Introduzione alle Reti Neurali

E.Mumolo

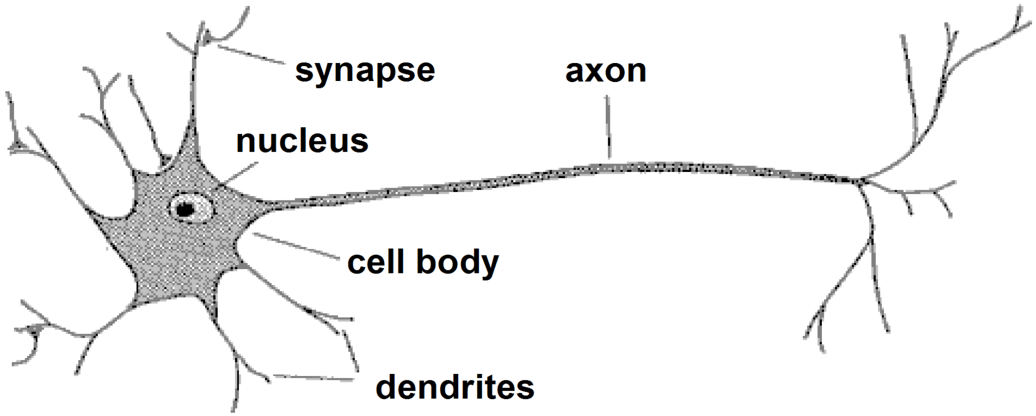
mumolo@units.it

Reti neurali artificiali

Modelli di calcolo parallelo vagamente ispirati dalla biologia.
Caratteristiche principali:

- Una rete non è programmata per realizzare un algoritmo ma apprende tramite esempi
- Capacità di apprendimento: fornendo esempi opportuni una rete riesce a organizzarsi per realizzare una funzione arbitraria y= f (x)
- Capacità di generalizzare: la rete correttamente addestrata può fornire una risposta anche a richieste non incluse negli esempi
- Tolleranza agli errori: indipendenza al rumore di ingresso
- Tolleranza ai guasti funzionamento ragionevole anche se qualche neurone non funziona correttamente

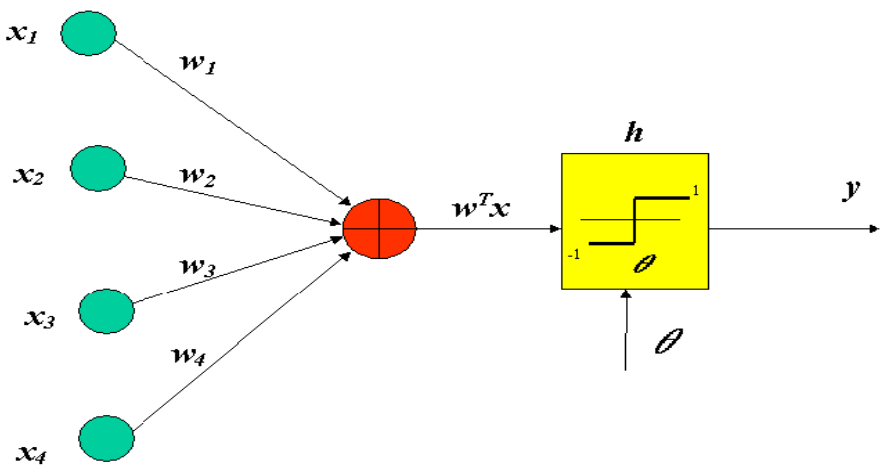
Neurone biologico



- Ingressi: Dendriti
- Uscita: Assone
- Connessione Assone-Dendriti mediante sinapsi
- I neuroni agiscono a soglia → quando la somma degli ingressi supera un certo valore il neurone produce una uscita

Neurone artificiale

- Funzione non lineare



$w_i \in R$ **pesi**

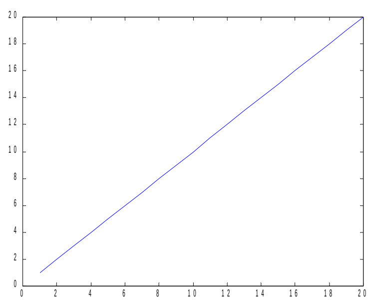
$\theta \in R$ **soglia**

h : **funzione di attivazione**

$$y(x) = h\,(\,w^T x - \theta\,)$$

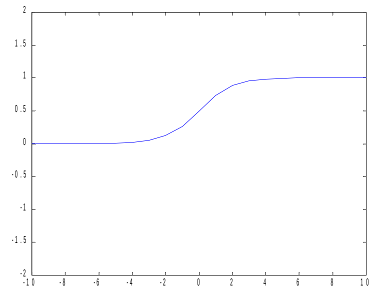
Uscita: 1 se somma pesata degli ingressi > soglia; -1 altrimenti

