

Reti Neurali Artificiali*

Corso di MOBD

Roma 25 Ottobre 2017

*Grippo, L., and M. Sciandrone. *Metodi di ottimizzazione per le reti neurali*. Rapporto Tecnico (2003): 09-03.

Perceptron

Perceptron

● Perceptron

● Addestramento I

● Addestramento II

● Limiti del perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

Il **Perceptron** esegue una trasformazione non lineare degli ingressi x e fornisce un'uscita scalare (nella versione di Mc Cullock e Pitts 43)

$$y(x) = \text{sgn}(w^T x - \theta)$$

Perceptron

Perceptron

● Perceptron

● Addestramento I

● Addestramento II

● Limiti del perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

Il **Perceptron** esegue una trasformazione non lineare degli ingressi x e fornisce un'uscita scalare (nella versione di Mc Cullock e Pitts 43)

$$y(x) = \text{sgn}(w^T x - \theta)$$

Scegliendo i pesi in modo appropriato è in grado di realizzare le funzioni and, or, not.

Perceptron

Perceptron

- Perceptron

- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Il **Perceptron** esegue una trasformazione non lineare degli ingressi x e fornisce un'uscita scalare (nella versione di Mc Cullock e Pitts 43)

$$y(x) = \text{sgn}(w^T x - \theta)$$

Scegliendo i pesi in modo appropriato è in grado di realizzare le funzioni and, or, not.

Nel '62 Rosenblatt lo reinterpreta come classificatore lineare: la funzione discriminante è la funzione lineare $w^T x - \theta$

Addestramento I

Perceptron

- Perceptron
- **Addestramento I**
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Sia dato un **training set**

$$\mathcal{T} = \{(x^p, y^p) : x \in \mathbb{R}^n, y^p \in \{-1, 1\}, p = 1, \dots, P\},$$

diviso in due insiemi

$$\mathcal{A} = \{x^p : y^p = 1\} \quad \mathcal{B} = \{x^p : y^p = -1\}$$

Addestramento I

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Sia dato un **training set**

$$\mathcal{T} = \{(x^p, y^p) : x \in \mathbb{R}^n, y^p \in \{-1, 1\}, p = 1, \dots, P\},$$

diviso in due insiemi

$$\mathcal{A} = \{x^p : y^p = 1\} \quad \mathcal{B} = \{x^p : y^p = -1\}$$

Addestrare il perceptron vuol dire individuare i pesi w e la soglia θ tali che:

$$\begin{aligned} w^T x^p - \theta &> 0 \quad \forall x^p \in \mathcal{A} \\ w^T x^p - \theta &< 0 \quad \forall x^p \in \mathcal{B} \end{aligned}$$

Addestramento I

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Sia dato un **training set**

$$\mathcal{T} = \{(x^p, y^p) : x \in \mathbb{R}^n, y^p \in \{-1, 1\}, p = 1, \dots, P\},$$

diviso in due insiemi

$$\mathcal{A} = \{x^p : y^p = 1\} \quad \mathcal{B} = \{x^p : y^p = -1\}$$

Addestrare il perceptron vuol dire individuare i pesi w e la soglia θ tali che:

$$\begin{aligned} w^T x^p - \theta &> 0 \quad \forall x^p \in \mathcal{A} \\ w^T x^p - \theta &< 0 \quad \forall x^p \in \mathcal{B} \end{aligned}$$

Si introduce un ingresso fittizio $x_0^p = -1, p = 1, \dots, P$ e si ridefinisce $w = (\theta \ w)$, il sistema diventa

$$\begin{aligned} w^T x^p &> 0 \quad \forall x^p \in \mathcal{A} \\ w^T x^p &< 0 \quad \forall x^p \in \mathcal{B} \end{aligned}$$

Addestramento II

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

```
Dati  $(x^p, y^p) \in \mathcal{T}, p = 1, \dots, P$   
Iniz  $w(0) = 0, k = 0, \text{nclass} = 0$   
While  $\text{nclass} < P$   
For  $p = 1, \dots, P,$   
  If  $\text{sgn}(w^T x^p) = y^p$  then  
     $\text{nclass} = \text{nclass} + 1$   
  else  
     $w(k + 1) = w(k) + y^p x^p$   
     $k = k + 1$   
  End if  
End for  
If  $\text{nclass} < P, \text{nclass} = 0$   
End do
```


Limiti del perceptron

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Se i due insiemi \mathcal{A} e \mathcal{B} sono **linearmente separabili**, l'algoritmo converge.

Limiti del perceptron

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Se i due insiemi \mathcal{A} e \mathcal{B} sono **linearmente separabili**, l'algoritmo converge.

Si potrebbe addestrare risolvendo il sistema di disequazioni con metodi ai punti interni.

Oppure sostituendo la funzione sgn con una qualunque funzione ρ monotona crescente tale che:

$$\lim_{t \rightarrow -\infty} \rho(t) = -1, \quad \lim_{t \rightarrow \infty} \rho(t) = 1$$

e minimizzare poi la funzione di errore quadratico sul training:

$$E(w) = \frac{1}{2} \sum_{p=1}^P (y^p - \rho(w^T x^p))^2$$

Limiti del perceptron

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Se i due insiemi \mathcal{A} e \mathcal{B} sono **linearmente separabili**, l'algoritmo converge.

Si potrebbe addestrare risolvendo il sistema di disequazioni con metodi ai punti interni.

Oppure sostituendo la funzione sgn con una qualunque funzione ρ monotona crescente tale che:

$$\lim_{t \rightarrow -\infty} \rho(t) = -1, \quad \lim_{t \rightarrow \infty} \rho(t) = 1$$

e minimizzare poi la funzione di errore quadratico sul training:

$$E(w) = \frac{1}{2} \sum_{p=1}^P (y^p - \rho(w^T x^p))^2$$

Non è in grado di rappresentare funzioni anche semplici come lo XOR (OR esclusivo o somma modulo 2, restituisce 1 se e solo se uno solo dei due operandi 1, mentre restituisce 0 in tutti gli altri casi).

Limiti del perceptron

Perceptron

- Perceptron
- Addestramento I
- Addestramento II
- Limiti del perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Se i due insiemi \mathcal{A} e \mathcal{B} sono **linearmente separabili**, l'algoritmo converge.

Si potrebbe addestrare risolvendo il sistema di disequazioni con metodi ai punti interni.

Oppure sostituendo la funzione sgn con una qualunque funzione ρ monotona crescente tale che:

$$\lim_{t \rightarrow -\infty} \rho(t) = -1, \quad \lim_{t \rightarrow \infty} \rho(t) = 1$$

e minimizzare poi la funzione di errore quadratico sul training:

$$E(w) = \frac{1}{2} \sum_{p=1}^P (y^p - \rho(w^T x^p))^2$$

Non è in grado di rappresentare funzioni anche semplici come lo XOR (OR esclusivo o somma modulo 2, restituisce 1 se e solo se uno solo dei due operandi 1, mentre restituisce 0 in tutti gli altri casi).

Si può solo garantire errore nullo sul training set, non ci sono garanzie sulla capacità di generalizzazione.

Definizione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Sono architetture costituite da più strati di neuroni connessi in cascata:

Definizione

Perceptron

Reti multistrato

● Definizione

● Struttura

● Esempio

● Funzioni di attivazione

● Rete a due strati

● Proprietà

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

Sono architetture costituite da più strati di neuroni connessi in cascata:

- consentono di approssimare qualunque funzione continua su un insieme compatto (sotto opportune ipotesi sulle funzioni di attivazione);

Definizione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Sono architetture costituite da più strati di neuroni connessi in cascata:

- consentono di approssimare qualunque funzione continua su un insieme compatto (sotto opportune ipotesi sulle funzioni di attivazione);
- risolvono problemi di classificazione anche di insieme non linearmente separabili;

Definizione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Sono architetture costituite da più strati di neuroni connessi in cascata:

- consentono di approssimare qualunque funzione continua su un insieme compatto (sotto opportune ipotesi sulle funzioni di attivazione);
- risolvono problemi di classificazione anche di insieme non linearmente separabili;
- l'addestramento è più complesso.

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

◁ insieme di n nodi in ingresso associati agli n ingressi della rete

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

- ◁ insieme di n nodi in ingresso associati agli n ingressi della rete
- ◁ un insieme di neuroni formali organizzati in $L \geq 2$ strati di cui:

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

- ◁ insieme di n nodi in ingresso associati agli n ingressi della rete
- ◁ un insieme di neuroni formali organizzati in $L \geq 2$ strati di cui:
 - $L - 1$ strati nascosti: le uscite fanno da ingresso ad altri neuroni della rete

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

- ◁ insieme di n nodi in ingresso associati agli n ingressi della rete
- ◁ un insieme di neuroni formali organizzati in $L \geq 2$ strati di cui:
 - $L - 1$ strati nascosti: le uscite fanno da ingresso ad altri neuroni della rete
 - uno strato di uscita costituito da $K \geq 1$ neuroni le cui uscite sono le uscite della rete

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

- ◁ insieme di n nodi in ingresso associati agli n ingressi della rete
- ◁ un insieme di neuroni formali organizzati in $L \geq 2$ strati di cui:
 - $L - 1$ strati nascosti: le uscite fanno da ingresso ad altri neuroni della rete
 - uno strato di uscita costituito da $K \geq 1$ neuroni le cui uscite sono le uscite della rete
- ◁ Un insieme di archi orientati e pesati che costituiscono le connessioni interneuroni e quelle con i nodi di ingresso.

