

Calcolo del gradiente con sigmoide

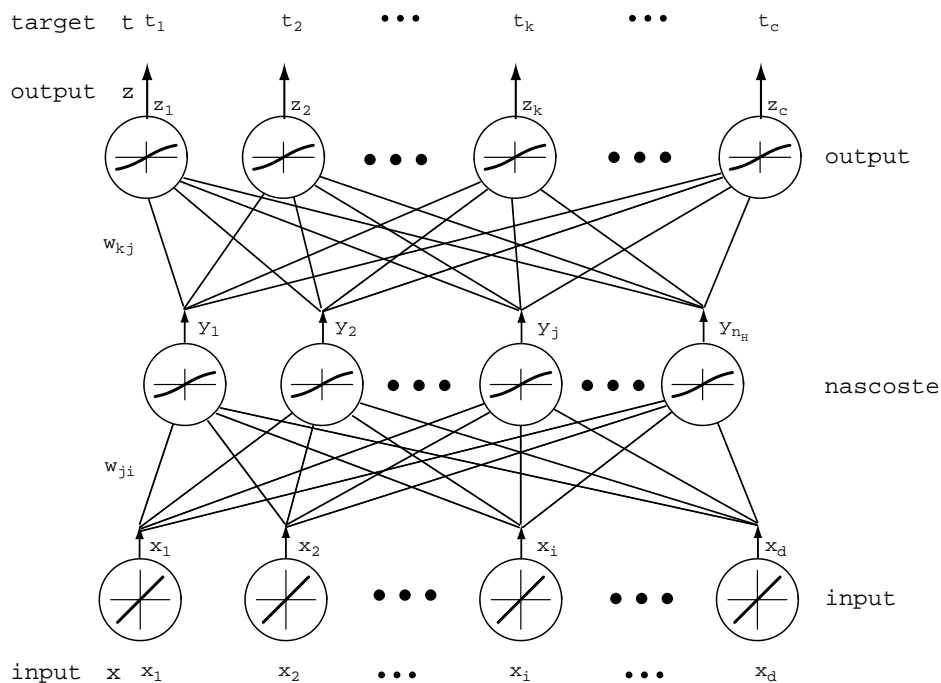
$$\begin{aligned}
 \frac{\partial E}{\partial w_i} &= \frac{\partial}{\partial w_i} \frac{1}{2N_{Tr}} \sum_{d \in Tr} (t^{(d)} - out^{(d)})^2 \\
 &= \frac{1}{2N_{Tr}} \sum_{d \in Tr} \frac{\partial}{\partial w_i} (t^{(d)} - out^{(d)})^2 \\
 &= \frac{1}{2N_{Tr}} \sum_{d \in Tr} 2(t^{(d)} - out^{(d)}) \frac{\partial}{\partial w_i} (t^{(d)} - out^{(d)}) \\
 &= \frac{1}{N_{Tr}} \sum_{d \in Tr} (t^{(d)} - out^{(d)}) \frac{\partial}{\partial w_i} (t^{(d)} - \sigma(\vec{w} \cdot \vec{x}^{(d)})) \\
 \frac{\partial E}{\partial w_i} &= \frac{1}{N_{Tr}} \sum_{d \in Tr} (t^{(d)} - out^{(d)}) \left(-\frac{\partial \sigma(\vec{w} \cdot \vec{x}^{(d)})}{\partial \vec{w} \cdot \vec{x}^{(d)}} \frac{\partial \vec{w} \cdot \vec{x}^{(d)}}{\partial w_i} \right) \\
 &= -\frac{1}{N_{Tr}} \sum_{d \in Tr} (t^{(d)} - out^{(d)}) \sigma(\vec{w} \cdot \vec{x}^{(d)}) (1 - \sigma(\vec{w} \cdot \vec{x}^{(d)})) x_i^{(d)}
 \end{aligned}$$

