

*IBM SPSS Neural Networks 32*

**IBM**

**Nota**

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in [“Informazioni particolari” a pagina 17.](#)

**Informazioni sul prodotto**

Questa edizione si applica alla versione 32, release 0, modifica 0 di IBM® SPSS Statistica e a tutte le release e modifiche successive se non diversamente indicato in nuove edizioni.

© **Copyright International Business Machines Corporation .**

---

# Indice

<b>Capitolo 1. Reti neurali.....</b>	<b>1</b>
Introduzione alle reti neurali.....	1
Cosa è una rete neurale?.....	1
Struttura di rete neurale.....	2
Percettore multistrato.....	3
Partizioni .....	5
Architettura .....	6
Formazione .....	7
Output .....	8
Salva .....	9
Export .....	10
Opzioni .....	10
Funzione di base radiale.....	11
Partizioni .....	13
Architettura .....	14
Output .....	14
Salva .....	15
Export .....	16
Opzioni .....	16
<b>Informazioni particolari.....</b>	<b>17</b>
Marchi.....	18
<b>Indice analitico.....</b>	<b>21</b>



---

# Capitolo 1. Reti neurali

Le seguenti funzioni di rete neurale sono incluse in SPSS Statistiche Premium Edition o l'opzione Neural Networks.

---

## Introduzione alle reti neurali

Le reti neurali sono lo strumento preferito per molte applicazioni di data mining predittive a causa della loro potenza, flessibilità e facilità d'uso. Le reti neurali predittive sono particolarmente utili nelle applicazioni in cui il processo sottostante è complesso, come ad esempio:

- Previsione della domanda dei consumatori per snellare i costi di produzione e di consegna.
- Previsione della probabilità di risposta al marketing di posta diretta per determinare quali famiglie su una mailing list deve essere inviata un'offerta.
- Punteggio di un richiedente per determinare il rischio di estendere il credito al richiedente.
- Rilevare le transazioni fraudolente in un database di crediti assicurativi.

Le reti neurali utilizzate in applicazioni predittive, come le reti multistrato di percettore (MLP) e le reti di base radiale (RBF), sono supervisionate nel senso che i risultati previsti dal modello possono essere confrontati contro valori noti delle variabili target. L'opzione Neural Networks consente di adattare le reti MLP e RBF e salvare i modelli risultanti per il punteggio.

## Cosa è una rete neurale?

Il termine **rete neurale** si applica a una famiglia di modelli lodati, caratterizzato da un ampio spazio di parametri e dalla struttura flessibile, discendendo da studi di funzionamento cerebrale. Mentre la famiglia crebbe, la maggior parte dei nuovi modelli furono progettati per applicazioni non biologiche, anche se gran parte della terminologia associata riflette la sua origine.

Definizioni specifiche delle reti neurali sono variegatae quanto i campi in cui vengono utilizzati. Mentre nessuna definizione unica copre correttamente l'intera famiglia di modelli, per ora, considera la seguente descrizione<sup>1</sup>:

Una rete neurale è un processore distribuito in modo massiccio e parallelo che ha una naturale propensione per la memorizzazione delle conoscenze esperienziali e renderla disponibile all'uso. Assomiglia al cervello sotto due aspetti:

- La conoscenza viene acquisita dalla rete attraverso un processo di apprendimento.
- I punti di forza di connessione Interneuron noti come pesi sinaptici sono utilizzati per memorizzare le conoscenze.

Per una discussione sul perché questa definizione è forse troppo restrittiva, cfr.<sup>2</sup>.

Al fine di differenziare le reti neurali dai metodi statistici tradizionali utilizzando questa definizione, ciò che è *non* detto è altrettanto significativo quanto il testo effettivo della definizione. Ad esempio, il modello di regressione lineare tradizionale può acquisire conoscenze attraverso il metodo dei minimi quadrati e memorizzare tale conoscenza nei coefficienti di regressione. In questo senso, si tratta di una rete neurale. Infatti si può argomentare che la regressione lineare è un caso particolare di alcune reti neurali. Tuttavia, la regressione lineare ha una struttura modello rigida e serie di assunzioni che vengono imposte prima di imparare dai dati.

Al contrario, la definizione sopra rende minime le richieste sulla struttura del modello e le ipotesi. Così una rete neurale può approssimare una vasta gamma di modelli statistici senza richiedere di ipotizzare in anticipo determinate relazioni tra le variabili dipendenti e indipendenti. Invece, la forma delle relazioni

---

<sup>1</sup> S. Haykin 1998. *Reti neurali: una guida completa*, 2a ed. New York: Macmillan College Publishing.

<sup>2</sup> B. D. Ripley 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

è determinata durante il processo di apprendimento. Se è opportuno un rapporto lineare tra le variabili dipendenti e indipendenti, i risultati della rete neurale dovrebbero approssimare strettamente quelli del modello di regressione lineare. Se una relazione non lineare risulta più appropriata, la rete neurale approssima automaticamente la struttura "corretta" del modello.

Il trade-off per questa flessibilità è che i pesi sinaptici di una rete neurale non sono facilmente interpretabili. Così, se si cerca di spiegare un processo sottostante che produce le relazioni tra le variabili dipendenti e indipendenti, sarebbe meglio utilizzare un modello statistico più tradizionale. Tuttavia, se l'interpretazione del modello non è importante, è spesso possibile ottenere buoni risultati modello più rapidamente utilizzando una rete neurale.

## Struttura di rete neurale

Sebbene le reti neurali impongano richieste minimali sulla struttura del modello e sulle ipotesi, è utile comprendere la generale **architettura di rete**. La rete multistrato percettore (MLP) o funzione di base radiale (RBF) è una funzione di predittori (chiamati anche input o variabili indipendenti) che minimizzano l'errore di previsione delle variabili target (chiamate anche outputs).

Si consideri il dataset *bankloan.sav* fornito con il prodotto, in cui si desidera poter identificare i possibili inadempienti tra un pool di richiedenti di prestiti. Una rete MLP o RBF applicata a questo problema è una funzione delle misurazioni che minimizzano l'errore nella previsione di default. La seguente figura è utile per correlare la forma di questa funzione.

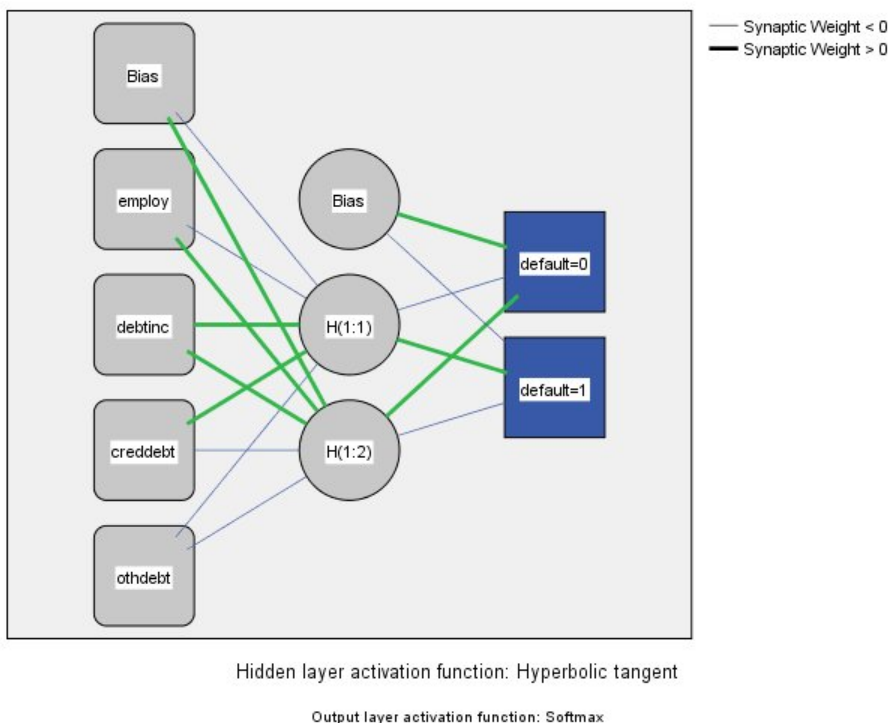


Figura 1. Architettura feedforward con uno strato nascosto

Questa struttura è nota come **architettura feedforward** perché le connessioni nel flusso di rete in avanti dallo strato di input allo strato di output senza alcun loop di feedback. In questa figura:

- Lo **strato di input** contiene i predittori.
- Lo **strato nascosto** contiene nodi non osservabili, o unità. Il valore di ogni unità nascosta è una qualche funzione dei predittori; la forma esatta della funzione dipende in parte dal tipo di rete e in parte dalle specifiche controllabili dall'utente.
- Lo **strato di output** contiene le risposte. Poiché la cronologia di default è una variabile categoriale con due categorie, viene recuperata come due variabili indicatore. Ogni unità di uscita è una qualche

funzione delle unità nascoste. Anche in questo caso, la forma esatta della funzione dipende in parte dal tipo di rete e in parte sulle specifiche controllabili dall'utente.