Funzioni di attivazione



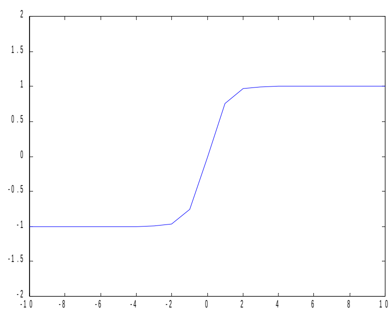
Lineare

$$y = x$$



Sigmoide

$$y = \frac{1}{1 + \exp(-x)}$$



Tangente iperbolica

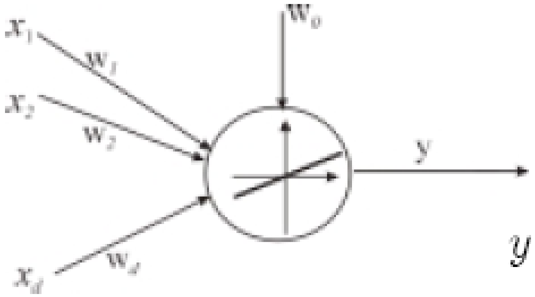
$$y = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$

Strutture a singolo strato → Percettrone

Percettroni per interpolazione e regressione

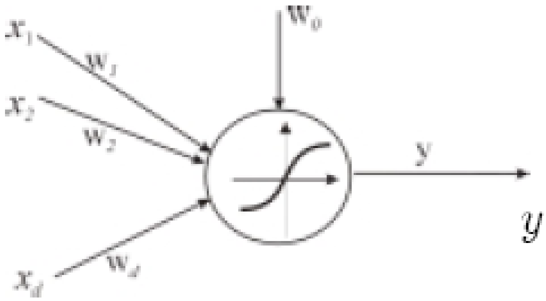
Percettrone lineare:

- implementa solo una funzione lineare



$$y = (\bar{x}' \cdot \bar{w} - w_0)\beta$$

Percettrone sigmoidale:



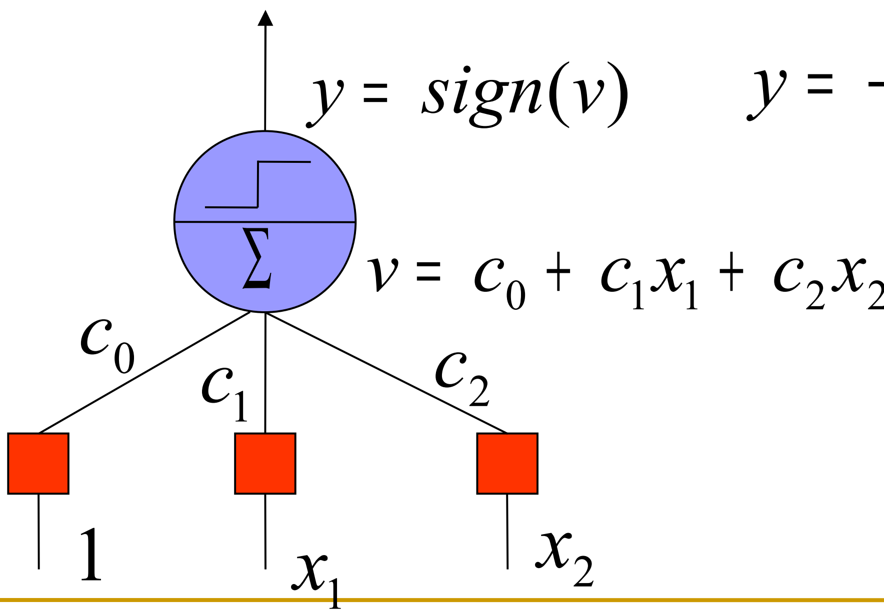
$$y = \sigma(\bar{x}' \cdot \bar{w} - w_0)$$

