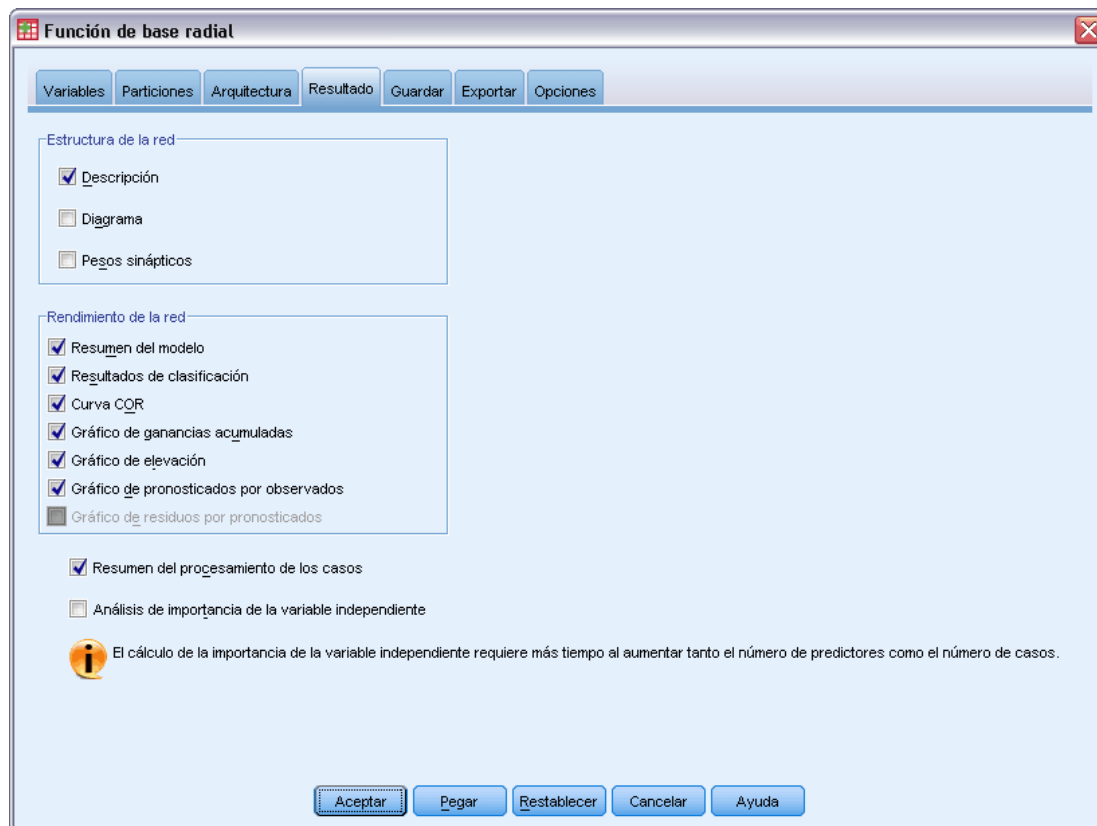
Función de activación para la capa oculta. La función de activación para la capa oculta es la función de base radial, que “enlaza” las unidades de una capa con los valores de unidades de la capa siguiente. Para la capa de resultado, la función de activación es la función de identidad, por lo que las unidades de resultado son simplemente las sumas ponderadas de las unidades ocultas.

- **Función de base radial normalizada.** Utiliza la función de activación softmax de modo que las activaciones de todas las unidades ocultas se normalizan para sumar uno.
- **Función de base radial ordinaria.** Utiliza la función de activación exponencial de modo que la activación de la unidad oculta es una “curva” de Gauss como una función de las entradas.

Superposición entre unidades ocultas. El factor de superposición es un multiplicador aplicado al ancho de las funciones de base radial. El valor calculado automáticamente del factor de superposición $1+0,1d$, donde d es el número de unidades de entrada (la suma del número de categorías de todos los factores y el número de covariables).

Resultado

Figura 3-5
Función de base radial: Pestaña Resultados



Estructura de red. Muestra información resumida sobre la red neuronal.

- **Descripción.** Muestra información sobre la red neuronal, incluyendo las variables dependientes, el número de unidades de entrada y de salida, el número de unidades y capas ocultas, y las funciones de activación.
- **Diagrama.** Muestra el diagrama de red como un gráfico que se puede editar. Tenga en cuenta que a medida que aumenta el número de covariables y niveles del factor, el diagrama se hace más difícil de interpretar.
- **Ponderaciones sinápticas.** Muestran las estimaciones de coeficiente que muestran la relación entre las unidades de una capa determinada con las unidades de la capa siguiente. Las ponderaciones sinápticas se basan en la muestra de entrenamiento incluso si el conjunto de datos activo se divide en datos de entrenamiento, comprobación y reservados. Tenga en cuenta que el número de ponderaciones sinápticas puede llegar a ser bastante elevado, y estas ponderaciones generalmente no se utilizan para interpretar los resultados de red.

Rendimiento de red. Muestra los resultados utilizados para determinar si el modelo es “bueno”.

Nota: los gráficos de este grupo se basan en las muestras de entrenamiento y comprobación combinadas, o sólo la muestra de entrenamiento si no existe muestra de comprobación.

- **Resumen de modelo.** Muestra un resumen de los resultados de la red neuronal por partición y global, incluyendo el error, el error relativo o el porcentaje de pronósticos incorrectos y el tiempo de entrenamiento.

El error es el error de sumas cuadráticas. Además, se muestran los errores relativos o porcentajes de pronósticos incorrectos dependiendo de los niveles de medida de la variable dependiente. Si alguna variable dependiente tiene un nivel de medida de escala, se muestra el error relativo global promedio (relativo al modelo de la media). Si todas las variables dependientes son categóricas, se muestra el porcentaje promedio de pronósticos incorrectos. También se muestran los errores relativos o porcentajes de pronósticos incorrectos para las variables dependientes individuales.

- **Resultados de clasificación.** Muestra una tabla de clasificación para cada variable dependiente categórica. Cada tabla proporciona el número de casos clasificados correctamente e incorrectamente para cada categoría de variable dependiente. También se informa del porcentaje de casos totales que se clasificaron correctamente.
- **Curva COR.** Muestra una curva COR (del inglés Receiver Operating Characteristic, características operativas del receptor) para cada variable dependiente categórica. También muestra una tabla que proporciona el área bajo cada curva. Para una variable dependiente dada, el gráfico COR muestra una curva para cada categoría. Si la variable dependiente tiene dos categorías, cada curva trata la categoría en cuestión como el estado positivo frente a la otra categoría. Si la variable dependiente tiene más de dos categorías, cada curva trata la categoría en cuestión como el estado positivo frente a la agregación de las demás categorías.
- **Gráfico de ganancias acumuladas.** Muestra un gráfico de ganancias acumuladas para cada variable dependiente categórica. La presentación de una curva para cada categoría de variable dependiente es la misma que para las curvas COR.
- **Gráfico de elevación.** Muestra un gráfico de elevación para cada variable dependiente categórica. La presentación de una curva para cada categoría de variable dependiente es la misma que para las curvas COR.
- **Gráfico de pronosticados por observados.** Muestra un gráfico de valores pronosticados por observados para cada variable dependiente. Para las variables dependientes categóricas, se muestran diagramas de caja agrupados de pseudoprobabilidades pronosticadas para cada categoría de respuesta, con la categoría de respuesta observada como la variable de conglomerado. Para las variables dependientes de escala se muestra un diagrama de dispersión.
- **Gráfico de residuos por pronosticados.** Muestra un gráfico de residuos por valores pronosticados para cada variable dependiente de escala. No debería haber patrones visibles entre los residuos y los valores pronosticados. Este gráfico sólo se genera para las variables dependientes de escala.

Resumen de procesamiento de casos. Muestra la tabla de resumen de procesamiento de casos, que resume el número de casos incluidos y excluidos en el análisis, en total y por muestras de entrenamiento, comprobación y reservadas.

Análisis de importancia de la variable independiente. Realiza un análisis de susceptibilidad, que calcula la importancia de cada predictor en la determinación de la red neuronal. El análisis se basa en las muestras de entrenamiento y comprobación combinadas, o sólo la muestra de entrenamiento si no existe muestra de comprobación. Así se crea una tabla y un gráfico que muestra la importancia y la importancia normalizada de cada predictor. Tenga en cuenta que el análisis de susceptibilidad supone un proceso de cálculo muy extenso que puede tardar bastante si el número de predictores o de casos es elevado.

Guardar

Figura 3-6
Función de base radial: Pestaña Guardar

Función de base radial

Variables Particiones Arquitectura Resultado Guardar Exportar Opciones

Guardar categoría o valor pronosticado para cada variable dependiente
 Guardar pseudoprobabilidad pronosticada para cada variable dependiente

Variables:

Variable dependiente	Categoría o valor pronosticado		Pseudoprobabilidad pronosticada	
	Nombre de la variable guardada	Nombre raíz de las variables guardadas	Categorías para guardar	
custcat	RBF_PredictedValue	RBF_PseudoProbability	25	

Nombres de las variables guardadas:

Generar automáticamente nombres únicos
 Seleccione esta opción si desea añadir un nuevo conjunto de variables guardadas al conjunto de datos cada vez que ejecute un modelo.

Nombres personalizados
 Especificar nombres de las variables. Si selecciona esta opción, se reemplazarán todas las variables existentes con el mismo nombre o nombre raíz cada vez que se ejecute un modelo.

Aceptar Pegar Restablecer Cancelar Ayuda

La pestaña Guardar se utiliza para guardar pronósticos como variables en el conjunto de datos.

- **Guardar categoría o valor pronosticado para cada variable dependiente.** Esta opción guarda el valor pronosticado para variables dependientes de escala y la categoría pronosticada para variables dependientes categóricas.
- **Guardar pseudoprobabilidad pronosticada para cada variable dependiente.** Esta opción guarda las pseudoprobabilidades pronosticadas para variables dependientes categóricas. Para cada una de las primeras n categorías se guarda una variable diferente, donde n se especifica en la columna *Categorías para guardar*.

Nombres de las variables guardadas. La generación automática de nombres garantiza que conserva todo su trabajo. Los nombres personalizados le permiten descartar/reemplazar los resultados de las ejecuciones anteriores sin eliminar antes las variables guardadas en el Editor de datos.

Probabilidades y pseudoprobabilidades

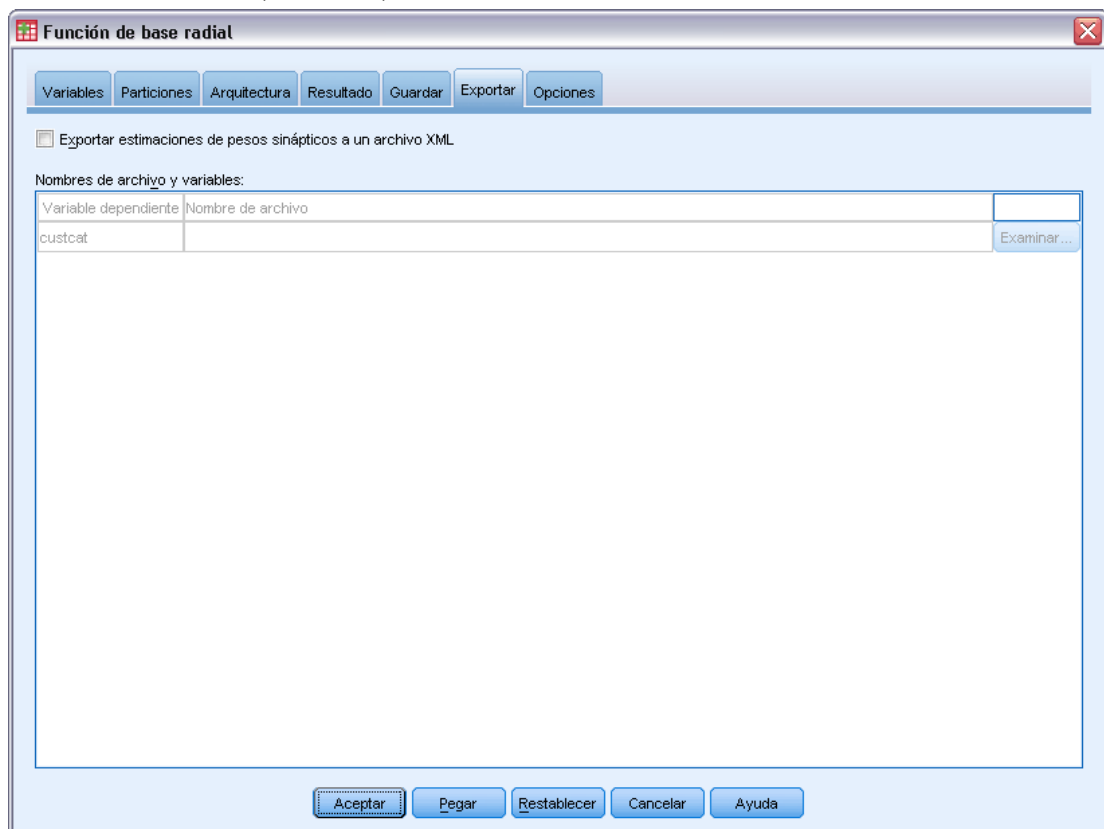
Las pseudoprobabilidades pronosticadas no se pueden interpretar como probabilidades porque el procedimiento Función de base radial utiliza el error de sumas cuadráticas y la función de activación de identidad para la capa de salida. El procedimiento guarda estas pseudoprobabilidades pronosticadas incluso si son menores que cero o mayores que uno, o la suma para una variable dependiente dada no es 1.

Los gráficos de curva COR, ganancias acumuladas y elevación (consulte [Resultado](#) el p. 29) se crean basándose en pseudoprobabilidades. En el caso de que alguna de las pseudoprobabilidades sea menor que 0 o mayor que 1, o la suma para una variable dada no sea 1, primero se reescalan para que estén entre 0 y 1 y para que sumen 1. Las pseudoprobabilidades se reescalan dividiéndolas por su suma. Por ejemplo, si un caso tiene pseudoprobabilidades pronosticadas de 0,50, 0,60 y 0,40 para una variable dependiente de tres categorías, cada pseudoprobabilidad se divide por la suma 1,50 para obtener 0,33, 0,40 y 0,27.

Si alguna de las pseudoprobabilidades es negativa, se suma el valor absoluto de la menor a todas las pseudoprobabilidades antes del cambio de escala anterior. Por ejemplo, si las pseudoprobabilidades son -0,30, 0,50 y 1,30, primero se añade 0,30 a cada valor para obtener 0,00, 0,80 y 1,60. A continuación, se divide cada nuevo valor por la suma 2,40 para obtener 0,00, 0,33 y 0,67.

Exportar

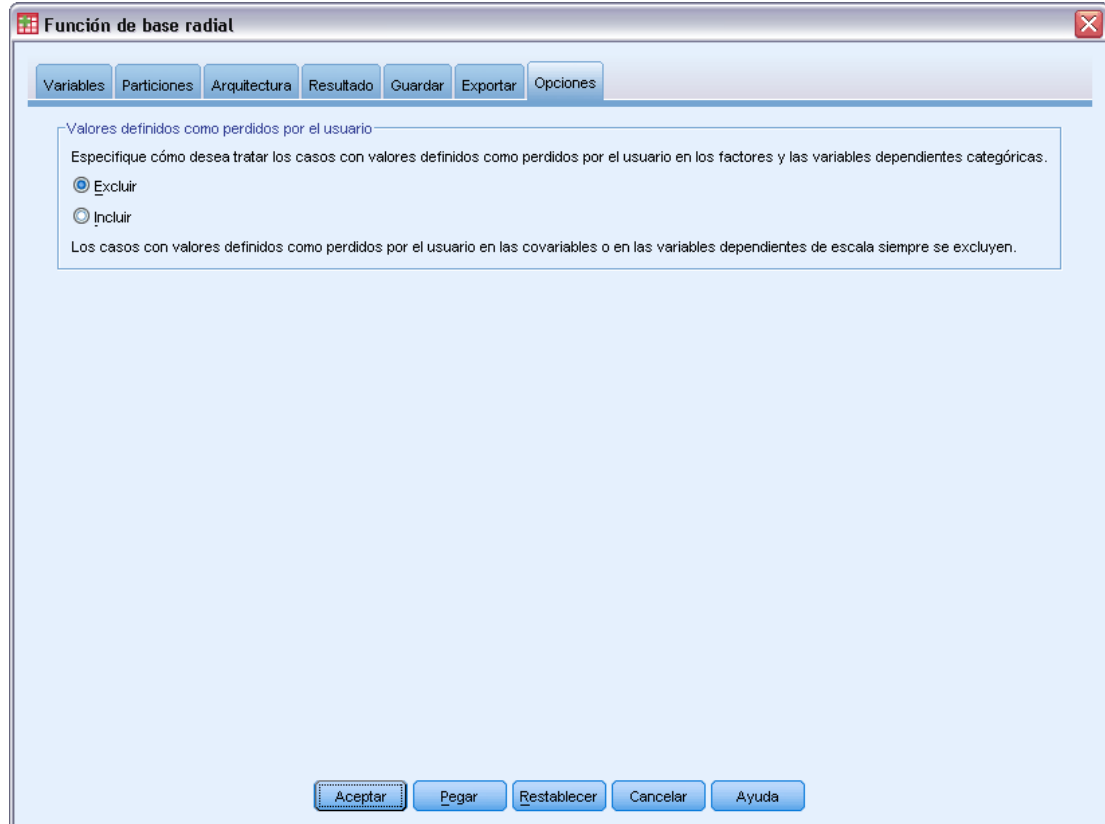
Figura 3-7
Función de base radial: pestaña Exportar



La pestaña Exportar se utiliza para guardar las estimaciones de ponderaciones sinápticas para cada variable dependiente en un archivo XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo. Esta opción no se encuentra disponible si se han definido archivos segmentados.

Opciones

Figura 3-8
Función de base radial: pestaña Opciones



Valores definidos como perdidos por el usuario. Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores definidos como perdidos por el usuario se deben tratar como válidos entre los factores y las variables dependientes categóricas.

Parte II: Ejemplos

Perceptrón multicapa

El procedimiento Perceptrón multicapa (MLP) genera un modelo predictivo para una o más variables dependientes (de destino) basada en los valores de las variables predictoras.

Uso del perceptrón multicapa para evaluar riesgos de crédito

El encargado de préstamos de un banco desea poder identificar características que sean indicativas de personas que puedan causar mora en los créditos y utilizar dichas características para identificar riesgos de crédito positivos y negativos.

Supongamos que los datos de 850 clientes anteriores y posibles clientes se encuentran en el archivo *bankloan.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 86](#). Los primeros 700 casos son clientes a los que anteriormente se les ha concedido un préstamo. Utilice una muestra aleatoria de estos 700 clientes para crear un perceptrón multicapa, reservando los clientes restantes para validar el análisis. A continuación, utilice el modelo para clasificar los 150 posibles clientes según los riesgos de crédito que presentan.

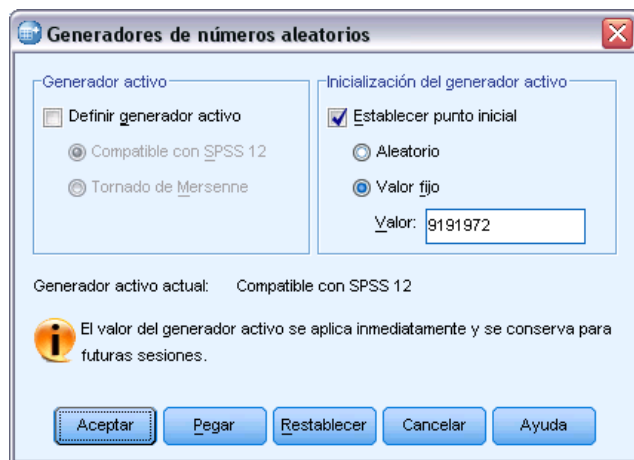
Además, el encargado de préstamos ha analizado anteriormente los datos mediante regresión logística (en la opción Regresión) y quiere evaluar el perceptrón multicapa como herramienta de clasificación.

Preparación de los datos para su análisis

Al establecer la semilla aleatoria, puede replicar el análisis de manera exacta.

- ▶ Para establecer la semilla aleatoria, elija en los menús:
Transformar > Generadores de números aleatorios...

Figura 4-1
Cuadro de diálogo Generadores de números aleatorios

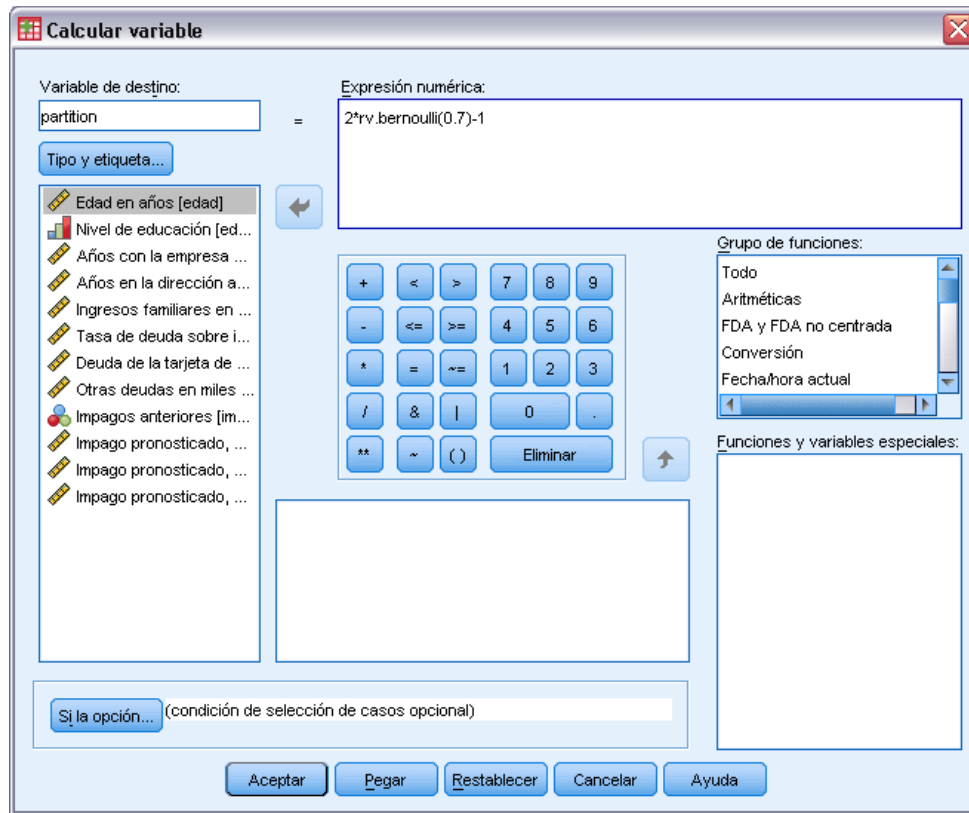


- ▶ Seleccione Establecer punto inicial.
- ▶ Seleccione Valor fijo y escriba 9191972 como el valor.
- ▶ Pulse en Aceptar.