La rete MLP consente un secondo strato nascosto; in quel caso ogni unità del secondo strato nascosto è funzione delle unità nel primo strato nascosto e ogni risposta è funzione delle unità nel secondo strato nascosto.

## Percettore multistrato

La Procedura MLP (Multilayer Perceptron) produce un modello predittivo per una o più variabili dipendenti (target) in base ai valori delle variabili predittive.

**Esempi.** Di seguito due scenari che utilizzano la procedura MLP:

Un funzionario di prestito presso una banca deve essere in grado di individuare caratteristiche indicanti le persone che rischiano di default sui prestiti e di utilizzare quelle caratteristiche per individuare buoni e cattivi rischi di credito. Utilizzando un campione di clienti passati, può addestrare un percettore multistrato, convalidare l'analisi utilizzando un campione holdout di clienti passati, e quindi utilizzare la rete per classificare i clienti prospettivi come buoni o cattivi rischi di credito.

Un sistema ospedaliero è interessato a tracciare costi e lunghezze di soggiorno per i pazienti ricoverati per il trattamento di infarto del miocardio (MI, o "attacco cardiaco"). Ottenere stime accurate di queste misure consente all'amministrazione di gestire correttamente lo spazio letto disponibile in quanto i pazienti sono trattati. Utilizzando i record di trattamento di un campione di pazienti che hanno ricevuto il trattamento per MI, l'amministratore può addestrare una rete per prevedere sia il costo che la durata del soggiorno.

Considerazioni sui dati

**Variabili dipendenti.** Le variabili dipendenti possono essere:

- *Nominale.* Una variabile può essere considerata nominale quando i suoi valori rappresentano categorie prive di classificazione intrinseca (ad esempio, il dipartimento della società in cui lavora un dipendente). Degli esempi di variabili nominali includono la regione, il codice postale e l'affiliazione religiosa.
- *Ordinale.* Una variabile può essere trattata come ordinale quando i suoi valori rappresentano categorie con una classificazione intrinseca (ad esempio, livelli di soddisfazione del servizio da molto insoddisfatti a molto soddisfatti). Degli esempi di variabili ordinali includono i punteggi di atteggiamento che rappresentano i gradi di soddisfazione o fiducia e i punteggi di classificazione delle preferenze.
- *Scala.* Una variabile può essere trattata come una scala (continua) quando i suoi valori rappresentano categorie ordinate con una metrica significativa, in modo che i confronti di distanza tra i valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

La procedura presuppone che il livello di misurazione appropriato sia stato assegnato a tutte le variabili dipendenti; tuttavia, è possibile modificare temporaneamente il livello di misurazione per una variabile facendo clic con il tasto destro del mouse sulla variabile nell'elenco delle variabili di origine e selezionando un livello di misurazione dal menu a comparsa.

L'icona accanto a ciascuna variabile nell'elenco delle variabili identifica il livello di misurazione e il tipo di dati.









Tabella 1. Icone del livello di misurazione				
	Numerico	Stringa	Dati	Ora
Scala (continuo)		n/d		
Ordinale				

Tabella 1. Icone del livello di misurazione (Continua)				
	Numerico	Stringa	Dati	Ora
Nominale				

**Variabili di previsione.** I predittori possono essere specificati come fattori (categoriali) o covariati (scala).

**Codifiche variabili categoriali.** La procedura ricodifica temporaneamente le variabili dipendenti e indipendenti categoriali utilizzando le codifiche one-of-c per la durata della procedura. Se ci sono le categorie  $c$  di una variabile, allora la variabile viene memorizzata come vettori  $c$ , con la prima categoria denotata  $(1,0, \dots, 0)$ , la categoria successiva  $(0,1,0, \dots, 0)$ , ..., e la categoria finale  $(0,0, \dots, 0, 1)$ .

Questo schema di codifica aumenta il numero di pesi sinaptici e può determinare una formazione più lenta; tuttavia, più metodi di codifica "compatti" di solito portano a reti neurali scarsamente adattate. Se la tua formazione di rete sta procedendo molto lentamente, prova a ridurre il numero di categorie nei tuoi predittori categoriali combinando categorie simili o eliminando i casi che hanno categorie estremamente rare.

Tutta la codifica di un  $-c$  si basa sui dati formativi, anche se viene definito un campione di test o holdout (vedere "Partizioni" a pagina 5). Così, se i campioni di prova o di holdout contengono casi con categorie predittrici che non sono presenti nei dati formativi, allora quei casi non sono utilizzati dalla procedura o dal punteggio. Se i campioni di prova o di holdout contengono casi con categorie variabili dipendenti che non sono presenti nei dati di formazione, allora quei casi non sono utilizzati dalla procedura, ma possono essere segnati.

**Modifica della scala.** Le variabili e le covariate dipendenti in scala vengono ridimensionate di default per migliorare la formazione di rete. Tutte le riscalings vengono eseguite in base ai dati di formazione, anche se viene definito un campione di test o holdout (vedere "Partizioni" a pagina 5). Cioè a seconda del tipo di ridimensionamento, si calcola la media, la deviazione standard, il valore minimo o il valore massimo di una variabile covariata o dipendente utilizzando solo i dati di allenamento. Se si specifica una variabile per definire le partizioni, è importante che queste covariate o variabili dipendenti abbiano distribuzioni simili attraverso i campioni di allenamento, testing e holdout.

**Peso della frequenza.** I pesi della frequenza vengono ignorati da questa procedura.

**Replica risultati.** Se si desidera replicare esattamente i propri risultati, utilizzare lo stesso valore di inizializzazione per il generatore di numeri casuali, lo stesso ordine di dati e lo stesso ordine variabile, oltre ad utilizzare le stesse impostazioni di procedura. Ulteriori dettagli su questo argomento seguono:

- **Generazione numero casuale.** La procedura utilizza la generazione di numeri casuali durante l'assegnazione casuale di partizioni, il sottocampionamento casuale per l'inizializzazione dei pesi sinaptici, il sottocampionamento casuale per la selezione automatica dell'architettura e l'algoritmo di annealing simulato utilizzato nell'inizializzazione del peso e nella selezione automatica dell'architettura. Per riprodurre gli stessi risultati randomizzati in futuro, utilizzare lo stesso valore di inizializzazione per il generatore di numeri casuali prima di ogni esecuzione della procedura Multilayer Percezione.
- **Ordine dei casi.** I metodi di allenamento online e Mini - batch (vedi "Formazione" a pagina 7) dipendono esplicitamente dall'ordine dei casi; tuttavia, anche la formazione Batch dipende dall'ordine dei casi in quanto l'inizializzazione dei pesi sinaptici comporta il sottocampionamento dal dataset.

Per ridurre al minimo gli effetti dell'ordine, disporre i casi in ordine casuale. Per verificare la stabilità di una determinata soluzione, è possibile ottenere diverse soluzioni con casi ordinati in diversi ordini casuali. Nelle situazioni con dimensioni file estremamente grandi, le esecuzioni multiple possono essere eseguite con un campione di casi ordinati in diversi ordini casuali.

- **Ordine variabile.** I risultati possono essere influenzati dall'ordine delle variabili nelle liste dei fattori e delle covariate a causa del diverso schema dei valori iniziali assegnati quando viene modificato l'ordine variabile. Come per gli effetti dell'ordine di caso, si potrebbero provare diversi ordini variabili (semplicemente trascinare e rilasciare all'interno delle liste dei fattori e delle covariate) per valutare la stabilità di una determinata soluzione.

Creazione di una rete di Multi Ayer Percezione

Dai menu, scegliere:

**Analizzare > Reti Neurali > Multiayer Percezione ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Selezionare almeno una variabile dipendente.
2. Selezionare almeno un fattore o covariato.

Opzionalmente, sulla scheda Variabili è possibile modificare il metodo per le covariate di ridimensionamento. Le scelte sono:

- **Standardizzato.** Sottrarre la media e dividere per la deviazione standard,  $(x - \text{mean}) / s$ .
- **Normalizzato.** Sottrarre il minimo e dividere per la gamma,  $(x - \text{min}) / (\text{max} - \text{min})$ . I valori normalizzati cadono tra il 0 e il 1.
- **Normalizzato regolato.** Versione regolata di sottrazione del minimo e dividendo per la gamma,  $[2 * (x - \text{min}) / (\text{max} - \text{min})] - 1$ . I valori normalizzati adattati sono compresi tra -1 e 1.
- **Messuno.** Nessun ridimensionamento delle covariate.

Campi con livello di misurazione sconosciuto

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) del dataset è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

**Esegui scansione dati.** Legge i dati del dataset attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con dataset di grandi dimensioni, questa operazione può richiedere del tempo.

**Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Vista variabile dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

## Partizioni

**Dataset di partizione.** Questo gruppo specifica il metodo di partizionamento del dataset attivo in campioni di formazione, verifica e holdout. Il **campione di formazione** comprende i record di dati utilizzati per formare la rete neurale; alcune percentuali di casi nel dataset devono essere assegnate al campione di formazione al fine di ottenere un modello. Il **campione di prova** è una serie indipendente di record di dati utilizzati per tenere traccia degli errori durante l'allenamento al fine di prevenire la sovrapposizione. Si consiglia vivamente di creare un campione di prova e la formazione in rete sarà generalmente più efficiente se il campione di prova è più piccolo rispetto al campione di allenamento. Il **campione holdout** è un'altra serie indipendente di record di dati utilizzati per valutare la rete neurale finale; l'errore per il campione holdout dà una stima "onesta" della capacità predittiva del modello perché i casi holdout non sono stati utilizzati per costruire il modello.

- **Randomly assegna i casi in base al numero relativo di casi.** Specificare il numero relativo (rapporto) dei casi assegnati casualmente ad ogni campione (allenamento, testing e holdout). La colonna % riporta la percentuale di casi che verranno assegnati a ciascun campione in base ai numeri relativi specificati.

Ad esempio, specificando 7, 3, 0 come i numeri relativi per la formazione, i test e i campioni holdout corrisponde a 70%, 30% e 0%. Specificando 2, 1, 1 come i numeri relativi corrisponde a 50%, 25% e 25%; 1, 1, 1 corrisponde a dividere il dataset in pari terzi tra allenamento, testing e holdout.

- **Utilizza variabile di partizionamento per assegnare i casi.** Specificare una variabile numerica che assegna ogni caso nel dataset attivo al campione di formazione, verifica o holdout. I casi con un valore

positivo sulla variabile sono assegnati al campione di formazione, i casi con un valore di 0, al campione di prova, e i casi con un valore negativo, al campione holdout. I casi con un valore mancante di sistema vengono esclusi dall'analisi. I valori mancanti definiti dall'utente per la variabile partizione sono sempre considerati validi.

*Nota:* l'utilizzo di una variabile di partizionamento non garantirà risultati identici nelle successive esecuzioni della procedura. Vedi "Risultati replicanti" nel topic principale [Multilayer Perceptron](#).

## Architettura

La scheda Architettura viene utilizzata per specificare la struttura della rete. La procedura può selezionare automaticamente l'architettura "migliore" oppure è possibile specificare un'architettura personalizzata.

La selezione automatica dell'architettura costruisce una rete con uno strato nascosto. Specificare il numero minimo e massimo di unità consentite nello strato nascosto e la selezione automatica dell'architettura calcola il numero "migliore" di unità nello strato nascosto. La selezione automatica dell'architettura utilizza le funzioni di attivazione predefinite per gli strati nascosti e di uscita.

La selezione di architettura personalizzata ti dà un controllo esperto sugli strati nascosti e di uscita e può essere più utile quando sai in anticipo quale architettura vuoi o quando hai bisogno di tweak i risultati della selezione dell'architettura automatica.

### Livelli nascosti

Lo strato nascosto contiene nodi di rete non osservabili (unità). Ogni unità nascosta è una funzione della somma pesata degli input. La funzione è quella di attivazione e i valori dei pesi vengono determinati dall'algoritmo di stima. Se la rete contiene un secondo strato nascosto, ogni unità nascosta nel secondo strato è una funzione della somma pesata delle unità nel primo strato nascosto. La stessa funzione di attivazione viene utilizzata in entrambi gli strati.

*Numero di strati nascosti.* Un percettore multi-livello può avere uno o due livelli nascosti.

*Funzione di attivazione.* La funzione di attivazione "collega" le somme ponderate delle unità in un livello ai valori delle unità nel livello successivo.

- **tangente Hyperbolica.** Questa funzione ha la forma:  $\gamma(c) = \tanh(c) = (e^c - e^{-c}) / (e^c + e^{-c})$ . Ci vogliono argomenti a valore reale e li trasforma in gamma (-1, 1). Quando viene utilizzata la selezione automatica dell'architettura, questa è la funzione di attivazione per tutte le unità negli strati nascosti.
- **Sigmoide.** Questa funzione ha il formato:  $\gamma(c) = 1 / (1 + e^{-c})$ . Ci vogliono argomenti a valore reale e li trasforma in gamma (0, 1).

*Numero di unità.* Il numero di unità in ogni livello nascosto può essere specificato esplicitamente oppure può essere determinato automaticamente dall'algoritmo di stima.

### Strato di output

Lo strato di output contiene le variabili target (dipendenti).

*Funzione di attivazione.* La funzione di attivazione "collega" le somme ponderate delle unità in un livello ai valori delle unità nel livello successivo.

- **Identità.** Questa funzione ha la forma:  $\gamma(c) = c$ . Ci vogliono argomenti a valore reale e li restituisce invariati. Quando viene utilizzata la selezione automatica dell'architettura, questa è la funzione di attivazione per le unità nello strato di output se ci sono variabili dipendenti in scala.
- **Softmax.** Questa funzione ha la forma:  $\gamma(c_k) = \exp(c_k) / \sum_j \exp(c_j)$ . Prende un vettore di argomenti a valore reale e lo trasforma in un vettore i cui elementi rientrano nella fascia (0, 1) e somma a 1. Softmax è disponibile solo se tutte le variabili dipendenti sono categoriali. Quando viene utilizzata la selezione automatica dell'architettura, questa è la funzione di attivazione per le unità nello strato di output se tutte le variabili dipendenti sono categoriali.
- **tangente Hyperbolica.** Questa funzione ha la forma:  $\gamma(c) = \tanh(c) = (e^c - e^{-c}) / (e^c + e^{-c})$ . Ci vogliono argomenti a valore reale e li trasforma in gamma (-1, 1).
- **Sigmoide.** Questa funzione ha il formato:  $\gamma(c) = 1 / (1 + e^{-c})$ . Ci vogliono argomenti a valore reale e li trasforma in gamma (0, 1).

**Rescaling di variabili dipendenti in scala.** Questi controlli sono disponibili solo se è stata selezionata almeno una variabile dipendente su scala.

- **Standardizzato.** Sottrarre la media e dividere per la deviazione standard,  $(x - \text{mean}) / s$ .
- **Normalizzato.** Sottrarre il minimo e dividere per la gamma,  $(x - \text{min}) / (\text{max} - \text{min})$ . I valori normalizzati cadono tra il 0 e il 1. Questo è il metodo di rescaling richiesto per le variabili dipendenti in scala se lo strato di output utilizza la funzione di attivazione sigmoideale. L'opzione di correzione specifica un piccolo numero  $\epsilon$  che viene applicato come correzione alla formula di rescaling; questa correzione garantisce che tutti i valori variabili dipendenti ridimensionati saranno all'interno dell'intervallo della funzione di attivazione. In particolare, i valori 0 e 1, che si verificano nella formula non corretta quando  $x$  assume il suo valore minimo e massimo, definiscono i limiti della gamma della funzione sigmoide ma non sono all'interno di tale intervallo. La formula corretta è  $[x - (\text{min} - \epsilon)] / [(\text{max} + \epsilon) - (\text{min} - \epsilon)]$ . Specificare un numero maggiore o uguale a 0.
- **Normalizzato regolato.** Versione regolata di sottrazione del minimo e dividendo per la gamma,  $[2 * (x - \text{min}) / (\text{max} - \text{min})] - 1$ . I valori normalizzati adattati sono compresi tra  $-1$  e  $1$ . Questo è il metodo di rescaling richiesto per le variabili dipendenti in scala se lo strato di output utilizza la funzione di attivazione della tangente iperbolica. L'opzione di correzione specifica un piccolo numero  $\epsilon$  che viene applicato come correzione alla formula di rescaling; questa correzione garantisce che tutti i valori variabili dipendenti ridimensionati saranno all'interno dell'intervallo della funzione di attivazione. In particolare, i valori  $-1$  e  $1$ , che si verificano nella formula non corretta quando  $x$  assume il suo valore minimo e massimo, definiscono i limiti della gamma della funzione di tangente iperbolica ma non sono all'interno di tale intervallo. La formula corretta è  $\{2 * [(x - (\text{min} - \epsilon)) / ((\text{max} + \epsilon) - (\text{min} - \epsilon))]\} - 1$ . Specificare un numero maggiore o uguale a 0.
- **Messuno.** Nessuna ridimensionamento di variabili dipendenti in scala.

## Formazione

La scheda Formazione viene utilizzata per specificare come deve essere addestrata la rete. Il tipo di allenamento e l'algoritmo di ottimizzazione determinano quali opzioni di formazione sono disponibili.