Alessandro Sperduti & Francesca Rossi

Sistemi di Elaborazione dell'Informazione

117

Reti Neurali Feed-forward: notazione



Alessandro Sperduti & Francesca Rossi

Reti Neurali Feed-forward: notazione

- d unità di ingresso, dimensione dei dati in ingresso $\vec{x} \equiv (x_1, \dots, x_d)$
($d + 1$ se si include la soglia nel vettore dei pesi $\vec{x}' \equiv (x_0, x_1, \dots, x_d)$)
- N_H unità nascoste (con output $\vec{y} \equiv (y_1, \dots, y_{N_H})$)
- c unità di output, dimensione dei dati in output $\vec{z} \equiv (z_1, \dots, z_c)$
- c , dimensione dei dati desiderati $\vec{t} \equiv (t_1, \dots, t_c)$
- w_{ji} peso dalla unità di ingresso i alla unità nascosta j
- w_{kj} peso dalla unità nascosta j alla unità di output k

La funzione errore, considerando che si hanno c unità di output, diventa

$$E[\vec{w}] = \frac{1}{2cN_{Tr}} \sum_{(\vec{x}^{(p)}, \vec{t}^{(p)}) \in Tr} \sum_{k=1}^c \left(t_k^{(p)} - z_k(\vec{x}^{(p)}) \right)^2$$

Alessandro Sperduti & Francesca Rossi

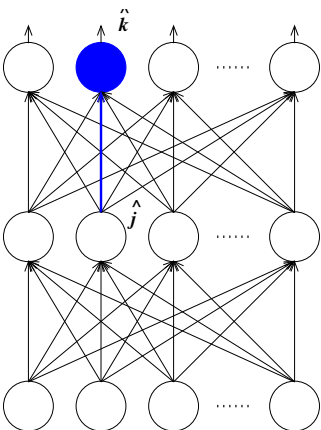
Sistemi di Elaborazione dell'Informazione

119

Calcolo del gradiente per i pesi di una unità di output

Fissiamo gli indici \hat{k} e \hat{j} :

$$\frac{\partial E}{\partial w_{\hat{k}\hat{j}}} = \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2$$

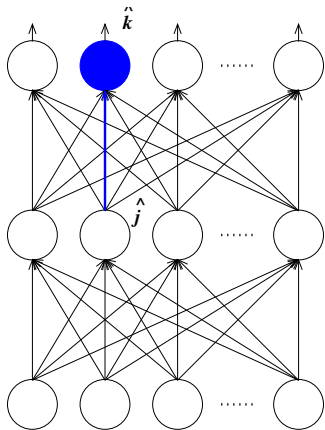


Alessandro Sperduti & Francesca Rossi

Calcolo del gradiente per i pesi di una unità di output

Fissiamo gli indici \hat{k} e \hat{j} :

$$\begin{aligned} \frac{\partial E}{\partial w_{\hat{k}\hat{j}}} &= \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \end{aligned}$$



Alessandro Sperduti & Francesca Rossi

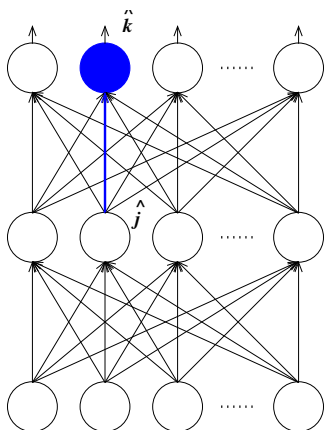
Sistemi di Elaborazione dell'Informazione

121

Calcolo del gradiente per i pesi di una unità di output

Fissiamo gli indici \hat{k} e \hat{j} :

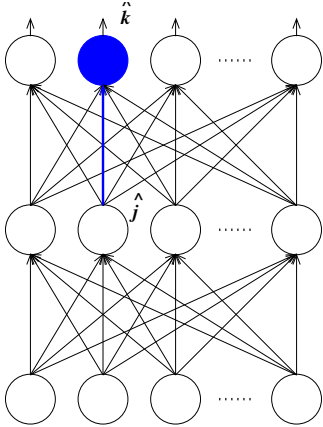
$$\begin{aligned} \frac{\partial E}{\partial w_{\hat{k}\hat{j}}} &= \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} 2(t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \frac{\partial}{\partial w_{\hat{k}\hat{j}}} (t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \end{aligned}$$