En el análisis de regresión logística anterior, aproximadamente el 70% de los clientes anteriores se asignaron a la muestra de entrenamiento y el 30% a una muestra reservada. Se necesita una variable de partición para volver a crear con exactitud las muestras utilizadas en esos análisis.

- ▶ Para crear la variable de partición, elija en los menús:
Transformar > Calcular variable...

Figura 4-2
Cuadro de diálogo *Calcular variable*



- ▶ Escriba *partición* en el cuadro de texto Variable de destino.
- ▶ Escriba $2*rv.bernoulli(0.7)-1$ en el cuadro de texto Expresión numérica.

Esto establece los valores de *partición* para ser variantes de **Bernoulli** generadas de forma aleatoria con un parámetro de probabilidad 0,7, modificado para tomar los valores 1 o -1, en lugar de 1 o 0. Recuerde que los casos con valores positivos de la variable de *partición* se asignan a la muestra de entrenamiento, los casos con valores negativos se asignan a la muestra reservada y los casos con el valor 0 se asignan a la muestra de prueba. Por ahora, no especificaremos una muestra de prueba.

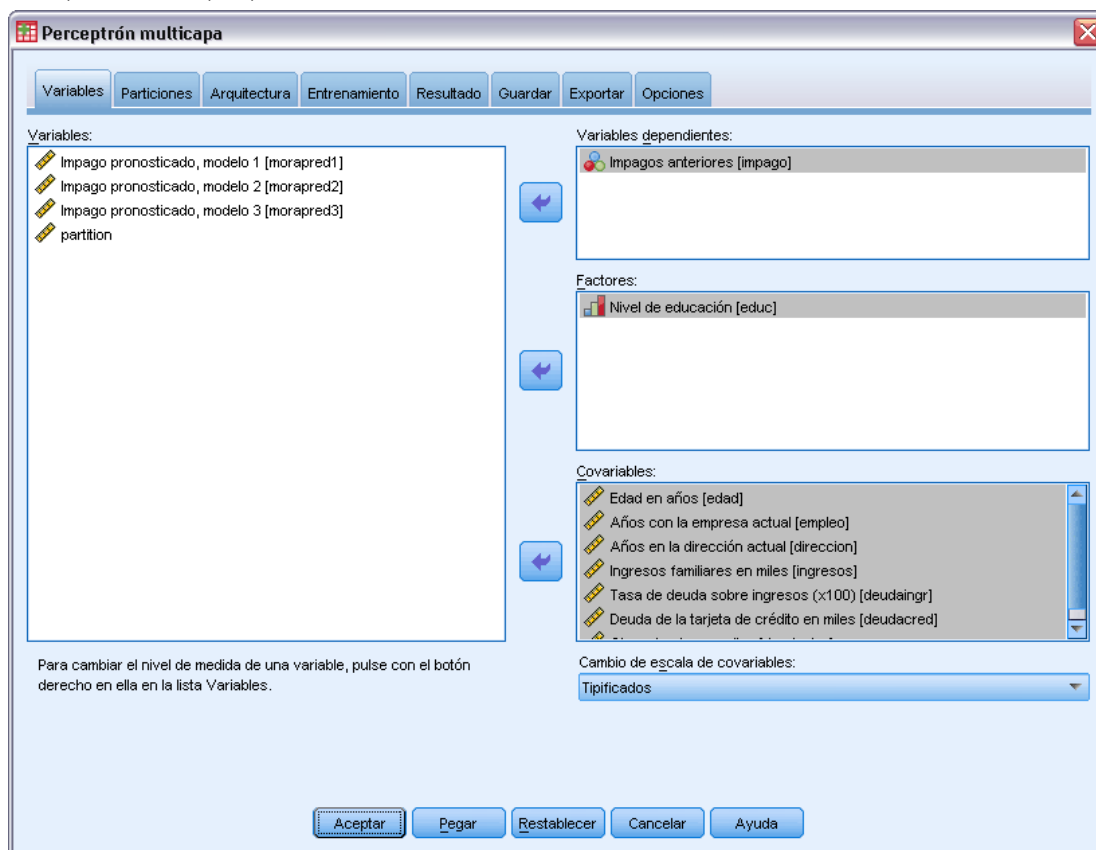
- ▶ Pulse Aceptar en el cuadro de diálogo *Calcular variable*.

Aproximadamente el 70% de los clientes que ya han obtenido créditos tendrán un valor de *partición* de 1. Estos clientes se utilizarán para crear el modelo. El resto de los clientes a quienes ya se habían concedido créditos se les asignará un valor de *partición* de -1 y se utilizarán para validar los resultados del modelo.

Ejecución del análisis

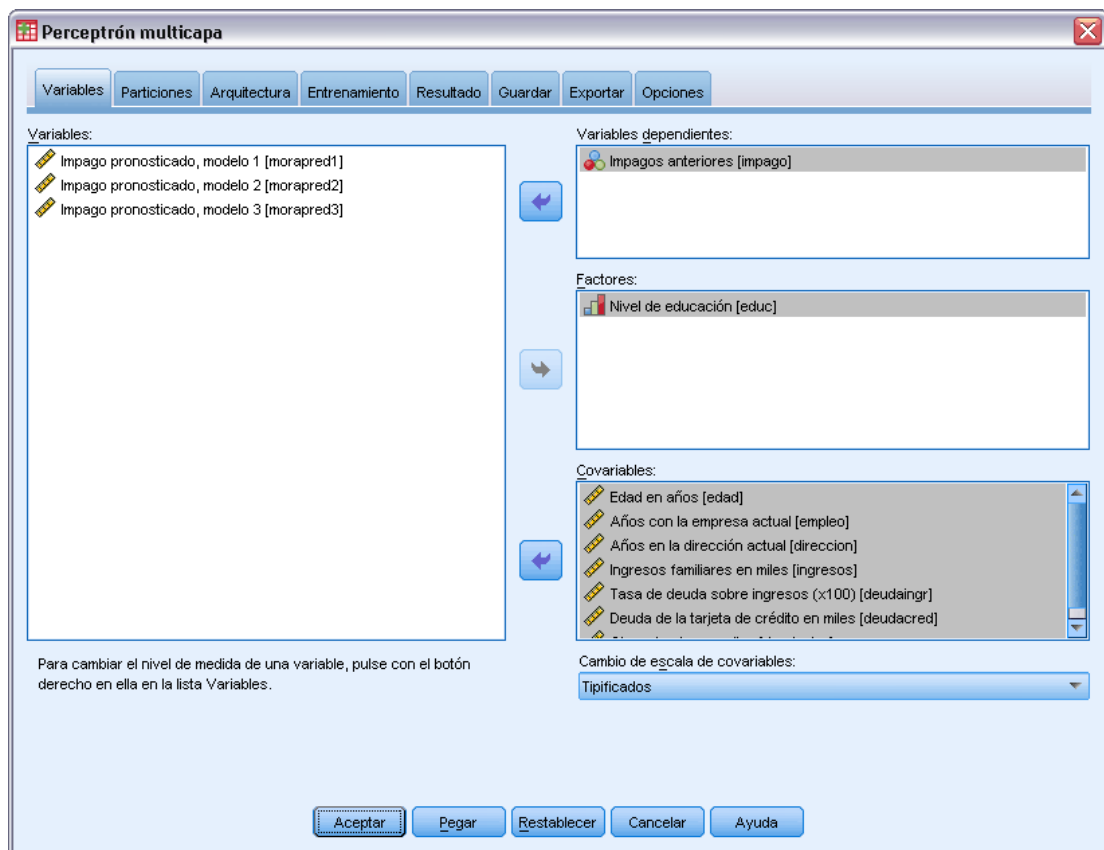
- ▶ Para ejecutar un análisis de perceptrón multicapa, elija en los menús:
Analizar > Neural Networks > Perceptrón multicapa...

Figura 4-3
Perceptrón multicapa: pestaña Variables



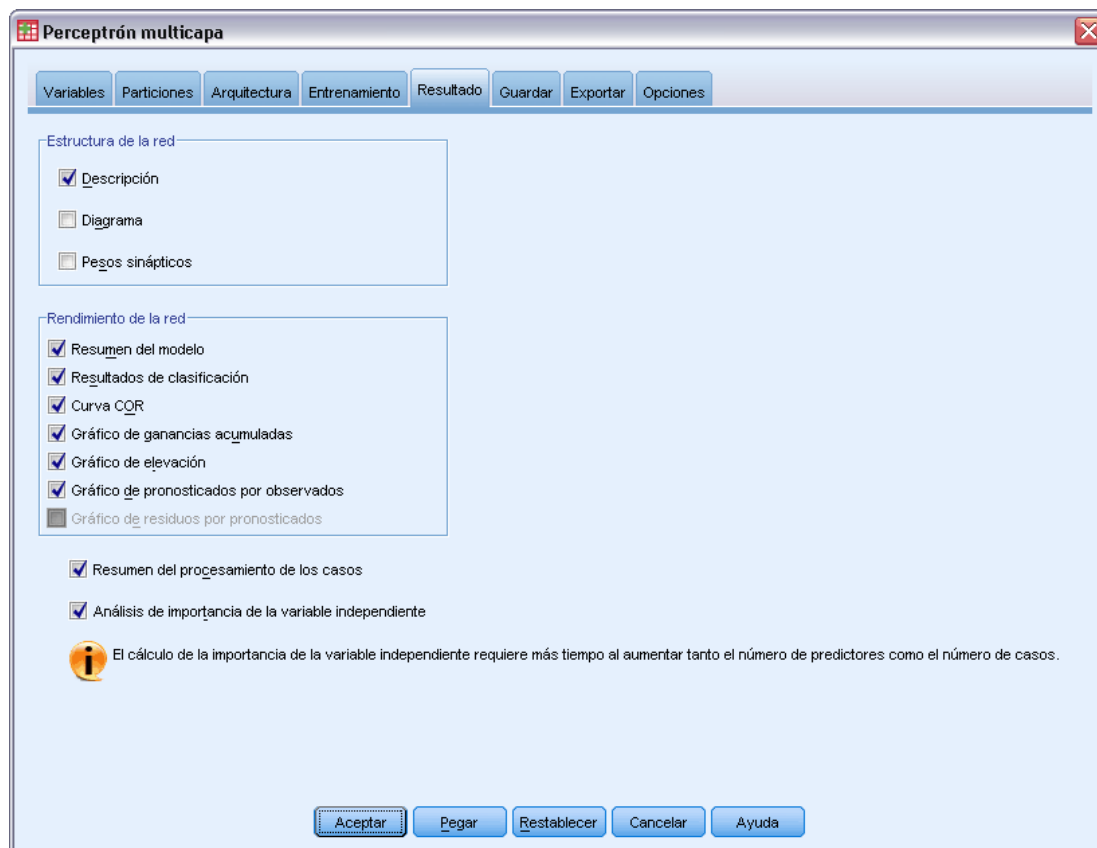
- ▶ Seleccione *Impagos anteriores [impago]* como la variable dependiente.
- ▶ Seleccione *Nivel educativo [educ]* como un factor.
- ▶ Seleccione *Edad en años [edad]* y *Otras deudas en miles [deudaotro]* como covariables.
- ▶ Pulse en la pestaña *Particiones*.

Figura 4-4
Perceptrón multicapa: Pestaña Particiones



- ▶ Seleccione Utilizar variable de partición para asignar los casos.
- ▶ Seleccione *partición* como la variable de división.
- ▶ Pulse en la pestaña Resultados.

Figura 4-5
Perceptrón multicapa: Pestaña Resultados



- ▶ Anule la selección en Diagrama en el grupo de estructura de red.
- ▶ Seleccione Curva COR, Gráfico de ganancias acumuladas, Gráfico de elevación (índice) y Gráfico de pronosticados por observados en el grupo de rendimiento de red. El gráfico de residuos por pronosticados no está disponible porque la variable dependiente no es de escala.
- ▶ Seleccione Análisis de importancia de la variable independiente.
- ▶ Pulse en Aceptar.

Resumen de procesamiento de casos

Figura 4-6
Resumen de procesamiento de casos

Muestra		N	Porcentaje
Entrenamiento		499	71,3%
Reserva		201	28,7%
Válido		700	100,0%
Excluido		150	
Total		850	

El resumen del procesamiento de los casos muestra que se asignaron 499 casos a la muestra de entrenamiento y 201 a la muestra reservada. Los 150 casos excluidos del análisis son posibles clientes.

Información sobre la red

Figura 4-7
Información de red

Capa de entrada	Factores	1	Nivel de educación
	Covariables	1	Edad en años
		2	Otras deudas en miles
	Número de unidades ^a		7
Método de cambio de escala de las covariables		Tipificados	
Capas ocultas	Número de capas ocultas		1
	Número de unidades de la capa oculta 1 ^a		3
Capa de salida	Función de activación		Tangente hiperbólica
	Variables dependientes	1	Impagos anteriores
	Número de unidades		2
	Función de activación		Softmax
	Función de error		Entropía cruzada
a. Sin incluir la unidad de sesgo			

La tabla de información de red muestra información sobre la red neuronal y resulta útil para garantizar que las especificaciones son correctas. Aquí se ve en particular que:

- El número de unidades en la capa de entrada es el número de covariables más el número total de niveles de factor. Se crea una unidad independiente para cada categoría de *Nivel de educación* y ninguna de las categorías se considera como una unidad “redundante”, como es habitual en muchos procedimientos de creación de modelos.
- De igual manera, se crea una unidad de resultado independiente para cada categoría de *Previously defaulted*, para un total de dos unidades en la capa de resultados.
- La selección de arquitectura automática ha elegido cuatro unidades en la capa oculta.
- El resto de la información de red se toma por defecto para el procedimiento.

Resumen del modelo

Figura 4-8
Resumen del modelo

Entrenamiento	Error de entropía cruzada	251,7%
	Porcentaje de pronósticos incorrectos	25,2%
	Regla de parada utilizada	Criterio de cambio relativo en el error de entrenamiento (,0001) alcanzado%
	Tiempo de entrenamiento	0:00:00.751%
Reserva	Porcentaje de pronósticos incorrectos	23,1%
Variable dependiente: Impagos anteriores		

El resumen del modelo muestra información sobre los resultados de entrenar y aplicar la red final a la muestra reservada.

- El error de entropía cruzada se muestra porque la capa de resultados usa la función de activación softmax. Ésta es la función de error que la red intenta minimizar durante el entrenamiento.
- El porcentaje de pronósticos incorrectos se toma de la tabla de clasificación, y se discutirá más adelante en ese tema.
- El algoritmo de estimación se ha detenido ya que se ha alcanzado el número máximo de épocas. Lo ideal es que el entrenamiento se detenga, puesto que el error ha convergido. Esto plantea cuestiones sobre si se ha producido algún error durante el entrenamiento, y se debe tener en cuenta al realizar una inspección adicional de los resultados.

Classification

Figura 4-9
Classification

Muestra	Observado	Pronosticado		
		No	Sí	Porcentaje correcto
Entrenamiento	No	328	22	93,7%
	Sí	101	37	26,8%
	Porcentaje global	87,9%	12,1%	74,8%
Reserva	No	155	12	92,8%
	Sí	37	8	17,8%
	Porcentaje global	90,6%	9,4%	76,9%

Variable dependiente: Impagos anteriores

La tabla de clasificación muestra los resultados prácticos de la utilización de la red. Para cada caso, la respuesta pronosticada es *Sí* si la pseudoprobabilidad pronosticada de ese caso es mayor que 0,5. Para cada muestra:

- Las casillas de la diagonal de la clasificación conjunta de los casos son los pronósticos correctos.
- Las casillas fuera de la diagonal de la clasificación conjunta de los casos son los pronósticos incorrectos.

De los casos utilizados para crear el modelo, 74 de las 124 personas que previamente han causado mora se han clasificado correctamente. 347 de las 375 personas que no han causado mora se han clasificado correctamente. En general, el 84,4% de los casos de entrenamiento se clasificaron correctamente, lo que corresponde al 15,6% de incorrectos mostrado en la tabla de resumen del modelo. Un modelo mejor debe identificar correctamente un mayor porcentaje de los casos.

Las clasificaciones basadas en los casos utilizados para crear el modelo tienden a ser demasiado “optimistas” en el sentido de que aumentan en exceso la tasa de clasificación. La muestra reservada ayuda a validar el modelo. En este caso, el modelo clasificó correctamente el 74,6% de casos. Esto sugiere que, en general, el modelo es correcto tres de cada cuatro veces.

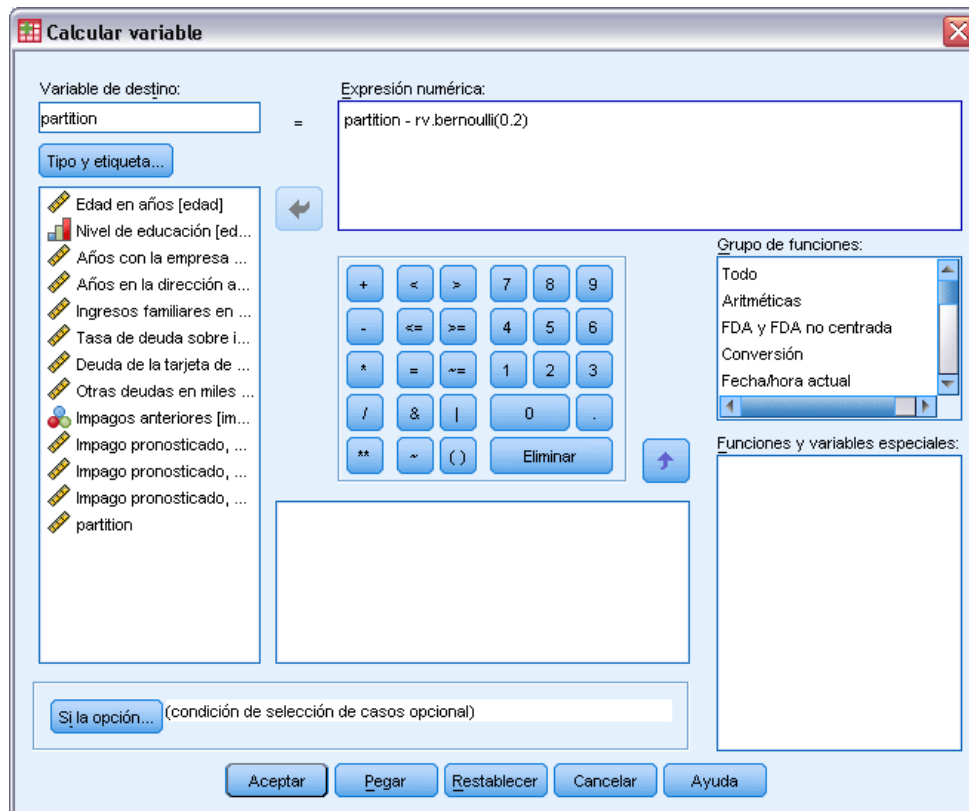
Corrección del sobreentrenamiento

Volviendo sobre el análisis de regresión logística realizado anteriormente, el encargado de préstamos recuerda que las muestras de entrenamiento y reservadas pronosticaron correctamente un porcentaje similar de casos, sobre el 80%. Por el contrario, la red neuronal obtuvo un porcentaje superior de casos correctos en la muestra de entrenamiento y peores resultados en la muestra reservada al pronosticar los clientes que realmente habían causado mora (el 45,8% correctos en la muestra reservada respecto al 59,7% en la muestra de entrenamiento). Combinado con la regla de parada indicada en la tabla de resumen del modelo, esto hace sospechar que la red puede estar **sobreentrenada**; es decir, está detectando falsos patrones que aparecen en los datos de entrenamiento por variación aleatoria.

Afortunadamente, la solución es relativamente sencilla: especificar una muestra de prueba que ayude a mantener la red bien ajustada. Creamos la variable de partición para que volviera a crear exactamente las muestras reservadas y de entrenamiento que se utilizaron en el análisis de regresión logística; sin embargo, la regresión logística no incluye el concepto de una muestra de “entrenamiento”. Tomaremos una porción de la muestra de entrenamiento y la volveremos a asignar a una muestra de prueba.

Creación de la muestra de prueba

Figura 4-10
Cuadro de diálogo *Calcular variable*

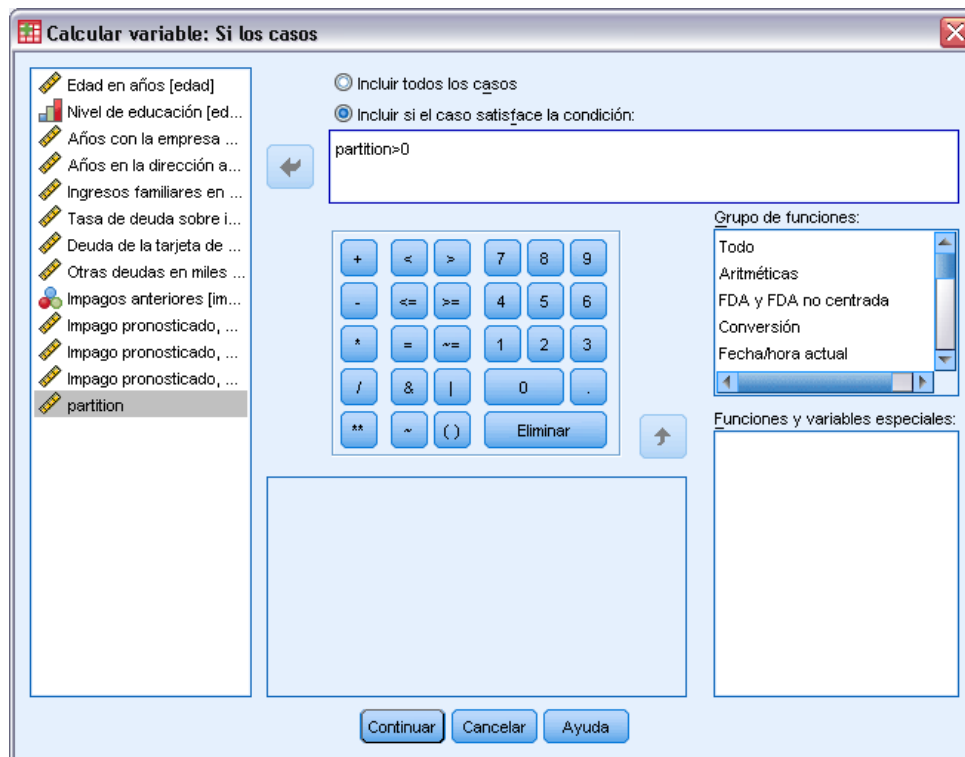


- Vuelva a abrir el cuadro de diálogo *Calcular variable*.