**Tipo di formazione.** Il tipo di formazione determina come la rete elabora i record. Seleziona uno dei seguenti tipi di allenamento:

- **Batch.** Aggiorna i pesi sinaptici solo dopo aver superato tutti i record di dati formativi; ovvero, la formazione batch utilizza le informazioni provenienti da tutti i record del dataset di formazione. La formazione batch viene spesso preferita perché minimizza direttamente l'errore totale; tuttavia, la formazione batch potrebbe dover aggiornare i pesi molte volte fino a quando una delle regole di arresto è soddisfatta e quindi potrebbe avere bisogno di molti passaggi di dati. È più utile per i datasetti "minori".
- **On line.** Aggiorna i pesi sinaptici dopo ogni singolo record di dati di allenamento; ovvero, la formazione online utilizza informazioni da un record alla volta. La formazione online ottiene continuamente un record e aggiorna i pesi fino a quando non viene soddisfatta una delle regole di arresto. Se tutti i record vengono utilizzati una volta e nessuna delle regole di arresto è soddisfatta, allora il processo continua riciclando i record di dati. La formazione online è superiore al lotto per i datasetti "più grandi" con i predittori associati; cioè se ci sono molti record e molti input, e i loro valori non sono indipendenti l'uno dall'altro, allora la formazione online può ottenere più rapidamente una risposta ragionevole rispetto alla formazione batch.
- **Mini - lotto.** Divide i record di dati formativi in gruppi di circa uguale dimensione, quindi aggiorna i pesi sinaptici dopo aver superato un gruppo; ovvero, la formazione mini - batch utilizza le informazioni provenienti da un gruppo di record. Poi il processo ricicla il gruppo di dati se necessario. La formazione mini - batch offre un compromesso tra il lotto e la formazione online, e potrebbe essere migliore per i datasetti "a medio raggio". La procedura può determinare automaticamente il numero di record di formazione per mini - batch oppure è possibile specificare un numero intero superiore a 1 e inferiore o uguale al numero massimo di casi da memorizzare in memoria. È possibile impostare il numero massimo di casi da memorizzare in memoria nella scheda Opzioni .

**Algoritmo di ottimizzazione.** Questo è il metodo utilizzato per stimare i pesi sinaptici.

- **gradiente di coniugato Scalato.** Le ipotesi che giustificano l'utilizzo dei metodi di gradimento coniugati si applicano solo ai tipi di allenamento dei lotti, quindi questo metodo non è disponibile per la formazione online o mini - batch.
- **discesa Gradiente.** Questo metodo deve essere utilizzato con formazione online o mini - batch; può essere utilizzato anche con allenamento batch.

**Opzioni di formazione.** Le opzioni di formazione consentono di affinare l'algoritmo di ottimizzazione. Generalmente non sarà necessario modificare queste impostazioni a meno che la rete non si esaurirà in problemi di stima.

Le opzioni di formazione per l'algoritmo gradiente di coniugato scalato includono:

- **Lambda iniziale.** Il valore iniziale del parametro lambda per l'algoritmo di gradiente coniugato scalato. Specificare un numero maggiore di 0 e minore di 0.000001.
- **Sigma iniziale.** Il valore iniziale del parametro sigma per l'algoritmo di gradiente coniugato scalato. Specificare un numero maggiore di 0 e minore di 0.0001.
- **Interval Center e Interval Offset.** L'intervallo centrale ( $a_0$ ) e l'intervallo offset ( $a$ ) definiscono l'intervallo  $[a_0 - a, a_0 + a]$ , in cui i vettori di peso vengono generati casualmente quando viene utilizzata la ricottura simulata. La annealing simulata viene utilizzata per spezzare un minimo locale, con l'obiettivo di trovare il minimo globale, durante l'applicazione dell'algoritmo di ottimizzazione. Questo approccio viene utilizzato nell'inizializzazione del peso e nella selezione automatica dell'architettura. Specificare un numero per il centro di intervallo e un numero maggiore di 0 per l'intervallo di intervallo.

Le opzioni di formazione per l'algoritmo di discesa gradiente includono:

- **Velocità di apprendimento iniziale.** Il valore iniziale del tasso di apprendimento per l'algoritmo di discesa gradiente. Un tasso di apprendimento superiore significa che la rete si allenerà più velocemente, possibilmente a costo di diventare instabile. Specificare un numero maggiore di 0.
- **Limite di apprendimento inferiore.** Il limite inferiore sul tasso di apprendimento per l'algoritmo di discesa gradiente. Questa impostazione si applica solo alla formazione online e mini - batch. Specificare un numero superiore a 0 e inferiore al tasso di apprendimento iniziale.
- **Slancio.** Il parametro iniziale di slancio per l'algoritmo di discesa gradiente. Il termine slancio aiuta a prevenire le instabilità causate da un tasso di apprendimento too-alto. Specificare un numero maggiore di 0.
- **riduzione della velocità di apprendimento, in Epochs.** Il numero di epoche ( $p$ ), o i passaggi di dati del campione di formazione, necessario per ridurre il tasso di apprendimento iniziale al limite inferiore del tasso di apprendimento quando si utilizza una discesa di gradiente con formazione online o mini - batch. Ciò fornisce il controllo del fattore di decadimento del tasso di apprendimento  $\beta = (1 / p K)^{\ln(\eta_0 / \eta_{\text{basso}})}$ , dove  $\eta_0$  è il tasso di apprendimento iniziale,  $\eta_{\text{basso}}$  è il limite inferiore del tasso di apprendimento e  $K$  è il numero totale di mini - batch (o il numero di record di addestramento per il training online) nel dataset di addestramento. Specificare un intero maggiore di 0.

## Output

**Struttura di rete.** Visualizza le informazioni di riepilogo sulla rete neurale.

- **Descrizione.** Visualizza le informazioni relative alla rete neurale, incluse le variabili dipendenti, il numero di unità di ingresso e di uscita, il numero di layer e unità nascosti e le funzioni di attivazione.
- **Diagramma.** Visualizza il diagramma di rete come grafico non modificabile. Da notare che come aumenta il numero di covariate e i livelli di fattore, il diagramma diventa più difficile da interpretare.
- **pesi sinaptici.** Visualizza le stime del coefficiente che mostrano il rapporto tra le unità in un dato strato alle unità nel seguente strato. I pesi sinaptici si basano sul campione di allenamento anche se il dataset attivo è partizionato in dati di formazione, verifica e holdout. Da notare che il numero di pesi sinaptici può diventare piuttosto grande e che questi pesi generalmente non vengono utilizzati per interpretare i risultati della rete.

**Performance di rete.** Visualizza i risultati utilizzati per determinare se il modello è "buono". *Nota:* i grafici di questo gruppo si basano sulla formazione combinata e sui campioni di prova o solo sul campione di formazione se non esiste un campione di prova.

- **Riepilogo del modello.** Visualizza un riepilogo dei risultati della rete neurale per partizione e nel complesso, incluso l'errore, l'errore relativo o la percentuale di previsioni errate, la regola di arresto utilizzata per fermare l'allenamento e il tempo di formazione.

L'errore è l'errore di somma - quadrati quando la funzione di attivazione dell'identità, sigmoid o di tangente iperbolica viene applicata allo strato di output. Si tratta dell'errore cross - entropico quando la funzione di attivazione softmax viene applicata allo strato di output.

Gli errori relativi o le percentuali di previsioni errate vengono visualizzati a seconda dei livelli di misurazione variabili dipendenti. Se qualsiasi variabile dipendente ha livello di misurazione della scala, viene visualizzato l'errore relativo medio complessivo (relativo al modello medio). Se tutte le variabili dipendenti sono categoriali, viene visualizzata la percentuale media di previsioni errate. Vengono visualizzati anche errori relativi o percentuali di previsioni errate per singole variabili dipendenti.

- **Risultati della classificazione.** Visualizza una tabella di classificazione per ogni variabile dipendente categoriale per partizione e complessiva. Ogni tabella dà il numero di casi classificati correttamente e incorrettamente per ogni categoria di variabile dipendente. Si segnala anche la percentuale dei casi totali classificati correttamente.
- **curva ROC.** Visualizza una curva ROC (Ricevitore Funzionamento Caratteristico) per ogni variabile dipendente categoriale. Visualizza anche una tabella che dà la zona sotto ogni curva. Per una determinata variabile dipendente, il grafico ROC visualizza una curva per ogni categoria. Se la variabile dipendente ha due categorie, allora ogni curva tratta la categoria in questione come lo stato positivo contro l'altra categoria. Se la variabile dipendente ha più di due categorie, allora ogni curva tratta la categoria in questione come lo stato positivo contro l'aggregato di tutte le altre categorie.
- **Grafico dei guadagni cumulativi.** Visualizza un grafico dei guadagni cumulativi per ogni variabile dipendente categoriale. La visualizzazione di una curva per ogni categoria variabile dipendente è la stessa per le curve ROC.
- **Grafico di sollevamento.** Visualizza un grafico di sollevamento per ogni variabile dipendente categoriale. La visualizzazione di una curva per ogni categoria variabile dipendente è la stessa per le curve ROC.
- **Previsto da grafico osservato.** Visualizza un grafico a valore previsto per ogni variabile dipendente. Per le variabili dipendenti categoriali, vengono visualizzati i boxplotti di pseudo - probabilità previsti per ogni categoria di risposta, con la categoria di risposta osservata come variabile cluster. Per le variabili dipendenti in scala viene visualizzata una trama di dispersione.
- **Residuo per grafico previsto.** Visualizza un grafico a valore residuo per ogni variabile dipendente da scala. Non ci dovrebbero essere pattern visibili tra residui e valori previsti. Questo grafico è prodotto solo per variabili dipendenti in scala.