Struttura

Perceptron

Reti multistrato

- Definizione
- **Struttura**
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

L'architettura si definisce tramite:

- ◁ insieme di n nodi in ingresso associati agli n ingressi della rete
- ◁ un insieme di neuroni formali organizzati in $L \geq 2$ strati di cui:
 - $L - 1$ strati nascosti: le uscite fanno da ingresso ad altri neuroni della rete
 - uno strato di uscita costituito da $K \geq 1$ neuroni le cui uscite sono le uscite della rete
- ◁ Un insieme di archi orientati e pesati che costituiscono le connessioni interneuroni e quelle con i nodi di ingresso.
- ◁ NON esistono connessioni tra neuroni dello stesso strato nè archi all'indietro

Esempio

Perceptron

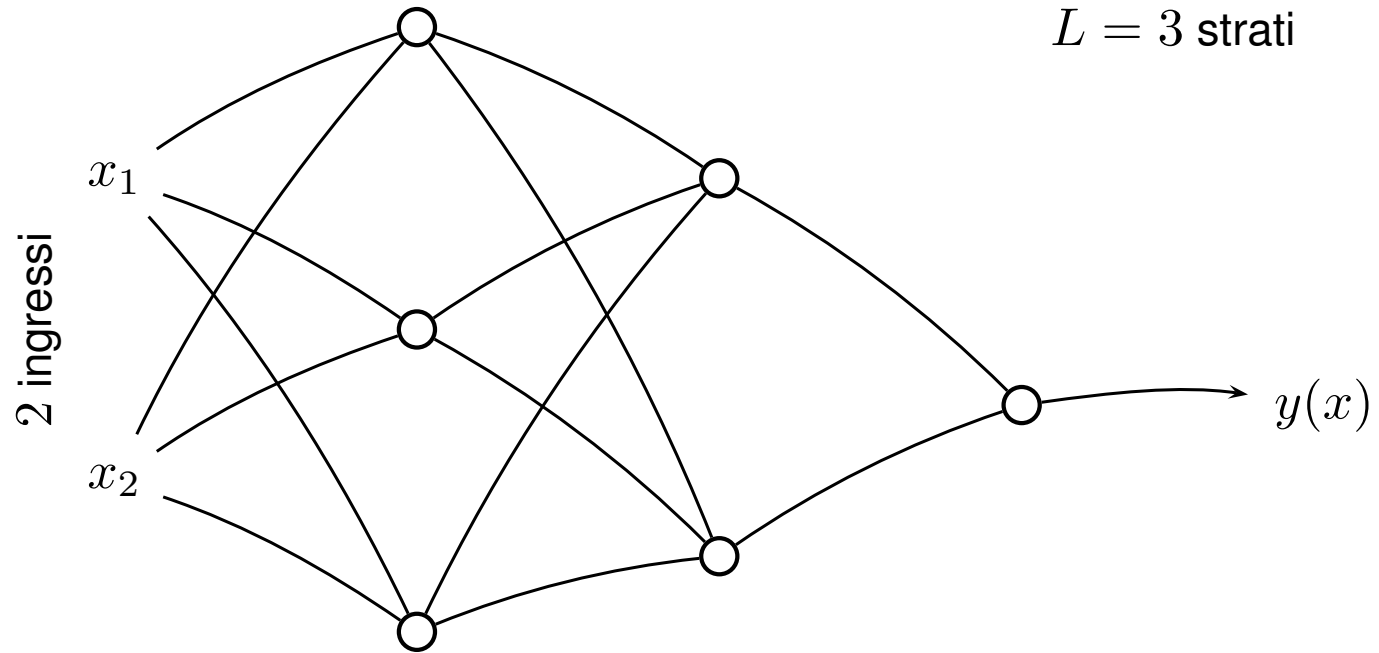
Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line



Esempio

Perceptron

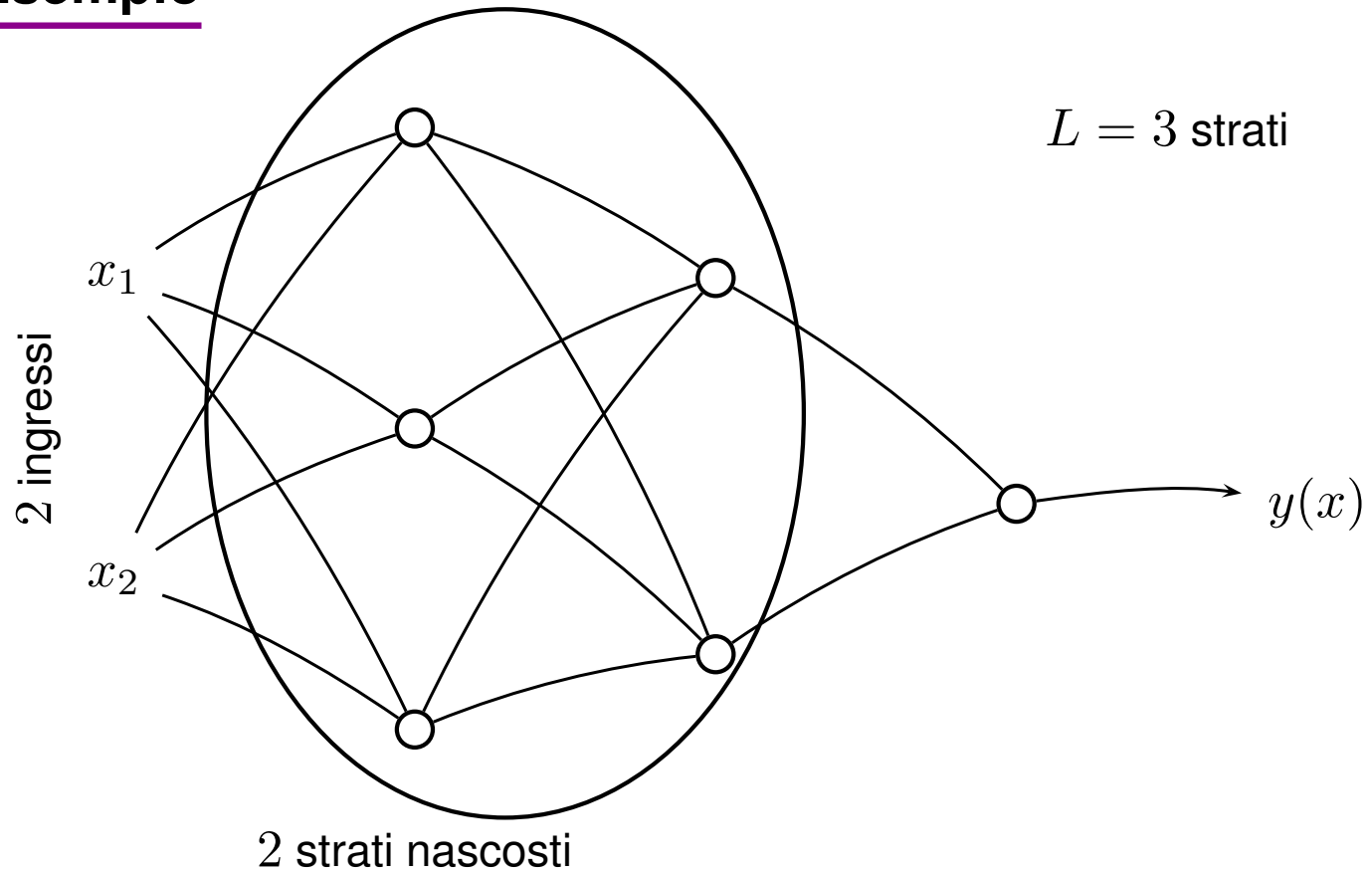
Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line