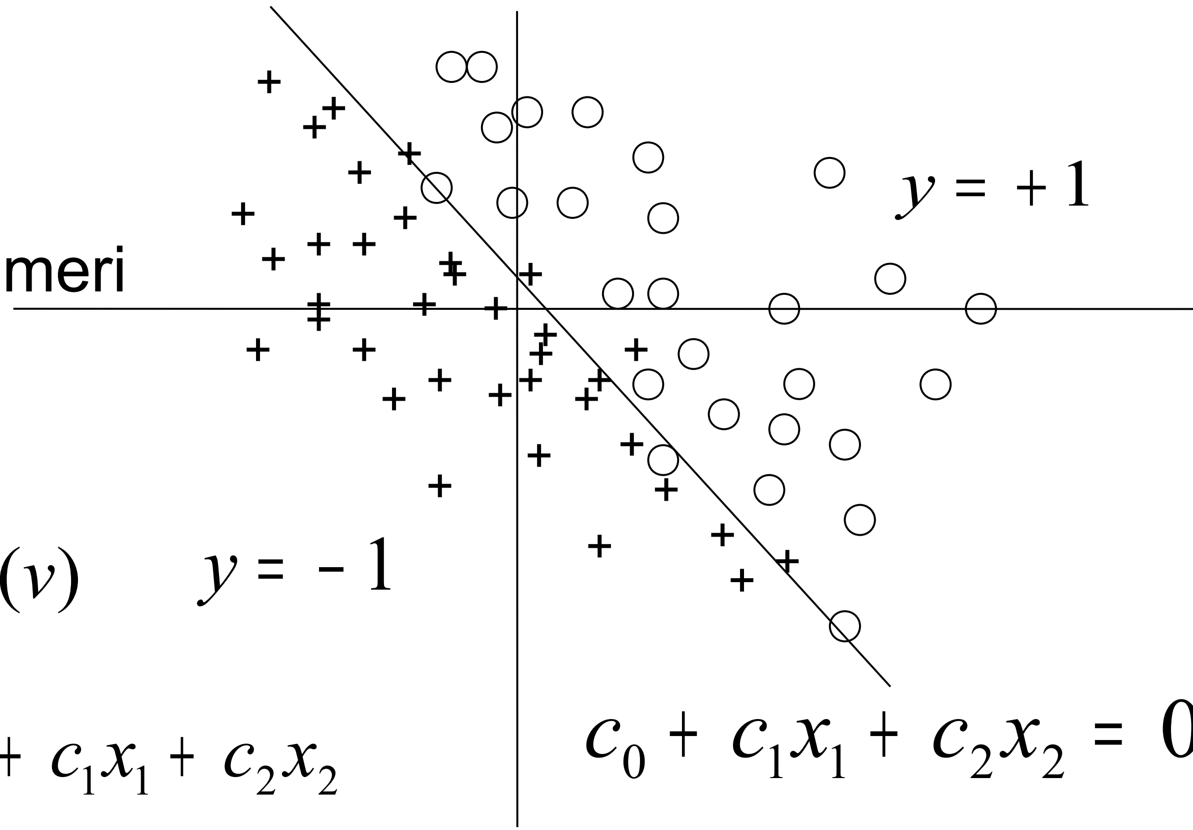
$\sigma(\cdot)$: funzione di attivazione sigmoidale, cioe' monotona, non decrescente tale che

$$\lim_{z \to +\infty} \sigma(z) = 1 \qquad \lim_{z \to -\infty} \sigma(z) = -1$$