Alessandro Sperduti & Francesca Rossi

Calcolo del gradiente per i pesi di una unità di output

Fissiamo gli indici \hat{k} e \hat{j} :



$$\begin{aligned}
 \frac{\partial E}{\partial w_{\hat{k}\hat{j}}} &= \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} 2(t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \frac{\partial}{\partial w_{\hat{k}\hat{j}}} (t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \\
 &= \frac{1}{cN_{Tr}} \sum_{p \in Tr} (t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \frac{\partial}{\partial w_{\hat{k}\hat{j}}} (t_{\hat{k}}^{(p)} - \sigma(\vec{w}_{\hat{k}} \cdot \vec{y}^{(p)}))
 \end{aligned}$$

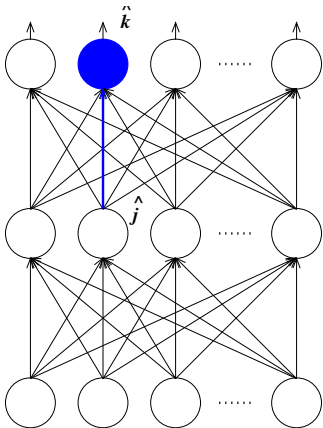
Alessandro Sperduti & Francesca Rossi

Sistemi di Elaborazione dell'Informazione

123

Calcolo del gradiente per i pesi di una unità di output

Fissiamo gli indici \hat{k} e \hat{j} :



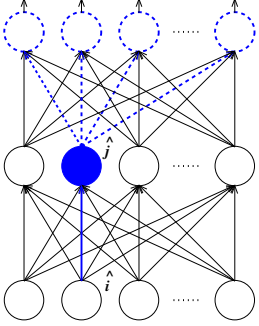
$$\begin{aligned}
 \frac{\partial E}{\partial w_{\hat{k}\hat{j}}} &= \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \frac{\partial}{\partial w_{\hat{k}\hat{j}}} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} 2(t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \frac{\partial}{\partial w_{\hat{k}\hat{j}}} (t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \\
 &= \frac{1}{cN_{Tr}} \sum_{p \in Tr} (t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \frac{\partial}{\partial w_{\hat{k}\hat{j}}} (t_{\hat{k}}^{(p)} - \sigma(\vec{w}_{\hat{k}} \cdot \vec{y}^{(p)})) \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} (t_{\hat{k}}^{(p)} - z_{\hat{k}}^{(p)}) \sigma'(\vec{w}_{\hat{k}} \cdot \vec{y}^{(p)}) y_{\hat{j}}^{(p)}
 \end{aligned}$$

Alessandro Sperduti & Francesca Rossi

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\frac{\partial E}{\partial w_{\hat{j}\hat{i}}} = \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2$$



Alessandro Sperduti & Francesca Rossi

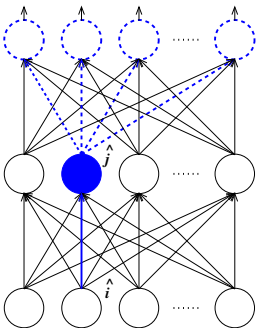
Sistemi di Elaborazione dell'Informazione

125

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned} \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (t_k^{(p)} - z_k^{(p)})^2 \end{aligned}$$

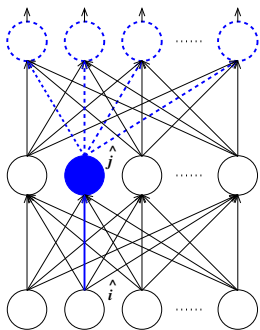


Alessandro Sperduti & Francesca Rossi

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned} \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c 2(t_k^{(p)} - z_k^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (-z_k^{(p)}) \end{aligned}$$