**Riepilogo elaborazione casi.** Visualizza la tabella di riepilogo delle elaborazioni dei casi, che sintetizza il numero di casi inclusi ed esclusi nell'analisi, in totale e mediante campioni di formazione, test e holdout.

**Analisi di importanza variabile indipendente.** Esegue un'analisi di sensibilità, che calcola l'importanza di ogni predittore nel determinare la rete neurale. L'analisi si basa sulla formazione combinata e sui campioni di prova o solo sul campione di formazione se non esiste un campione di prova. Questo crea una tabella e un grafico che visualizzano importanza e importanza normalizzata per ogni predittore. Da notare che l'analisi di sensibilità è computazionalmente costosa e dispensa nel tempo se ci sono grandi numeri di predittori o casi.

## Salva

La scheda Salva viene utilizzata per salvare le previsioni come variabili nel dataset.

- **Salva valore previsto o categoria per ogni variabile dipendente.** Questo salva il valore previsto per le variabili dipendenti in scala e la categoria prevista per le variabili dipendenti categoriali.

- **Salva con pseudo - probabilità o categoria prevista per ogni variabile dipendente.** Questo salva le pseudo probabilità previste per le variabili dipendenti categoriali. Una variabile separata viene salvata per ciascuna delle prime categorie  $n$ , dove  $n$  viene specificato nella colonna **Categorie alla Salva**.

**Nomi delle variabili salvate.** La generazione automatica del nome assicura il mantenimento di tutto il lavoro. I nomi personalizzati consentono di eliminare/sostituire i risultati di precedenti esecuzioni senza dover prima eliminare le variabili salvate nell'Editor dei dati.

Probabilità e Pseudo - Probabilità

Le variabili dipendenti categoriali con attivazione softmax e errore cross - entropico avranno un valore previsto per ogni categoria, dove ogni valore previsto è la probabilità che il caso appartenga alla categoria.

Le variabili dipendenti categoriali con errore di somma - quadrato avranno un valore previsto per ogni categoria, ma i valori previsti non possono essere interpretati come probabilità. La procedura salva queste pseudo - probabilità previste anche se le eventuali sono inferiori a 0 o superiori a 1, oppure la somma per una determinata variabile dipendente non è 1.

I ROC, i guadagni cumulativi e i grafici di sollevamento (vedi "Output" a pagina 8) vengono create in base a pseudo - probabilità. Nel caso in cui una qualsiasi delle pseudo - probabilità siano inferiori a 0 o superiori a 1, oppure la somma per una determinata variabile non è di 1, vengono prima ridimensionate ad essere comprese tra 0 e 1 e a somma a 1. Le probabilità pseudo - probabili sono ridimensionate dividendo per la loro somma. Ad esempio, se un caso ha previsto pseudo - probabilità di 0.50, 0.60 e 0.40 per una variabile dipendente a tre categorie, ciascuna pseudo - probabilità viene divisa per la somma 1.50 per ottenere 0.33, 0.40 e 0.27.

Se una qualsiasi delle pseudo probabilità sono negative, allora il valore assoluto del minimo viene aggiunto a tutte le pseudo probabilità prima del ridimensionamento sopra riportato. Ad esempio, se le pseudoprobabilità sono -0.30, 0.50 e 1.30, aggiungere prima 0.30 a ciascun valore per ottenere 0.00, 0.80 e 1.60. Quindi, dividere ogni nuovo valore per la somma 2.40 per ottenere 0.00, 0.33 e 0.67.

## Export

La scheda Export viene utilizzata per salvare le stime di peso sinaptico per ogni variabile dipendente in un file XML (PMML). È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio. Questa opzione non è disponibile se sono stati definiti file suddivisi.

## Opzioni

**Valori mancanti definiti dall'utente.** I fattori devono avere valori validi per un caso da inserire nell'analisi. Questi controlli consentono di decidere se i valori mancanti degli utenti sono trattati come validi tra i fattori e le variabili dipendenti categoriali.

**Regole di arresto.** Queste sono le regole che determinano quando smettere di allenare la rete neurale. La formazione procede attraverso almeno un passaggio di dati. La formazione può quindi essere fermata secondo i seguenti criteri, che vengono controllati nell'ordine elencato. Nelle definizioni delle regole di arresto che seguono, un passo corrisponde ad un passaggio di dati per i metodi online e mini - batch e un'iterazione per il metodo batch.

- **Massimo passi senza diminuzione di errore.** Il numero di passaggi da consentire prima di verificare una diminuzione di errore. Se non vi è alcuna diminuzione di errore dopo il numero specificato di passi, poi la formazione si arresta. Specificare un intero maggiore di 0. È inoltre possibile specificare quale esempio di dati viene utilizzato per calcolare l'errore. **Scegli automaticamente** utilizza il campione di prova se esiste e utilizza il campione di formazione altrimenti. Si noti che la formazione batch garantisce una diminuzione dell'errore del campione di formazione dopo ogni passaggio di dati; quindi, questa opzione si applica solo alla formazione batch se esiste un campione di verifica. **Entrambi i dati di formazione e di test** verificano l'errore per ciascuno di questi campioni; questa opzione si applica solo se un campione di test esce.

*Nota:* Dopo ogni passaggio di dati completo, la formazione online e mini - batch richiede un passaggio di dati extra per poter calcolare l'errore di formazione. Questo passaggio extra dati può rallentare

notevolmente la formazione, quindi generalmente si consiglia di fornire un campione di test e selezionare **Scegli automaticamente** in ogni caso.

- **Tempo di formazione massimo.** Scegliere se specificare un numero massimo di minuti per l'algoritmo da eseguire. Specificare un numero maggiore di 0.
- **Epoche di formazione massima.** Il numero massimo di epoche (dati passa) consentito. Se il numero massimo di epoche è superato, allora si allenano le fermate. Specificare un intero maggiore di 0.
- **Modifica minima relativa nell'errore di formazione.** La formazione si arresta se la variazione relativa dell'errore di formazione rispetto al passo precedente è inferiore al valore di criterio. Specificare un numero maggiore di 0. Per la formazione online e mini - batch questo criterio viene ignorato se solo i dati di prova vengono utilizzati per calcolare l'errore.
- **Variazione minima relativa del rapporto di errore di formazione.** La formazione si arresta se il rapporto dell'errore di formazione all'errore del modello nullo è inferiore al valore di criterio. Il modello nullo prevede il valore medio per tutte le variabili dipendenti. Specificare un numero maggiore di 0. Per la formazione online e mini - batch questo criterio viene ignorato se solo i dati di prova vengono utilizzati per calcolare l'errore.

**Massimo casi da memorizzare in memoria.** Questo controlla le seguenti impostazioni all'interno degli algoritmi di percezione multistrato. Specificare un numero intero maggiore di 1.

- Nella selezione automatica dell'architettura, la dimensione del campione utilizzato per determinare l'architettura della rete è  $\min(1000, memsize)$ , dove *memsize* è il numero massimo di casi da memorizzare in memoria.
- Nell'addestramento mini - batch con calcolo automatico del numero di mini - batch, il numero di mini - batch è  $\min(\max(M/10, 2), memsize)$ , dove *M* è il numero di casi nel campione di addestramento.

## Funzione di base radiale

---

La Procedura Radial Basis Function (RBF) produce un modello predittivo per una o più variabili dipendenti (target) in base ai valori delle variabili predittive.

**Esempio.** Un provider di telecomunicazioni ha segmentato la base clienti per modelli di utilizzo del servizio, suddividendo i clienti in quattro categorie. Una rete RBF che utilizza i dati demografici per prevedere l'appartenenza al gruppo permette all'azienda di personalizzare le offerte per i singoli clienti prospettati.

Considerazioni sui dati












**Variabili dipendenti.** Le variabili dipendenti possono essere:

- **Nominale.** Una variabile può essere considerata nominale quando i suoi valori rappresentano categorie prive di classificazione intrinseca (ad esempio, il dipartimento della società in cui lavora un dipendente). Degli esempi di variabili nominali includono la regione, il codice postale e l'affiliazione religiosa.
- **Ordinale.** Una variabile può essere trattata come ordinale quando i suoi valori rappresentano categorie con una classificazione intrinseca (ad esempio, livelli di soddisfazione del servizio da molto insoddisfatti a molto soddisfatti). Degli esempi di variabili ordinali includono i punteggi di atteggiamento che rappresentano i gradi di soddisfazione o fiducia e i punteggi di classificazione delle preferenze.
- **Scala.** Una variabile può essere trattata come una scala (continua) quando i suoi valori rappresentano categorie ordinate con una metrica significativa, in modo che i confronti di distanza tra i valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

La procedura presuppone che il livello di misurazione appropriato sia stato assegnato a tutte le variabili dipendenti, sebbene sia possibile modificare temporaneamente il livello di misurazione per una variabile facendo clic con il tasto destro del mouse sulla variabile nell'elenco delle variabili di origine e selezionando un livello di misurazione dal menu a comparsa.