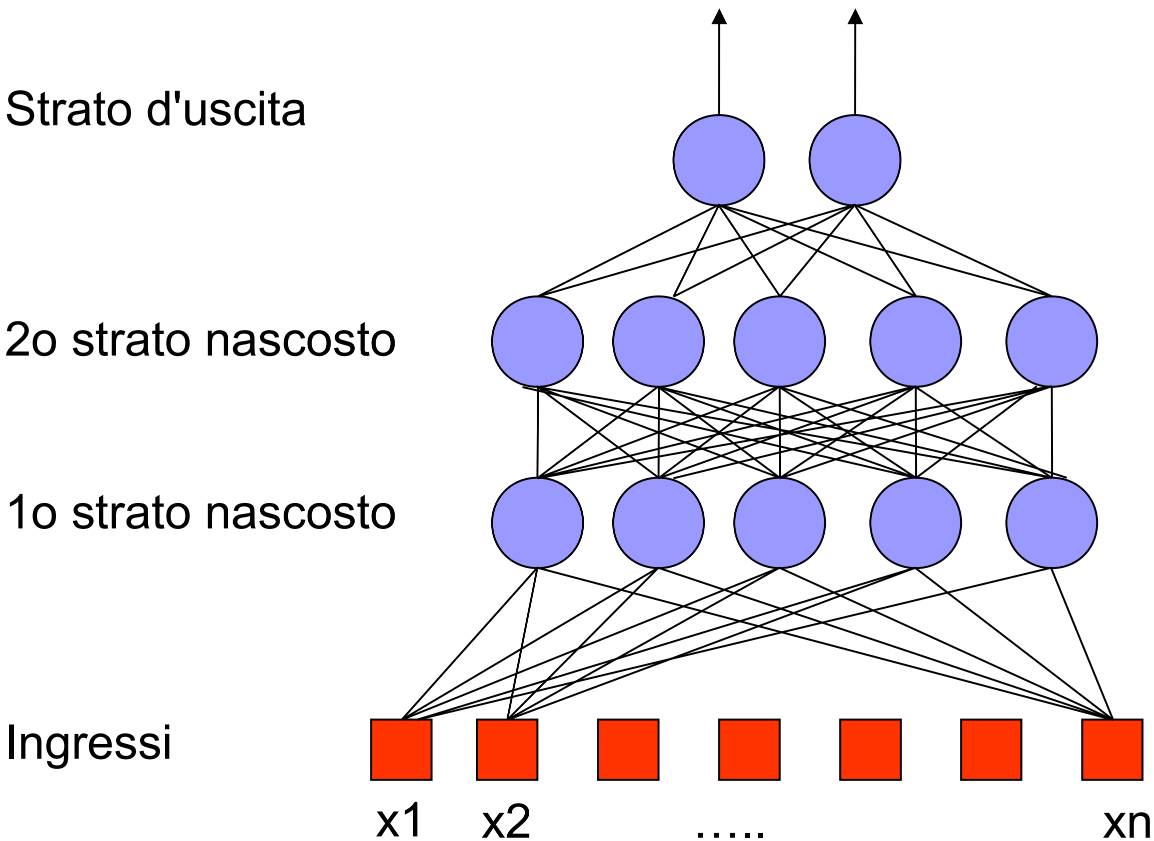
Percettrone

- Rosenblatt (1962)
- Separatore lineare
- Ingressi: vettori di numeri reali
- Uscita :1 or -1





Strutture MultiStrato
→ Rete Neurale Feed Forward

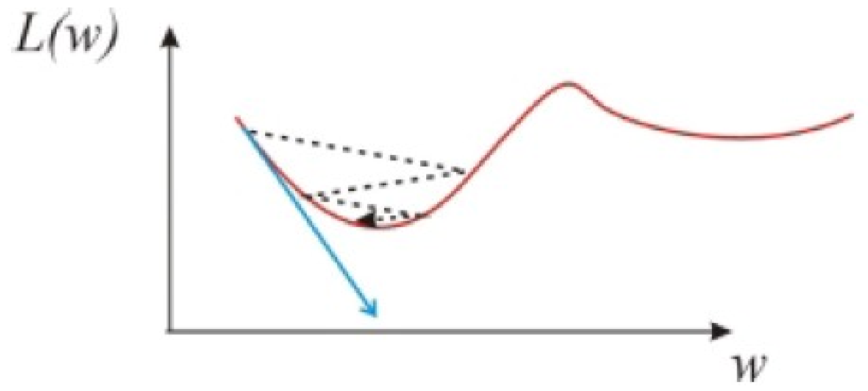


Proprietà fondamentale delle reti

- Approssimatori universali
- Teorema: Ogni funzione limitata può essere approssimata con una rete neurale con un numero finito di strati nascosti con precisione arbitraria
- Tipi di approssimatori;
  - Lineari: il numero di parametri cresce esponenzialmente con le variabili
  - Non lineari: il numero di parametri cresce linearmente con le variabili

Addestramento della rete (Apprendimento)