Alessandro Sperduti & Francesca Rossi

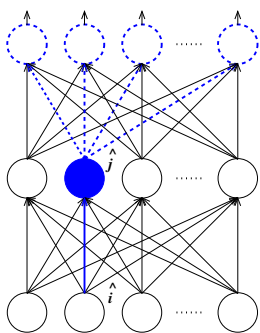
Sistemi di Elaborazione dell'Informazione

127

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned} \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (t_k^{(p)} - z_k^{(p)})^2 \\ &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c 2(t_k^{(p)} - z_k^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (-z_k^{(p)}) \\ &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \vec{w}_k \cdot \vec{y}^{(p)} \end{aligned}$$

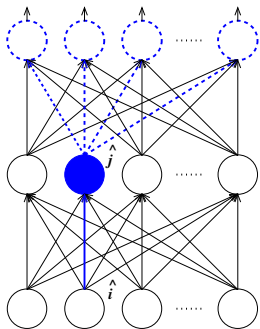


Alessandro Sperduti & Francesca Rossi

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned}
 \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c 2(t_k^{(p)} - z_k^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (-z_k^{(p)}) \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \sum_{j=1}^{N_H} w_{kj} y_j^{(p)}
 \end{aligned}$$



Alessandro Sperduti & Francesca Rossi

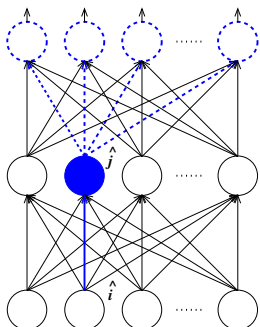
Sistemi di Elaborazione dell'Informazione

129

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned}
 \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (t_k^{(p)} - z_k^{(p)})^2 \\
 &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c 2(t_k^{(p)} - z_k^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (-z_k^{(p)}) \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \sum_{j=1}^{N_H} w_{kj} y_j^{(p)} \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) w_{k\hat{j}} \frac{\partial}{\partial w_{\hat{j}\hat{i}}} y_{\hat{j}}^{(p)}
 \end{aligned}$$

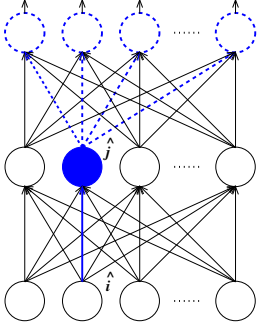


Alessandro Sperduti & Francesca Rossi

Calcolo del gradiente per i pesi di una unità nascosta

Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned}
 \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c 2(t_k^{(p)} - z_k^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (-z_k^{(p)}) \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \sum_{j=1}^{N_H} w_{kj} y_j^{(p)} \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) w_{k\hat{j}} \frac{\partial}{\partial w_{\hat{j}\hat{i}}} y_{\hat{j}}^{(p)} \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) w_{k\hat{j}} \sigma'(\vec{w}_{\hat{j}} \cdot \vec{x}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (\vec{w}_{\hat{j}} \cdot \vec{x}^{(p)})
 \end{aligned}$$



Alessandro Sperduti & Francesca Rossi

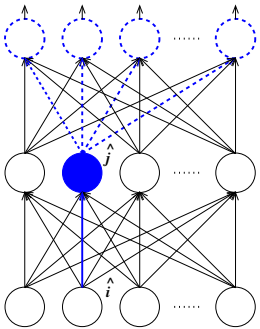
Sistemi di Elaborazione dell'Informazione