- ▶ Escriba `partition - rv.bernoulli(0.2)` en el cuadro de texto Expresión numérica.
- ▶ Pulse en Si.

Figura 4-11

Calcular variable: Cuadro de diálogo Si los casos



- ▶ Seleccione Incluir si el caso satisface la condición.
- ▶ Escriba `partición>0` en el cuadro de texto.
- ▶ Pulse en Continuar.
- ▶ Pulse Aceptar en el cuadro de diálogo Calcular variable.

Esto restablece los valores de *partición* que eran mayores que 0 de forma que aproximadamente el 20% tome el valor de 0 y el 80% permanezca con el valor 1. En general, aproximadamente el $100*(0,7*0,8)=56\%$ de los clientes a quienes se habían concedido créditos estarán en la muestra de entrenamiento y el 14% en la muestra de prueba. Los clientes que se habían asignado originalmente a la muestra reservada permanecen allí.

Ejecución del análisis

- ▶ Recupere el cuadro de diálogo Perceptrón multicapa y pulse en la pestaña Guardar.
- ▶ Seleccione Guardar categoría o valor pronosticado para cada variable dependiente.
- ▶ Pulse en Aceptar.

Resumen de procesamiento de casos

Figura 4-12

Resumen del procesamiento de casos para el modelo con muestra de prueba

		N	Porcentaje
Muestra	Entrenamiento	395	56,4%
	Prueba	93	13,3%
	Reserva	212	30,3%
	Válido	700	100,0%
	Excluido	150	
Total		850	

De los 499 casos que se habían asignado originalmente a la muestra de prueba, 101 se han vuelto a asignar a la muestra de prueba.

Información sobre la red

Figura 4-13

Información de red

Capa de entrada	Factores	1	Nivel de educación Edad en años Otras deudas en miles 7 Tipificados
	Covariables	1	
		2	
	Número de unidades ^a		
	Método de cambio de escala de las covariables		
Capas ocultas	Número de capas ocultas		1
	Número de unidades de la capa oculta 1 ^a		3
Capa de salida	Función de activación		Tangente hiperbólica
	Variables dependientes	1	Impagos anteriores
	Número de unidades		2
	Función de activación		Softmax
	Función de error		Entropía cruzada
a. Sin incluir la unidad de sesgo			

El único cambio de la tabla de información de red es que la selección de arquitectura automática ha elegido siete unidades de la capa oculta.

Resumen del modelo

Figura 4-14
Resumen del modelo

Entrenamiento	Error de entropía cruzada	211,1%
	Porcentaje de pronósticos incorrectos	28,4%
	Regla de parada utilizada	1 pasos consecutivos sin disminución del error% ^a
	Tiempo de entrenamiento	0:00:00.961%
Prueba	Error de entropía cruzada	51,4%
	Porcentaje de pronósticos incorrectos	26,9%
Reserva	Porcentaje de pronósticos incorrectos	21,7%
Variable dependiente: Impagos anteriores		
a. Los cálculos del error se basan en la muestra de prueba.		

El resumen del modelo muestra un par de signos positivos:

- El porcentaje de pronósticos incorrectos es prácticamente igual en las muestras reservadas, de entrenamiento y de prueba.
- El algoritmo de estimación se detuvo debido a que el error no disminuyó después de realizar un paso en el algoritmo.

Esto sugiere que, en efecto, el modelo original pudiera estar sobreentrenado y el problema se resolviera al agregar una muestra de prueba. Por supuesto, el tamaño de las muestras es relativamente pequeño y puede que no debamos sacar demasiadas conclusiones a partir de unos cuantos puntos de porcentaje.

Classification

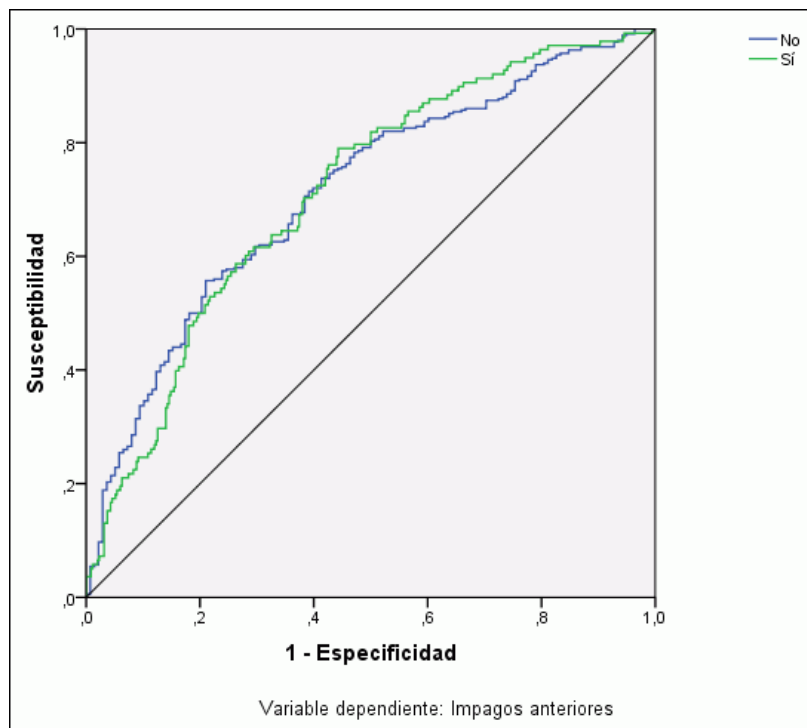
Figura 4-15
Classification

Muestra	Observado	Pronosticado		
		No	Sí	Porcentaje correcto
Entrenamiento	No	276	6	97,9%
	Sí	106	7	6,2%
	Porcentaje global	96,7%	3,3%	71,6%
Prueba	No	66	2	97,1%
	Sí	23	2	8,0%
	Porcentaje global	95,7%	4,3%	73,1%
Reserva	No	164	3	98,2%
	Sí	43	2	4,4%
	Porcentaje global	97,6%	2,4%	78,3%
Variable dependiente: Impagos anteriores				

La tabla de clasificación muestra que, usando 0,5 con el corte de pseudoprobabilidad para la clasificación, la red funciona considerablemente mejor pronosticando las personas que no han causado mora que las que han causado mora. Desafortunadamente, el valor de corte único proporciona una vista muy limitada de la capacidad predictora de la red, por lo que no es necesariamente muy útil para comparar redes que compiten. En su lugar, observe la curva ROC.

Curva COR

Figura 4-16
curva COR



La curva COR proporciona una representación visual de la **susceptibilidad** y **especificidad** para todos los cortes posibles en un único gráfico, que es mucho más limpio y potente que una serie de tablas. El gráfico que se muestra aquí presenta dos curvas, una para la categoría *No* y una para la categoría *Sí*. Debido a que sólo hay dos categorías, las curvas son simétricas con respecto a una línea de 45 grados (no se muestra) que va de la esquina superior izquierda del gráfico a la esquina inferior derecha.

Tenga en cuenta que este gráfico se basa en las muestras de prueba y entrenamiento combinadas. Para generar un gráfico COR para la muestra reservada, divida el archivo de la variable de partición y ejecute el procedimiento Curva COR en las pseudoprobabilidades pronosticadas guardadas.

Figura 4-17
Área bajo la curva

		Área
Impagos anteriores	No	,712
	Sí	,712

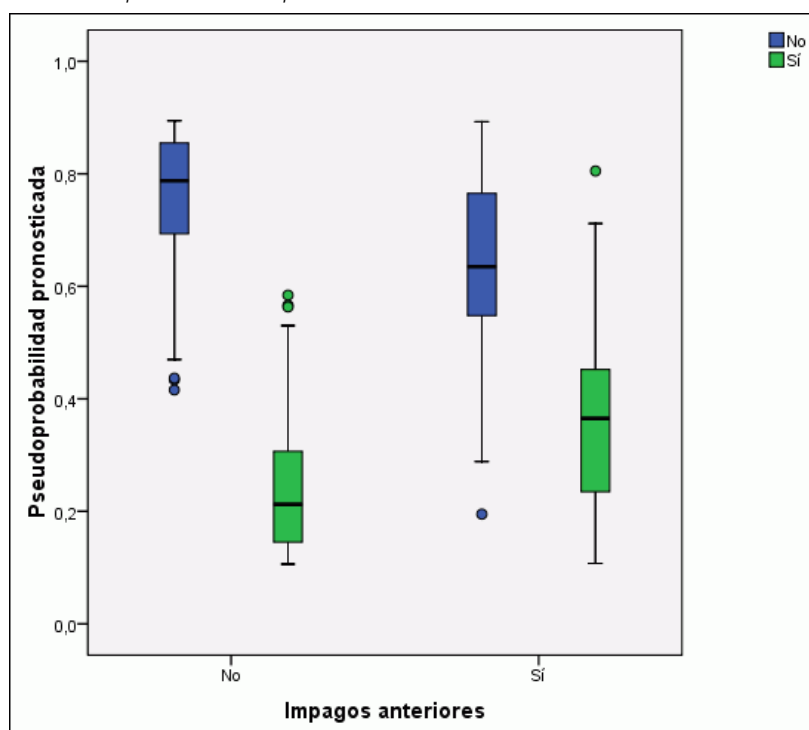
El área bajo la curva es un resumen numérico de la curva COR, y los valores de la tabla representan, para cada categoría, la probabilidad de que la pseudoprobabilidad pronosticada de estar en esa categoría sea mayor para un caso elegido aleatoriamente en esa categoría que para un caso elegido aleatoriamente que no pertenezca a esa categoría. Por ejemplo, en el caso de un

individuo que cause mora seleccionado aleatoriamente y otro que no cause mora seleccionado aleatoriamente, existe una probabilidad del 0,853 de que la pseudoprobabilidad pronosticada por el modelo de impago sea mayor para el caso que causa mora que para el que no la causa.

Mientras que el área bajo la curva es un resumen de una estadística de la precisión de la red útil, debe poder elegir un criterio específico por el que clasificar a los clientes. El gráfico de pronosticados por observados proporciona un punto de partida visual sobre este proceso.

Gráfico de pronosticados por observados

Figura 4-18
Gráfico de pronosticados por observados



En el caso de las variables dependientes categóricas, el gráfico de pronosticados por observados muestra diagramas de caja agrupados de pseudoprobabilidades pronosticadas para las muestras de entrenamiento y de prueba combinadas. El eje X corresponde a las categorías de respuesta observadas, y la leyenda corresponde a categorías pronosticadas.

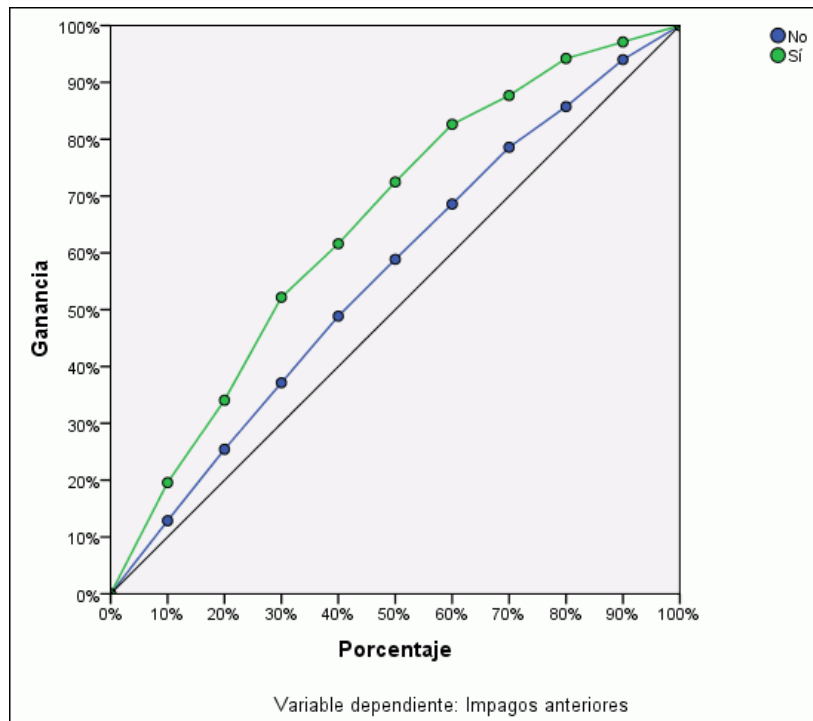
- El diagrama de caja situado más a la izquierda muestra, para los casos que tienen la categoría observada No, la pseudoprobabilidad pronosticada de la categoría No. La parte del diagrama de caja situada por encima de la marca de 0,5 del eje Y representa los pronósticos correctos que aparecen en la tabla de clasificación. La parte situada por debajo de la marca de 0,5 representa los pronósticos incorrectos. Recuerde de la tabla de clasificación que la red es muy eficaz al pronosticar casos con la categoría No mediante el corte de 0,5, por lo que sólo se clasifica incorrectamente una parte del bigote inferior y algunos casos atípicos.

- El siguiente diagrama de caja de la derecha muestra, para los casos que tienen la categoría observada No, la pseudoprobabilidad pronosticada de la categoría Sí. Como sólo hay dos categorías en la variable de destino, los primeros dos diagramas de caja son simétricos con respecto a la línea horizontal de 0,5.
- El tercer diagrama de caja muestra, para los casos que tienen la categoría observada Sí, la pseudoprobabilidad pronosticada de la categoría No. Éste y el último diagrama de caja son simétricos con respecto a la línea horizontal de 0,5.
- El último diagrama de caja muestra, para los casos que tienen la categoría observada Sí, la pseudoprobabilidad pronosticada de la categoría Sí. La parte del diagrama de caja situada por encima de la marca de 0,5 del eje *Y* representa los pronósticos correctos que aparecen en la tabla de clasificación. La parte situada por debajo de la marca de 0,5 representa los pronósticos incorrectos. En la tabla de clasificación, recuerde que la red pronostica algo más de la mitad de los casos con la categoría Sí mediante el corte de 0,5, por lo que buena parte del diagrama se clasifica incorrectamente.

Al observar el diagrama, parece que, al bajar el corte para clasificar un caso como Sí desde 0,5 a aproximadamente 0,3 (es prácticamente el valor donde se encuentran la parte superior del segundo diagrama y la parte inferior del cuarto diagrama) puede aumentar la posibilidad de detectar correctamente personas que pueden causar mora sin perder muchos clientes potenciales que no causarán mora. Es decir, entre 0,5 y 0,3 en el segundo diagrama, se vuelven a clasificar incorrectamente relativamente pocos clientes que no causarán mora en el bigote como clientes que causarán mora pronosticados, mientras que en el cuarto diagrama, este desplazamiento vuelve a clasificar correctamente muchos clientes que causarán mora dentro del diagrama como clientes que causarán mora pronosticados.

Gráficos de elevación y ganancias acumuladas

Figura 4-19
Gráfico de ganancias acumuladas



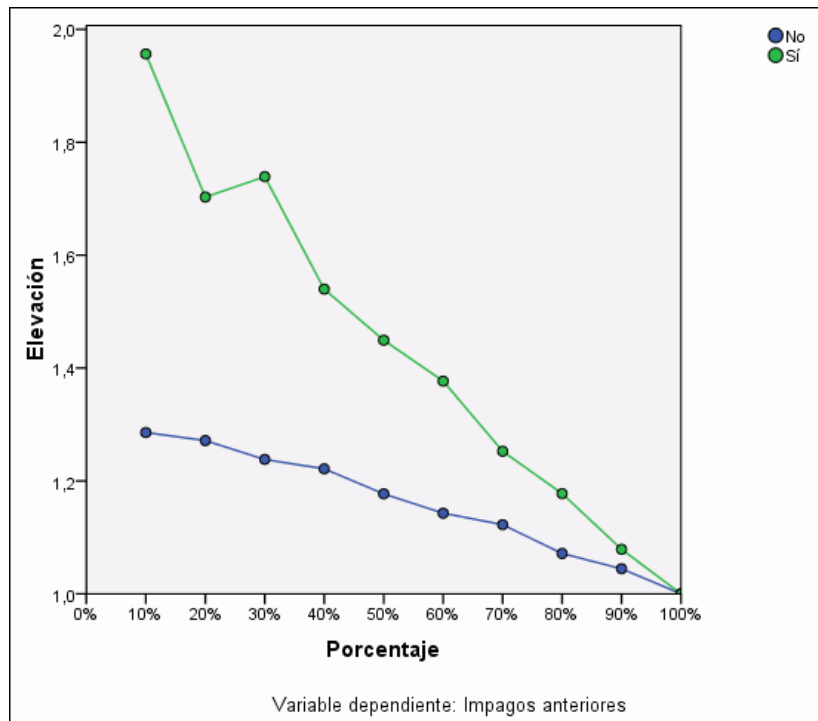
El gráfico de ganancias acumuladas muestra el porcentaje del número total de casos de una categoría dada “ganada” al dirigirse a un porcentaje del número total de casos. Por ejemplo, el primer punto de la curva de la categoría *Sí* está en (10%, 30%), lo que significa que si puntúa un conjunto de datos con la red y clasifica todos los casos por la pseudoprobabilidad pronosticada de *Sí*, esperaría que el 10% superior contuviera aproximadamente el 30% de todos los casos que realmente toman la categoría *Sí* (clientes que causarán mora). Del mismo modo, el 20% superior contendría aproximadamente el 50% de los clientes que causan mora, el 30% superior de los casos, el 70% de los clientes que causan mora, y así sucesivamente. Si selecciona el 100% del conjunto de datos puntuados, obtendrá todos los clientes que causan mora del conjunto de datos.

La línea diagonal es la curva de “línea base”, si selecciona el 10% de los casos del conjunto de datos puntuados de forma aleatoria, esperaría “ganar” aproximadamente el 10% de todos los casos que realmente toman la categoría *Sí*. Cuanto más arriba está la curva de línea base, mayor es la ganancia. Puede usar el gráfico de ganancias acumuladas para seleccionar un corte de clasificación al seleccionar un porcentaje que corresponde a una ganancia deseada y, a continuación, asignar ese porcentaje al valor de corte adecuado.

La definición de ganancia “deseada” depende del coste de los errores de Tipo I y Tipo II. Es decir, ¿cuál es el coste de clasificar una persona susceptible de causar mora como una persona que no va a causar mora (Tipo I)? ¿Cuál es el coste de clasificar una persona que no va a causar mora como una persona susceptible de causar mora (Tipo II)? Si las deudas incobrables son la preocupación principal, deberá disminuir el error de Tipo I. En el gráfico de ganancias acumuladas, esto puede corresponder a rechazar créditos a solicitantes del 40% superior de la

pseudoprobabilidad pronosticada de *Sí*, que captura cerca del 90% de los posibles clientes que causan mora pero elimina cerca de la mitad de la combinación de solicitantes. Si la prioridad es aumentar la base de clientes, entonces será deseable reducir el error de Tipo II. En el gráfico, esto puede corresponder a rechazar el 10% superior, que captura el 30% de los clientes que causan mora y deja la mayor parte de la combinación de solicitantes intacta. Normalmente, ambas son cuestiones importantes, así que se deberá elegir una regla de decisión para clasificar los clientes que ofrezcan la mejor combinación de susceptibilidad y especificidad.

Figura 4-20
Gráfico de elevación



El gráfico de elevación se deriva del gráfico de ganancias acumuladas. Los valores del eje Y corresponden a la relación de la ganancia acumulada de cada curva con respecto a la línea base. Así, la elevación de 10% para la categoría Sí es $30\%/10\% = 3,0$. Proporciona otra forma de ver la información en el gráfico de ganancias acumuladas.

Nota: los gráficos de elevación y de ganancias acumuladas se basan en las muestras de prueba y entrenamiento combinadas.

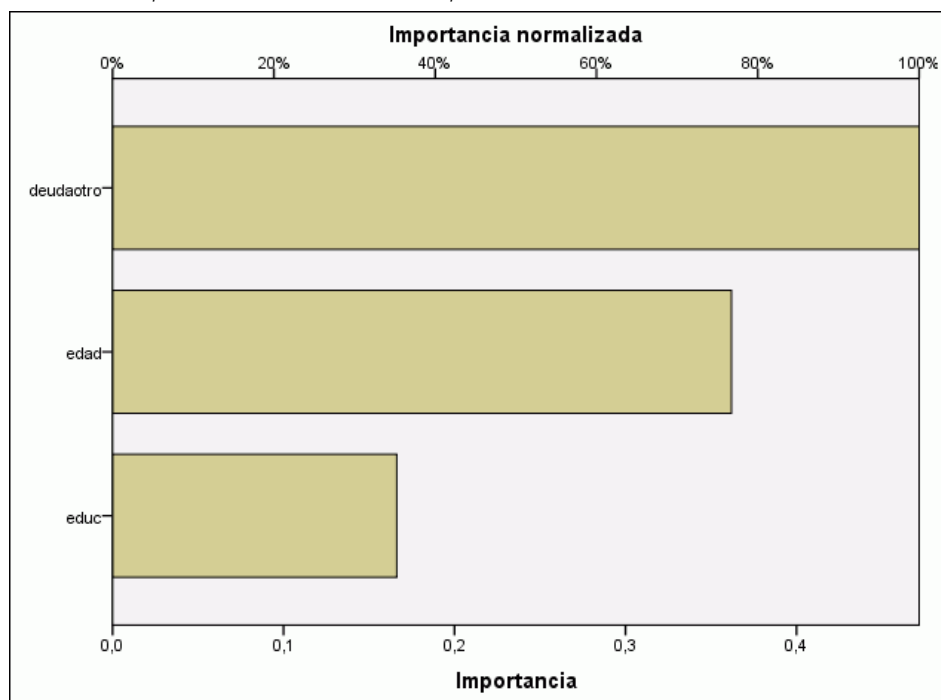
Importancia de las variables independientes

Figura 4-21
Importancia de la variable independiente

	Importancia	Importancia normalizada
Nivel de educación	,166	35,2%
Edad en años	,362	76,7%
Otras deudas en miles	,472	100,0%

La importancia de una variable independiente es una medida que indica cuánto cambia el valor pronosticado por el modelo de la red para diferentes valores de la variable independiente. La importancia normalizada es el resultado de los valores de importancia divididos por los valores de importancia mayores expresados como porcentajes.