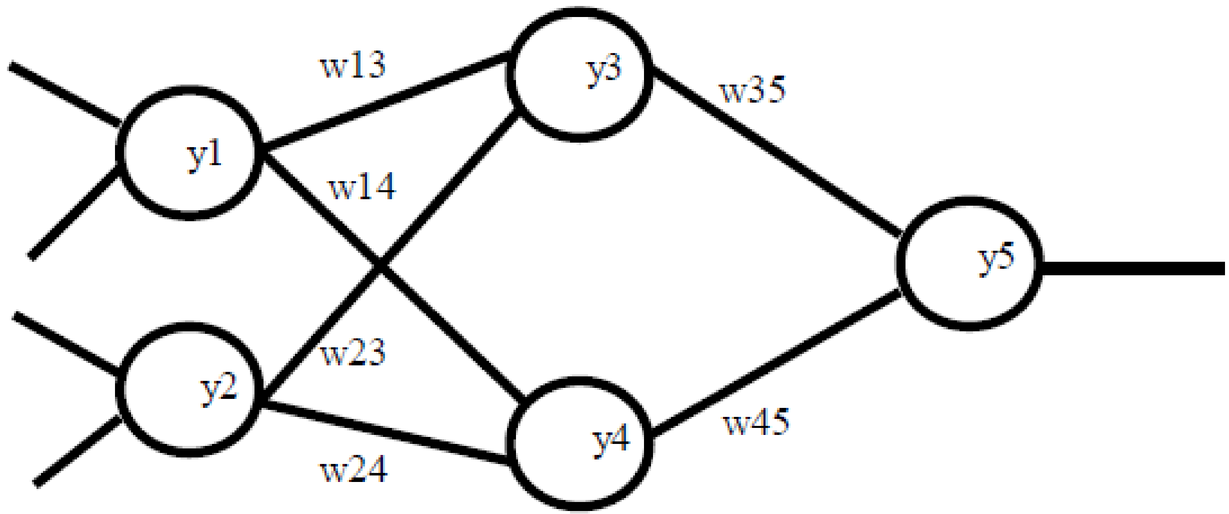
Metodo del gradiente



- Update: $w(k+1) = w(k) - \eta \left.\frac{\partial L(w)}{\partial w}\right|_{w=w(k)} \qquad \eta > 0$
- Condizione di terminazione (Es. errore relativo $\frac{|L(w(k+1)) - L(w(k))|}{|L(w(k))|} \leq \epsilon$)

Backpropagation

Supponiamodi considerare questa porzione semplificata di rete



dove wij sono i pesi delle connessioni e i valori calcolati sono i seguenti:

$$y_3=\frac{1}{1+e^{-(w_{13}y_1+w_{23}y_2)}};\ y_4=\frac{1}{1+e^{-(w_{14}y_1+w_{24}y_2)}};\ y_5=\frac{1}{1+e^{-(w_{35}y_3+w_{45}y_4)}}$$

Backpropagation

L'aggiornamento dei pesi col metodo del gradiente procede come segue:

Definizione della funzione d'errore: $E=\frac{(d-y)^2}{2}$ dove d è l'uscita desiderata mentre y è l'uscita della rete, y5. Per cui:

$$\frac{\partial E}{\partial w_{ij}}=\frac{\partial E}{\partial y}\frac{\partial y}{\partial w_{ij}}=-(d-y)\frac{\partial y}{\partial w_{ij}}=-\delta\frac{\partial y}{\partial w_{ij}}$$

A questo punto calcoliamo le derivate parziali di y rispetto ai pesi.
Cominciamo a porre le uscita in questa forma:

$$y=y_5=\frac{1}{1+e^{-(w_{35}y_3+w_{45}y_4)}}=\frac{1}{1+e^{-a}},\ a=(w_{35}y_3+w_{45}y_4)$$

$$y_3=\frac{1}{1+e^{-(w_{13}y_1+w_{23}y_2)}}=\frac{1}{1+e^{-b}},\ b=(w_{13}y_1+w_{23}y_2)$$

$$y_4=\frac{1}{1+e^{-(w_{14}y_1+w_{24}y_2)}}=\frac{1}{1+e^{-c}},\ c=(w_{14}y_1+w_{24}y_2)$$