L'icona accanto a ciascuna variabile nell'elenco delle variabili identifica il livello di misurazione e il tipo di dati.

Tabella 2. Icone del livello di misurazione

	Numerico	Stringa	Dati	Ora
Scala (continuo)		n/d		
Ordinale				
Nominale				

**Variabili di previsione.** I predittori possono essere specificati come fattori (categoriali) o covariati (scala).

**Codifiche variabili categoriali.** La procedura ricodifica temporaneamente le variabili dipendenti e indipendenti categoriali utilizzando le codifiche one-of-c per la durata della procedura. Se ci sono le categorie  $c$  di una variabile, allora la variabile viene memorizzata come vettori  $c$ , con la prima categoria denotata  $(1,0, \dots, 0)$ , la categoria successiva  $(0,1,0, \dots, 0)$ , ..., e la categoria finale  $(0,0, \dots, 0, 1)$ .

Questo schema di codifica aumenta il numero di pesi sinaptici e può determinare una formazione più lenta, ma i metodi di codifica più "compatti" di solito portano a reti neurali scarsamente adattate. Se la tua formazione di rete sta procedendo molto lentamente, prova a ridurre il numero di categorie nei tuoi predittori categoriali combinando categorie simili o eliminando i casi che hanno categorie estremamente rare.

Tutta la codifica *one*-of-*c* si basa sui dati di addestramento, anche se viene definito un campione di test o di holdout (vedi "Partizioni" a pagina 13). Così, se i campioni di prova o di holdout contengono casi con categorie predittrici che non sono presenti nei dati formativi, allora quei casi non sono utilizzati dalla procedura o dal punteggio. Se i campioni di prova o di holdout contengono casi con categorie variabili dipendenti che non sono presenti nei dati di formazione, allora quei casi non sono utilizzati dalla procedura ma possono essere segnati.

**Modifica della scala.** Le variabili dipendenti in scala e le covariate vengono ridimensionate di default per migliorare la formazione di rete. Tutte le operazioni di riscalatura vengono eseguite sulla base dei dati di addestramento, anche se è definito un campione di test o di holdout (vedere "Partizioni" a pagina 13). Ovvero a seconda del tipo di ridimensionamento, la media, la deviazione standard, il valore minimo o il valore massimo di una variabile covariata o dipendente vengono calcolati utilizzando solo i dati di formazione. Se si specifica una variabile per definire le partizioni, è importante che queste covariate o variabili dipendenti abbiano distribuzioni simili attraverso i campioni di allenamento, testing e holdout.

**Peso della frequenza.** I pesi della frequenza vengono ignorati da questa procedura.

**Replica risultati.** Se si desidera esattamente replicare i propri risultati, utilizzare lo stesso valore di inizializzazione per il generatore di numeri casuali e lo stesso ordine di dati, oltre ad utilizzare le stesse impostazioni di procedura. Ulteriori dettagli su questo argomento seguono:

- **Generazione numero casuale.** La procedura utilizza la generazione di numeri casuali durante l'assegnazione casuale di partizioni. Per riprodurre gli stessi risultati randomizzati in futuro, utilizzare lo stesso valore di inizializzazione per il generatore di numeri casuali prima di ogni esecuzione della procedura Radial Basis Function.
- **Ordine dei casi.** I risultati dipendono anche dall'ordine dei dati in quanto l'algoritmo cluster a due passi viene utilizzato per determinare le funzioni di base radiale.

Per ridurre al minimo gli effetti dell'ordine, disporre i casi in ordine casuale. Per verificare la stabilità di una determinata soluzione, è possibile ottenere diverse soluzioni con casi ordinati in diversi ordini casuali. Nelle situazioni con dimensioni file estremamente grandi, le esecuzioni multiple possono essere eseguite con un campione di casi ordinati in diversi ordini casuali.

Creazione di una rete di funzioni Radial Basis

Dai menu, scegliere:

## Analizzare > Reti Neurali > Funzione Base Radiale ...

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Selezionare almeno una variabile dipendente.
2. Selezionare almeno un fattore o covariato.

Opzionalmente, sulla scheda Variabili è possibile modificare il metodo per le covariate di ridimensionamento. Le scelte sono:

- **Standardizzato.** Sottrarre la media e dividere per la deviazione standard,  $(x - \text{mean}) / s$ .
- **Normalizzato.** Sottrarre il minimo e dividere per la gamma,  $(x - \text{min}) / (\text{max} - \text{min})$ . I valori normalizzati cadono tra il 0 e il 1.
- **Normalizzato regolato.** Versione regolata di sottrazione del minimo e dividendo per la gamma,  $[2 * (x - \text{min}) / (\text{max} - \text{min})] - 1$ . I valori normalizzati adattati sono compresi tra -1 e 1.
- **Messuno.** Nessun ridimensionamento delle covariate.

Campi con livello di misurazione sconosciuto

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) del dataset è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

**Esegui scansione dati.** Legge i dati del dataset attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con dataset di grandi dimensioni, questa operazione può richiedere del tempo.

**Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Vista variabile dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

## Partizioni

**Dataset di partizione.** Questo gruppo specifica il metodo di partizionamento del dataset attivo in campioni di formazione, verifica e holdout. Il **campione di formazione** comprende i record di dati utilizzati per formare la rete neurale; alcune percentuali di casi nel dataset devono essere assegnate al campione di formazione al fine di ottenere un modello. Il **campione di prova** è una serie indipendente di record di dati utilizzati per tenere traccia degli errori durante l'allenamento al fine di prevenire la sovrapposizione. Si consiglia vivamente di creare un campione di prova e la formazione in rete sarà generalmente più efficiente se il campione di prova è più piccolo rispetto al campione di allenamento. Il **campione holdout** è un'altra serie indipendente di record di dati utilizzati per valutare la rete neurale finale; l'errore per il campione holdout dà una stima "onesta" della capacità predittiva del modello perché i casi holdout non sono stati utilizzati per costruire il modello.

- **Randomly assegna i casi in base al numero relativo di casi.** Specificare il numero relativo (rapporto) dei casi assegnati casualmente ad ogni campione (allenamento, testing e holdout). La colonna % riporta la percentuale di casi che verranno assegnati a ciascun campione in base ai numeri relativi specificati.

Ad esempio, specificando 7, 3, 0 come i numeri relativi per la formazione, i test e i campioni holdout corrisponde a 70%, 30% e 0%. Specificando 2, 1, 1 come i numeri relativi corrisponde a 50%, 25% e 25%; 1, 1, 1 corrisponde a dividere il dataset in pari terzi tra allenamento, testing e holdout.

- **Utilizza variabile di partizionamento per assegnare i casi.** Specificare una variabile numerica che assegna ogni caso nel dataset attivo al campione di formazione, verifica o holdout. I casi con un valore positivo sulla variabile sono assegnati al campione di formazione, i casi con un valore di 0, al campione di prova, e i casi con un valore negativo, al campione holdout. I casi con un valore mancante di sistema

vengono esclusi dall'analisi. I valori mancanti definiti dall'utente per la variabile partizione sono sempre considerati validi.

## Architettura

La scheda Architettura viene utilizzata per specificare la struttura della rete. La procedura crea una rete neurale con uno strato "funzione di base radiale" nascosta; in generale, non sarà necessario modificare queste impostazioni.

**Numero di Unità in Strato Nascosto.** Ci sono tre modi di scegliere il numero di unità nascoste.

1. **Trova il numero migliore di unità all'interno di una gamma automaticamente calcolata.** La procedura calcola automaticamente i valori minimi e massimi della gamma e trova il numero migliore di unità nascoste all'interno della gamma.  
  
Se viene definito un campione di verifica, allora la procedura utilizza il criterio dei dati di verifica: Il numero migliore di unità nascoste è quello che restituisce l'errore più piccolo nei dati di verifica. Se un campione di prova non è definito, allora la procedura utilizza il criterio di informazione Bayesiana (BIC): Il miglior numero di unità nascoste è quello che restituisce il BIC più piccolo in base ai dati formativi.
2. **Trova il numero migliore di unità all'interno di un intervallo specificato.** È possibile fornire la propria gamma e la procedura troverà il numero "migliore" di unità nascoste all'interno di tale intervallo. Come in precedenza, il numero migliore di unità nascoste dalla gamma è determinato utilizzando il criterio dei dati di prova o il BIC.
3. **Utilizzare un numero specificato di unità.** È possibile sovrascrivere l'utilizzo di un intervallo e specificare direttamente un determinato numero di unità.

