

IBM SPSS Neural Networks 23

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, asegúrese de leer la información que encontrará en el apartado "Avisos" en la página 21.

Información sobre el producto

Esta edición se aplica a la versión 23, release 0, modificación 0 de IBM® SPSS Statistics y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

Contenido

Capítulo 1. Introducción a las redes

neuronales 1

¿Qué es una red neuronal? 1

Estructura de red neuronal 2

Capítulo 2. Perceptrón multicapa 5

Particiones 7

Arquitectura 8

Entrenamiento 10

Salida 11

Guardar 12

Exportar 13

Opciones 13

Capítulo 3. Función de base radial 15

Particiones. 17

Arquitectura 18

Resultado 18

Guardar 19

Exportar 20

Opciones 20

Avisos 21

Marcas comerciales 23

Índice 25

Capítulo 1. 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 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 ¹:

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 ².

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.

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

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

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 mediciones que minimiza el error al pronosticar la posibilidad de mora. La figura siguiente resulta muy útil para ilustrar la forma de esta función.

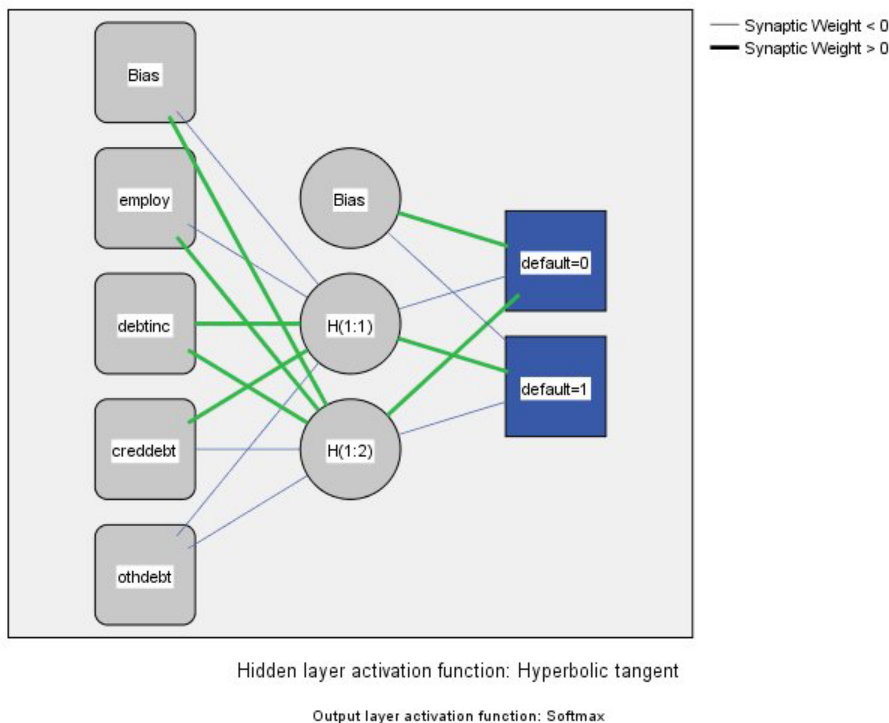


Figura 1. Arquitectura Feedforward con una capa oculta

Esta estructura se denomina **arquitectura feedforward** porque las conexiones de la red fluyen unidimensionalmente 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.

Capítulo 2. 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.

Consideraciones de los datos

Variables dependientes. Las variables dependientes pueden ser:

- *Nominal.* Una variable puede ser tratada como nominal cuando sus valores representan categorías que no obedecen a una clasificación intrínseca. Por ejemplo, el departamento de la compañía en el que trabaja un empleado. Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- *Ordinal.* Una variable puede ser tratada como ordinal cuando sus valores representan categorías con alguna clasificación intrínseca. Por ejemplo, los niveles de satisfacción con un servicio, que abarquen 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.
- *Escalas.* 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 medición adecuado a todas las variables dependientes. No obstante, puede cambiar temporalmente el nivel de medición para una variable pulsando con el botón derecho en la variable en la lista de variables de origen y seleccionar un nivel de medición en el menú emergente.

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

Tabla 1. Iconos de nivel de medición












	Numérico	Cadena	Fecha	Hora
Escala (Continuo)		n/a		

Tabla 1. Iconos de nivel de medición (continuación)

	Numérico	Cadena	Fecha	Hora
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" en la página 7). 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. De forma predeterminada, 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 en base a los datos de entrenamiento, incluso si se define una muestra reservada o de comprobación (consulte "Particiones" en la página 7). Es decir, dependiendo del tipo de cambio de escala, la media, la desviación estándar, 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.

- **Orden de casos.** Los métodos de entrenamiento en línea y por mini lotes (“Entrenamiento” en la página 10) 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

En los menús, elija:

Analizar > Redes neuronales > Perceptrón multicapa...

1. Seleccione al menos una variable dependiente.
2. 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 estándar, $(x-\text{media})/s$.
- **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 ajustada de la resta del mínimo y división por el rango, $[2*(x-\text{mín})/(\text{máx}-\text{mín})]-1$. Los valores normalizados ajustados quedan comprendidos entre -1 y 1.
- **Ninguno.** 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.