Figura 4-22
Gráfico de importancia de la variable independiente



El gráfico de importancia es un gráfico de barras de los valores de la tabla de importancia, clasificado en valor de importancia descendente. Parece que las variables relacionadas con la estabilidad de un cliente (*empleo, dirección*) y la deuda (*creddebt, debtinc*) tienen el efecto mayor sobre cómo clasifica la red a los clientes; lo que no se puede saber es la “dirección” de las relaciones entre estas variables y la probabilidad pronosticada de causar mora. Se podría decir que una cantidad mayor de deudas indica una mayor probabilidad de causar mora, pero está claro que se necesitaría usar un modelo con parámetros más fáciles de interpretar.

Resumen

Mediante el procedimiento Perceptrón multicapa, se ha construido una red para pronosticar la probabilidad de que un cliente dado cause mora en un crédito. Los resultados del modelo son comparables a los obtenidos mediante la regresión logística o el análisis discriminante, por lo que se puede estar bastante seguro de que los datos no contienen relaciones que no puedan capturar estos modelos, y por lo tanto, puede utilizarlos para seguir investigando la relación entre las variables dependientes e independientes.

Uso del perceptrón multicapa para estimar los costes de asistencia sanitaria y las duraciones de las estancias

Un sistema hospitalario está interesado en realizar un seguimiento de los costes y las duraciones de las estancias de los pacientes ingresados para tratamientos de infartos de miocardio (IM o “ataque al corazón”). La obtención de estimaciones precisas de estas medidas permite a la administración gestionar adecuadamente las camas disponibles a medida que se trata a los pacientes.

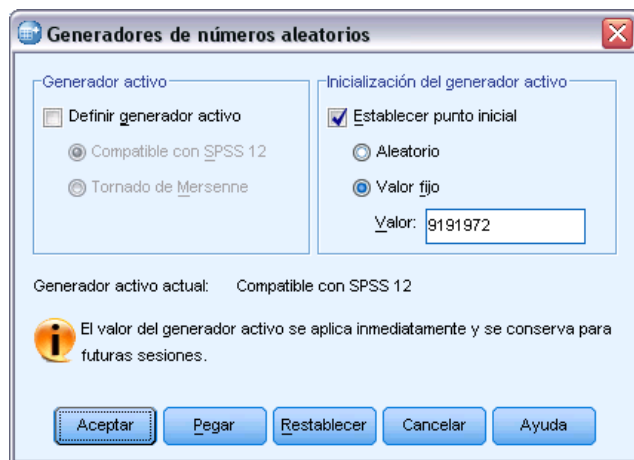
El archivo de datos *patient_los.sav* contiene los registros de tratamiento de una muestra de pacientes que han recibido tratamiento por infarto de miocardio. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 86.](#) Utilice el procedimiento Perceptrón multicapa para generar una red que pronostique los costes y las duraciones de las estancias.

Preparación de los datos para su análisis

Al establecer la semilla aleatoria, puede replicar el análisis de manera exacta.

- Para establecer la semilla aleatoria, elija en los menús:
Transformar > Generadores de números aleatorios...

Figura 4-23
Cuadro de diálogo Generadores de números aleatorios

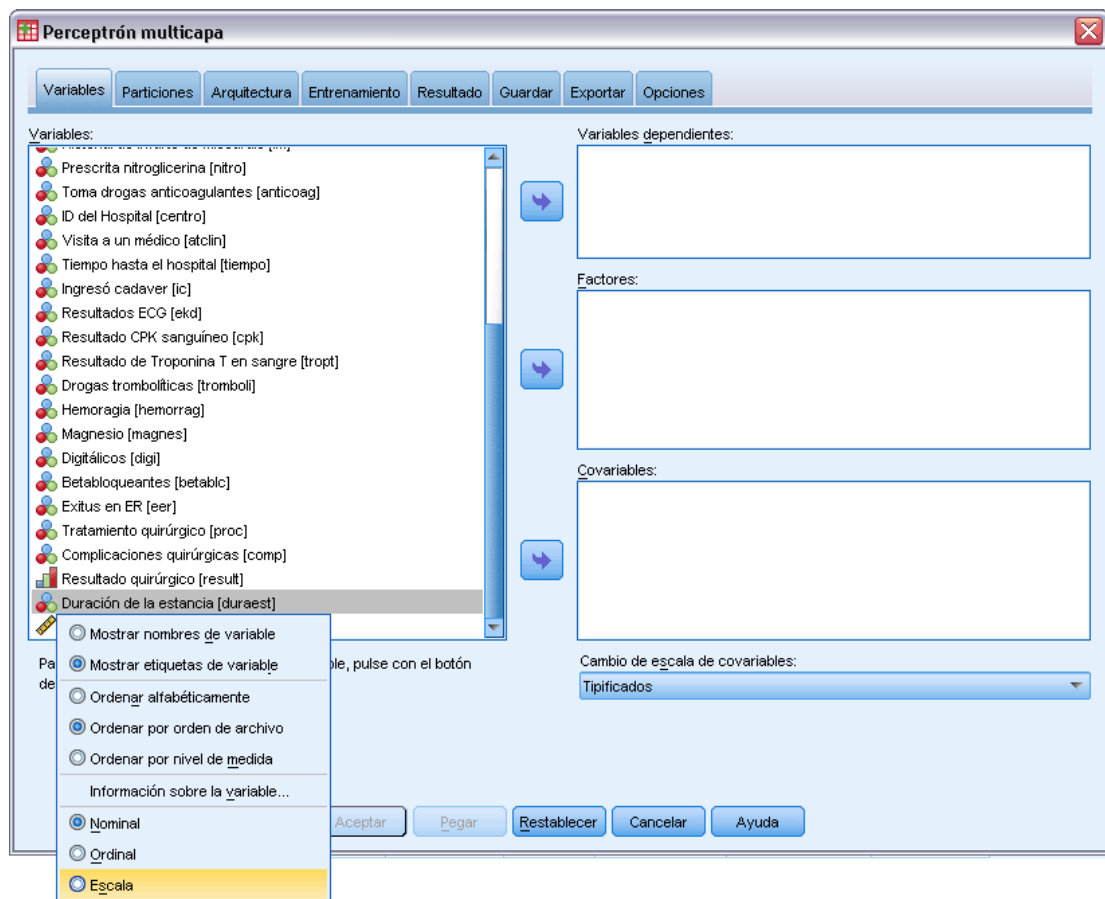


- ▶ Seleccione Establecer punto inicial.
- ▶ Seleccione Valor fijo y escriba 9191972 como el valor.
- ▶ Pulse en Aceptar.

Ejecución del análisis

- ▶ Para ejecutar un análisis de perceptrón multicapa, elija en los menús:
Analizar > Neural Networks > Perceptrón multicapa...

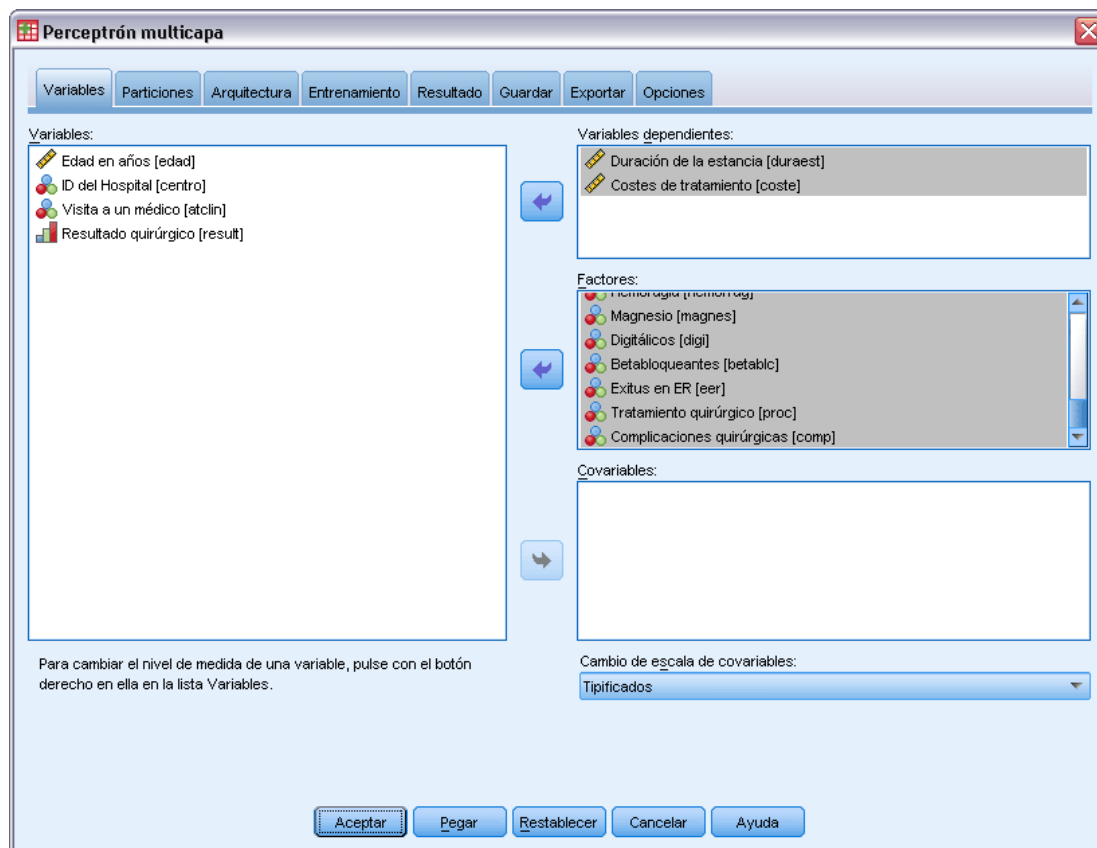
Figura 4-24
Perceptrón multicapa: pestaña Variables y menú contextual para la duración de la estancia



Duración de la estancia [duraest] tiene un nivel de medida ordinal, pero quiere que la red trate esta variable como si fuera de escala.

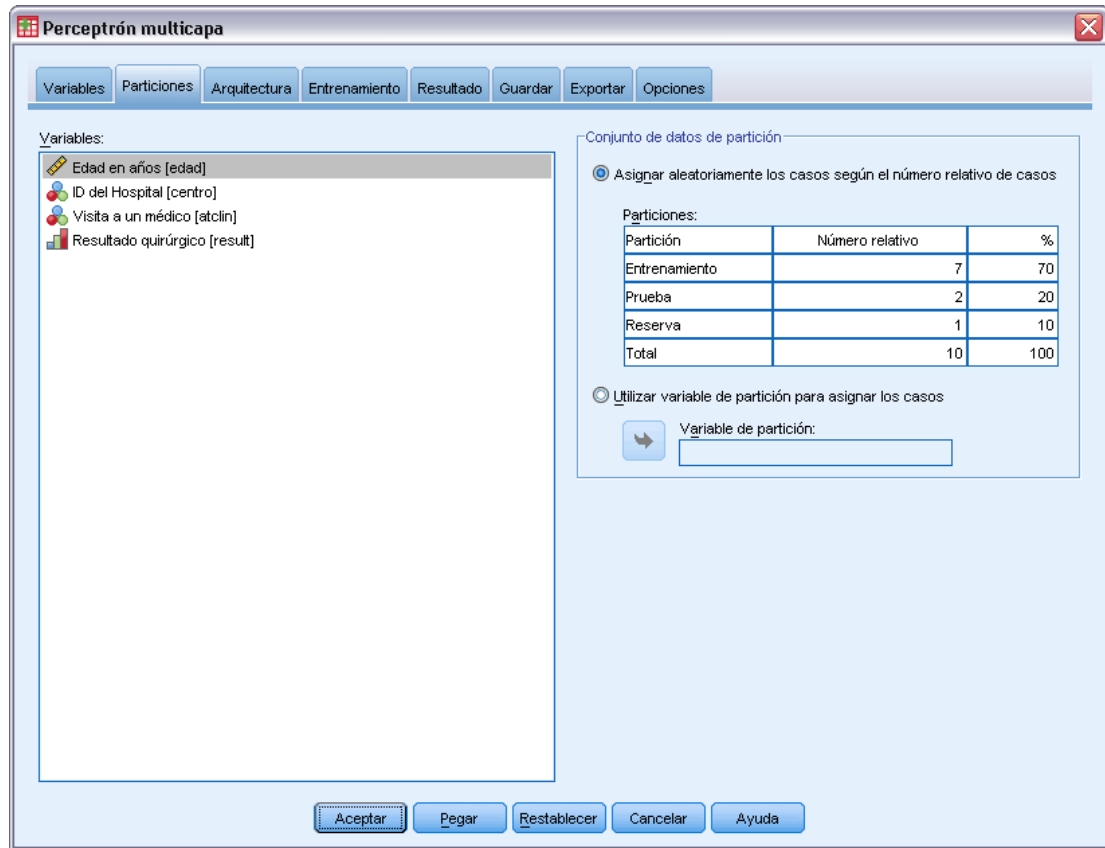
- Pulse con el botón derecho en *Duración de la estancia [duraest]* y seleccione Escala en el menú contextual.

Figura 4-25
Perceptrón multicapa: Pestaña Variables con variables dependientes y factores seleccionados



- ▶ Seleccione *Duración de la estancia [duraest]* y *Costes de tratamiento [coste]* como variables dependientes.
- ▶ Seleccione desde *Categoría de edad [edadcat]* hasta *Toma drogas anticoagulantes [anticoag]* y desde *Tiempo hasta el hospital [tiempo]* hasta *Complicaciones quirúrgicas [comp]* como factores. Para garantizar la replicación exacta de los resultados del modelo que se incluyen a continuación, asegúrese de respetar el orden de las variables en la lista de factores. Para ello, tal vez le resulte útil seleccionar cada conjunto de predictores y utilizar el botón que permite moverlos a la lista de factores, en vez de arrastrarlos y soltarlos. También puede cambiar el orden de las variables para evaluar la estabilidad de la solución.
- ▶ Pulse en la pestaña *Particiones*.

Figura 4-26
Perceptrón multicapa: Pestaña Particiones



- ▶ Escriba 2 como número relativo de casos que asignará a la muestra de prueba.
- ▶ Escriba 1 como número relativo de casos que asignará a la muestra reservada.
- ▶ Pulse en la pestaña Arquitectura.

Figura 4-27
Perceptrón multicapa: Pestaña Arquitectura

Perceptrón multicapa

Variables Particiones **Arquitectura** Entrenamiento Resultado Guardar Exportar Opciones

Selección automática de arquitectura
 Número mínimo de unidades de la capa oculta: 1
 Número máximo de unidades de la capa oculta: 50

Arquitectura personalizada

Capas ocultas

Número de capas ocultas

Prueba
 Dos

Función de activación

Tangente hiperbólica
 Sigmoide

Número de unidades

Calcular automáticamente
 Personalizado
 Capa oculta 1:
 Capa oculta 2:

Capa de salida

Función de activación

Identidad
 Softmax
 Tangente hiperbólica
 Sigmoide

Cambio de escala de variables dependientes de escala

Tipificados
 Normalizada
 Corrección: 0.02
 Corregida normalizada
 Corrección: 0.02
 Ninguna

La función de activación elegida para la capa de salida determina los métodos de cambio de escala que están disponibles.

Aceptar Pegar Restablecer Cancelar Ayuda

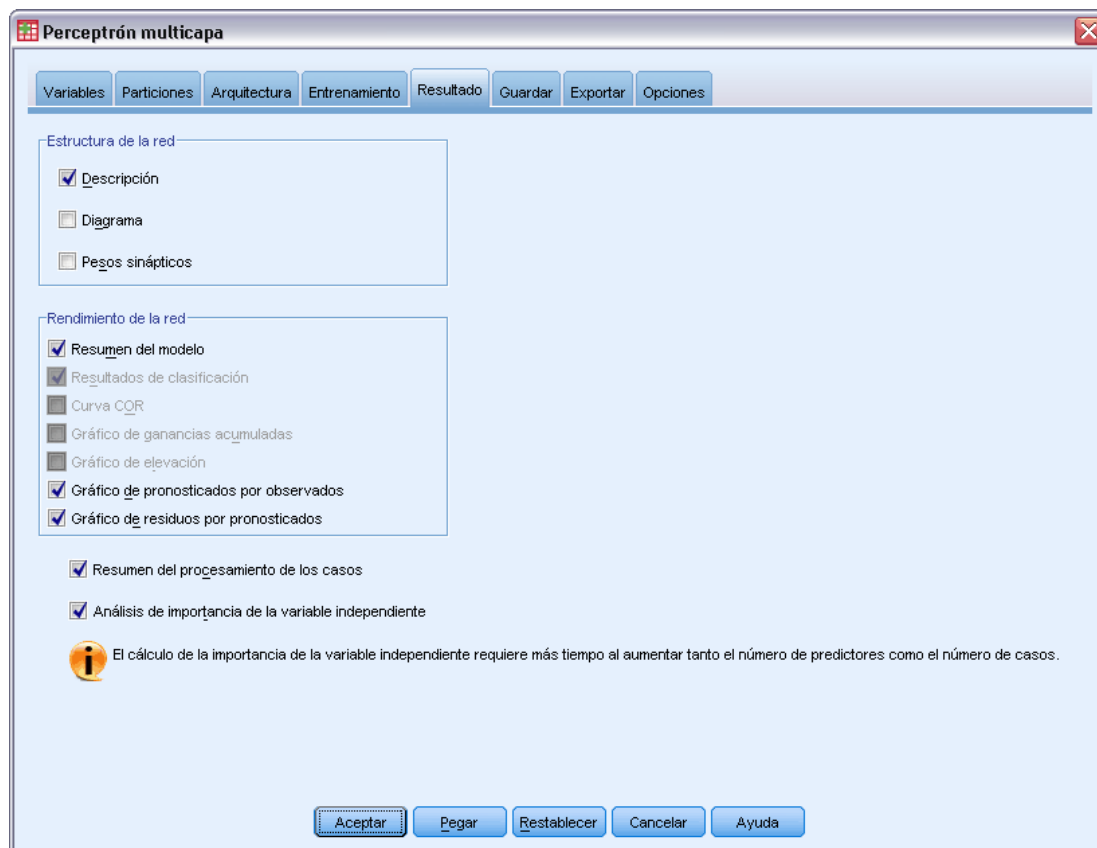
- ▶ Seleccione Arquitectura personalizada.
- ▶ Seleccione Dos como número de capas ocultas.
- ▶ Seleccione Tangente hiperbólica como función de activación de la capa de salida. Tenga en cuenta que se establecerá automáticamente el método de cambio de escala de las variables dependientes en Corregida normalizada.
- ▶ Pulse en la pestaña Entrenamiento.

Figura 4-28
Perceptrón multicapa: Pestaña Entrenamiento

Opción	Valor
Tasa de aprendizaje inicial	0.4
Límite inferior de la tasa de aprendizaje	0.001
Reducción de la tasa de aprendizaje, en Épocas	10
Impulso	0.9
Centro de intervalo	0
Desplazamiento de intervalo	± 0.5

- ▶ Seleccione En línea como tipo de entrenamiento. El entrenamiento en línea se supone que funciona bien en conjuntos de datos “más grandes” con predictores correlacionados. Tenga en cuenta que se establece automáticamente Pendiente de gradiente como algoritmo de optimización con las correspondientes opciones por defecto.
- ▶ Pulse en la pestaña Resultados.

Figura 4-29
Perceptrón multicapa: Pestaña Resultados



- ▶ Anule la selección de Diagrama; hay muchas entradas y el diagrama resultante sería confuso y difícil de interpretar.
- ▶ Seleccione Gráfico de pronosticados por observados y Gráfico de residuos por pronosticados en el grupo Rendimiento de la red. Los resultados de clasificación, la curva ROC, el gráfico de ganancias acumuladas y el gráfico de elevación no están disponibles ya que ninguna de las variables dependientes se trata como categórica (ordinal o nominal).
- ▶ Seleccione Análisis de importancia de la variable independiente.
- ▶ Pulse en la pestaña Opciones.

Figura 4-30
Pestaña Opciones

- ▶ Elija Incluir las variables con valores definidos como perdidos por el usuario. Los pacientes que no han sido sometidos a un procedimiento quirúrgico tienen valores definidos como perdidos por el usuario en la variable *Complicaciones quirúrgicas*. Esta selección garantiza que dichos pacientes se incluyen en el análisis.
- ▶ Pulse en Aceptar.

Advertencias

Figura 4-31
Advertencias

Las siguientes variables independientes son constantes en la muestra de entrenamiento y se excluirán del análisis: *ic*, *eer*.

La tabla de advertencias señala que las variables *ic* y *eer* son constantes en la muestra de entrenamiento. Los pacientes que ingresaron cadáveres o que fallecieron en la sala de emergencias tienen valores definidos como perdidos por el usuario en *Duración de la estancia*. Como estamos tratando *Duración de la estancia* como una variable de escala para este análisis y se excluyen

los casos con valores definidos como perdidos por el usuario en las variables de escala, sólo se incluirán aquellos pacientes que siguieran vivos tras pasar por la sala de emergencias.

Resumen de procesamiento de casos

Figura 4-32
Resumen de procesamiento de casos

	N	Porcentaje
Muestra Entrenamiento	5647	70,6%
Prueba	1570	19,6%
Reserva	781	9,8%
Válido	7998	100,0%
Excluido	2002	
Total	10000	

El resumen del procesamiento de los casos muestra que se asignaron 5647 casos a la muestra de entrenamiento, 1570 a la muestra de prueba y 781 a la muestra reservada. Los 2002 casos excluidos del análisis son pacientes que fallecieron en el trayecto hasta el hospital o en la sala de emergencias.

Información sobre la red

Figura 4-33
Información de red

Capa de entrada	Factors	1	Categoría de edad
		2	Género
		3	Historial de diabétes
		4	Presión sanguínea
		5	Fumador
		6	Colesterol
		7	Físicamente activo
		8	Obesidad
		9	Historial de angina
		10	Historial de infarto de miocardio
		11	Prescrita nitroglicerina
		12	Toma drogas anticoagulantes
		13	Tiempo hasta el hospital
		14	Resultados ECG
		15	Resultado CPK sanguíneo
		16	Resultado de Troponina T en sangre
		17	Drogas trombolíticas
		18	Hemorragia
		19	Magnesio
		20	Digitálicos
		21	Betabloqueantes
		22	Tratamiento quirúrgico
		23	Complicaciones quirúrgicas
Capas ocultas	Número de unidades ^a		63
	Número de capas ocultas		2
	Número de unidades de la capa oculta 1 ^a		12
	Number of Units in Hidden Layer 2 ^a		9
Capa de salida	Función de activación		Tangente hiperbólica
	Dependent Variables	1	Duración de la estancia
		2	Costes de tratamiento
	Número de unidades		2
	Rescaling Method for Scale Dependents		Adjusted Normalized
Función de activación		Tangente hiperbólica	
Función de error		Suma de cuadrados	

a. Sin incluir la unidad de sesgo

La tabla de información de red muestra información sobre la red neuronal y resulta útil para garantizar que las especificaciones son correctas. Aquí se ve en particular que:

- El número de unidades de la capa de entrada es el número total de niveles de factores (no hay ninguna covariable).
- Se han solicitado dos capas ocultas y el procedimiento ha elegido 12 unidades en la primera capa oculta y 9 en la segunda.