Backpropagation

quindi

$$\frac{\partial y}{\partial w_{35}}=\frac{\partial}{\partial a}\left(\frac{1}{1+e^{-a}}\right)\frac{\partial a}{\partial w_{35}}=\frac{-e^{-a}}{(1+e^{-a})^2}y_3=y(1-y)y_3$$

$$\frac{\partial y}{\partial w_{45}}=\frac{\partial}{\partial a}\left(\frac{1}{1+e^{-a}}\right)\frac{\partial a}{\partial w_{45}}=\frac{-e^{-a}}{(1+e^{-a})^2}y_4=y(1-y)y_4$$

$$\frac{\partial y}{\partial w_{13}}=\frac{\partial y}{\partial a}\frac{\partial a}{\partial y_3}\frac{\partial y_3}{\partial b}\frac{\partial b}{\partial w_{13}}=\frac{-e^{-a}}{(1+e^{-a})^2}w_{35}\frac{-e^{-b}}{(1+e^{-b})^2}y_1=y(1-y)w_{35}y_3(1-y_3)y_1$$

$$\frac{\partial y}{\partial w_{14}}=\frac{\partial y}{\partial a}\frac{\partial a}{\partial y_4}\frac{\partial y_4}{\partial c}\frac{\partial c}{\partial w_{14}}=\frac{-e^{-a}}{(1+e^{-a})^2}w_{45}\frac{-e^{-c}}{(1+e^{-c})^2}y_1=y(1-y)w_{45}y_4(1-y_4)y_1$$

$$\frac{\partial y}{\partial w_{23}}=\frac{\partial y}{\partial a}\frac{\partial a}{\partial y_3}\frac{\partial y_3}{\partial b}\frac{\partial b}{\partial w_{23}}=\frac{-e^{-a}}{(1+e^{-a})^2}w_{35}\frac{-e^{-b}}{(1+e^{-b})^2}y_2=y(1-y)w_{45}y_3(1-y_3)y_2$$

$$\frac{\partial y}{\partial w_{24}}=\frac{\partial y}{\partial a}\frac{\partial a}{\partial y_4}\frac{\partial y_4}{\partial c}\frac{\partial c}{\partial w_{24}}=\frac{-e^{-a}}{(1+e^{-a})^2}w_{45}\frac{-e^{-c}}{(1+e^{-c})^2}y_2=y(1-y)w_{45}y_4(1-y_3)y_2$$

Backpropagation

L'aggiornamento dei pesi con il metodo del gradiente procedono come segue:

$$w'_{35} = w_{35} - \eta \frac{\partial E}{\partial w_{35}} = w_{35} + \eta \delta y (1 - y) y_3$$

$$w'_{45} = w_{35} - \eta \frac{\partial E}{\partial w_{35}} = w_{35} + \eta \delta y (1 - y) y_4$$

$$w'_{13} = w_{35} - \eta \frac{\partial E}{\partial w_{13}} = w_{35} + \eta w_{35} \delta y (1 - y) y_3 (1 - y_3) y_1$$

$$w'_{14} = w_{14} - \eta \frac{\partial E}{\partial w_{14}} = w_{14} + \eta w_{45} \delta y (1 - y) y_4 (1 - y_4) y_1$$

$$w'_{23} = w_{23} - \eta \frac{\partial E}{\partial w_{23}} = w_{23} + \eta w_{35} \delta y (1 - y) y_3 (1 - y_3) y_2$$