Esempio

Perceptron

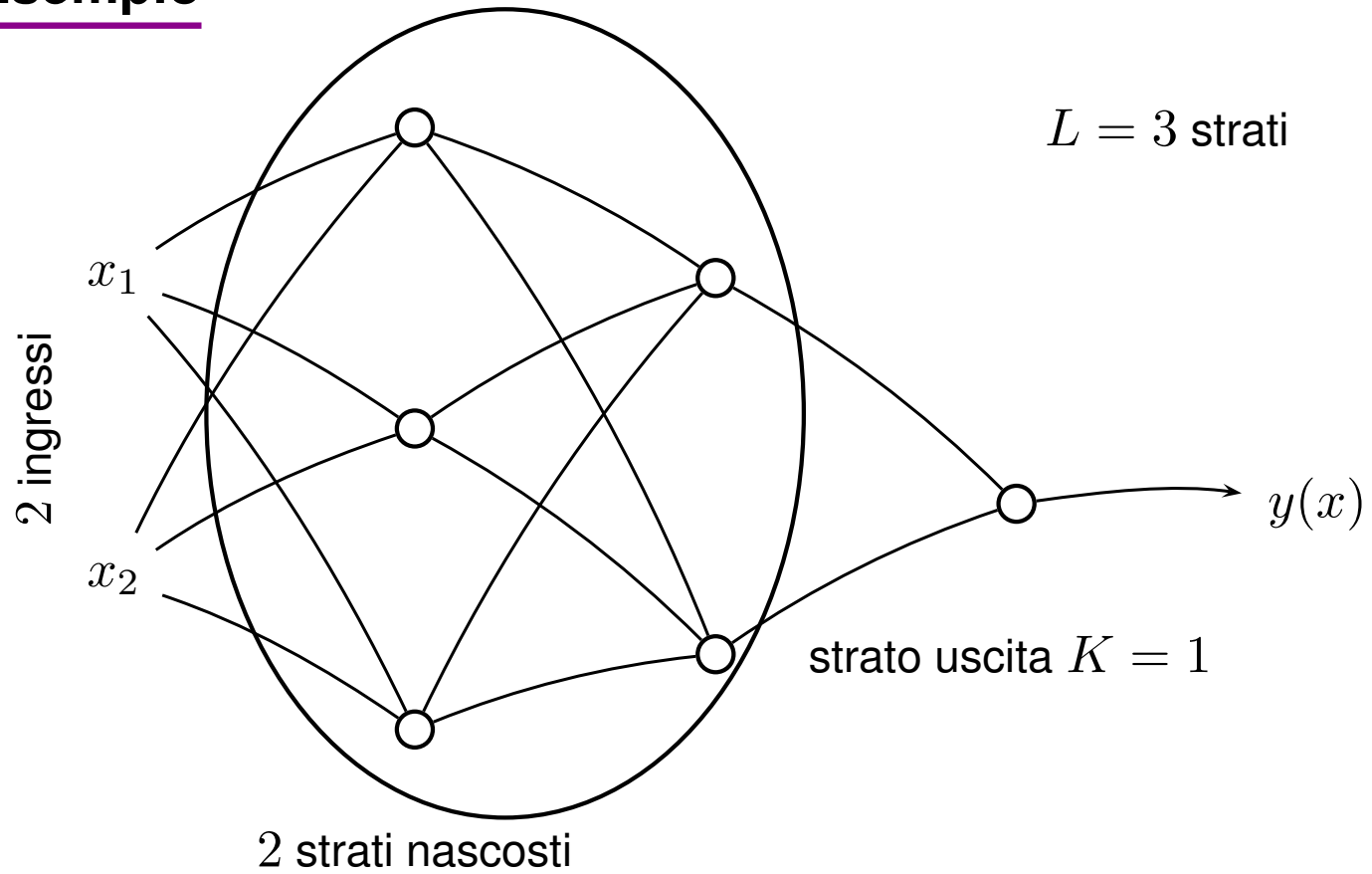
Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line



Esempio

Perceptron

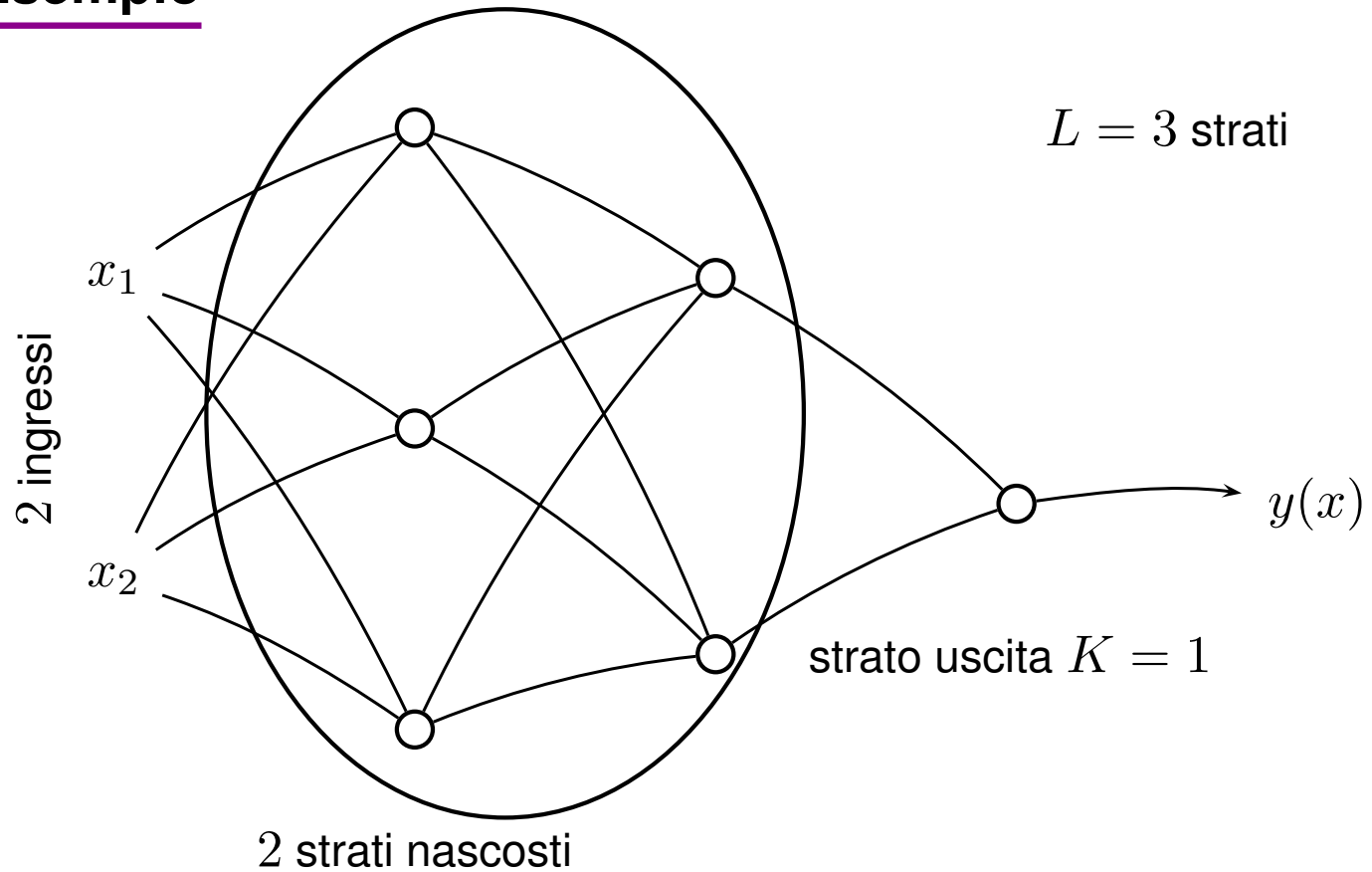
Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

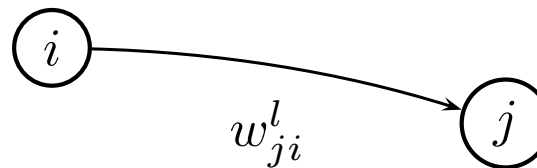
Creazione di una rete multistrato

Metodi Offline

Metodi on-line



Dati due neuroni i dello strato $l - 1$ e j dello strato l si ha



Funzioni di attivazione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

Ogni neurone j ha una funzione di attivazione $g_j^{(l)} : \mathbb{R} \Rightarrow \mathbb{R}$ che prende in ingresso la combinazione pesata degli ingressi a_j , mentre w_{j0} è la soglia.

Funzioni di attivazione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Ogni neurone j ha una funzione di attivazione $g_j^{(l)} : \mathbb{R} \Rightarrow \mathbb{R}$ che prende in ingresso la combinazione pesata degli ingressi a_j , mentre w_{j0} è la soglia.

Per il primo strato si ha:

$$a_j^{(1)} = \sum_{k=1}^n w_{jk} x_k - w_{j0} \quad z_j^1 = g_j(a_j^{(1)}),$$

Funzioni di attivazione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Ogni neurone j ha una funzione di attivazione $g_j^{(l)} : \mathbb{R} \Rightarrow \mathbb{R}$ che prende in ingresso la combinazione pesata degli ingressi a_j , mentre w_{j0} è la soglia.