- Se crea una unidad de salida distinta para cada una de las variables dependientes de escala. Se cambia su escala según el método corregido normalizado, que requiere el uso de la función de activación de tangente hiperbólica para la capa de salida.
- Se informa de un error de suma de cuadrados ya que las variables dependientes son de escala.

Resumen del modelo

Figura 4-34
Resumen del modelo

Resumen del modelo			
Entrenamiento	Suma de errores cuadráticos		91,812
	Error relativo global promedio		,083
	Error relativo para dependientes de escala	Duración de la estancia	,131
		Costes de tratamiento	,033
	Regla de parada utilizada		1 pasos consecutivos sin disminución del error ^a
	Tiempo de entrenamiento		00:00:18.055
Prueba	Suma de errores cuadráticos		26,798
	Error relativo global promedio		,088
	Error relativo para dependientes de escala	Duración de la estancia	,141
		Costes de tratamiento	,033
Reserva	Error relativo global promedio		,099
	Error relativo para dependientes de escala	Duración de la estancia	,154
		Costes de tratamiento	,041

a. Los cálculos del error se basan en la muestra de prueba.

El resumen del modelo muestra información sobre los resultados de entrenar y aplicar la red final a la muestra reservada.

- Se muestra un error de suma de cuadrados, ya que la capa de salida tiene variables dependientes de escala. Ésta es la función de error que la red intenta minimizar durante el entrenamiento. Observe que las sumas de los cuadrados y todos los valores de error siguientes se calculan para los valores tras el cambio de escala de las variables dependientes.
- El error relativo de cada variable dependiente de escala es el cociente entre la suma de los errores cuadráticos de la variable dependiente y la suma de los errores cuadráticos del modelo “nulo”, en el que se utiliza el valor medio de la variable dependiente como valor pronosticado para todos los casos. El error parece ser mayor para los pronósticos de *Duración de la estancia* que para *Costes de tratamiento*.
- El error global promedio es el cociente entre la suma de los errores cuadráticos de todas las variables dependientes y la suma de los errores cuadráticos del modelo “nulo”, en el que se utilizan los valores medios de las variables dependientes como valores pronosticados para todos los casos. En este ejemplo, el error global promedio resulta estar próximo al promedio de los errores relativos, pero no siempre tiene que ser así.

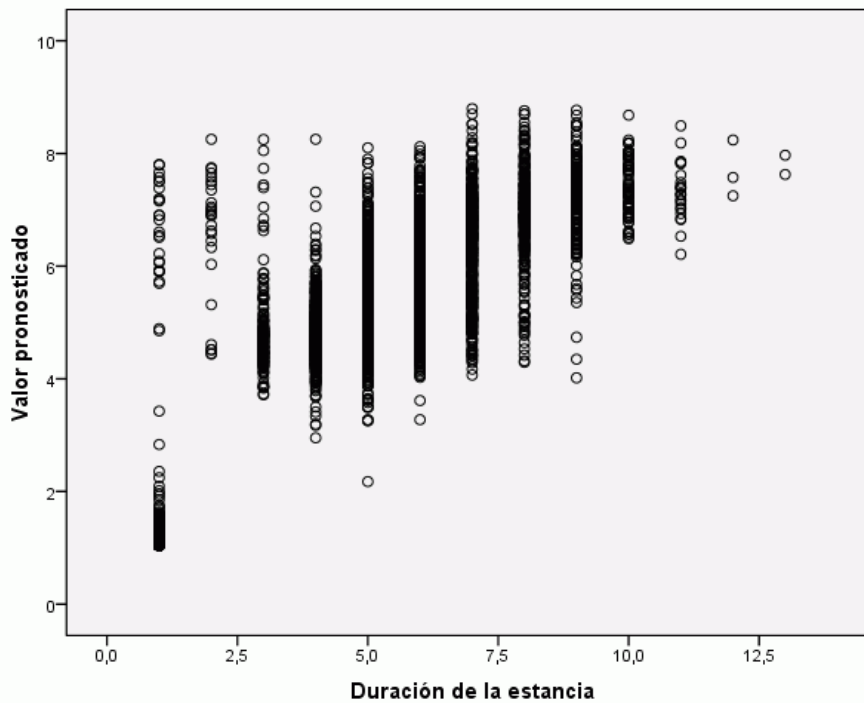
El error relativo global promedio y los errores relativos son bastante constantes en las muestras de entrenamiento, prueba y reserva, lo que le ofrece cierta seguridad de que el modelo no está sobreentrenado y el error en los casos futuros puntuados por la red será similar al error indicado en esta tabla.

- El algoritmo de estimación se detuvo debido a que el error no disminuyó después de realizar un paso en el algoritmo.

Gráficos de pronosticados por observados

Figura 4-35

Gráfico de pronosticados por observados para la duración de la estancia



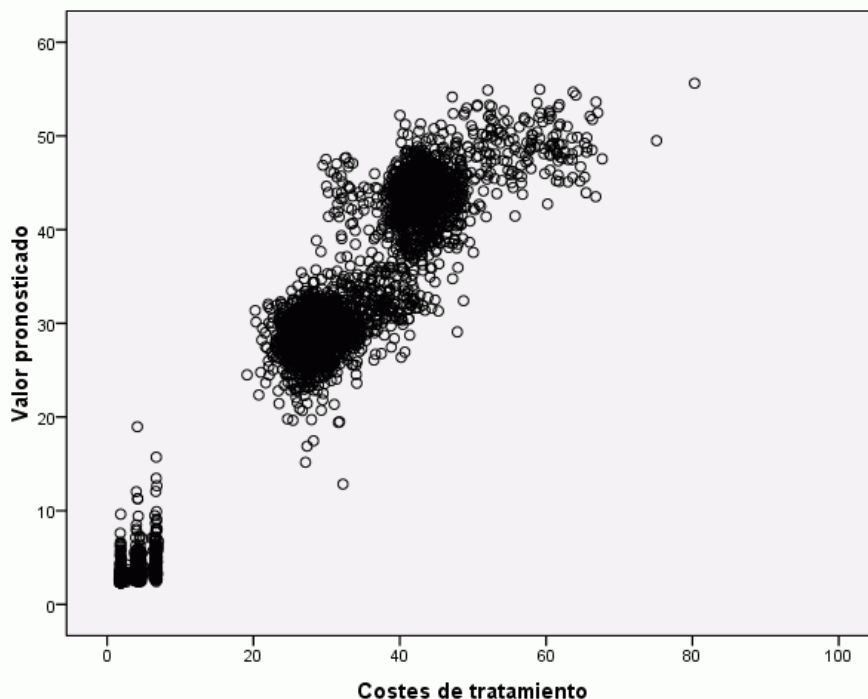
En el caso de las variables dependientes de escala, el gráfico de pronosticados por observados muestra un diagrama de dispersión con los valores pronosticados en el eje Y y los valores observados en el eje X para las muestras de entrenamiento y prueba combinadas. Lo ideal sería que los valores estuvieran situados a lo largo de una línea de 45 grados que partiese del origen. Los puntos de este gráfico forman líneas verticales en cada número de días de *Duración de la estancia* observado.

Al examinar el gráfico, vemos que la red pronostica razonablemente bien *Duración de la estancia*. La tendencia general del gráfico se aparta de la línea ideal de 45 grados en el sentido de que los pronósticos de las duraciones de las estancias observadas por debajo de cinco días tienden a estimar al alza la duración de la estancia, mientras que los pronósticos de las duraciones de las estancias observadas por encima de los seis días tienden a estimar a la baja la duración de la estancia.

El conglomerado de pacientes situados en la parte inferior izquierda del gráfico es probable que sean pacientes que no han sido sometidos a cirugía. Hay también un conglomerado de pacientes en la parte superior izquierda del gráfico, donde la duración de la estancia observada es de uno a tres días y cuyos valores pronosticados son mucho mayores. Es probable que estos casos correspondan a pacientes que fallecieron en el hospital tras ser sometidos a cirugía.

Figura 4-36

Gráfico de pronosticados por observados para los costes de tratamiento



La red también parece pronosticar razonablemente bien *Costes de tratamiento*. Parece haber tres conglomerados principales de pacientes:

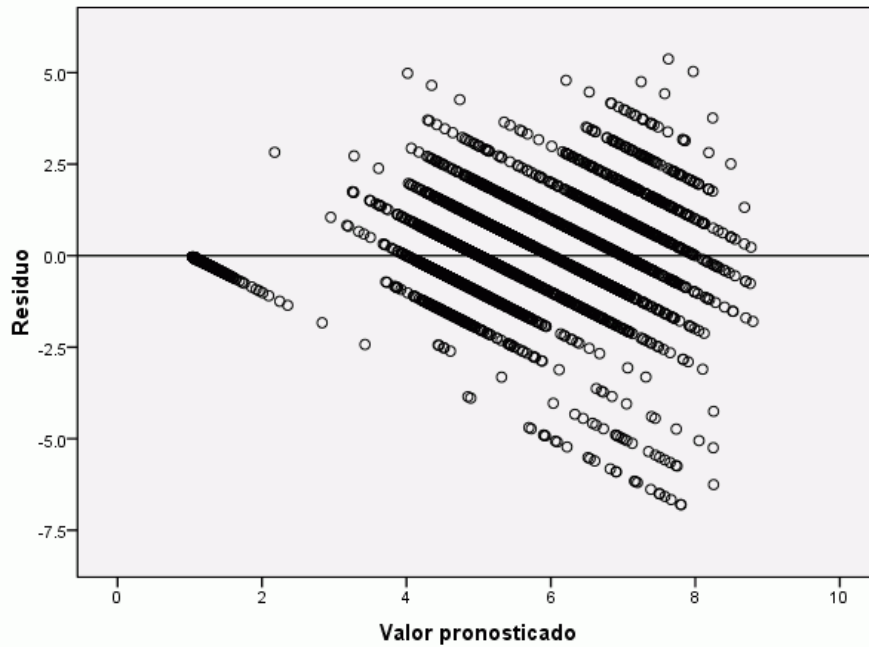
- En la parte inferior izquierda se encuentran principalmente pacientes que no han sido sometidos a cirugía. Sus costes son relativamente bajos y se diferencian por el tipo de *Drogas trombolíticas [tromboli]* administradas en la sala de emergencias.
- El siguiente conglomerado de pacientes tienen costes de tratamiento de aproximadamente 30.000 dólares. Son pacientes que han sido sometidos a una angioplastia coronaria transluminal percutánea (ACTP).
- El conglomerado final tienen costes de tratamiento que superan los 40.000 dólares. Se trata de pacientes que han sido sometidos a una intervención de bypass de la arteria coronaria (BACI). Esta intervención es algo más cara que la ACTP y los pacientes tienen que pasar un mayor tiempo de recuperación en el hospital, lo que a su vez aumenta aún más los costes.

También hay varios casos cuyos costes superan los 50.000 dólares que la red no pronostica demasiado bien. Se trata de pacientes que han sufrido complicaciones durante la cirugía, lo que puede aumentar el coste de la cirugía y la duración de la estancia.

Gráficos de residuos por pronosticados

Figura 4-37

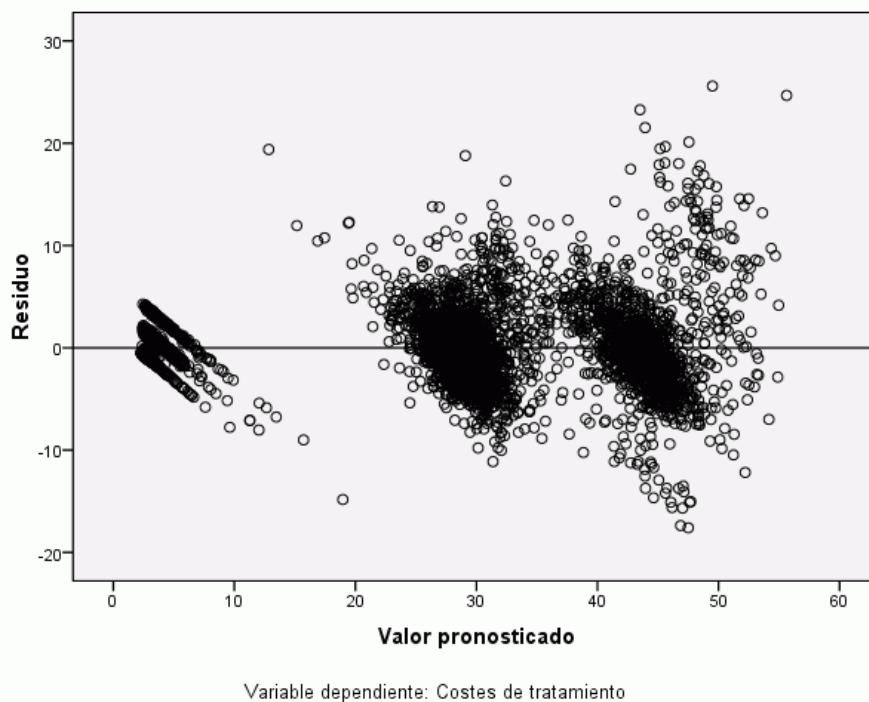
Gráfico de residuos por pronosticados para la duración de la estancia



Variable dependiente: Duración de la estancia

El gráfico de residuos por pronosticados muestra un diagrama de dispersión de los residuos (el valor observado menos el valor pronosticado) en el eje Y y los valores pronosticados en el eje X . Cada línea diagonal de este gráfico corresponde a una línea vertical del gráfico de pronosticados por observados y se puede ver con mayor claridad la progresión desde la predicción al alza hasta la predicción a la baja de la duración de la estancia a medida que aumenta la duración de la estancia observada.

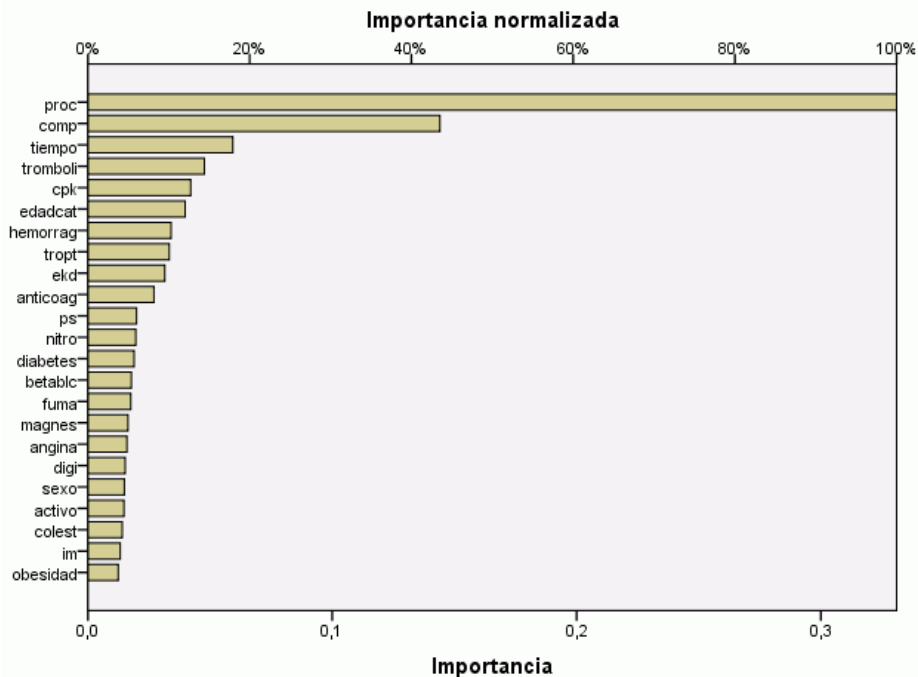
Figura 4-38
Gráfico de pronosticados por observados para los costes de tratamiento



Igualmente, para cada uno de los tres conglomerados de pacientes observados en el gráfico de pronosticados por observados para *Costes de tratamiento*, el gráfico de residuos por pronosticados muestra una progresión desde la predicción al alza hasta la predicción a la baja de los costes, a medida que aumentan los costes observados. Los pacientes con complicaciones durante BACI siguen siendo claramente visibles, pero también es más fácil ver los pacientes que han sufrido complicaciones durante ACTP, que aparecen como un subconglomerado situado ligeramente a la derecha y por encima del grupo principal de pacientes de ACTP alrededor de la marca de 30.000 dólares en el eje X.

Importancia de las variables independientes

Figura 4-39
Gráfico de importancia de la variable independiente



El gráfico de importancia muestra que los resultados están dominados por el procedimiento quirúrgico realizado, seguido de si han surgido complicaciones, seguidos a distancia por otros predictores. La importancia del procedimiento quirúrgico puede apreciarse claramente en los gráficos de *Costes de tratamiento*; algo menos en *Duración de la estancia*, aunque el efecto de las complicaciones sobre *Duración de la estancia* parece ser visible en los pacientes con las mayores duraciones de las estancias observadas.

Resumen

La red parece funcionar bien para pronosticar los valores de los pacientes “habituales”, pero no detecta los pacientes que mueren tras la intervención quirúrgica. Una manera posible de solucionar esto sería crear varias redes. Una red pronosticaría el resultado del paciente, tal vez sencillamente si el paciente ha sobrevivido o ha fallecido y las otras redes pronosticarían *Costes de tratamiento* y *Duración de la estancia* con la condición de que el paciente haya sobrevivido. A continuación, podría combinar los resultados de las redes con el objetivo de obtener mejores pronósticos. Podría utilizar un método para resolver el problema del pronóstico a la baja de los costes y las duraciones de las estancias de los pacientes que han sufrido complicaciones durante la intervención quirúrgica.

Lecturas recomendadas

Consulte los siguientes textos para obtener más información sobre redes neuronales y perceptrones multicapa:

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*, 3rd ed. Oxford: Oxford University Press.

Fine, T. L. 1999. *Feedforward Neural Network Methodology*, 3rd ed. Nueva York: Springer-Verlag.

Haykin, S. 1998. *Neural Networks: A Comprehensive Foundation*, 2nd ed. Nueva York: Macmillan College Publishing.

Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Función de base radial

El procedimiento Función de base radial (RBF) genera un modelo predictivo para una o más variables dependientes (de destino) basado en los valores de las variables predictoras.

Uso de la función de base radial para la clasificación de clientes de telecomunicaciones

Un proveedor de telecomunicaciones ha segmentado su base de clientes por patrones de uso de servicio, categorizando los clientes en cuatro grupos. Si los datos demográficos se pueden utilizar para predecir la pertenencia a un grupo, se pueden personalizar las ofertas para cada uno de los posibles clientes.

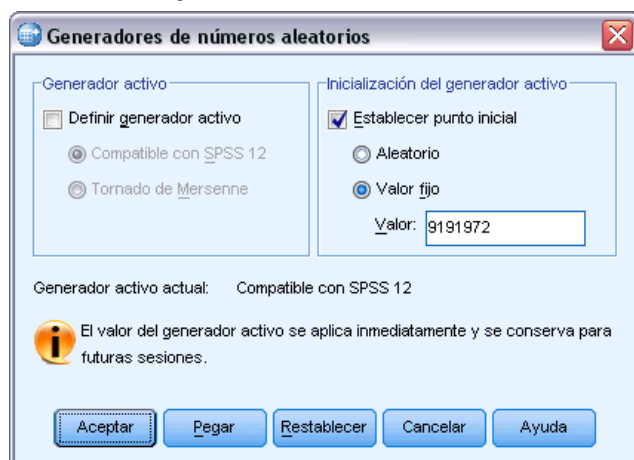
Supongamos que la información sobre los clientes actuales se encuentra en *telco.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 86.](#) Uso de la función de base radial para la clasificación de clientes de telecomunicaciones

Preparación de los datos para su análisis

Al establecer la semilla aleatoria, puede replicar el análisis de manera exacta.

- ▶ Para establecer la semilla aleatoria, elija en los menús:
Transformar > Generadores de números aleatorios...

Figura 5-1
Cuadro de diálogo Generadores de números aleatorios

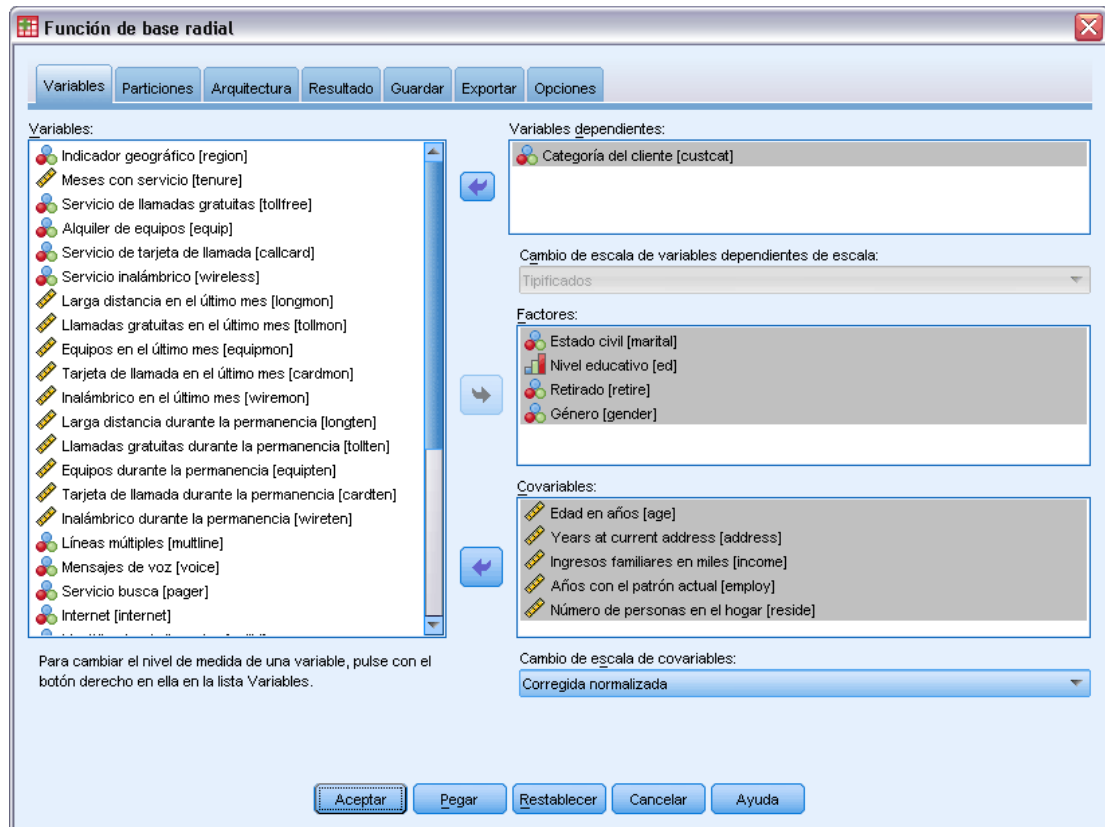


- ▶ Seleccione Establecer punto inicial.
- ▶ Seleccione Valor fijo y escriba 9191972 como el valor.
- ▶ Pulse en Aceptar.