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

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 "Replicación de los resultados" en el tema principal Perceptrón multicapa.

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 predeterminadas 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 de 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 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 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 de 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 forma: $\gamma(c) = c$. Acepta argumentos con 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 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 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 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 estándar, $(x-\text{media})/s$.
- **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 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 correcta es $[x-(\text{mín}-\epsilon)]/[(\text{máx}+\epsilon)-(\text{mín}-\epsilon)]$. Especifique un número mayor o igual que 0.
- **Normalizados corregidos.** Versión ajustada de la resta del mínimo y división por el rango, $[2*(x-\text{mín})/(\text{máx}-\text{mín})]-1$. Los valores ajustados normalizados 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 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

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 a almacenar 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. Simulated annealing is used to break out of a local minimum, with the goal of finding the global minimum, during application of the optimization

algorithm. 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 $\beta = (1/p K) * \ln(\eta_0/\eta_{low})$, 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

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 predicciones incorrectas, 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 predicciones incorrectas se muestran dependiendo de los niveles de medición de la variable dependiente. Si alguna variable dependiente tiene un nivel de medición 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 predicciones incorrectas. También se muestran los errores relativos o porcentajes de predicciones incorrectas 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 predichos por observación.** 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 clúster. 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 sensibilidad, 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 sensibilidad supone un proceso de cálculo muy extenso que puede tardar bastante si el número de predictores o de casos es elevado.

Guardar

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

- **Guardar categoría o valor predicho para cada variable dependiente.** Esta opción guarda el valor predicho 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 predicho para cada categoría, donde cada valor predicho es la probabilidad de que el caso pertenezca a la categoría.

Las variables dependientes categóricas con error de sumas cuadráticas tendrá un valor predicho 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" en la página 11) se crean en base a 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

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

Valores perdidos del usuario. Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores perdidos del 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. Especifique 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.

Capítulo 3. 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.

Consideraciones de los datos

Variables dependientes. Las variables dependientes pueden ser:

- *Nominal.* Una variable puede ser tratada como nominal cuando sus valores representan categorías que no obedecen a una clasificación intrínseca. Por ejemplo, el departamento de la compañía en el que trabaja un empleado. Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- *Ordinal.* Una variable puede ser tratada como ordinal cuando sus valores representan categorías con alguna clasificación intrínseca. Por ejemplo, los niveles de satisfacción con un servicio, que abarquen 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.
- *Escalas.* 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 medición adecuado se ha asignado a todas las variables dependientes, aunque puede cambiar temporalmente el nivel de medición para una variable si pulsa el botón derecho la variable en la lista de variables de origen y selecciona un nivel de medición en el menú emergente.

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

Tabla 2. Iconos de nivel de medición

	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" en la página 17). 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. De forma predeterminada, 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 en base a los datos de entrenamiento, incluso si se define una muestra reservada o de comprobación (consulte "Particiones" en la página 17). Es decir, dependiendo del tipo de cambio de escala, la media, la desviación estándar, 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 para obtener instrucciones específicas.
- **Orden de casos.** Los resultados dependen también del orden de los datos porque se utiliza el algoritmo de agrupación en clústeres 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

En los menús, elija:

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

1. Seleccione al menos una variable dependiente.
2. 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 estándar, $(x-\text{media})/s$.

- **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 ajustada de la resta del mínimo y división por el rango, $[2*(x-\text{mín})/(\text{máx}-\text{mín})]-1$. Los valores normalizados ajustados quedan comprendidos entre -1 y 1.
- **Ninguno.** 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.

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

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

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

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 predicciones incorrectas y el tiempo de entrenamiento.

El error es el error de sumas cuadráticas. Además, se muestran los errores relativos o porcentajes de predicciones incorrectas dependiendo de los niveles de medición de la variable dependiente. Si alguna variable dependiente tiene un nivel de medición 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 predicciones incorrectas. También se muestran los errores relativos o porcentajes de predicciones incorrectas 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 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 predichos por observación.** 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 clúster. 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 sensibilidad, 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 sensibilidad supone un proceso de cálculo muy extenso que puede tardar bastante si el número de predictores o de casos es elevado.

Guardar

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

- **Guardar categoría o valor predicho para cada variable dependiente.** Esta opción guarda el valor predicho para variables dependientes de escala y la categoría pronosticada para variables dependientes categóricas.

- **Guardar pseudoprobabilidad predicha 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” en la página 18) se crean en base a 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

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

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

Avisos

Esta información se ha escrito para los productos y servicios ofrecidos en EE.UU.

Es posible que IBM no ofrezca los productos, servicios o características que se tratan en este documento en otros países. Póngase en contacto con el representante local de IBM si desea obtener información sobre los productos y servicios disponibles en su zona. Cualquier referencia a un producto, programa o servicio de IBM no pretende afirmar ni implicar que solamente se pueda utilizar ese producto, programa o servicio de IBM. En su lugar, puede utilizarse cualquier producto, programa o servicio equivalente que no vulnere los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patentes pendientes que cubran la materia descrita en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing
IBM National Language Solution Center
North Castle Drive
Armonk, NY 10504-1785
EE.UU.