Per il primo strato si ha:

$$a_j^{(1)} = \sum_{k=1}^n w_{jk} x_k - w_{j0} \quad z_j^1 = g_j(a_j^{(1)}),$$

Per lo strato l generico si ha:

$$a_j^{(l)} = \sum_{k=1}^{N^{l-1}} w_{jk}^l z_k^{l-1} - w_{j0}^l \quad z_j^l = g_j^{(l)}(a_j^{(l)}),$$

dove N^l è il numero di neuroni dello strato nascosto l .

Funzioni di attivazione

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Ogni neurone j ha una funzione di attivazione $g_j^{(l)} : \mathbb{R} \Rightarrow \mathbb{R}$ che prende in ingresso la combinazione pesata degli ingressi a_j , mentre w_{j0} è la soglia.

Per il primo strato si ha:

$$a_j^{(1)} = \sum_{k=1}^n w_{jk} x_k - w_{j0} \quad z_j^1 = g_j(a_j^{(1)}),$$

Per lo strato l generico si ha:

$$a_j^{(l)} = \sum_{k=1}^{N^{l-1}} w_{jk}^l z_k^{l-1} - w_{j0}^l \quad z_j^l = g_j^{(l)}(a_j^{(l)}),$$

dove N^l è il numero di neuroni dello strato nascosto l .

Tipici esempi di funzioni di attivazione (che sono di solito differenziabili e sigmoidali) sono:

$$g(t) = \frac{1}{1 + e^{-ct}}, \quad c > 0 \quad g(t) = \frac{1 - e^{-t}}{1 + e^{-t}}$$

Rete a due strati

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Consideriamo la rete a 2 strati con uno strato nascosto e uscita lineare:

Rete a due strati

Perceptron

Reti multistrato

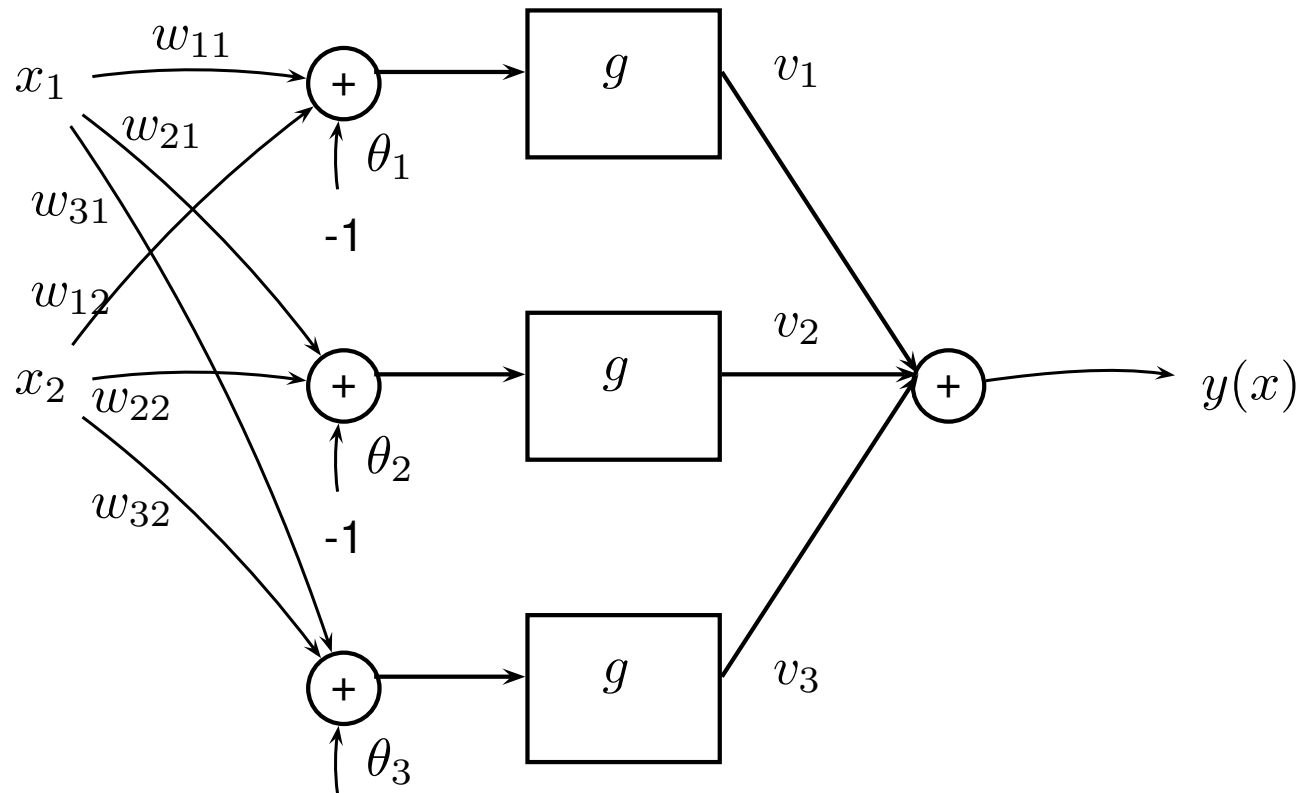
- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Consideriamo la rete a 2 strati con uno strato nascosto e uscita lineare:



$$y(x) = \sum_{j=1}^3 v_j g\left(\sum_{k=1}^2 (w_{jk} x_k - \theta_j)\right)$$

Proprietà

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Una rete multistrato a 2 strati con uno strato nascosto e uscita lineare è un approssimatore universale: approssima qualunque funzione continua su un insieme compatto (sotto l'ipotesi che g sia continua e non sia un polinomio)

Proprietà

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Una rete multistrato a 2 strati con uno strato nascosto e uscita lineare è un approssimatore universale: approssima qualunque funzione continua su un insieme compatto (sotto l'ipotesi che g sia continua e non sia un polinomio) Approssimare una funzione $f(x)$ su un insieme compatto Ω vuol dire che, per un qualunque $\epsilon > 0$,

$$\max_{x \in \Omega} |f(x) - y(x)| < \epsilon$$

Proprietà

Perceptron

Reti multistrato

- Definizione
- Struttura
- Esempio
- Funzioni di attivazione
- Rete a due strati
- Proprietà

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

Una rete multistrato a 2 strati con uno strato nascosto e uscita lineare è un approssimatore universale: approssima qualunque funzione continua su un insieme compatto (sotto l'ipotesi che g sia continua e non sia un polinomio) Approssimare una funzione $f(x)$ su un insieme compatto Ω vuol dire che, per un qualunque $\epsilon > 0$,

$$\max_{x \in \Omega} |f(x) - y(x)| < \epsilon$$

Le reti multistrato possono essere usate sia per problemi di regressione che per problemi di classificazione.

Struttura e addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● **Struttura e addestramento**

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Determinare una rete multistrato con n ingressi e K uscite implica due passi:

Struttura e addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● **Struttura e addestramento**

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Determinare una rete multistrato con n ingressi e K uscite implica due passi:

1. Scegliere l'architettura della rete: numero di strati, numero di neuroni per ogni strato

Struttura e addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● **Struttura e addestramento**

● Scelta dell'architettura
● Stabilizzazione strutturale

● Tecnica di Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Determinare una rete multistrato con n ingressi e K uscite implica due passi:

1. Scegliere l'architettura della rete: numero di strati, numero di neuroni per ogni strato
2. addestrare la rete: scegliere w e θ per ogni strato.

Struttura e addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● **Struttura e addestramento**

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Determinare una rete multistrato con n ingressi e K uscite implica due passi:

1. Scegliere l'architettura della rete: numero di strati, numero di neuroni per ogni strato
2. addestrare la rete: scegliere w e θ per ogni strato.

Addestrare la rete vuol dire risolvere il problema:

$$\min_{w \in \mathbb{R}^m} E(w) = \sum_{p=1}^P E_p(w)$$

dove $E_p(w)$ è il termine di errore relativo al campione p del training set, che misura la distanza tra l'uscita desiderata y^p e l'uscita prodotta dalla rete, ad es.:

$$E_p = \frac{1}{2} \|y(x^p, w) - y^p\|^2$$

Struttura e addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● **Struttura e addestramento**

● Scelta dell'architettura
● Stabilizzazione strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Determinare una rete multistrato con n ingressi e K uscite implica due passi:

1. Scegliere l'architettura della rete: numero di strati, numero di neuroni per ogni strato
2. addestrare la rete: scegliere w e θ per ogni strato.

Addestrare la rete vuol dire risolvere il problema:

$$\min_{w \in \mathbb{R}^m} E(w) = \sum_{p=1}^P E_p(w)$$

dove $E_p(w)$ è il termine di errore relativo al campione p del training set, che misura la distanza tra l'uscita desiderata y^p e l'uscita prodotta dalla rete, ad es.:

$$E_p = \frac{1}{2} \|y(x^p, w) - y^p\|^2$$

Supponiamo $E_p \in \mathcal{C}^1$

Scelta dell'architettura

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione strutturale

● Tecnica di Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Scopo dell'addestramento non è interpolare i dati del training set, ma ottenere una buona capacità di generalizzazione.