Ejecución del análisis

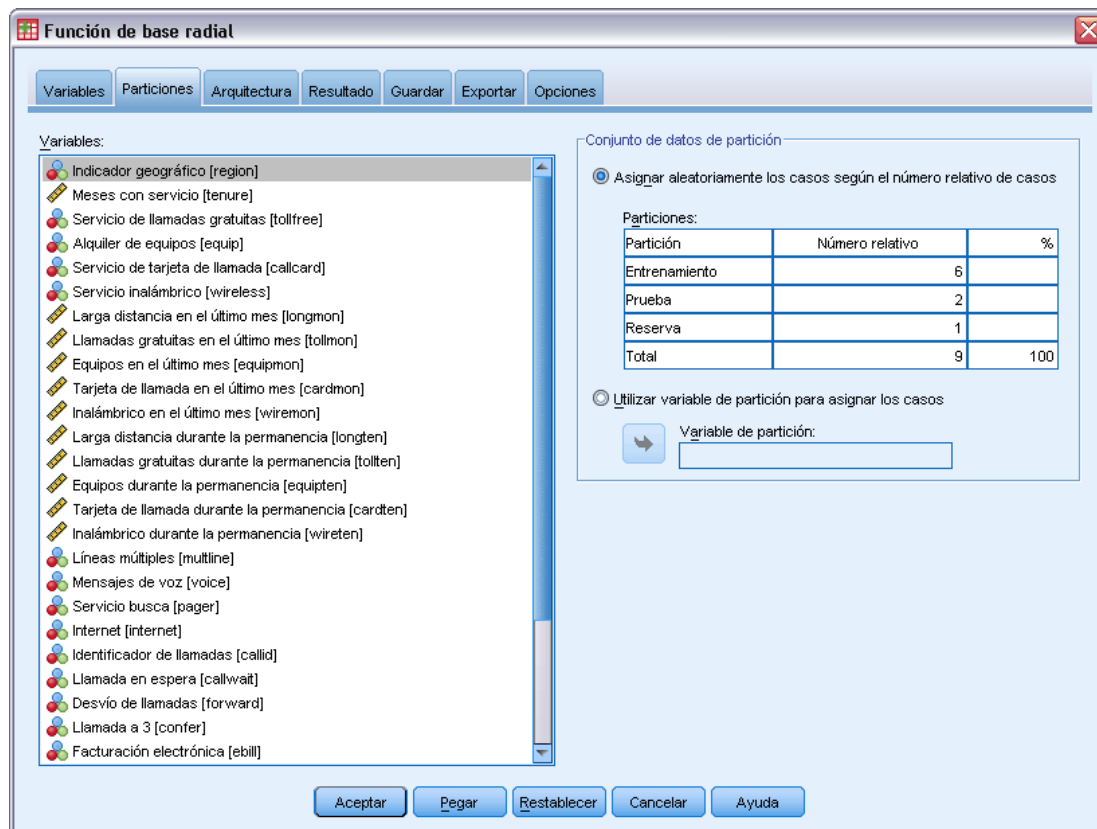
- ▶ Para ejecutar un análisis de función de base radial, elija en los menús: Analizar > Neural Networks > Función de base radial...

Figura 5-2
Función de base radial: pestaña Variables



- ▶ Seleccione *Categoría del cliente [custcat]* como variable dependiente.
- ▶ Seleccione *Estado civil [marital]*, *Nivel educativo [ed]*, *Retirado [retire]* y *Género [gender]* como factores.
- ▶ Seleccione desde *Edad en años [age]* hasta *Número de personas en el hogar [reside]* como covariables.
- ▶ Seleccione *Corregida normalizada* como método para cambiar la escala de las covariables.
- ▶ Pulse en la pestaña *Particiones*.

Figura 5-3
Función de base radial: Pestaña Particiones



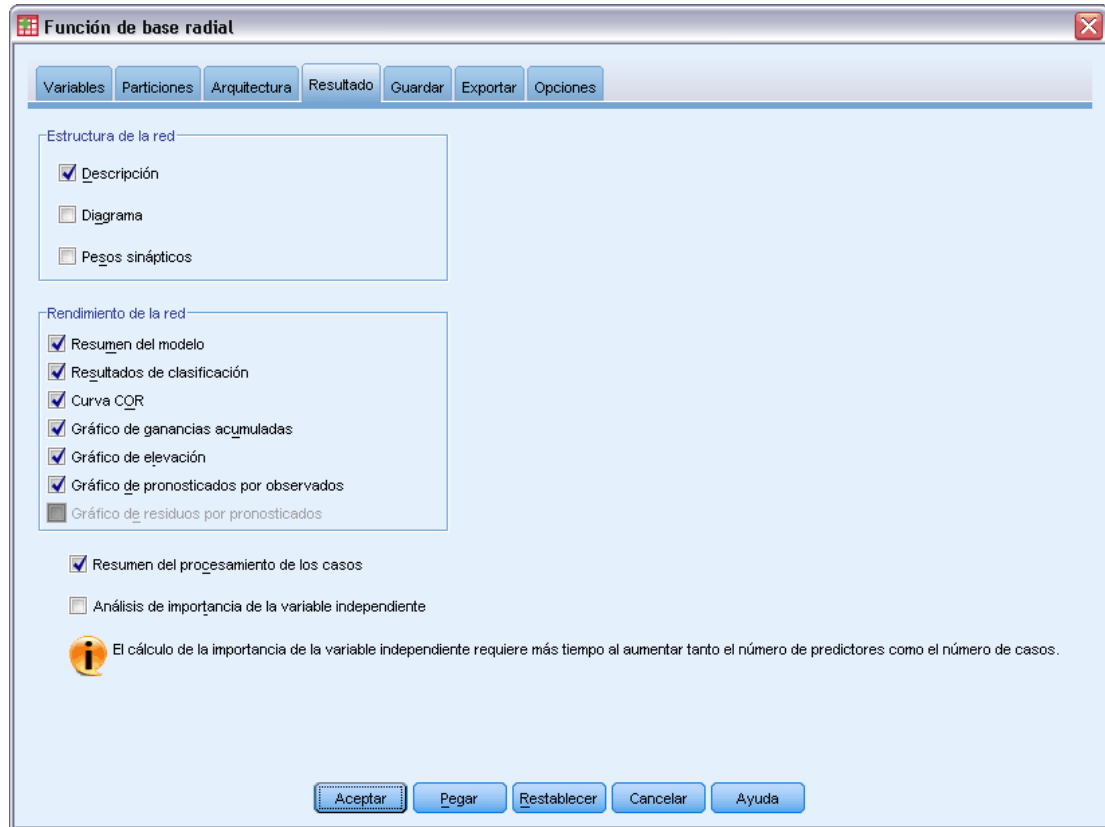
Mediante la especificación de números relativos de casos, es fácil crear particiones fraccionales cuyos porcentajes serían más difíciles de especificar. Imaginemos que desea asignar $2/3$ del conjunto de datos a la muestra de entrenamiento y $2/3$ de los casos restantes a las pruebas.

- ▶ Escriba 6 como número relativo de la muestra de entrenamiento.
- ▶ Escriba 2 como número relativo de la muestra de prueba.
- ▶ Escriba 1 como número relativo de la muestra reservada.

Se han especificado un total de 9 casos relativos. $6/9 = 2/3$, o aproximadamente el 66,67%, se asignarán a la muestra de entrenamiento; $2/9$, o aproximadamente el 22,22%, se asignarán a la prueba; $1/9$, a aproximadamente el 11,11% se asignarán a la muestra reservada.

- ▶ Pulse en la pestaña Resultados.

Figura 5-4
Función de base radial: Pestaña Resultados



- ▶ Anule la selección en Diagrama en el grupo de estructura de red.
- ▶ Seleccione Curva COR, Gráfico de ganancias acumuladas, Gráfico de elevación (índice) y Gráfico de pronosticados por observados en el grupo de rendimiento de red.
- ▶ Pulse en la pestaña Guardar.

Figura 5-5
Función de base radial: Pestaña Guardar

Función de base radial

Variables Particiones Arquitectura Resultado Guardar Exportar Opciones

Guardar categoría o valor pronosticado para cada variable dependiente
 Guardar pseudoprobabilidad pronosticada para cada variable dependiente

Variables:

Variable dependiente	Categoría o valor pronosticado		Pseudoprobabilidad pronosticada	
	Nombre de la variable guardada	Nombre raíz de las variables guardadas	Categorías para guardar	
custcat	RBF_PredictedValue	RBF_PseudoProbability	25	

Nombres de las variables guardadas:

Generar automáticamente nombres únicos
 Seleccione esta opción si desea añadir un nuevo conjunto de variables guardadas al conjunto de datos cada vez que ejecute un modelo.

Nombres personalizados
 Especificar nombres de las variables. Si selecciona esta opción, se reemplazarán todas las variables existentes con el mismo nombre o nombre raíz cada vez que se ejecute un modelo.

Aceptar Pegar Restablecer Cancelar Ayuda

- ▶ Seleccione Guardar categoría o valor pronosticado para cada variable dependiente y Guardar pseudoprobabilidad pronosticada para cada variable dependiente.
- ▶ Pulse en Aceptar.

Resumen de procesamiento de casos

Figura 5-6
Resumen de procesamiento de casos

	N	Porcentaje
Muestra Entrenamiento	665	66,5%
Prueba	224	22,4%
Reserva	111	11,1%
Válido	1000	100,0%
Excluido	0	
Total	1000	

El resumen del procesamiento de los casos muestra que se asignaron 665 casos a la muestra de entrenamiento, 224 a la muestra de prueba y 111 a la muestra reservada. No se excluyó ningún caso del análisis.

Información sobre la red

Figura 5-7
Información de red

Información sobre la red				
Capa de entrada	Factors	1	Estado civil	
		2	Nivel educativo	
		3	Retirado	
		4	Género	
	Covariates	1	Edad en años	
		2	Número de personas en el hogar	
		3	Ingresos familiares en miles	
		4	Años con el patrón actual	
		Número de unidades	15	
		Rescaling Method for Covariates		Tipificados
Capa oculta	Número de unidades		5 ^a	
	Función de activación		Softmax	
Capa de salida	Dependent Variables	1	Categoría del cliente	
	Número de unidades		4	
	Función de activación		Identidad	
	Función de error		Suma de cuadrados	

Determinado por el criterio de datos de prueba: El número "óptimo" de unidades ocultas es el que produce el menor error en los datos de prueba.

La tabla de información de red muestra información sobre la red neuronal y resulta útil para garantizar que las especificaciones son correctas. Aquí se ve en particular que:

- El número de unidades en la capa de entrada es el número de covariables más el número total de niveles de factor. Se crea una unidad independiente para cada categoría de *Estado civil*, *Nivel educativo*, *Retirado* y *Género* y ninguna de las categorías se considera como una unidad "redundante", como es habitual en muchos procedimientos de modelado.
- De igual manera, se crea una unidad de resultado independiente para cada categoría de *Categoría del cliente*, para un total de 4 unidades en la capa de salida.
- Se cambia la escala de las covariables utilizando el método normalizado corregido.
- La selección de arquitectura automática ha elegido 9 unidades en la capa oculta.
- El resto de la información de red se toma por defecto para el procedimiento.

Resumen del modelo

Figura 5-8
Resumen del modelo

Entrenamiento	Suma de errores cuadráticos	241.111
	Porcentaje de pronósticos incorrectos	63.9%
	Tiempo de entrenamiento	00:00:04.594
Prueba	Suma de errores cuadráticos	81.873 ^a
	Porcentaje de pronósticos incorrectos	67.0%
	Reserva	Porcentaje de pronósticos incorrectos

Variable dependiente: Categoría del cliente

El número de unidades ocultas está determinado por el criterio de datos de prueba:
El número "óptimo" de unidades ocultas es el que produce el menor error en los datos de prueba.

El resumen del modelo muestra información sobre los resultados del entrenamiento, la prueba y la aplicación de la red final a la muestra reservada.

- Se muestra la suma de errores cuadráticos, ya que siempre se utiliza para redes RBF. Ésta es la función de error que la red intenta minimizar durante el entrenamiento y la prueba.
- El porcentaje de pronósticos incorrectos se toma de la tabla de clasificación, y se discutirá más adelante en ese tema.

Classification

Figura 5-9
Classification

Muestra	Observado	Pronosticado				Percent Correct
		Servicio básico	E-Servicio	Servicio Plus	Servicio Total	
Entrenamiento	Servicio básico	64	0	66	45	36,6%
	E-Servicio	22	1	57	61	,7%
	Servicio Plus	47	0	104	34	56,2%
	Servicio Total	29	1	49	85	51,8%
	Overall Percent	24,4%	,3%	41,5%	33,8%	38,2%
Prueba	Servicio básico	18	0	26	15	30,5%
	E-Servicio	15	0	16	22	,0%
	Servicio Plus	11	0	39	15	60,0%
	Servicio Total	4	0	17	26	55,3%
	Overall Percent	21,4%	,0%	43,8%	34,8%	37,1%
Reserva	Servicio básico	11	0	11	10	34,4%
	E-Servicio	4	0	9	10	,0%
	Servicio Plus	10	0	19	2	61,3%
	Servicio Total	5	0	5	15	60,0%
	Overall Percent	27,0%	,0%	39,6%	33,3%	40,5%

Variable dependiente: Categoría del cliente

La tabla de clasificación muestra los resultados prácticos de la utilización de la red. Para cada caso, la respuesta pronosticada es la categoría con la mayor pseudoprobabilidad pronosticada por el modelo.

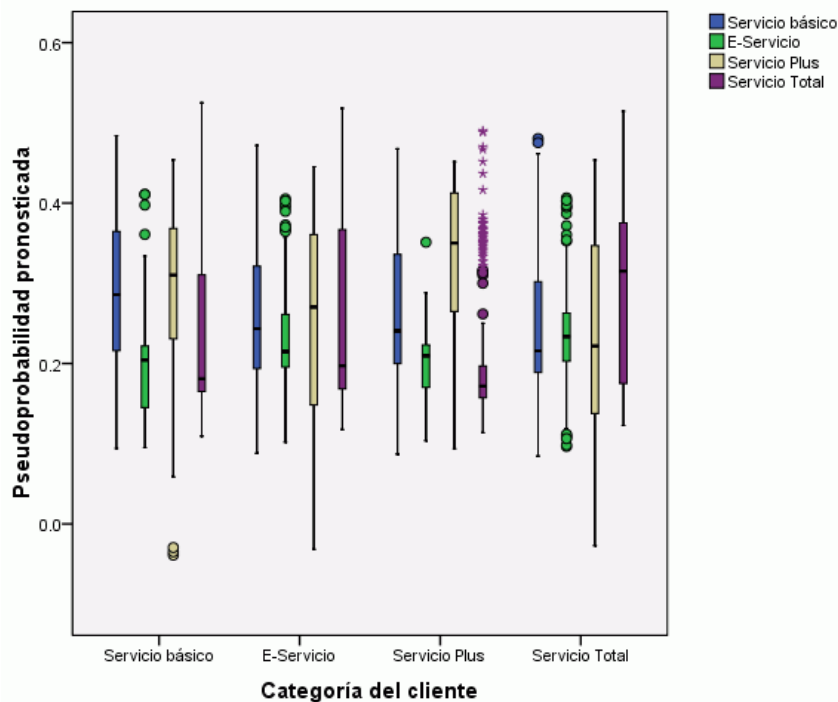
- Las casillas de la diagonal son los pronósticos correctos.
- Las casillas fuera de la diagonal son los pronósticos incorrectos.

Dados los datos observados, el modelo “nulo” (es decir, el que no incluye ningún predictor) clasificaría a todos los clientes en el grupo modal, *Servicio plus*. Por tanto, el modelo nulo sería correcto $281/1000 = 28,1\%$ de las veces. La red RBF consigue un $10,1\%$ más, es decir, un $38,2\%$ de los clientes. En concreto, el modelo es particularmente bueno para identificar los clientes de *Servicio Plus* y *Servicio total*. Sin embargo, funciona excepcionalmente mal para clasificar los clientes de *Servicio electrónico*. Tal vez necesite encontrar otro predictor para separar estos clientes. Otra posibilidad, dado que estos clientes se suelen clasificar erróneamente como clientes *Servicio plus* y *Servicio total*, es que la compañía intentase sencillamente vender un producto superior a los posibles clientes que normalmente figurarían en la categoría *E-servicio*.

Las clasificaciones basadas en los casos utilizados para crear el modelo tienden a ser demasiado “optimistas” en el sentido de que aumentan en exceso la tasa de clasificación. La muestra reservada ayuda a validar el modelo. En este caso, el modelo clasificó correctamente el $40,2\%$ de casos. Aunque la muestra reservada es bastante pequeña, eso sugiere que el modelo es correcto aproximadamente dos de cada cinco veces.

Gráfico de pronosticados por observados

Figura 5-10
Gráfico de pronosticados por observados



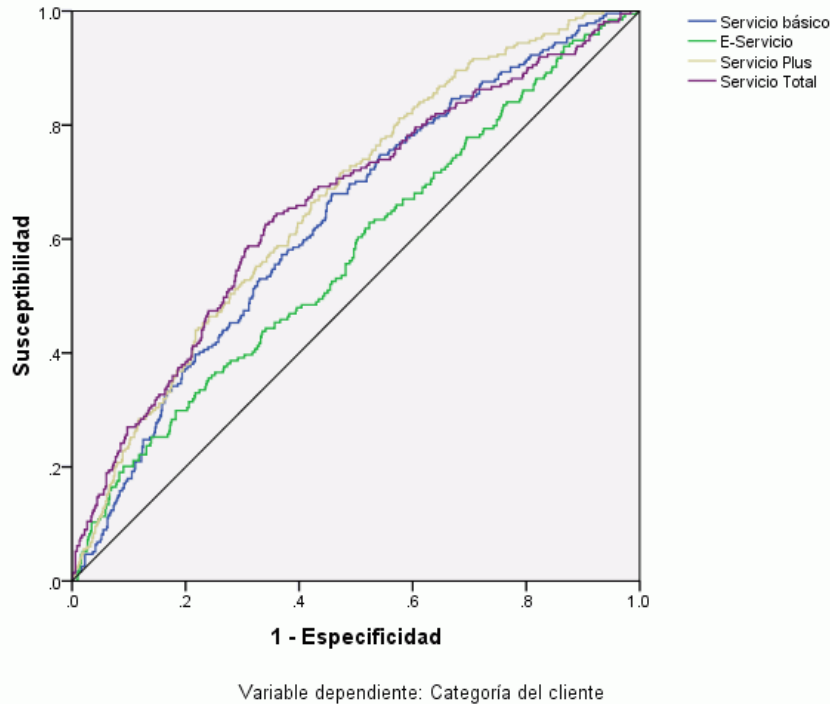
En el caso de las variables dependientes categóricas, el gráfico de pronosticados por observados muestra diagramas de caja agrupados de pseudoprobabilidades pronosticadas para las muestras de entrenamiento y de prueba combinadas. El eje X corresponde a las categorías de respuesta observadas, y la leyenda corresponde a categorías pronosticadas. Por tanto:

- El diagrama de caja situado más a la izquierda muestra, para los casos que tienen la categoría observada *Servicio básico*, la pseudoprobabilidad pronosticada de la categoría *Servicio básico*.
- El siguiente diagrama de caja de la derecha muestra, para los casos que tienen la categoría observada *Servicio básico*, a pseudoprobabilidad pronosticada de la categoría *E-servicio*.
- El tercer diagrama de caja muestra, para los casos que tienen la categoría observada *Servicio básico*, la pseudoprobabilidad pronosticada de la categoría *Servicio plus*. Hay que recordar que en la tabla de clasificación había aproximadamente tantos clientes de *Servicio básico* que se habían clasificado incorrectamente como *Servicio plus* como clasificados correctamente como *Servicio básico*; por lo que este cuadro de caja es equivalente aproximadamente al que aparece más a la izquierda.
- El cuarto diagrama de caja muestra, para los casos que tienen la categoría observada *Servicio básico*, la pseudoprobabilidad pronosticada de la categoría *Servicio total*

Como hay más de dos categorías en la variable de destino, los primeros cuatro diagramas de caja no son simétricos con respecto a la línea horizontal de 0,5, ni de ninguna otra manera. Como consecuencia, puede resultar difícil interpretar este gráfico para destinos con más de dos categorías, ya que es imposible determinar, examinando la parte de casos de un diagrama de caja, la ubicación correspondiente de dichos casos en otro diagrama de caja.

Curva COR

Figura 5-11
curva COR



Una curva COR proporciona una representación visual de la **susceptibilidad** por **especificidad** de todos los puntos de corte de clasificación posibles. El gráfico que aquí se incluye muestra cuatro curvas, una para cada categoría de la variable de destino.

Tenga en cuenta que este gráfico se basa en las muestras de prueba y entrenamiento combinadas. Para generar un gráfico COR para la muestra reservada, divida el archivo de la variable de partición y ejecute el procedimiento Curva COR en las pseudoprobabilidades pronosticadas guardadas.