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país donde estas disposiciones sean incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO PERO NO LIMITÁNDOSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se aplica a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede efectuar mejoras y/o cambios en los productos y/o programas descritos en esta publicación en cualquier momento y sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales de esos sitios web no forman parte de los materiales de este producto de IBM y la utilización de esos sitios web será responsabilidad del usuario.

IBM puede utilizar o distribuir la información que el usuario le suministre en el modo que considere apropiado sin incurrir en ninguna obligación con el usuario.

Los propietarios de licencia de este programa que deseen tener información sobre el mismo con el fin de: (i) intercambiar información entre programas creados de forma independiente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deberán ponerse en contacto con:

Tel. 900 100 400
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
Estados Unidos

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

IBM proporciona el programa bajo licencia descrito en este documento y todo el material con licencia disponible para el mismo bajo los términos del Acuerdo de cliente de IBM, el Acuerdo de licencia de programa internacional de IBM o cualquier acuerdo equivalente entre las dos partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas mediciones se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas mediciones sean las mismas en los sistemas comerciales. Además, es posible que algunas mediciones hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado esos productos y no puede confirmar la precisión del rendimiento, la compatibilidad ni ninguna otra afirmación relacionada con los productos no IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

LICENCIA DE COPYRIGHT:

Esta información contiene programas de aplicación de muestra en lenguaje fuente, que ilustran técnicas de programación en las distintas plataformas operativas. Puede copiar, modificar y distribuir estos programas de muestra de cualquier forma, sin pagar nada a IBM, con los fines de desarrollar, utilizar, comercializar o distribuir programas de aplicación de acuerdo con la interfaz de programación de aplicaciones para la plataforma operativa para la cual se han escrito los programas de muestra. Estos ejemplos no se han probado de forma exhaustiva bajo todas las condiciones. Por consiguiente, IBM no puede garantizar ni confirmar la fiabilidad, capacidad de servicio o funcionamiento de estos programas. Los programas de muestra se proporcionan "TAL CUAL" sin garantías de ningún tipo. IBM no se hará responsable de los daños que surjan por el uso de los programas de muestra.

Cada copia de cualquier parte de estos programas de ejemplo o de cualquier trabajo que derive de éstos, debe incluir un aviso de copyright, tal como se indica a continuación:

© (nombre de la empresa) (año). Algunas partes de este código se derivan de programas de ejemplo de IBM Corp.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas registradas o marcas comerciales registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de producto y servicio podrían ser marcas registradas de IBM u otras compañías. Encontrará una lista actual de marcas registradas de IBM en la Web en "Información de copyright y marca registrada" en www.ibm.com/legal/copytrade.shtml.

Adobe, el logotipo Adobe, PostScript y el logotipo PostScript son marcas registradas o marcas comerciales de Adobe Systems Incorporated en Estados Unidos y/o otros países.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas registradas y logotipos basados en Java son marcas registradas de Oracle o sus filiales.

Índice

A

- arquitectura
 - redes neuronales 2
- arquitectura de red
 - en función de base radial 18
 - en perceptrón multicapa 8

C

- capa de salida
 - en función de base radial 18
 - en perceptrón multicapa 8
- capa oculta
 - en función de base radial 18
 - en perceptrón multicapa 8
- Curva COR
 - en función de base radial 18
 - en perceptrón multicapa 11

D

- diagrama de red
 - en función de base radial 18
 - en perceptrón multicapa 11

E

- entrenamiento de red
 - en perceptrón multicapa 10
- entrenamiento en línea
 - en perceptrón multicapa 10
- entrenamiento por lotes
 - en perceptrón multicapa 10
- entrenamiento por mini lotes
 - en perceptrón multicapa 10

F

- función de activación
 - en función de base radial 18
 - en perceptrón multicapa 8
- Función de base radial 15
 - arquitectura de red 18
 - exportación del modelo 20
 - guardar variables en el conjunto de datos activo 19
 - opciones 20
 - particiones 17
 - resultado 18

G

- gráfico de elevación
 - en función de base radial 18
 - en perceptrón multicapa 11
- gráfico de ganancias
 - en función de base radial 18
 - en perceptrón multicapa 11

M

- muestra de comprobación
 - en función de base radial 17
 - en perceptrón multicapa 7
- muestra de entrenamiento
 - en función de base radial 17
 - en perceptrón multicapa 7
- muestra reservada
 - en función de base radial 17
 - en perceptrón multicapa 7

P

- Perceptrón multicapa 5
 - arquitectura de red 8
 - entrenamiento 10
 - exportación del modelo 13
 - guardar variables en el conjunto de datos activo 12
 - opciones 13
 - particiones 7
 - resultado 11

R

- redes neuronales
 - arquitectura 2
- reglas de parada
 - en perceptrón multicapa 13

V

- valores perdidos
 - en perceptrón multicapa 13



Impreso en España