**Funzione di attivazione per Strato Nascosto.** La funzione di attivazione per lo strato nascosto è la funzione di base radiale, che "collega" le unità in uno strato ai valori delle unità nello strato successivo. Per lo strato di uscita, la funzione di attivazione è la funzione di identità; quindi le unità di uscita sono semplicemente somme ponderate delle unità nascoste.

- **Funzione di base radiale normalizzata.** Utilizza la funzione di attivazione softmax così le attivazioni di tutte le unità nascoste sono normalizzate a somma a 1.
- **Funzione di base radiale ordinaria.** Utilizza la funzione di attivazione esponenziale così l'attivazione dell'unità nascosta è un "bump" gaussiano in funzione degli input.

**Sovrapposizione Tra Unità Nascoste.** Il fattore di sovrapposizione è un moltiplicatore applicato alla larghezza delle funzioni di base radiale. Il valore calcolato automaticamente del fattore di sovrapposizione è  $1+0,1d$ , dove  $d$  è il numero di unità di input (la somma del numero di categorie in tutti i fattori e il numero di covariate).

## Output

**Struttura di rete.** Visualizza le informazioni di riepilogo sulla rete neurale.

- **Descrizione.** Visualizza le informazioni relative alla rete neurale, incluse le variabili dipendenti, il numero di unità di ingresso e di uscita, il numero di layer e unità nascosti e le funzioni di attivazione.
- **Diagramma.** Visualizza il diagramma di rete come grafico non modificabile. Da notare che come aumenta il numero di covariate e i livelli di fattore, il diagramma diventa più difficile da interpretare.
- **pesi sinaptici.** Visualizza le stime del coefficiente che mostrano il rapporto tra le unità in un dato strato alle unità nel seguente strato. I pesi sinaptici si basano sul campione di allenamento anche se il dataset attivo è partizionato in dati di formazione, verifica e holdout. Da notare che il numero di pesi sinaptici può diventare piuttosto grande e questi pesi generalmente non vengono utilizzati per interpretare i risultati della rete.

**Performance di rete.** Visualizza i risultati utilizzati per determinare se il modello è "buono". *Nota:* i grafici di questo gruppo si basano sulla formazione combinata e sui campioni di prova o solo il campione di formazione se non esiste un campione di prova.

- **Riepilogo del modello.** Visualizza un riepilogo dei risultati della rete neurale per partizione e nel complesso, incluso l'errore, l'errore relativo o la percentuale di previsioni errate e il tempo di formazione.

L'errore è l'errore di somma - quadrati. Inoltre, vengono visualizzati errori relativi o percentuali di previsioni errate, a seconda dei livelli di misurazione variabili dipendenti. Se qualsiasi variabile dipendente ha livello di misurazione della scala, viene visualizzato l'errore relativo medio complessivo (relativo al modello medio). Se tutte le variabili dipendenti sono categoriali, viene visualizzata la percentuale media di previsioni errate. Vengono visualizzati anche errori relativi o percentuali di previsioni errate per singole variabili dipendenti.

- **Risultati della classificazione.** Visualizza una tabella di classificazione per ogni variabile dipendente categoriale. Ogni tabella dà il numero di casi classificati correttamente e incorrettamente per ogni categoria di variabile dipendente. Si segnala anche la percentuale dei casi totali classificati correttamente.
- **curva ROC.** Visualizza una curva ROC (Ricevitore Funzionamento Caratteristico) per ogni variabile dipendente categoriale. Visualizza anche una tabella che dà la zona sotto ogni curva. Per una determinata variabile dipendente, il grafico ROC visualizza una curva per ogni categoria. Se la variabile dipendente ha due categorie, allora ogni curva tratta la categoria in questione come lo stato positivo contro l'altra categoria. Se la variabile dipendente ha più di due categorie, allora ogni curva tratta la categoria in questione come lo stato positivo contro l'aggregato di tutte le altre categorie.
- **Grafico dei guadagni cumulativi.** Visualizza un grafico dei guadagni cumulativi per ogni variabile dipendente categoriale. La visualizzazione di una curva per ogni categoria variabile dipendente è la stessa per le curve ROC.
- **Grafico di sollevamento.** Visualizza un grafico di sollevamento per ogni variabile dipendente categoriale. La visualizzazione di una curva per ogni categoria variabile dipendente è la stessa per le curve ROC.
- **Previsto da grafico osservato.** Visualizza un grafico a valore previsto per ogni variabile dipendente. Per le variabili dipendenti categoriali, vengono visualizzati i boxplotti di pseudo - probabilità previsti per ogni categoria di risposta, con la categoria di risposta osservata come variabile cluster. Per le variabili dipendenti in scala viene visualizzata una trama di dispersione.
- **Residuo per grafico previsto.** Visualizza un grafico a valore residuo per ogni variabile dipendente di scala. Non ci dovrebbero essere pattern visibili tra residui e valori previsti. Questo grafico è prodotto solo per variabili dipendenti in scala.

**Riepilogo elaborazione casi.** Visualizza la tabella di riepilogo delle elaborazioni dei casi, che sintetizza il numero di casi inclusi ed esclusi nell'analisi, in totale e mediante campioni di formazione, test e holdout.

**Analisi di importanza variabile indipendente.** Esegue un'analisi di sensibilità, che calcola l'importanza di ogni predittore nel determinare la rete neurale. L'analisi si basa sulla formazione combinata e sui campioni di prova o solo il campione di formazione se non esiste un campione di prova. Questo crea una tabella e un grafico che visualizzano importanza e importanza normalizzata per ogni predittore. Da notare che l'analisi di sensibilità è computazionalmente costosa e dispensa nel tempo se ci sono un gran numero di predittori o casi.

## Salva

La scheda Salva viene utilizzata per salvare le previsioni come variabili nel dataset.

- **Salva valore previsto o categoria per ogni variabile dipendente.** Questo salva il valore previsto per le variabili dipendenti in scala e la categoria prevista per le variabili dipendenti categoriali.
- **Salva prevista pseudo - probabilità per ogni variabile dipendente.** Questo salva le pseudo probabilità previste per le variabili dipendenti categoriali. Una variabile separata viene salvata per ciascuna delle prime categorie  $n$ , dove  $n$  viene specificato nella colonna *Categorie alla Salva*.

**Nomi delle variabili salvate.** La generazione automatica del nome assicura il mantenimento di tutto il lavoro. I nomi personalizzati consentono di scartare o sostituire i risultati delle esecuzioni precedenti senza prima cancellare le variabili salvate nell'editor dei dati.

## Probabilità e Pseudo - Probabilità

Le pseudo - probabilità previste non possono essere interpretate come probabilità perché la procedura Radial Basis Function utilizza l'errore di somma - quadrati e la funzione di attivazione dell'identità per lo strato di output. La procedura salva queste pseudo probabilità previste anche se le eventuali sono inferiori a 0 o superiori a 1 o la somma per una determinata variabile dipendente non è 1.

I ROC, i guadagni cumulativi e i grafici di sollevamento (vedi "Output" a pagina 14) vengono create in base a pseudo - probabilità. Nel caso in cui una qualsiasi delle pseudo - probabilità siano inferiori a 0 o superiori a 1 o la somma per una determinata variabile non è di 1, vengono prima ridimensionate ad essere comprese tra 0 e 1 e a somma a 1. Le probabilità pseudo - probabili sono ridimensionate dividendo per la loro somma. Ad esempio, se un caso ha previsto pseudo - probabilità di 0.50, 0.60 e 0.40 per una variabile dipendente a tre categorie, ciascuna pseudo - probabilità viene divisa per la somma 1.50 per ottenere 0.33, 0.40 e 0.27.

Se una qualsiasi delle pseudo probabilità sono negative, allora il valore assoluto del minimo viene aggiunto a tutte le pseudo probabilità prima del ridimensionamento sopra riportato. Ad esempio, se le pseudoprobabilità sono -0.30, .50 e 1.30, aggiungere prima 0.30 a ciascun valore per ottenere 0.00, 0.80 e 1.60. Quindi, dividere ogni nuovo valore per la somma 2.40 per ottenere 0.00, 0.33 e 0.67.

## Export

La scheda Export viene utilizzata per salvare le stime di peso sinaptico per ogni variabile dipendente in un file XML (PMML). È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio. Questa opzione non è disponibile se sono stati definiti file suddivisi.

## Opzioni

**Valori mancanti definiti dall'utente.** I fattori devono avere valori validi per un caso da inserire nell'analisi. Questi controlli consentono di decidere se i valori mancanti degli utenti sono trattati come validi tra i fattori e le variabili dipendenti categoriali.

## Informazioni particolari

---

Queste informazioni sono state sviluppate per prodotti e servizi offerti negli Stati Uniti. Questo materiale potrebbe essere disponibile da IBM in altre lingue. Tuttavia, può essere necessario possedere una copia del prodotto o versione del prodotto in tale lingua per potervi accedere.