Scelta dell'architettura

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento

- Scelta dell'architettura

- Stabilizzazione strutturale

- Tecnica di

Regolarizzazione

- Early stopping

- Addestramento

- Classi di algoritmi

- Metodi batch

Metodi Offline

Metodi on-line

Scopo dell'addestramento non è interpolare i dati del training set, ma ottenere una buona capacità di generalizzazione.

Uno dei problemi più difficili è quello di stabilire la **complessità** del modello, e di conseguenza il numero di parametri liberi, sulla base dei dati disponibili per l'addestramento.

Scelta dell'architettura

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Scopo dell'addestramento non è interpolare i dati del training set, ma ottenere una buona capacità di generalizzazione.

Uno dei problemi più difficili è quello di stabilire la **complessità** del modello, e di conseguenza il numero di parametri liberi, sulla base dei dati disponibili per l'addestramento.

Per le reti multistrato esistono delle stime teoriche del numero di campioni del training set necessari per garantire una buona capacità di generalizzazione. Ma sono spesso inadeguate, si usano quindi tecniche euristiche per scegliere la struttura della rete e dimensionare il training set.

Scelta dell'architettura

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Scopo dell'addestramento non è interpolare i dati del training set, ma ottenere una buona capacità di generalizzazione.

Uno dei problemi più difficili è quello di stabilire la **complessità** del modello, e di conseguenza il numero di parametri liberi, sulla base dei dati disponibili per l'addestramento.

Per le reti multistrato esistono delle stime teoriche del numero di campioni del training set necessari per garantire una buona capacità di generalizzazione. Ma sono spesso inadeguate, si usano quindi tecniche euristiche per scegliere la struttura della rete e dimensionare il training set.

Esistono due strategie fondamentali:

Scelta dell'architettura

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Scopo dell'addestramento non è interpolare i dati del training set, ma ottenere una buona capacità di generalizzazione.

Uno dei problemi più difficili è quello di stabilire la **complessità** del modello, e di conseguenza il numero di parametri liberi, sulla base dei dati disponibili per l'addestramento.

Per le reti multistrato esistono delle stime teoriche del numero di campioni del training set necessari per garantire una buona capacità di generalizzazione. Ma sono spesso inadeguate, si usano quindi tecniche euristiche per scegliere la struttura della rete e dimensionare il training set.

Esistono due strategie fondamentali:

1. Stabilizzazione strutturale

Scelta dell'architettura

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Scopo dell'addestramento non è interpolare i dati del training set, ma ottenere una buona capacità di generalizzazione.

Uno dei problemi più difficili è quello di stabilire la **complessità** del modello, e di conseguenza il numero di parametri liberi, sulla base dei dati disponibili per l'addestramento.

Per le reti multistrato esistono delle stime teoriche del numero di campioni del training set necessari per garantire una buona capacità di generalizzazione. Ma sono spesso inadeguate, si usano quindi tecniche euristiche per scegliere la struttura della rete e dimensionare il training set.

Esistono due strategie fondamentali:

1. **Stabilizzazione strutturale**
2. **Tecnica di regolarizzazione**

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si sceglie il numero di unità facendo crescere/diminuire il numero di neuroni

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si sceglie il numero di unità facendo crescere/diminuire il numero di neuroni
3. Per ogni numero di neuroni, si addestra la rete corrispondente usando sempre il training set.

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- **Stabilizzazione strutturale**
- Tecnica di Regularizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si sceglie il numero di unità facendo crescere/diminuire il numero di neuroni
3. Per ogni numero di neuroni, si addestra la rete corrispondente usando sempre il training set.
4. Si valuta l'architettura scelta tramite il validation set (**cross validation**) e si sceglie quella che dà l'errore più basso.

Stabilizzazione strutturale

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:
training set
validation set
test set
2. Si sceglie il numero di unità facendo crescere/diminuire il numero di neuroni
3. Per ogni numero di neuroni, si addestra la rete corrispondente usando sempre il training set.
4. Si valuta l'architettura scelta tramite il validation set (**cross validation**) e si sceglie quella che dà l'errore più basso.
5. La migliore architettura viene valutata tramite il test set.

Tecnica di Regolarizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- **Tecnica di Regolarizzazione**
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

Tecnica di Regolarizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- **Tecnica di Regolarizzazione**
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

Tecnica di Regularizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- **Tecnica di Regularizzazione**
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

Tecnica di Regularizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- **Tecnica di Regularizzazione**
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

Tecnica di Regolarizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale

● **Tecnica di Regolarizzazione**

- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si aggiunge alla funzione di errore un termine di penalità nella norma di w (equivale a fare ipotesi di regolarità sulle funzioni realizzate dalla rete):

$$E(w) = \sum_{p=1}^P E_p(w) + \gamma \|w\|^2$$

Tecnica di Regolarizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale

● **Tecnica di Regolarizzazione**

- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si aggiunge alla funzione di errore un termine di penalità nella norma di w (equivale a fare ipotesi di regolarità sulle funzioni realizzate dalla rete):

$$E(w) = \sum_{p=1}^P E_p(w) + \gamma \|w\|^2$$

3. Per diversi valori di γ , si addestra la rete usando sempre il training set.

Tecnica di Regolarizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale

● **Tecnica di Regolarizzazione**

- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si aggiunge alla funzione di errore un termine di penalità nella norma di w (equivale a fare ipotesi di regolarità sulle funzioni realizzate dalla rete):

$$E(w) = \sum_{p=1}^P E_p(w) + \gamma \|w\|^2$$

3. Per diversi valori di γ , si addestra la rete usando sempre il training set.
4. Si sceglie il valore migliore di γ tramite **cross validation**.

Tecnica di Regolarizzazione

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale

● **Tecnica di Regolarizzazione**

- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

1. Si dividono i dati in 3 insiemi:

training set

validation set

test set

2. Si aggiunge alla funzione di errore un termine di penalità nella norma di w (equivale a fare ipotesi di regolarità sulle funzioni realizzate dalla rete):

$$E(w) = \sum_{p=1}^P E_p(w) + \gamma \|w\|^2$$

3. Per diversi valori di γ , si addestra la rete usando sempre il training set.
4. Si sceglie il valore migliore di γ tramite **cross validation**.
5. La rete risultante viene valutata tramite il test set.

Early stopping

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- **Early stopping**
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Un'alternativa alla regolarizzazione è l'**early stopping**:

Early stopping

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- **Early stopping**
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Un'alternativa alla regolarizzazione è l'**early stopping**:

1. Si interrompe prematuramente la minimizzazione della funzione di errore.

Early stopping

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- **Early stopping**
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Un'alternativa alla regolarizzazione è l'**early stopping**:

1. Si interrompe prematuramente la minimizzazione della funzione di errore.
2. Si valuta la rete sul validation set.

Early stopping

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione

● **Early stopping**

- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Un'alternativa alla regolarizzazione è l'**early stopping**:

1. Si interrompe prematuramente la minimizzazione della funzione di errore.
2. Si valuta la rete sul validation set.
3. All'inizio l'errore sul validation set diminuisce al migliorare dell'errore sul training.

Early stopping

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- **Early stopping**
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Un'alternativa alla regolarizzazione è l'**early stopping**:

1. Si interrompe prematuramente la minimizzazione della funzione di errore.
2. Si valuta la rete sul validation set.
3. All'inizio l'errore sul validation set diminuisce al migliorare dell'errore sul training.
4. Se si ha il fenomeno dell'overfitting, l'errore sul validation set sale. A questo punto si interrompe il processo.