131

Calcolo del gradiente per i pesi di una unità nascosta

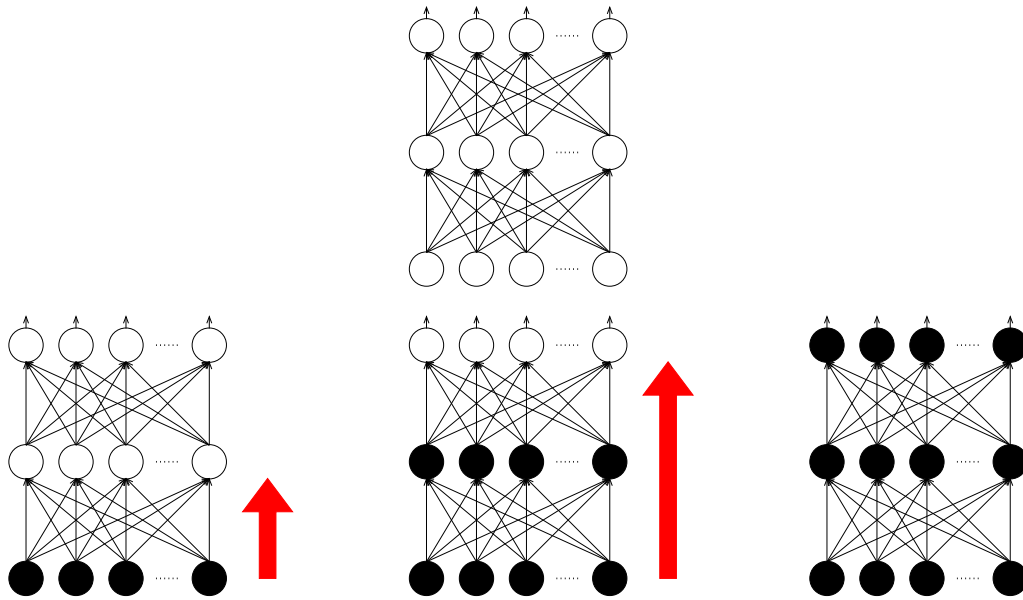
Fissiamo gli indici \hat{j} e \hat{i} :

$$\begin{aligned}
 \frac{\partial E}{\partial w_{\hat{j}\hat{i}}} &= \frac{1}{2cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c 2(t_k^{(p)} - z_k^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (-z_k^{(p)}) \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} \sum_{j=1}^{N_H} w_{kj} y_j^{(p)} \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) w_{k\hat{j}} \frac{\partial}{\partial w_{\hat{j}\hat{i}}} y_{\hat{j}}^{(p)} \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) w_{k\hat{j}} \sigma'(\vec{w}_{\hat{j}} \cdot \vec{x}^{(p)}) \frac{\partial}{\partial w_{\hat{j}\hat{i}}} (\vec{w}_{\hat{j}} \cdot \vec{x}^{(p)}) \\
 &= -\frac{1}{cN_{Tr}} \sum_{p \in Tr} \sum_{k=1}^c (t_k^{(p)} - z_k^{(p)}) \sigma'(\vec{w}_k \cdot \vec{y}^{(p)}) w_{k\hat{j}} \sigma'(\vec{w}_{\hat{j}} \cdot \vec{x}^{(p)}) x_{\hat{i}}^{(p)}
 \end{aligned}$$