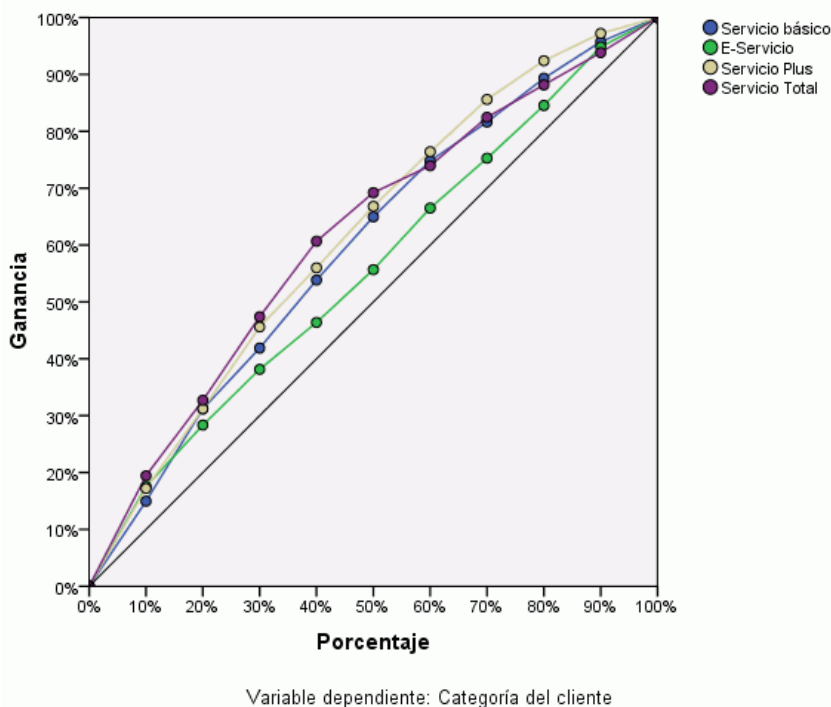
Figura 5-12
Área bajo la curva

Categoría del cliente	Área
Servicio básico	,635
E-Servicio	,573
Servicio Plus	,668
Servicio Total	,659

El área bajo la curva es un resumen numérico de la curva COR, y los valores de la tabla representan, para cada categoría, la probabilidad de que la pseudoprobabilidad pronosticada de estar en esa categoría sea mayor para un caso elegido aleatoriamente en esa categoría que para un caso elegido aleatoriamente que no pertenezca a esa categoría. Por ejemplo, para un cliente seleccionado de forma aleatoria en *Servicio plus* y un cliente seleccionado de forma aleatoria en *Servicio básico*, *E-servicio* o *Servicio total*, hay una probabilidad de 0,668 de que la pseudoprobabilidad pronosticada por el modelo de que cause mora sea superior para el cliente en *Servicio plus*.

Gráficos de elevación y ganancias acumuladas

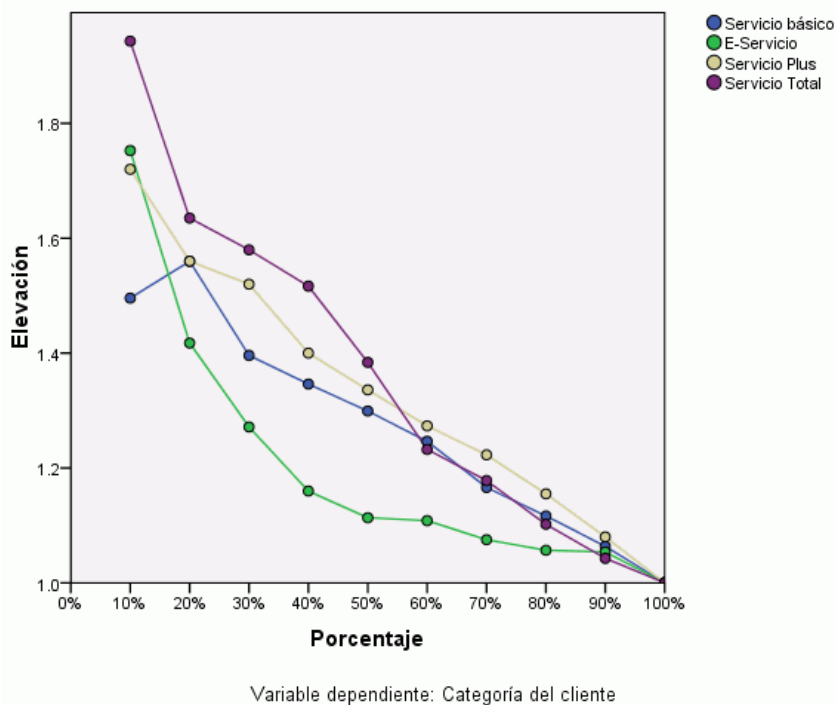
Figura 5-13
Gráfico de ganancias acumuladas



El gráfico de ganancias acumuladas muestra el porcentaje del número total de casos de una categoría dada “ganada” al dirigirse a un porcentaje del número total de casos. Por ejemplo, el primer punto de la curva de la categoría *Servicio total* es de aproximadamente (10%, 20%), lo que significa que si puntúa un conjunto de datos con la red y ordena todos los casos por la pseudoprobabilidad pronosticada de *Servicio total*, esperaríamos que el 10% superior contuviera aproximadamente el 20% de todos los casos que realmente corresponden a la categoría *Servicio total*. Del mismo modo, el 20% superior contendría aproximadamente el 30% de los clientes que causan mora, el 30% superior de los casos, el 50% de los clientes que causan mora, y así sucesivamente. Si selecciona el 100% del conjunto de datos puntuados, obtendrá todos los clientes que causan mora del conjunto de datos.

La línea diagonal es la curva de “línea base”, si selecciona el 10% de los casos del conjunto de datos puntuados de forma aleatoria, esperaríamos “ganar” aproximadamente el 10% de todos los casos que realmente corresponden a cualquier categoría. Cuanto más arriba está la curva de línea base, mayor es la ganancia.

Figura 5-14
Gráfico de elevación



El gráfico de elevación se deriva del gráfico de ganancias acumuladas. Los valores del eje Y corresponden a la relación de la ganancia acumulada de cada curva con respecto a la línea base. Así, la elevación de 10% para la categoría *Servicio total* es de aproximadamente $20\%/10\% = 2,0$. Proporciona otra forma de ver la información en el gráfico de ganancias acumuladas.

Nota: los gráficos de elevación y de ganancias acumuladas se basan en las muestras de prueba y entrenamiento combinadas.

Lecturas recomendadas

Consulte los siguientes textos si desea obtener más información sobre la función de base radial:

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*, 3rd ed. Oxford: Oxford University Press.

Fine, T. L. 1999. *Feedforward Neural Network Methodology*, 3rd ed. Nueva York: Springer-Verlag.

Haykin, S. 1998. *Neural Networks: A Comprehensive Foundation*, 2nd ed. Nueva York: Macmillan College Publishing.

Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Tao, K. K. 1993. A closer look at the radial basis function (RBF) networks. En: *Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems, and Computers*, A. Singh, ed. Los Alamitos, Calif.: IEEE Comput. Soc. Press.

Uykan, Z., C. Guzelis, M. E. Celebi, y H. N. Koivo. 2000. Analysis of input-output clustering for determining centers of RBFN. *IEEE Transactions on Neural Networks*, 11, .

Archivos muestrales

Los archivos muestrales instalados con el producto se encuentran en el subdirectorio *Samples* del directorio de instalación. Hay una carpeta independiente dentro del subdirectorio *Samples* para cada uno de los siguientes idiomas: Inglés, francés, alemán, italiano, japonés, coreano, polaco, ruso, chino simplificado, español y chino tradicional.

No todos los archivos muestrales están disponibles en todos los idiomas. Si un archivo muestral no está disponible en un idioma, esa carpeta de idioma contendrá una versión en inglés del archivo muestral.

Descripciones

A continuación, se describen brevemente los archivos muestrales usados en varios ejemplos que aparecen a lo largo de la documentación.

- **accidents.sav.** Archivo de datos hipotéticos sobre una compañía de seguros que estudia los factores de riesgo de edad y género que influyen en los accidentes de automóviles de una región determinada. Cada caso corresponde a una clasificación cruzada de categoría de edad y género.
- **adl.sav.** Archivo de datos hipotéticos relativo a los esfuerzos para determinar las ventajas de un tipo propuesto de tratamiento para pacientes que han sufrido un derrame cerebral. Los médicos dividieron de manera aleatoria a pacientes (mujeres) que habían sufrido un derrame cerebral en dos grupos. El primer grupo recibió el tratamiento físico estándar y el segundo recibió un tratamiento emocional adicional. Tres meses después de los tratamientos, se puntuaron las capacidades de cada paciente para realizar actividades cotidianas como variables ordinales.
- **advert.sav.** Archivo de datos hipotéticos sobre las iniciativas de un minorista para examinar la relación entre el dinero invertido en publicidad y las ventas resultantes. Para ello, se recopilaron las cifras de ventas anteriores y los costes de publicidad asociados.
- **aflatoxin.sav.** Archivo de datos hipotéticos sobre las pruebas realizadas en las cosechas de maíz con relación a la aflatoxina, un veneno cuya concentración varía ampliamente en los rendimientos de cultivo y entre los mismos. Un procesador de grano ha recibido 16 muestras de cada uno de los 8 rendimientos de cultivo y ha medido los niveles de aflatoxinas en partes por millón (PPM).
- **aflatoxin20.sav.** Este archivo de datos contiene las medidas de aflatoxina de cada una de las 16 muestras de los rendimientos 4 y 8 procedentes del archivo de datos *aflatoxin.sav*.
- **anorectic.sav.** Mientras trabajaban en una sintomatología estandarizada del comportamiento anoréxico/bulímico, los investigadores (Van der Ham, Meulman, Van Strien, y Van Engeland, 1997) realizaron un estudio de 55 adolescentes con trastornos de la alimentación conocidos. Cada paciente fue examinado cuatro veces durante cuatro años, lo que representa un total

de 220 observaciones. En cada observación, se puntuó a los pacientes por cada uno de los 16 síntomas. Faltan las puntuaciones de los síntomas para el paciente 71 en el tiempo 2, el paciente 76 en el tiempo 2 y el paciente 47 en el tiempo 3, lo que nos deja 217 observaciones válidas.

- **autoaccidents.sav.** Archivo de datos hipotéticos sobre las iniciativas de un analista de seguros para elaborar un modelo del número de accidentes de automóvil por conductor teniendo en cuenta la edad y el género del conductor. Cada caso representa un conductor diferente y registra el sexo, la edad en años y el número de accidentes de automóvil del conductor en los últimos cinco años.
- **band.sav** Este archivo de datos contiene las cifras de ventas semanales hipotéticas de CD de música de una banda. También se incluyen datos para tres variables predictoras posibles.
- **bankloan.sav.** Archivo de datos hipotéticos sobre las iniciativas de un banco para reducir la tasa de moras de créditos. El archivo contiene información financiera y demográfica de 850 clientes anteriores y posibles clientes. Los primeros 700 casos son clientes a los que anteriormente se les ha concedido un préstamo. Al menos 150 casos son posibles clientes cuyos riesgos de crédito el banco necesita clasificar como positivos o negativos.
- **bankloan_binning.sav.** Archivo de datos hipotéticos que contiene información financiera y demográfica sobre 5.000 clientes anteriores.
- **behavior.sav.** En un ejemplo clásico (Price y Bouffard, 1974), se pidió a 52 estudiantes que valoraran las combinaciones de 15 situaciones y 15 comportamientos en una escala de 10 puntos que oscilaba entre 0 =“extremadamente apropiado” y 9=“extremadamente inapropiado”. Los valores promediados respecto a los individuos se toman como disimilaridades.
- **behavior_ini.sav.** Este archivo de datos contiene una configuración inicial para una solución bidimensional de *behavior.sav*.
- **brakes.sav.** Archivo de datos hipotéticos sobre el control de calidad de una fábrica que produce frenos de disco para automóviles de alto rendimiento. El archivo de datos contiene las medidas del diámetro de 16 discos de cada una de las 8 máquinas de producción. El diámetro objetivo para los frenos es de 322 milímetros.
- **breakfast.sav.** En un estudio clásico (Green y Rao, 1972), se pidió a 21 estudiantes de administración de empresas de la Wharton School y sus cónyuges que ordenaran 15 elementos de desayuno por orden de preferencia, de 1=“más preferido” a 15=“menos preferido”. Sus preferencias se registraron en seis escenarios distintos, de “Preferencia global” a “Aperitivo, con bebida sólo”.
- **breakfast-overall.sav.** Este archivo de datos sólo contiene las preferencias de elementos de desayuno para el primer escenario, “Preferencia global”.
- **broadband_1.sav** Archivo de datos hipotéticos que contiene el número de suscriptores, por región, a un servicio de banda ancha nacional. El archivo de datos contiene números de suscriptores mensuales para 85 regiones durante un período de cuatro años.
- **broadband_2.sav** Este archivo de datos es idéntico a *broadband_1.sav* pero contiene datos para tres meses adicionales.
- **car_insurance_claims.sav.** Un conjunto de datos presentados y analizados en otro lugar (McCullagh y Nelder, 1989) estudia las reclamaciones por daños en vehículos. La cantidad de reclamaciones media se puede modelar como si tuviera una distribución Gamma, mediante

una función de enlace inversa para relacionar la media de la variable dependiente con una combinación lineal de la edad del asegurado, el tipo de vehículo y la antigüedad del vehículo. El número de reclamaciones presentadas se puede utilizar como una ponderación de escalamiento.

- **car_sales.sav.** Este archivo de datos contiene estimaciones de ventas, precios de lista y especificaciones físicas hipotéticas de varias marcas y modelos de vehículos. Los precios de lista y las especificaciones físicas se han obtenido de *edmunds.com* y de sitios de fabricantes.
- **car_sales_uprepared.sav.** Ésta es una versión modificada de *car_sales.sav* que no incluye ninguna versión transformada de los campos.
- **carpet.sav** En un ejemplo muy conocido (Green y Wind, 1973), una compañía interesada en sacar al mercado un nuevo limpiador de alfombras desea examinar la influencia de cinco factores sobre la preferencia del consumidor: diseño del producto, marca comercial, precio, sello de *buen producto para el hogar* y garantía de devolución del importe. Hay tres niveles de factores para el diseño del producto, cada uno con una diferente colocación del cepillo del aplicador; tres nombres comerciales (*K2R*, *Glory* y *Bissell*); tres niveles de precios; y dos niveles (no o sí) para los dos últimos factores. Diez consumidores clasificaron 22 perfiles definidos por estos factores. La variable *Preferencia* contiene el rango de las clasificaciones medias de cada perfil. Las clasificaciones inferiores corresponden a preferencias elevadas. Esta variable refleja una medida global de la preferencia de cada perfil.
- **carpet_prefs.sav** Este archivo de datos se basa en el mismo ejemplo que el descrito para *carpet.sav*, pero contiene las clasificaciones reales recogidas de cada uno de los 10 consumidores. Se pidió a los consumidores que clasificaran los 22 perfiles de los productos empezando por el menos preferido. Las variables desde *PREF1* hasta *PREF22* contienen los ID de los perfiles asociados, como se definen en *carpet_plan.sav*.
- **catalog.sav** Este archivo de datos contiene cifras de ventas mensuales hipotéticas de tres productos vendidos por una compañía de venta por catálogo. También se incluyen datos para cinco variables predictoras posibles.
- **catalog_seasons.sav** Este archivo de datos es igual que *catalog.sav*, con la excepción de que incluye un conjunto de factores estacionales calculados a partir del procedimiento Descomposición estacional junto con las variables de fecha que lo acompañan.
- **cellular.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía de telefonía móvil para reducir el abandono de clientes. Las puntuaciones de propensión al abandono de clientes se aplican a las cuentas, oscilando de 0 a 100. Las cuentas con una puntuación de 50 o superior pueden estar buscando otros proveedores.
- **ceramics.sav.** Archivo de datos hipotéticos sobre las iniciativas de un fabricante para determinar si una nueva aleación de calidad tiene una mayor resistencia al calor que una aleación estándar. Cada caso representa una prueba independiente de una de las aleaciones; la temperatura a la que registró el fallo del rodamiento.
- **cereal.sav.** Archivo de datos hipotéticos sobre una encuesta realizada a 880 personas sobre sus preferencias en el desayuno, teniendo también en cuenta su edad, sexo, estado civil y si tienen un estilo de vida activo o no (en función de si practican ejercicio al menos dos veces a la semana). Cada caso representa un encuestado diferente.
- **clothing_defects.sav.** Archivo de datos hipotéticos sobre el proceso de control de calidad en una fábrica de prendas. Los inspectores toman una muestra de prendas de cada lote producido en la fábrica, y cuentan el número de prendas que no son aceptables.

- **coffee.sav.** Este archivo de datos pertenece a las imágenes percibidas de seis marcas de café helado (Kennedy, Riquier, y Sharp, 1996). Para cada uno de los 23 atributos de imagen de café helado, los encuestados seleccionaron todas las marcas que quedaban descritas por el atributo. Las seis marcas se denotan AA, BB, CC, DD, EE y FF para mantener la confidencialidad.
- **contacts.sav.** Archivo de datos hipotéticos sobre las listas de contactos de un grupo de representantes de ventas de ordenadores de empresa. Cada uno de los contactos está categorizado por el departamento de la compañía en el que trabaja y su categoría en la compañía. Además, también se registran los importes de la última venta realizada, el tiempo transcurrido desde la última venta y el tamaño de la compañía del contacto.
- **creditpromo.sav.** Archivo de datos hipotéticos sobre las iniciativas de unos almacenes para evaluar la eficacia de una promoción de tarjetas de crédito reciente. Para este fin, se seleccionaron aleatoriamente 500 titulares. La mitad recibieron un anuncio promocionando una tasa de interés reducida sobre las ventas realizadas en los siguientes tres meses. La otra mitad recibió un anuncio estacional estándar.
- **customer_dbase.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía para usar la información de su almacén de datos para realizar ofertas especiales a los clientes con más probabilidades de responder. Se seleccionó un subconjunto de la base de clientes aleatoriamente a quienes se ofrecieron las ofertas especiales y sus respuestas se registraron.
- **customer_information.sav.** Archivo de datos hipotéticos que contiene la información de correo del cliente, como el nombre y la dirección.
- **customer_subset.sav.** Un subconjunto de 80 casos de *customer_dbase.sav*.
- **customers_model.sav.** Este archivo contiene datos hipotéticos sobre los individuos a los que va dirigida una campaña de marketing. Estos datos incluyen información demográfica, un resumen del historial de compras y si cada individuo respondió a la campaña. Cada caso representa un individuo diferente.
- **customers_new.sav.** Este archivo contiene datos hipotéticos sobre los individuos que son candidatos potenciales para una campaña de marketing. Estos datos incluyen información demográfica y un resumen del historial de compras de cada individuo. Cada caso representa un individuo diferente.
- **debate.sav.** Archivos de datos hipotéticos sobre las respuestas emparejadas de una encuesta realizada a los asistentes a un debate político antes y después del debate. Cada caso corresponde a un encuestado diferente.
- **debate_aggregate.sav.** Archivo de datos hipotéticos que agrega las respuestas de *debate.sav*. Cada caso corresponde a una clasificación cruzada de preferencias antes y después del debate.
- **demo.sav.** Archivos de datos hipotéticos sobre una base de datos de clientes adquirida con el fin de enviar por correo ofertas mensuales. Se registra si el cliente respondió a la oferta, junto con información demográfica diversa.
- **demo_cs_1.sav.** Archivo de datos hipotéticos sobre el primer paso de las iniciativas de una compañía para recopilar una base de datos de información de encuestas. Cada caso corresponde a una ciudad diferente, y se registra la identificación de la ciudad, la región, la provincia y el distrito.
- **demo_cs_2.sav.** Archivo de datos hipotéticos sobre el segundo paso de las iniciativas de una compañía para recopilar una base de datos de información de encuestas. Cada caso corresponde a una unidad familiar diferente de las ciudades seleccionadas en el primer paso, y

se registra la identificación de la unidad, la subdivisión, la ciudad, el distrito, la provincia y la región. También se incluye la información de muestreo de las primeras dos etapas del diseño.

- **demo_cs.sav.** Archivo de datos hipotéticos que contiene información de encuestas recopilada mediante un diseño de muestreo complejo. Cada caso corresponde a una unidad familiar distinta, y se recopila información demográfica y de muestreo diversa.
- **dmdata.sav.** Éste es un archivo de datos hipotéticos que contiene información demográfica y de compras para una empresa de marketing directo. *dmdata2.sav* contiene información para un subconjunto de contactos que recibió un envío de prueba, y *dmdata3.sav* contiene información sobre el resto de contactos que no recibieron el envío de prueba.
- **dietstudy.sav.** Este archivo de datos hipotéticos contiene los resultados de un estudio sobre la “dieta Stillman” (Rickman, Mitchell, Dingman, y Dalen, 1974). Cada caso corresponde a un sujeto distinto y registra sus pesos antes y después de la dieta en libras y niveles de triglicéridos en mg/100 ml.
- **dvdplayer.sav.** Archivo de datos hipotéticos sobre el desarrollo de un nuevo reproductor de DVD. El equipo de marketing ha recopilado datos de grupo de enfoque mediante un prototipo. Cada caso corresponde a un usuario encuestado diferente y registra información demográfica sobre los encuestados y sus respuestas a preguntas acerca del prototipo.
- **german_credit.sav.** Este archivo de datos se toma del conjunto de datos “German credit” de las Repository of Machine Learning Databases (Blake y Merz, 1998) de la Universidad de California, Irvine.
- **grocery_1month.sav.** Este archivo de datos hipotéticos es el archivo de datos *grocery_coupons.sav* con las compras semanales “acumuladas” para que cada caso corresponda a un cliente diferente. Algunas de las variables que cambiaban semanalmente desaparecen de los resultados, y la cantidad gastada registrada se convierte ahora en la suma de las cantidades gastadas durante las cuatro semanas del estudio.
- **grocery_coupons.sav.** Archivo de datos hipotéticos que contiene datos de encuestas recopilados por una cadena de tiendas de alimentación interesada en los hábitos de compra de sus clientes. Se sigue a cada cliente durante cuatro semanas, y cada caso corresponde a un cliente-semana distinto y registra información sobre dónde y cómo compran los clientes, incluida la cantidad que invierten en comestibles durante esa semana.
- **guttman.sav.** Bell (Bell, 1961) presentó una tabla para ilustrar posibles grupos sociales. Guttman (Guttman, 1968) utilizó parte de esta tabla, en la que se cruzaron cinco variables que describían elementos como la interacción social, sentimientos de pertenencia a un grupo, proximidad física de los miembros y grado de formalización de la relación con siete grupos sociales teóricos, incluidos multitudes (por ejemplo, las personas que acuden a un partido de fútbol), espectadores (por ejemplo, las personas que acuden a un teatro o de una conferencia), públicos (por ejemplo, los lectores de periódicos o los espectadores de televisión), muchedumbres (como una multitud pero con una interacción mucho más intensa), grupos primarios (íntimos), grupos secundarios (voluntarios) y la comunidad moderna (confederación débil que resulta de la proximidad cercana física y de la necesidad de servicios especializados).
- **health_funding.sav.** Archivo de datos hipotéticos que contiene datos sobre inversión en sanidad (cantidad por 100 personas), tasas de enfermedad (índice por 10.000 personas) y visitas a centros de salud (índice por 10.000 personas). Cada caso representa una ciudad diferente.