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- **Addestramento**
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Un algoritmo efficiente è quindi uno strumento essenziale, e deve tenere conto delle difficoltà intrinseche del problema (1):

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Un algoritmo efficiente è quindi uno strumento essenziale, e deve tenere conto delle difficoltà intrinseche del problema (1):

1. forti non linearità di $E(w)$

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Un algoritmo efficiente è quindi uno strumento essenziale, e deve tenere conto delle difficoltà intrinseche del problema (1):

1. forti non linearità di $E(w)$
2. dimensioni di w elevate e numero di campioni P molto grande

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura
● Stabilizzazione strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Un algoritmo efficiente è quindi uno strumento essenziale, e deve tenere conto delle difficoltà intrinseche del problema (1):

1. forti non linearità di $E(w)$
2. dimensioni di w elevate e numero di campioni P molto grande
3. possibile mal condizionamento dell'hessiana di $E(w)$

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Un algoritmo efficiente è quindi uno strumento essenziale, e deve tenere conto delle difficoltà intrinseche del problema (1):

1. forti non linearità di $E(w)$
2. dimensioni di w elevate e numero di campioni P molto grande
3. possibile mal condizionamento dell'hessiana di $E(w)$
4. presenza di minimi locali non globali

Addestramento

Perceptron

Reti multistrato

Creazione di una rete multistrato

● Struttura e addestramento

● Scelta dell'architettura

● Stabilizzazione

strutturale

● Tecnica di

Regolarizzazione

● Early stopping

● Addestramento

● Classi di algoritmi

● Metodi batch

Metodi Offline

Metodi on-line

Addestrare una rete multistrato richiede in ogni caso di risolvere (più volte) il problema di ottimizzazione non vincolata (con o senza termine di regolarizzazione):

$$\min_{w \in \mathbb{R}} E(w) \quad (1)$$

Un algoritmo efficiente è quindi uno strumento essenziale, e deve tenere conto delle difficoltà intrinseche del problema (1):

1. forti non linearità di $E(w)$
2. dimensioni di w elevate e numero di campioni P molto grande
3. possibile mal condizionamento dell'hessiana di $E(w)$
4. presenza di minimi locali non globali
5. insiemi di livello non compatti (salvo caso in cui si ha il termine di regolarizzazione)

Classi di algoritmi

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- **Classi di algoritmi**
- Metodi batch

Metodi Offline

Metodi on-line

Esistono due classi di algoritmi di addestramento:

Classi di algoritmi

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Esistono due classi di algoritmi di addestramento:

1. metodi **batch**: ad ogni passo i pesi vengono aggiornati utilizzando tutti i campioni del training set (metodi di ottimizzazione non vincolata)

Classi di algoritmi

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

Esistono due classi di algoritmi di addestramento:

1. metodi **batch**: ad ogni passo i pesi vengono aggiornati utilizzando tutti i campioni del training set (metodi di ottimizzazione non vincolata)
2. metodi **on-line**: ad ogni passo i pesi vengono aggiornati tenendo conto di un solo campione del training set (corrispondono ai metodi incrementali, che utilizzano cioè solo un'informazione parziale sulla funzione obiettivo e le sue derivate).

Metodi batch

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- **Metodi batch**

Metodi Offline

Metodi on-line

◁ Sono riconducibili a tecniche di ottimizzazione non vincolata, garantiscono convergenza a punti stazionari (buona riduzione dell'errore rispetto alla stima iniziale)

Metodi batch

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regolarizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- Metodi batch

Metodi Offline

Metodi on-line

- ◁ Sono riconducibili a tecniche di ottimizzazione non vincolata, garantiscono convergenza a punti stazionari (buona riduzione dell'errore rispetto alla stima iniziale)
- ◁ Tipicamente utilizzano soltanto le derivate prime (o troppo dispendiosi)

Metodi batch

Perceptron

Reti multistrato

Creazione di una rete multistrato

- Struttura e addestramento
- Scelta dell'architettura
- Stabilizzazione strutturale
- Tecnica di Regularizzazione
- Early stopping
- Addestramento
- Classi di algoritmi
- **Metodi batch**

Metodi Offline

Metodi on-line

- ◁ Sono riconducibili a tecniche di ottimizzazione non vincolata, garantiscono convergenza a punti stazionari (buona riduzione dell'errore rispetto alla stima iniziale)
- ◁ Tipicamente utilizzano soltanto le derivate prime (o troppo dispendiosi)
- ◁ Sono utilizzabili solo fuori linea, e si deve avere a disposizione l'intero training set

Backpropagation

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

● Backpropagation

● Formule

backpropagation

● Propagazione degli Ingressi

● Errore strato di uscita

● Errore strato nascosto

● Costo computazionale

● Backpropagation

Metodi on-line

1. Backpropagation: $d_k = -\nabla_w E(w)$ (metodo del gradiente): di solito combinato con line search di tipo Armijo (o a passo costante, chiamato **learning rate**). Converge globalmente, ma lentamente. Si è cercato di velocizzarlo con euristiche quali (**momentum updating rule** corrispondente all'heavy ball method):

$$w_{k+1} = w_k - \eta \nabla E(w_k) + \beta(w_k - w_{k-1})$$

Backpropagation

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

● Backpropagation

● Formule

backpropagation

● Propagazione degli Ingressi

● Errore strato di uscita

● Errore strato nascosto

● Costo computazionale

● Backpropagation

Metodi on-line

1. Backpropagation: $d_k = -\nabla_w E(w)$ (metodo del gradiente): di solito combinato con line search di tipo Armijo (o a passo costante, chiamato **learning rate**). Converge globalmente, ma lentamente. Si è cercato di velocizzarlo con euristiche quali (**momentum updating rule** corrispondente all'heavy ball method):

$$w_{k+1} = w_k - \eta \nabla E(w_k) + \beta(w_k - w_{k-1})$$

2. Come calcolare efficientemente $\nabla E(w)$?

$$\min_{w \in \mathbb{R}^m} E(w) = \sum_{p=1}^P E_p(w) = \sum_{p=1}^P \frac{1}{2} \|y(x^p, w) - y^p\|^2$$

dove $E_p(w)$ è il termine di errore relativo al campione p del training set, che misura la distanza tra l'uscita desiderata y^p e l'uscita prodotta dalla rete,

Formule backpropagation

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

● Backpropagation

● Formule

backpropagation

● Propagazione degli Ingressi

● Errore strato di uscita

● Errore strato nascosto

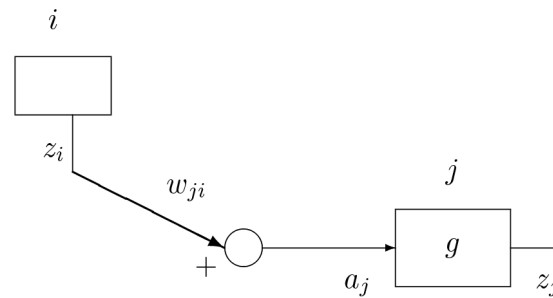
● Costo computazionale

● Backpropagation

Metodi on-line

Consideriamo una rete multistrato con L strati nascosti e definiamo

$$z_i^{(0)} = x_i, \quad i = 1, \dots, n \quad z_i^L = y_i, \quad i = 1, \dots, n$$



Formule backpropagation

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

● Backpropagation

● Formule

backpropagation

● Propagazione degli Ingressi

● Errore strato di uscita

● Errore strato nascosto

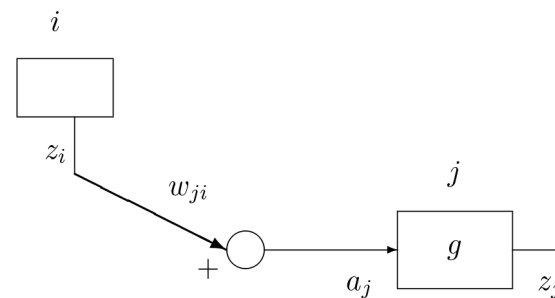
● Costo computazionale

● Backpropagation

Metodi on-line

Consideriamo una rete multistrato con L strati nascosti e definiamo

$$z_i^{(0)} = x_i, \quad i = 1, \dots, n \quad z_i^L = y_i, \quad i = 1, \dots, n$$



Utilizziamo le regole di derivazione si può porre

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}}$$

Propagazione degli Ingressi

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

- Backpropagation

Metodi on-line

Poichè $a_j = \sum_h w_{jh} z_h$, ponendo

$$\delta_j = \frac{\partial E_p}{\partial a_j}$$

si ha

$$\frac{\partial E_p}{\partial w_{ji}} = \delta_j z_j$$

Propagazione in avanti dell'ingresso

For $\ell = 1, \dots, L$

For $j = 1, \dots, N^{(\ell)}$

calcola $a_j^{(\ell)} = \sum_{i=1}^{N^{(\ell-1)}+1} w_{ji}^{(\ell)} z_i^{(\ell-1)}, \quad z_j^{(\ell)} = g_j(a_j^{(\ell)})$

End For

Poni $z_{N^{(\ell)}+1}^{(\ell)} = -1$

End For

For $i = 1, \dots, K$

poni $e_i = z_i^{(L)} - y_i^p$

End For

Propagazione degli Ingressi

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

● Backpropagation

● Formule

backpropagation

● Propagazione degli Ingressi

● Errore strato di uscita

● Errore strato nascosto

● Costo computazionale

● Backpropagation

Metodi on-line

Poichè $a_j = \sum_h w_{jh} z_h$, ponendo

$$\delta_j = \frac{\partial E_p}{\partial a_j}$$

si ha

$$\frac{\partial E_p}{\partial w_{ji}} = \delta_j z_j$$

Le quantità z_j si ottengono propagando in avanti gli ingressi:

Propagazione in avanti dell'ingresso

For $\ell = 1, \dots, L$

For $j = 1, \dots, N^{(\ell)}$

calcola $a_j^{(\ell)} = \sum_{i=1}^{N^{(\ell-1)}+1} w_{ji}^{(\ell)} z_i^{(\ell-1)}, \quad z_j^{(\ell)} = g_j(a_j^{(\ell)})$