Alessandro Sperduti & Francesca Rossi

Reti Neurali Feed-forward: Fase Forward

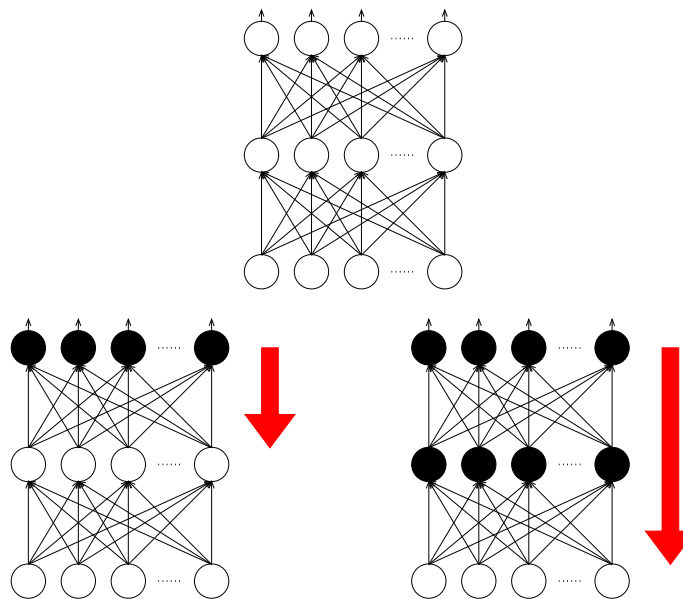


Alessandro Sperduti & Francesca Rossi

Sistemi di Elaborazione dell'Informazione

133

Reti Neurali Feed-forward: Fase Backward



Alessandro Sperduti & Francesca Rossi

Algoritmo Back-Propagation (uno strato nascosto, stocastico)

Back-Propagation-1hl-stocastico($Tr, \eta, \text{topologia rete}$)

- Inizializza tutti i pesi a valori random piccoli
- **Finché** la condizione di terminazione non è verificata, **fai**
 - Per ogni (\vec{x}, \vec{t}) in Tr , **fai**
 1. presenta \vec{x} alla rete e calcola il corrispondente output
 2. Per ogni unità di output k

$$\delta_k \leftarrow o_k(1 - o_k)(t_k - o_k)$$

3. Per ogni unità nascosta j

$$\delta_j \leftarrow o_j(1 - o_j) \sum_{k \in \text{outputs}} w_{k,j} \delta_k$$

4. aggiorna tutti i pesi $w_{p,q}$ della rete

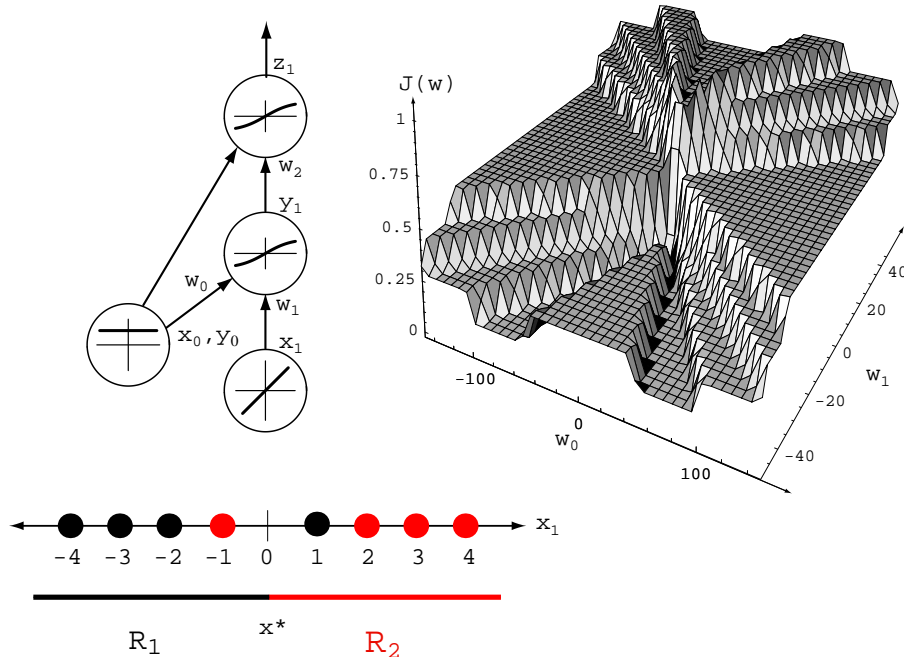
$$w_{s,q} \leftarrow w_{s,q} + \eta \Delta w_{s,q} \quad \text{dove } \Delta w_{s,q} = \begin{cases} \delta_s x_q & \text{se } s \in \text{nascoste} \\ \delta_s y_q & \text{se } s \in \text{outputs} \end{cases}$$

Alessandro Sperduti & Francesca Rossi

Sistemi di Elaborazione dell'Informazione

135

Esempio di Funzione Errore



Alessandro Sperduti & Francesca Rossi