IBM può non offrire i prodotti, i servizi o le funzioni presentati in questo documento in altri paesi. Consultare il proprio rappresentante locale IBM per informazioni sui prodotti ed i servizi attualmente disponibili nella propria zona. Qualsiasi riferimento ad un prodotto, programma o servizio IBM non implica o intende dichiarare che solo quel prodotto, programma o servizio IBM può essere utilizzato. Qualsiasi prodotto funzionalmente equivalente al prodotto, programma o servizio che non violi alcun diritto di proprietà intellettuale IBM può essere utilizzato. È comunque responsabilità dell'utente valutare e verificare la possibilità di utilizzare altri prodotti, programmi o servizi non IBM.

IBM può avere applicazioni di brevetti o brevetti in corso relativi all'argomento descritto in questo documento. La fornitura del presente documento non concede alcuna licenza a tali brevetti. È possibile inviare per iscritto richieste di licenze a:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
it*

Per richieste di licenze relative ad informazioni double-byte (DBCS), contattare il Dipartimento di Proprietà Intellettuale IBM nel proprio paese o inviare richieste per iscritto a:

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan*

IBM (INTERNATIONAL BUSINESS MACHINES CORPORATION) FORNISCE LA PRESENTE PUBBLICAZIONE "NELLO STATO IN CUI SI TROVA" SENZA GARANZIE DI ALCUN TIPO, ESPRESSE O IMPLICITE, IVI INCLUSE, A TITOLO DI ESEMPIO, GARANZIE IMPLICITE DI NON VIOLAZIONE, DI COMMERCIALIZZABILITÀ E DI IDONEITÀ PER UNO SCOPO PARTICOLARE. Alcune giurisdizioni non consentono la rinuncia ad alcune garanzie espresse o implicite in determinate transazioni, pertanto, la presente dichiarazione può non essere applicabile.

Queste informazioni potrebbero contenere imprecisioni tecniche o errori tipografici. Le modifiche alle presenti informazioni vengono effettuate periodicamente; tali modifiche saranno incorporate nelle nuove pubblicazioni della pubblicazione. IBM si riserva il diritto di apportare miglioramenti e/o modifiche al prodotto o al programma descritto nel manuale in qualsiasi momento e senza preavviso.

I riferimenti in queste informazioni a siti Web non IBM vengono forniti solo per comodità e non implicano in alcun modo l'approvazione di tali siti web. I materiali presenti in tali siti web non sono parte dei materiali per questo prodotto IBM e l'utilizzo di tali siti è a proprio rischio.

IBM può utilizzare o distribuire qualsiasi informazione fornita in qualsiasi modo ritenga appropriato senza incorrere in alcun obbligo verso l'utente.

Coloro che detengano la licenza su questo programma e desiderano avere informazioni su di esso allo scopo di consentire: (i) uno scambio di informazioni tra programmi indipendenti ed altri (compreso questo) e (ii) l'utilizzo reciproco di tali informazioni, dovrebbe rivolgersi a:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119*

Tali informazioni possono essere disponibili, in base ad appropriate clausole e condizioni, includendo in alcuni casi, il pagamento di una tassa.

Il programma concesso in licenza descritto nel presente documento e tutto il materiale concesso in licenza disponibile sono forniti da IBM in base alle clausole dell'Accordo per Clienti IBM (IBM Customer Agreement), dell'IBM IPLA (IBM International Program License Agreement) o qualsiasi altro accordo equivalente tra le parti.

Gli esempi client e dati delle prestazioni citati sono presentati esclusivamente a scopo illustrativo. Gli effettivi risultati delle prestazioni possono variare in base alle configurazioni e alle condizioni operative specifiche.

Le informazioni relative a prodotti non IBM sono ottenute dai fornitori di quei prodotti, dagli annunci pubblicati o da altre fonti disponibili al pubblico. IBM non ha testato quei prodotti e non può confermarne la precisione della prestazione, la compatibilità o qualsiasi altro reclamo relativo ai prodotti non IBM. Le domande sulle funzionalità dei prodotti non IBM devono essere indirizzate ai fornitori di tali prodotti.

Le dichiarazioni relative all'orientamento o alle intenzioni future di IBM sono soggette a modifica o a ritiro senza preavviso e rappresentano solo mete e obiettivi.

Questa pubblicazione contiene esempi di dati e prospetti utilizzati quotidianamente nelle operazioni aziendali. Pertanto, per maggiore completezza, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti i nomi contenuti nel manuale sono fittizi e ogni riferimento a persone o aziende reali è puramente casuale.

#### LICENZA DI COPYRIGHT:

Queste informazioni contengono programmi applicativi di esempio in linguaggio sorgente, che illustrano tecniche di programmazione su varie piattaforme operative. È possibile copiare, modificare e distribuire questi programmi di esempio sotto qualsiasi forma senza alcun pagamento a IBM, allo scopo di sviluppare, utilizzare, commercializzare o distribuire i programmi applicativi in conformità alle API (application programming interface) a seconda della piattaforma operativa per cui i programmi di esempio sono stati scritti. Questi esempi non sono stati testati approfonditamente tenendo conto di tutte le condizioni possibili. IBM, quindi, non può garantire o sottintendere l'affidabilità, l'utilità o il funzionamento di questi programmi. I programmi di esempio sono forniti "NELLO STATO IN CUI SI TROVANO", senza alcun tipo di garanzia. IBM non intende essere responsabile per alcun danno derivante dall'uso dei programmi di esempio.

Ogni copia o qualsiasi parte di questi programmi di esempio o qualsiasi lavoro derivato, devono contenere le seguenti informazioni relative alle leggi sul diritto d'autore:

© Copyright IBM Corp. 2021. Le porzioni di questo codice derivano da IBM Corp. Programmi di esempio.

© Copyright IBM Corp. 1989 - 2021. Tutti i diritti riservati.

## Marchi

---

IBM, il logo IBM e [ibm.com](http://ibm.com) sono marchi o marchi registrati di International Business Machines Corp., registrati in molte giurisdizioni in tutto il mondo. Altri nomi di prodotti e servizi possono essere marchi di IBM o di altre società. Un elenco corrente dei marchi IBM è disponibile sul web in "Copyright and trademark information" all'indirizzo [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, il logo Adobe, PostScript e il logo PostScript sono marchi o marchi registrati di Adobe Systems Incorporated negli Stati Uniti e/o in altri paesi.

Intel, il logo Intel, Intel Inside, il logo Intel Inside, Intel Centrino, il logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium sono marchi o marchi registrati di Intel Corporation o delle sue consociate negli Stati Uniti e in altri paesi.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o negli altri paesi.

Microsoft, Windows, Windows NT e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o negli altri paesi.

UNIX è un marchio della The Open Group negli Stati Uniti e/o negli altri paesi.

Java e tutti i marchi e i logo basati su Java sono marchi o marchi registrati di Oracle e/o associate.



---

# Indice analitico

## A

- addestramento in linea
  - in Multilayer Percezione [7](#)
- architettura di rete
  - in Multilayer Percezione [6](#)
  - in Radial Basis Funzione [14](#)

## C

- campione di addestramento
  - in Multilayer Percezione [5](#)
  - in Radial Basis Funzione [13](#)
- campione holdout
  - in Multilayer Percezione [5](#)
  - in Radial Basis Funzione [13](#)
- Curva ROC
  - in Multilayer Percezione [8](#)
  - in Radial Basis Funzione [14](#)

## D

- diagramma di rete
  - in Multilayer Percezione [8](#)
  - in Radial Basis Funzione [14](#)

## E

- Esempio di test
  - in Multilayer Percezione [5](#)
  - in Radial Basis Funzione [13](#)

## F

- formazione batch
  - in Multilayer Percezione [7](#)
- formazione di rete
  - in Multilayer Percezione [7](#)
- formazione mini - batch
  - in Multilayer Percezione [7](#)
- funzione di attivazione
  - in Multilayer Percezione [6](#)
  - in Radial Basis Funzione [14](#)
- Funzione di base radiale
  - architettura di rete [14](#)
  - Esportazione del modello [16](#)
  - opzioni [16](#)
  - output [14](#)
  - partizioni [13](#)
  - Salva variabili nel dataset attivo [15](#)

## G

- grafico dei guadagni
  - in Multilayer Percezione [8](#)
  - in Radial Basis Funzione [14](#)

- grafico lift
  - in Multilayer Percezione [8](#)
  - in Radial Basis Funzione [14](#)

## L

- livello nascosto
  - in Multilayer Percezione [6](#)
  - in Radial Basis Funzione [14](#)

## M

- multicloud
  - reti neurali [2](#)

## P

- Percettore multistrato
  - addestramento [7](#)
  - architettura di rete [6](#)
  - Esportazione del modello [10](#)
  - opzioni [10](#)
  - output [8](#)
  - partizioni [5](#)
  - Salva variabili nel dataset attivo [9](#)

## R

- regole di arresto
  - in Multilayer Percezione [10](#)
- reti neurali
  - multicloud [2](#)

## S

- Strato di output
  - in Multilayer Percezione [6](#)
  - in Radial Basis Funzione [14](#)

## V

- valori mancanti
  - in Multilayer Percezione [10](#)