End For

Poni $z_{N^{(\ell)}+1}^{(\ell)} = -1$

End For

For $i = 1, \dots, K$

poni $e_i = z_i^{(L)} - y_i^p$

End For

Errore strato di uscita

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

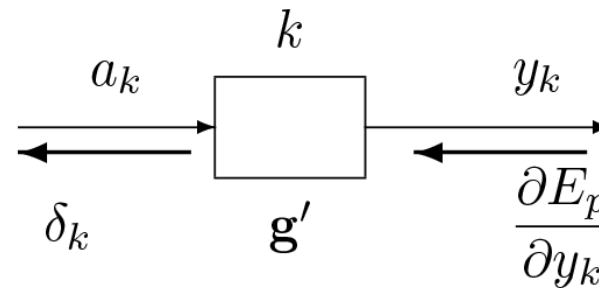
- Costo computazionale

- Backpropagation

Metodi on-line

La quantità δ_j viene chiamata **errore** e si differenzia a seconda se il neurone appartiene allo strato di uscita o a uno strato nascosto

Se il neurone k appartiene allo strato di uscita, si ha $y_k = g(a_k)$:



Errore strato di uscita

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

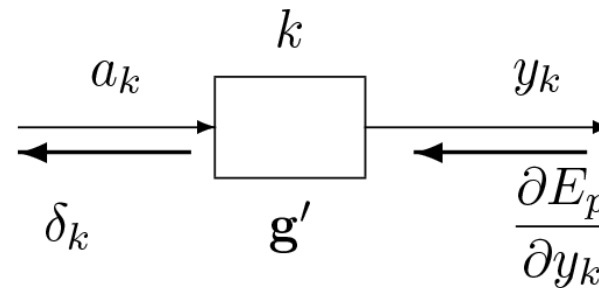
- Costo computazionale

- Backpropagation

Metodi on-line

La quantità δ_j viene chiamata **errore** e si differenzia a seconda se il neurone appartiene allo strato di uscita o a uno strato nascosto

Se il neurone k appartiene allo strato di uscita, si ha $y_k = g(a_k)$:



e quindi si ha

$$\delta_k = \frac{\partial E_p}{\partial a_k} = g(a_k)' \frac{\partial E_p}{\partial y_k}$$

Errore strato nascosto

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

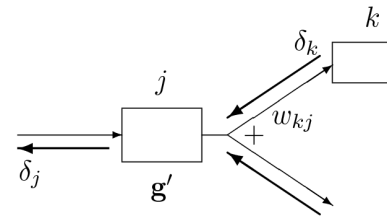
- Backpropagation

Metodi on-line

Se il neurone j appartiene allo strato nascosto, si ha

$$\delta_j = \frac{\partial E_p}{\partial a_j} = \sum_k \frac{\partial E_p}{\partial a_k} \frac{\partial a_k}{\partial a_j},$$

dove la sommatoria è estesa a tutti i neuroni che ricevono segnali dal neurone j



Errore strato nascosto

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

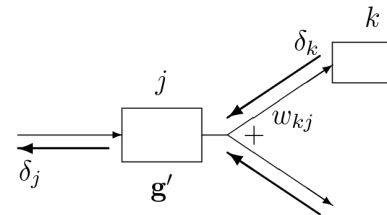
- Backpropagation

Metodi on-line

Se il neurone j appartiene allo strato nascosto, si ha

$$\delta_j = \frac{\partial E_p}{\partial a_j} = \sum_k \frac{\partial E_p}{\partial a_k} \frac{\partial a_k}{\partial a_j},$$

dove la sommatoria è estesa a tutti i neuroni che ricevono segnali dal neurone j



Poichè $a_k = \dots + w_{kj}z_j + \dots = \dots + w_{kj}g(a_j) + \dots$ si ha

$$\frac{\partial a_k}{\partial a_j} = g'(a_j)w_{kj}$$

Errore strato nascosto

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

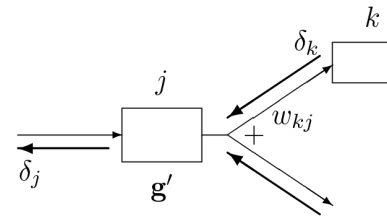
- Backpropagation

Metodi on-line

Se il neurone j appartiene allo strato nascosto, si ha

$$\delta_j = \frac{\partial E_p}{\partial a_j} = \sum_k \frac{\partial E_p}{\partial a_k} \frac{\partial a_k}{\partial a_j},$$

dove la sommatoria è estesa a tutti i neuroni che ricevono segnali dal neurone j



Poichè $a_k = \dots + w_{kj}z_j + \dots = \dots + w_{kj}g(a_j) + \dots$ si ha

$$\frac{\partial a_k}{\partial a_j} = g'(a_j)w_{kj}$$

e quindi

$$\delta_j = g'(a_j) \sum_k \delta_k w_{kj}$$

Costo computazionale

Perceptron
Reti multistrato
Creazione di una rete multistrato
Metodi Offline
● Backpropagation
● Formule
backpropagation
● Propagazione degli Ingressi
● Errore strato di uscita
● Errore strato nascosto
● **Costo computazionale**
● Backpropagation
Metodi on-line

In conclusione, gli ingressi si propagano in avanti, mentre gli errori si propagano all'indietro lungo la rete gli errori relativi ai neuroni dello strato successivo, a partire dallo strato di uscita.

Costo computazionale

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

- Backpropagation

Metodi on-line

In conclusione, gli ingressi si propagano in avanti, mentre gli errori si propagano all'indietro lungo la rete gli errori relativi ai neuroni dello strato successivo, a partire dallo strato di uscita.

La tecnica di backpropagation consente di calcolare $\nabla E = \sum_p \nabla E_p$ con un costo $O(P \times W)$, dove W è il numero di parametri, P il numero di pattern nel training set. È possibile fornire formule di backpropagation anche per il calcolo dell'Hessiana e per il calcolo del prodotto Hessiana per direzione

Costo computazionale

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

- Backpropagation

Metodi on-line

In conclusione, gli ingressi si propagano in avanti, mentre gli errori si propagano all'indietro lungo la rete gli errori relativi ai neuroni dello strato successivo, a partire dallo strato di uscita.

La tecnica di backpropagation consente di calcolare $\nabla E = \sum_p \nabla E_p$ con un costo $O(P \times W)$, dove W è il numero di parametri, P il numero di pattern nel training set. È possibile fornire formule di backpropagation anche per il calcolo dell'Hessiana e per il calcolo del prodotto Hessiana per direzione

Complessivamente si ottiene per una rete con L strati, N ingressi, K uscite e una funzione di errore

$$E_p(w) = \frac{1}{2} \|y(x^p; w) - y^p\|^2$$

Backpropagation

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

- Backpropagation

- Formule

backpropagation

- Propagazione degli Ingressi

- Errore strato di uscita

- Errore strato nascosto

- Costo computazionale

- **Backpropagation**

Metodi on-line

Procedura di Backpropagation per il calcolo di $\nabla E_p(w)$

Dati. Input $x^p \in R^n$, Target $y^p \in R^K$.

Poni $z_i^{(0)} = x_i^p, i = 1, \dots, n, z_{n+1}^{(0)} = -1$

Propagazione in avanti dell'ingresso

For $\ell = 1, \dots, L$

For $j = 1, \dots, N^{(\ell)}$

calcola $a_j^{(\ell)} = \sum_{i=1}^{N^{(\ell-1)+1}} w_{ji}^{(\ell)} z_i^{(\ell-1)}, \quad z_j^{(\ell)} = g_j(a_j^{(\ell)})$

End For

Poni $z_{N^{(\ell)+1}}^{(\ell)} = -1$

End For

For $i = 1, \dots, K$

poni $e_i = z_i^{(L)} - y_i^p$

End For

Retro-propagazione dell'errore

For $j = 1, \dots, K$

calcola $\delta_j^{(L)} = e_j g_j'(a_j^{(L)})$

poni $\frac{\partial E_p}{\partial w_{ji}^{(L)}} = \delta_j^{(L)} z_i^{(L-1)}$

End For

For $\ell = L - 1, \dots, 1$

For $j = 1, \dots, N^{(\ell)}$

calcola $\delta_j^\ell = g_j'(a_j^{(\ell)}) \sum_{k=1}^{N^{(\ell+1)}} \delta_k^{(\ell+1)} w_{kj}^{(\ell+1)}$

For $i = 1, \dots, N^{(\ell-1)} + 1$

poni $\frac{\partial E_p}{\partial w_{ji}^{(\ell)}} = \delta_j^\ell z_i^{(\ell-1)}$

End For

End For

End For

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).
- ◁ La convergenza può essere caratterizzata solo in termini probabilistici.