- **hivassay.sav.** Archivo de datos hipotéticos sobre las iniciativas de un laboratorio farmacéutico para desarrollar un ensayo rápido para detectar la infección por VIH. Los resultados del ensayo son ocho tonos de rojo con diferentes intensidades, donde los tonos más oscuros indican una mayor probabilidad de infección. Se llevó a cabo una prueba de laboratorio de 2.000 muestras de sangre, de las cuales una mitad estaba infectada con el VIH y la otra estaba limpia.
- **hourlywagedata.sav.** Archivo de datos hipotéticos sobre los salarios por horas de enfermeras de puestos de oficina y hospitales y con niveles distintos de experiencia.
- **insurance_claims.sav.** Éste es un archivo de datos hipotéticos sobre una compañía de seguros que desea generar un modelo para etiquetar las reclamaciones sospechosas y potencialmente fraudulentas. Cada caso representa una reclamación diferente.
- **insure.sav.** Archivo de datos hipotéticos sobre una compañía de seguros que estudia los factores de riesgo que indican si un cliente tendrá que hacer una reclamación a lo largo de un contrato de seguro de vida de 10 años. Cada caso del archivo de datos representa un par de contratos (de los que uno registró una reclamación y el otro no), agrupados por edad y sexo.
- **judges.sav.** Archivo de datos hipotéticos sobre las puntuaciones concedidas por jueces cualificados (y un aficionado) a 300 actuaciones gimnásticas. Cada fila representa una actuación diferente; los jueces vieron las mismas actuaciones.
- **kinship_dat.sav.** Rosenberg y Kim (Rosenberg y Kim, 1975) comenzaron a analizar 15 términos de parentesco [tía, hermano, primos, hija, padre, nieta, abuelo, abuela, nieto, madre, sobrino, sobrina, hermana, hijo, tío]. Le pidieron a cuatro grupos de estudiantes universitarios (dos masculinos y dos femeninos) que ordenaran estos grupos según las similitudes. A dos grupos (uno masculino y otro femenino) se les pidió que realizaran la ordenación dos veces, pero que la segunda ordenación la hicieran según criterios distintos a los de la primera. Así, se obtuvo un total de seis “fuentes“. Cada fuente se corresponde con una matriz de proximidades de 15×15 cuyas casillas son iguales al número de personas de una fuente menos el número de veces que se partitionaron los objetos en esa fuente.
- **kinship_ini.sav.** Este archivo de datos contiene una configuración inicial para una solución tridimensional de *kinship_dat.sav*.
- **kinship_var.sav.** Este archivo de datos contiene variables independientes *sexo*, *gener(ación)*, y *grado* (de separación) que se pueden usar para interpretar las dimensiones de una solución para *kinship_dat.sav*. Concretamente, se pueden usar para restringir el espacio de la solución a una combinación lineal de estas variables.
- **marketvalues.sav.** Archivo de datos sobre las ventas de casas en una nueva urbanización de Algonquin, Ill., durante los años 1999 y 2000. Los datos de estas ventas son públicos.
- **nhis2000_subset.sav.** La National Health Interview Survey (NHIS, encuesta del Centro Nacional de Estadísticas de Salud de EE.UU.) es una encuesta detallada realizada entre la población civil de Estados Unidos. Las encuestas se realizaron en persona a una muestra representativa de las unidades familiares del país. Se recogió tanto la información demográfica como las observaciones acerca del estado y los hábitos de salud de los integrantes de cada unidad familiar. Este archivo de datos contiene un subconjunto de información de la encuesta de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Archivo de datos y documentación de uso público. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Fecha de acceso: 2003.

- **ozono.sav.** Los datos incluyen 330 observaciones de seis variables meteorológicas para pronosticar la concentración de ozono a partir del resto de variables. Los investigadores anteriores (Breiman y Friedman, 1985), (Hastie y Tibshirani, 1990) han encontrado que no hay linealidad entre estas variables, lo que dificulta los métodos de regresión típica.
- **pain_medication.sav.** Este archivo de datos hipotéticos contiene los resultados de una prueba clínica sobre medicación antiinflamatoria para tratar el dolor artrítico crónico. Resulta de particular interés el tiempo que tarda el fármaco en hacer efecto y cómo se compara con una medicación existente.
- **patient_los.sav.** Este archivo de datos hipotéticos contiene los registros de tratamiento de pacientes que fueron admitidos en el hospital ante la posibilidad de sufrir un infarto de miocardio (IM o “ataque al corazón”). Cada caso corresponde a un paciente distinto y registra diversas variables relacionadas con su estancia hospitalaria.
- **patlos_sample.sav.** Este archivo de datos hipotéticos contiene los registros de tratamiento de una muestra de pacientes que recibieron trombolíticos durante el tratamiento del infarto de miocardio (IM o “ataque al corazón”). Cada caso corresponde a un paciente distinto y registra diversas variables relacionadas con su estancia hospitalaria.
- **polishing.sav.** Archivo de datos “Nambeware Polishing Times” (Tiempo de pulido de metal) de la biblioteca de datos e historiales. Contiene datos sobre las iniciativas de un fabricante de cuberterías de metal (Nambe Mills, Santa Fe, N. M.) para planificar su programa de producción. Cada caso representa un artículo distinto de la línea de productos. Se registra el diámetro, el tiempo de pulido, el precio y el tipo de producto de cada artículo.
- **poll_cs.sav.** Archivo de datos hipotéticos sobre las iniciativas de los encuestadores para determinar el nivel de apoyo público a una ley antes de una asamblea legislativa. Los casos corresponden a votantes registrados. Cada caso registra el condado, la población y el vecindario en el que vive el votante.
- **poll_cs_sample.sav.** Este archivo de datos hipotéticos contiene una muestra de los votantes enumerados en *poll_cs.sav*. La muestra se tomó según el diseño especificado en el archivo de plan *poll_csplan* y este archivo de datos registra las probabilidades de inclusión y las ponderaciones muestrales. Sin embargo, tenga en cuenta que debido a que el plan muestral hace uso de un método de probabilidad proporcional al tamaño (PPS), también existe un archivo que contiene las probabilidades de selección conjunta (*poll_jointprob.sav*). Las variables adicionales que corresponden a los datos demográficos de los votantes y sus opiniones sobre la propuesta de ley se recopilaron y añadieron al archivo de datos después de tomar la muestra.
- **property_assess.sav.** Archivo de datos hipotéticos sobre las iniciativas de un asesor del condado para mantener actualizada la evaluación de los valores de las propiedades utilizando recursos limitados. Los casos corresponden a las propiedades vendidas en el condado el año anterior. Cada caso del archivo de datos registra la población en que se encuentra la propiedad, el último asesor que visitó la propiedad, el tiempo transcurrido desde la última evaluación, la valoración realizada en ese momento y el valor de venta de la propiedad.
- **property_assess_cs.sav.** Archivo de datos hipotéticos sobre las iniciativas de un asesor de un estado para mantener actualizada la evaluación de los valores de las propiedades utilizando recursos limitados. Los casos corresponden a propiedades del estado. Cada caso del archivo de datos registra el condado, la población y el vecindario en el que se encuentra la propiedad, el tiempo transcurrido desde la última evaluación y la valoración realizada en ese momento.

- **property_assess_cs_sample.sav** Este archivo de datos hipotéticos contiene una muestra de las propiedades recogidas en *property_assess_cs.sav*. La muestra se tomó en función del diseño especificado en el archivo de plan *property_assess.csplan*, y este archivo de datos registra las probabilidades de inclusión y las ponderaciones muestrales. La variable adicional *Valor actual* se recopiló y añadió al archivo de datos después de tomar la muestra.
- **recidivism.sav**. Archivo de datos hipotéticos sobre las iniciativas de una agencia de orden público para comprender los índices de reincidencia en su área de jurisdicción. Cada caso corresponde a un infractor anterior y registra su información demográfica, algunos detalles de su primer delito y, a continuación, el tiempo transcurrido desde su segundo arresto, si ocurrió en los dos años posteriores al primer arresto.
- **recidivism_cs_sample.sav**. Archivo de datos hipotéticos sobre las iniciativas de una agencia de orden público para comprender los índices de reincidencia en su área de jurisdicción. Cada caso corresponde a un delincuente anterior, puesto en libertad tras su primer arresto durante el mes de junio de 2003 y registra su información demográfica, algunos detalles de su primer delito y los datos de su segundo arresto, si se produjo antes de finales de junio de 2006. Los delincuentes se seleccionaron de una muestra de departamentos según el plan de muestreo especificado en *recidivism_cs.csplan*. Como este plan utiliza un método de probabilidad proporcional al tamaño (PPS), también existe un archivo que contiene las probabilidades de selección conjunta (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav**. Archivo de datos hipotéticos que contiene datos de transacciones de compra, incluida la fecha de compra, los artículos adquiridos y el importe de cada transacción.
- **salesperformance.sav**. Archivo de datos hipotéticos sobre la evaluación de dos nuevos cursos de formación de ventas. Sesenta empleados, divididos en tres grupos, reciben formación estándar. Además, el grupo 2 recibe formación técnica; el grupo 3, un tutorial práctico. Cada empleado se sometió a un examen al final del curso de formación y se registró su puntuación. Cada caso del archivo de datos representa a un alumno distinto y registra el grupo al que fue asignado y la puntuación que obtuvo en el examen.
- **satisf.sav**. Archivo de datos hipotéticos sobre una encuesta de satisfacción llevada a cabo por una empresa minorista en cuatro tiendas. Se encuestó a 582 clientes en total y cada caso representa las respuestas de un único cliente.
- **screws.sav** Este archivo de datos contiene información acerca de las características de tornillos, pernos, clavos y tacos (Hartigan, 1975).
- **shampoo_ph.sav**. Archivo de datos hipotéticos sobre el control de calidad en una fábrica de productos para el cabello. Se midieron seis lotes de resultados distintos en intervalos regulares y se registró su pH. El intervalo objetivo es de 4,5 a 5,5.
- **ships.sav**. Un conjunto de datos presentados y analizados en otro lugar (McCullagh et al., 1989) sobre los daños en los cargueros producidos por las olas. Los recuentos de incidentes se pueden modelar como si ocurrieran con una tasa de Poisson dado el tipo de barco, el período de construcción y el período de servicio. Los meses de servicio agregados para cada casilla de la tabla formados por la clasificación cruzada de factores proporcionan valores para la exposición al riesgo.
- **site.sav**. Archivo de datos hipotéticos sobre las iniciativas de una compañía para seleccionar sitios nuevos para sus negocios en expansión. Se ha contratado a dos consultores para evaluar los sitios de forma independiente, quienes, además de un informe completo, han resumido cada sitio como una posibilidad “buena”, “media” o “baja”.

- **smokers.sav.** Este archivo de datos es un resumen de la encuesta sobre toxicomanía 1998 National Household Survey of Drug Abuse y es una muestra de probabilidad de unidades familiares americanas. (<http://dx.doi.org/10.3886/ICPSR02934>) Así, el primer paso de un análisis de este archivo de datos debe ser ponderar los datos para reflejar las tendencias de población.
- **stroke_clean.sav.** Este archivo de datos hipotéticos contiene el estado de una base de datos médica después de haberla limpiado mediante los procedimientos de la opción Preparación de datos.
- **stroke_invalid.sav.** Este archivo de datos hipotéticos contiene el estado inicial de una base de datos médica que incluye contiene varios errores de entrada de datos.
- **stroke_survival.** Este archivo de datos hipotéticos registra los tiempos de supervivencia de los pacientes que finalizan un programa de rehabilitación tras un ataque isquémico. Tras el ataque, la ocurrencia de infarto de miocardio, ataque isquémico o ataque hemorrágico se anotan junto con el momento en el que se produce el evento registrado. La muestra está truncada a la izquierda ya que únicamente incluye a los pacientes que han sobrevivido al final del programa de rehabilitación administrado tras el ataque.
- **stroke_valid.sav.** Este archivo de datos hipotéticos contiene el estado de una base de datos médica después de haber comprobado los valores mediante el procedimiento Validar datos. Sigue conteniendo casos potencialmente anómalos.
- **survey_sample.sav.** Este archivo de datos contiene datos de encuestas, incluyendo datos demográficos y diferentes medidas de actitud. Se basa en un subconjunto de variables de NORC General Social Survey de 1998, aunque algunos valores de datos se han modificado y que existen variables ficticias adicionales se han añadido para demostraciones.
- **telco.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía de telecomunicaciones para reducir el abandono de clientes en su base de clientes. Cada caso corresponde a un cliente distinto y registra diversa información demográfica y de uso del servicio.
- **telco_extra.sav.** Este archivo de datos es similar al archivo de datos *telco.sav*, pero las variables de meses con servicio y gasto de clientes transformadas logarítmicamente se han eliminado y sustituido por variables de gasto del cliente transformadas logarítmicamente tipificadas.
- **telco_missing.sav.** Este archivo de datos es un subconjunto del archivo de datos *telco.sav*, pero algunos valores de datos demográficos se han sustituido con valores perdidos.
- **testmarket.sav.** Archivo de datos hipotéticos sobre los planes de una cadena de comida rápida para añadir un nuevo artículo a su menú. Hay tres campañas posibles para promocionar el nuevo producto, por lo que el artículo se presenta en ubicaciones de varios mercados seleccionados aleatoriamente. Se utiliza una promoción diferente en cada ubicación y se registran las ventas semanales del nuevo artículo durante las primeras cuatro semanas. Cada caso corresponde a una ubicación semanal diferente.
- **testmarket_1month.sav.** Este archivo de datos hipotéticos es el archivo de datos *testmarket.sav* con las ventas semanales “acumuladas” para que cada caso corresponda a una ubicación diferente. Como resultado, algunas de las variables que cambiaban semanalmente desaparecen y las ventas registradas se convierten en la suma de las ventas realizadas durante las cuatro semanas del estudio.
- **tree_car.sav.** Archivo de datos hipotéticos que contiene datos demográficos y de precios de compra de vehículos.

- **tree_credit.sav** Archivo de datos hipotéticos que contiene datos demográficos y de historial de créditos bancarios.
- **tree_missing_data.sav** Archivo de datos hipotéticos que contiene datos demográficos y de historial de créditos bancarios con un elevado número de valores perdidos.
- **tree_score_car.sav.** Archivo de datos hipotéticos que contiene datos demográficos y de precios de compra de vehículos.
- **tree_textdata.sav.** Archivo de datos sencillos con dos variables diseñadas principalmente para mostrar el estado por defecto de las variables antes de realizar la asignación de nivel de medida y etiquetas de valor.
- **tv-survey.sav.** Archivo de datos hipotéticos sobre una encuesta dirigida por un estudio de TV que está considerando la posibilidad de ampliar la emisión de un programa de éxito. Se preguntó a 906 encuestados si verían el programa en distintas condiciones. Cada fila representa un encuestado diferente; cada columna es una condición diferente.
- **ulcer_recurrence.sav.** Este archivo contiene información parcial de un estudio diseñado para comparar la eficacia de dos tratamientos para prevenir la reaparición de úlceras. Constituye un buen ejemplo de datos censurados por intervalos y se ha presentado y analizado en otro lugar (Collett, 2003).
- **ulcer_recurrence_recoded.sav.** Este archivo reorganiza la información de *ulcer_recurrence.sav* para permitir modelar la probabilidad de eventos de cada intervalo del estudio en lugar de sólo la probabilidad de eventos al final del estudio. Se ha presentado y analizado en otro lugar (Collett et al., 2003).
- **verd1985.sav.** Archivo de datos sobre una encuesta (Verdegaal, 1985). Se han registrado las respuestas de 15 sujetos a 8 variables. Se han dividido las variables de interés en tres grupos. El conjunto 1 incluye *edad* y *ecivil*, el conjunto 2 incluye *mascota* y *noticia*, mientras que el conjunto 3 incluye *música* y *vivir*. Se escala *mascota* como nominal múltiple y *edad* como ordinal; el resto de variables se escalan como nominal simple.
- **virus.sav.** Archivo de datos hipotéticos sobre las iniciativas de un proveedor de servicios de Internet (ISP) para determinar los efectos de un virus en sus redes. Se ha realizado un seguimiento (aproximado) del porcentaje de tráfico de correos electrónicos infectados en sus redes a lo largo del tiempo, desde el momento en que se descubre hasta que la amenaza se contiene.
- **wheeze_steubenville.sav.** Subconjunto de un estudio longitudinal de los efectos sobre la salud de la polución del aire en los niños (Ware, Dockery, Spiro III, Speizer, y Ferris Jr., 1984). Los datos contienen medidas binarias repetidas del estado de las sibilancias en niños de Steubenville, Ohio, con edades de 7, 8, 9 y 10 años, junto con un registro fijo de si la madre era fumadora durante el primer año del estudio.
- **workprog.sav.** Archivo de datos hipotéticos sobre un programa de obras del gobierno que intenta colocar a personas desfavorecidas en mejores trabajos. Se siguió una muestra de participantes potenciales del programa, algunos de los cuales se seleccionaron aleatoriamente para entrar en el programa, mientras que otros no siguieron esta selección aleatoria. Cada caso representa un participante del programa diferente.

Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing,

using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



Bibliografía

- Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. Nueva York: Harper & Row.
- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*, 3rd ed. Oxford: Oxford University Press.
- Blake, C. L., y C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- Breiman, L., y J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, .
- Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Fine, T. L. 1999. *Feedforward Neural Network Methodology*, 3rd ed. Nueva York: Springer-Verlag.
- Green, P. E., y V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.
- Green, P. E., y Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, .
- Hartigan, J. A. 1975. *Clustering algorithms*. Nueva York: John Wiley and Sons.
- Hastie, T., y R. Tibshirani. 1990. *Generalized additive models*. Londres: Chapman and Hall.
- Haykin, S. 1998. *Neural Networks: A Comprehensive Foundation*, 2nd ed. Nueva York: Macmillan College Publishing.
- Kennedy, R., C. Riquier, y B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, .
- McCullagh, P., y J. A. Nelder. 1989. *Modelos lineales generalizados*, 2nd ed. Londres: Chapman & Hall.
- Price, R. H., y D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .
- Rickman, R., N. Mitchell, J. Dingman, y J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .
- Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
- Rosenberg, S., y M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .
- Tao, K. K. 1993. A closer look at the radial basis function (RBF) networks. En: *Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems, and Computers*, A. Singh, ed. Los Alamitos, Calif.: IEEE Comput. Soc. Press.

Uykan, Z., C. Guzelis, M. E. Celebi, y H. N. Koivo. 2000. Analysis of input-output clustering for determining centers of RBFN. *IEEE Transactions on Neural Networks*, 11, .

Van der Ham, T., J. J. Meulman, D. C. Van Strien, y H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .

Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (en neerlandés)*. Leiden: Department of Data Theory, University of Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, y B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .

Índice

- advertencias
 - en perceptrón multicapa, 62
- algo
 - en función de base radial, 72
- archivos de ejemplo
 - posición, 86
- arquitectura
 - redes neuronales, 2
- arquitectura de red
 - en función de base radial, 27
 - en perceptrón multicapa, 10

- capa de salida
 - en función de base radial, 27
 - en perceptrón multicapa, 10
- capa oculta
 - en función de base radial, 27
 - en perceptrón multicapa, 10
- clasificación
 - en función de base radial, 79
 - en perceptrón multicapa, 43, 47
- curva COR
 - en función de base radial, 82
 - en perceptrón multicapa, 48
- Curva COR
 - en función de base radial, 29
 - en perceptrón multicapa, 15

- diagrama de red
 - en función de base radial, 29
 - en perceptrón multicapa, 15

- entrenamiento de red
 - en perceptrón multicapa, 13
- entrenamiento en línea
 - en perceptrón multicapa, 13
- entrenamiento por lotes
 - en perceptrón multicapa, 13
- entrenamiento por mini lotes
 - en perceptrón multicapa, 13

- función de activación
 - en función de base radial, 27
 - en perceptrón multicapa, 10
- Función de base radial, 22, 72
 - algo, 72
 - arquitectura de red, 27
 - clasificación, 79
 - curva COR, 82
 - exportación del modelo, 33
 - gráfico de elevación, 83
 - gráfico de ganancias acumuladas, 83
 - gráfico de pronosticados por observados, 80
 - guardar variables en el conjunto de datos activo, 31
 - información de red, 78
 - opciones, 34
 - particiones, 26
 - resumen del modelo, 79
 - resumen del procesamiento de los casos, 77
 - salida, 29

- gráfico de elevación
 - en función de base radial, 29, 83
 - en perceptrón multicapa, 15, 51
- gráfico de ganancias
 - en función de base radial, 29
 - en perceptrón multicapa, 15
- gráfico de ganancias acumuladas
 - en función de base radial, 83
 - en perceptrón multicapa, 51
- gráfico de pronosticados por observados
 - en función de base radial, 80

- importancia
 - en perceptrón multicapa, 53, 70
- información de red
 - en función de base radial, 78
 - en perceptrón multicapa, 42, 46, 64

- legal notices, 96

- muestra de comprobación
 - en función de base radial, 26
 - en perceptrón multicapa, 9
- muestra de entrenamiento
 - en función de base radial, 26
 - en perceptrón multicapa, 9
- muestra reservada
 - en función de base radial, 26
 - en perceptrón multicapa, 9

- Perceptrón multicapa, 4, 36
 - advertencias, 62
 - arquitectura de red, 10
 - clasificación, 43, 47
 - curva COR, 48
 - entrenamiento, 13
 - exportación del modelo, 19
 - gráfico de elevación, 51

-
- gráfico de ganancias acumuladas, 51
 - gráfico de pronosticados por observados, 49, 66
 - gráfico de residuos por pronosticados, 68
 - guardar variables en el conjunto de datos activo, 18
 - importancia de la variable independiente, 53, 70
 - información de red, 42, 46, 64
 - opciones, 20
 - particiones, 9
 - resumen del modelo, 42, 47, 65
 - resumen del procesamiento de los casos, 41, 46, 63
 - salida, 15
 - sobreentrenamiento, 44
 - variable de partición, 37
-
- redes neuronales
 - arquitectura, 2
 - definición, 1
 - reglas de parada
 - en perceptrón multicapa, 20
 - resumen del procesamiento de los casos
 - en función de base radial, 77
 - en perceptrón multicapa, 41, 46, 63
-
- sobreentrenamiento
 - en perceptrón multicapa, 44
-
- trademarks, 97
-
- valores perdidos
 - en perceptrón multicapa, 20
 - variable de partición
 - en perceptrón multicapa, 37