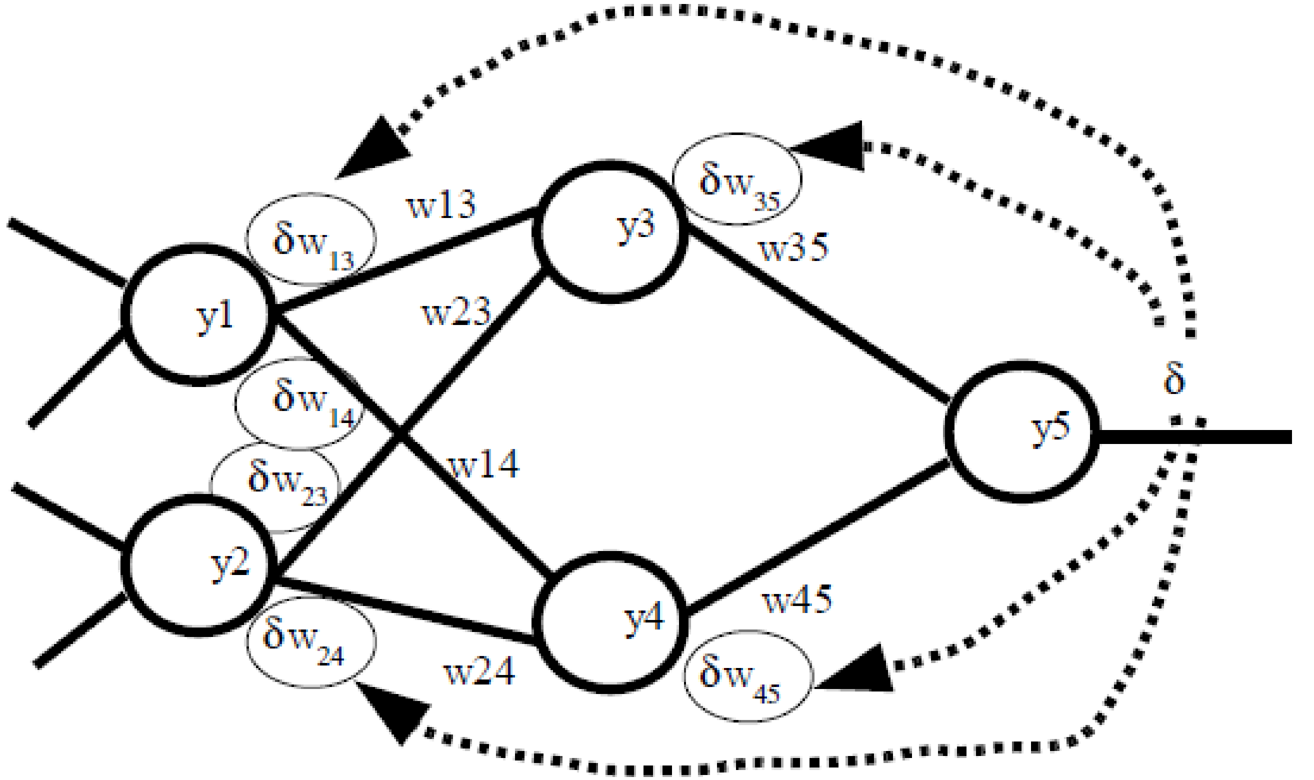
$$w'_{24} = \text{¿} w_{24} - \eta \frac{\partial E}{\partial w_{24}} = w_{24} + \eta w_{45} \delta y (1 - y) y_4 (1 - y_4) y_2$$

Perchè 'Backpropagation'?
Perchè dalle formule si vede che l'errore δ viene propagato all'indietro, dall'ultimo nodo ai nodi precedenti, con il peso della connessione precedente.

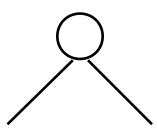
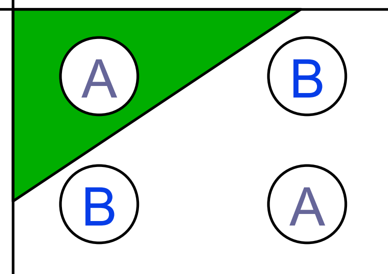
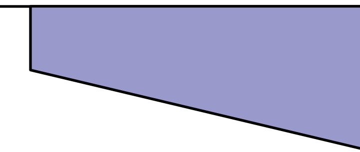
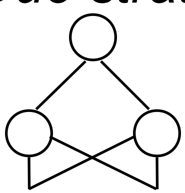
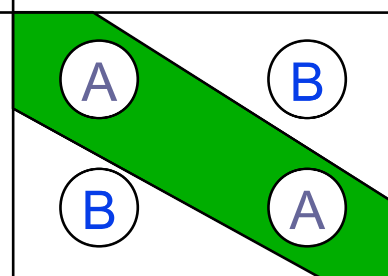
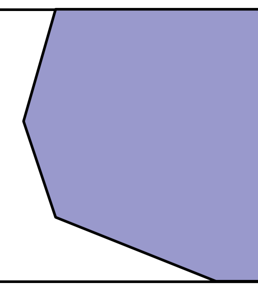
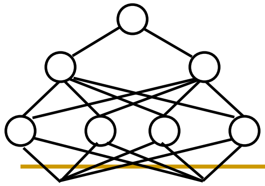
In altri termini i pesi w_{ij} vengono aggiustati con la relazione $w_{ij}' = w_{ij} + \eta\ \Delta w_{ij}$ dove Δw_{ij} si calcola come segue:

$$\Delta w_{ij} = \delta w_{j,j+1}\ y_i y_j (1-y)$$

Propagazione dell'errore:



Problemi non separabili linearmente

Strutture	*Tipi di regioni*	*Problemi XOR*	*Configurazioni generiche*
Singolo strato 	*Semipiano*		
Due strati 	*Regioni convesse*		
Tre strati 	Regioni di arbitraria complessità