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).
- ◁ La convergenza può essere caratterizzata solo in termini probabilistici.
- ◁ Presentano alcuni vantaggi anche nell'addestramento fuori linea:

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).
- ◁ La convergenza può essere caratterizzata solo in termini probabilistici.
- ◁ Presentano alcuni vantaggi anche nell'addestramento fuori linea:
 1. a grande distanza dalla soluzione può non essere conveniente calcolare tutta la funzione e le sue derivate: conviene ridurre rapidamente un singolo termine di errore

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).
- ◁ La convergenza può essere caratterizzata solo in termini probabilistici.
- ◁ Presentano alcuni vantaggi anche nell'addestramento fuori linea:
 1. a grande distanza dalla soluzione può non essere conveniente calcolare tutta la funzione e le sue derivate: conviene ridurre rapidamente un singolo termine di errore
 2. se il training set e il numero di campioni sono molto elevati il calcolo di E e ∇E può essere troppo dispendioso

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

● Metodi on-line

● Backpropagation on-line

● Backpropagation on-line

● Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).
- ◁ La convergenza può essere caratterizzata solo in termini probabilistici.
- ◁ Presentano alcuni vantaggi anche nell'addestramento fuori linea:
 1. a grande distanza dalla soluzione può non essere conveniente calcolare tutta la funzione e le sue derivate: conviene ridurre rapidamente un singolo termine di errore
 2. se il training set e il numero di campioni sono molto elevati il calcolo di E e ∇E può essere troppo dispendioso
 3. il training set potrebbe essere ridondante

Metodi on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

● Metodi on-line

● Backpropagation on-line

● Backpropagation on-line

● Metodi globali

- ◁ Il vettore w viene aggiornato iterativamente in corrispondenza a ogni singolo termine $E_p(w)$ della funzione di errore senza formare la funzione complessiva $E(w)$.
- ◁ Per problemi di apprendimento in tempo reale sono l'unica opzione possibile (on-line learning).
- ◁ La convergenza può essere caratterizzata solo in termini probabilistici.
- ◁ Presentano alcuni vantaggi anche nell'addestramento fuori linea:
 1. a grande distanza dalla soluzione può non essere conveniente calcolare tutta la funzione e le sue derivate: conviene ridurre rapidamente un singolo termine di errore
 2. se il training set e il numero di campioni sono molto elevati il calcolo di E e ∇E può essere troppo dispendioso
 3. il training set potrebbe essere ridondante
 4. la componente di randomicità di un algoritmo on-line può permettere di evitare minimi locali

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- **Backpropagation
on-line**
- Backpropagation
on-line
- Metodi globali

Il primo esempio di metodo on-line è l'addestramento del perceptron

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- **Backpropagation on-line**
- Backpropagation on-line
- Metodi globali

Il primo esempio di metodo on-line è l'addestramento del perceptron

Per le reti multistrato il metodo più noto è la versione on-line della backpropagation, ed è descritto da un'iterazione del tipo

$$w_{k+1} = w_k - \eta_k \nabla E_{p(k)}(w_k)$$

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

Il primo esempio di metodo on-line è l'addestramento del perceptron

Per le reti multistrato il metodo più noto è la versione on-line della backpropagation, ed è descritto da un'iterazione del tipo

$$w_{k+1} = w_k - \eta_k \nabla E_{p(k)}(w_k)$$

Per ottenere la convergenza, tutti i termini E_p devono essere considerati ciclicamente, ovvero $p(k) = k(\text{mod}P) + 1$, e il passo η_k deve essere scelto in modo da soddisfare:

$$\sum_{k=0}^{\infty} \eta_k = \infty, \quad \sum_{k=0}^{\infty} (\eta_k)^2 < \infty$$

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

Il primo esempio di metodo on-line è l'addestramento del perceptron

Per le reti multistrato il metodo più noto è la versione on-line della backpropagation, ed è descritto da un'iterazione del tipo

$$w_{k+1} = w_k - \eta_k \nabla E_{p(k)}(w_k)$$

Per ottenere la convergenza, tutti i termini E_p devono essere considerati ciclicamente, ovvero $p(k) = k(\text{mod}P) + 1$, e il passo η_k deve essere scelto in modo da soddisfare:

$$\sum_{k=0}^{\infty} \eta_k = \infty, \quad \sum_{k=0}^{\infty} (\eta_k)^2 < \infty$$

Un esempio di scelta di η_k è $\eta_k = c/k$ con $c > 0$.

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

Il primo esempio di metodo on-line è l'addestramento del perceptron

Per le reti multistrato il metodo più noto è la versione on-line della backpropagation, ed è descritto da un'iterazione del tipo

$$w_{k+1} = w_k - \eta_k \nabla E_{p(k)}(w_k)$$

Per ottenere la convergenza, tutti i termini E_p devono essere considerati ciclicamente, ovvero $p(k) = k(\text{mod}P) + 1$, e il passo η_k deve essere scelto in modo da soddisfare:

$$\sum_{k=0}^{\infty} \eta_k = \infty, \quad \sum_{k=0}^{\infty} (\eta_k)^2 < \infty$$

Un esempio di scelta di η_k è $\eta_k = c/k$ con $c > 0$.

Esiste anche la versione on-line della momentum updating rule (con proprietà di convergenza analoghe):

$$w_{k+1} = w_k - \eta_k \nabla E_{p(k)}(w_k) + \beta(w_k - w_{k-1})$$

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- **Backpropagation on-line**
- Metodi globali

La principale limitazione è la lentezza (dovuta alla scelta del passo che deve tendere a zero con una legge prefissata)

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- **Backpropagation on-line**
- Metodi globali

La principale limitazione è la lentezza (dovuta alla scelta del passo che deve tendere a zero con una legge prefissata)

Nell'addestramento fuori-linea un compromesso è l'uso di tecniche ibride on-line-batch chiamate **bold driver methods**: la backpropagation on-line è usata per uno o più cicli di iterazioni e il passo è ricalcolato ciclicamente usando la funzione di errore complessiva (esistono risultati di convergenza). Queste tecniche però richiedono il calcolo di E e ∇E .

Backpropagation on-line

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- **Backpropagation on-line**
- Metodi globali

La principale limitazione è la lentezza (dovuta alla scelta del passo che deve tendere a zero con una legge prefissata)

Nell'addestramento fuori-linea un compromesso è l'uso di tecniche ibride on-line-batch chiamate **bold driver methods**: la backpropagation on-line è usata per uno o più cicli di iterazioni e il passo è ricalcolato ciclicamente usando la funzione di errore complessiva (esistono risultati di convergenza). Queste tecniche però richiedono il calcolo di E e ∇E .

Un altro tipo di compromesso si ottiene passando gradualmente dalla backpropagation on-line a quella batch, per accelerare la convergenza.

Metodi globali

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

La presenza di minimi locali non globali della funzione di errore è stata provata per alcuni specifici problemi test e rappresenta una delle difficoltà dei problemi di addestramento

Metodi globali

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation
on-line
- Backpropagation
on-line
- Metodi globali

La presenza di minimi locali non globali della funzione di errore è stata provata per alcuni specifici problemi test e rappresenta una delle difficoltà dei problemi di addestramento

NB: non esistono minimi locali non globali nel caso di problemi di classificazione di insiemi linearmente separabili.

Metodi globali

Perceptron

Reti multistrato

Creazione di una rete
multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

La presenza di minimi locali non globali della funzione di errore è stata provata per alcuni specifici problemi test e rappresenta una delle difficoltà dei problemi di addestramento

NB: non esistono minimi locali non globali nel caso di problemi di classificazione di insiemi linearmente separabili.

Spesso la difficoltà maggiore è la presenza di plateau.

Metodi globali

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

La presenza di minimi locali non globali della funzione di errore è stata provata per alcuni specifici problemi test e rappresenta una delle difficoltà dei problemi di addestramento

NB: non esistono minimi locali non globali nel caso di problemi di classificazione di insiemi linearmente separabili.

Spesso la difficoltà maggiore è la presenza di plateau.

In realtà quello che si vuole è un errore ragionevolmente basso sul training set che garantisca una buona capacità di generalizzazione.

Metodi globali

Perceptron

Reti multistrato

Creazione di una rete multistrato

Metodi Offline

Metodi on-line

- Metodi on-line
- Backpropagation on-line
- Backpropagation on-line
- Metodi globali

La presenza di minimi locali non globali della funzione di errore è stata provata per alcuni specifici problemi test e rappresenta una delle difficoltà dei problemi di addestramento

NB: non esistono minimi locali non globali nel caso di problemi di classificazione di insiemi linearmente separabili.

Spesso la difficoltà maggiore è la presenza di plateau.

In realtà quello che si vuole è un errore ragionevolmente basso sul training set che garantisca una buona capacità di generalizzazione.

Per questo i metodi globali (troppo onerosi a grandi dimensioni) non sono molto usati. A volte si utilizza una prima fase stocastica per eliminare dei minimi non interessanti e poi si usano metodi locali.