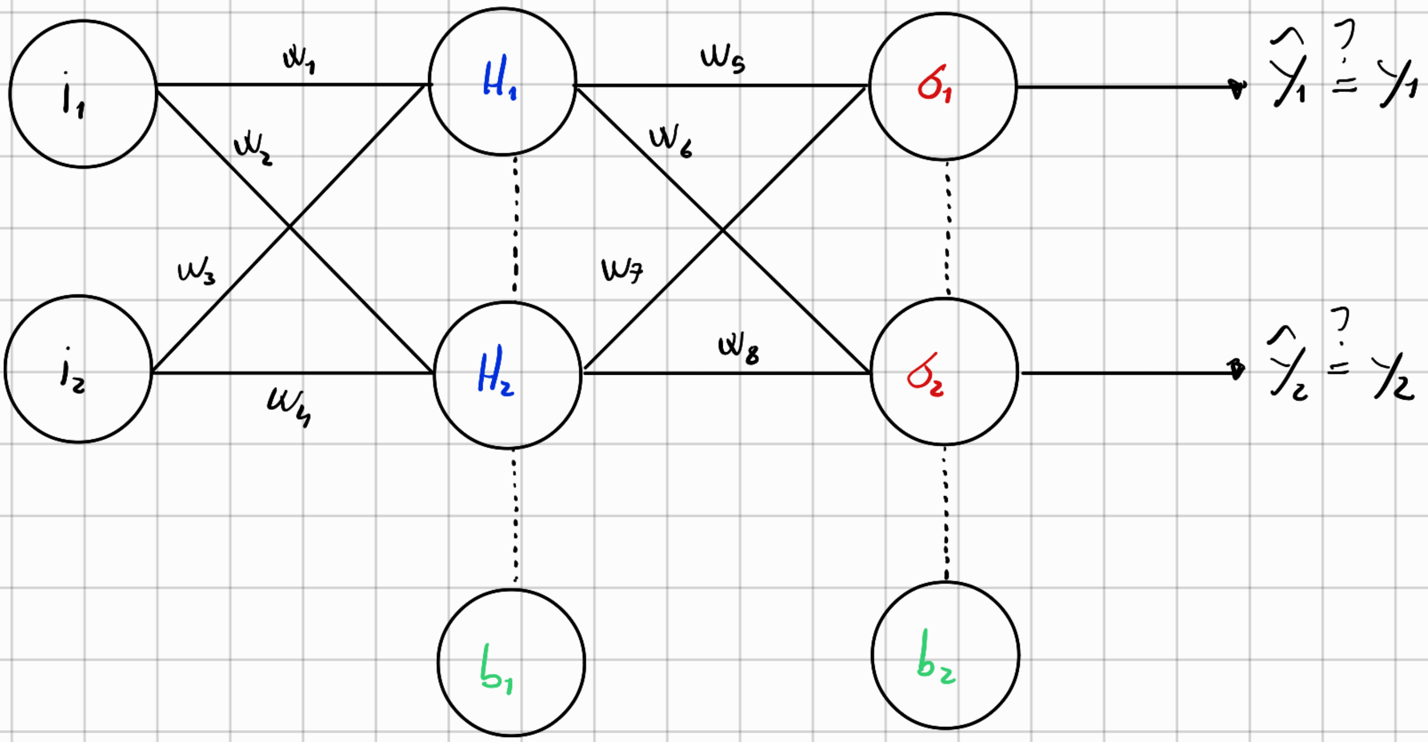
ESEMPIO BACKPROPAGATION



- i_1, i_2 : VALORI IN INPUT
- H_1, H_2 : HIDDEN NODES DELL'UNICO HIDDEN LAYER
- $w_1, \ldots, w_8$: PESI
- b_1, b_2 : BIAS
- o_1, o_2 : OUTPUT LAYER
- $\hat{y}_1, \hat{y}_2$: PREDIZIONI DEL MODELLO

---

1. SI INIZIALIZZANO I VALORI DEI PESI $w_1, \ldots, w_8$ AD UN VALORE RANDOM INIZIALE E I BIAS

---------- FORWARD PROPAGATION ----------

2.

SI CALCOLANO I VALORI DI CIASCUN HIDDEN NODE H_1, H_2 SULLA BASE DEI VALORI IN INPUT E DEI PESI $w_1, \dots, w_4$:

- $H_1 = w_1 i_1 + w_2 i_2 + b_1$
- $H_2 = w_3 i_1 + w_4 i_2 + b_2$

3. SI APPLICA LA FUNZIONE DI ATTIVAZIONE f_1, f_2 PER CIASCUN VALORE RICAVATO DEI NODI DELL' HIDDEN LAYER

- $$f(H_1) = \frac{1}{1+e^{-H_1}}$$
- $$f(H_2) = \frac{1}{1+e^{-H_2}}$$

4. SI CALCOLANO I VALORI DELL' OUTPUT LAYER O_1, O_2 SULLA BASE DEI PESI $w_5, \dots, w_8$ E DEL RISULTATO DELL'APPLICAZIONE DELLA FUNZIONE DI ATTIVAZIONE $f(H_1), f(H_2)$

- $O_1 = w_5 f(H_1) + w_6 f(H_2)$
- $O_2 = w_7 f(H_1) + w_8 f(H_2)$

5. SI CALCOLANO I VALORI DI OUTPUT $\hat{y}_1$ E $\hat{y}_2$ APPLICANDO NUOVAMENTE

LA FUNZIONE DI ATTIVAZIONE:

$$\hat{y}_1 = f(\sigma_1) = \frac{1}{1+e^{-\sigma_1}}$$

$$\hat{y}_2 = f(\sigma_2) = \frac{1}{1+e^{-\sigma_2}}$$

6 SI CALCOLA L'ERRORE COMMESSO PER CIASCUN NEURONE:

$$E_{\sigma_1} = \frac{1}{2}\left(y_1 - \hat{y}_1\right)^2$$

$$E_{\sigma_2} = \frac{1}{2}\left(y_2 - \hat{y}_2\right)^2$$

$$E_{TOT} = E_{\sigma_1} + E_{\sigma_2}$$

BACKPROPAGATION

7 VOGLIAMO ADESSO MISURARE L'IMPATTO DI CIASCUN PESO DELL' OUTPUT LAYER SULL' ERRORE TOTALE. CALCOLIAMO QUINDI IL GRADIENTE DELL'ERRORE TOTALE RISPETTO A CIASCUN PESO DELL' OUTPUT LAYER

CHAIN RULE

$$\frac{\partial E_{TOT}}{\partial w_8} = \frac{\partial E_{TOT}}{\partial \hat{y}_2} \cdot \frac{\partial \hat{y}_2}{\partial \sigma_2} \cdot \frac{\partial \sigma_2}{\partial w_8}$$

- $$\frac{\partial E_{TOT}}{\partial w_7} = \frac{\partial E_{TOT}}{\partial \hat{y}_2} \cdot \frac{\partial \hat{y}_2}{\partial \sigma_2} \cdot \frac{\partial \sigma_2}{\partial w_7}$$

- $$\frac{\partial E_{TOT}}{\partial w_6} = \frac{\partial E_{TOT}}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial \sigma_1} \cdot \frac{\partial \sigma_1}{\partial w_6}$$

- $$\frac{\partial E_{TOT}}{\partial w_5} = \frac{\partial E_{TOT}}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial \sigma_1} \cdot \frac{\partial \sigma_1}{\partial w_5}$$

8) SI AGGIORNANO I NUOVI VALORI DEI PESI $w_5, \ldots, w_8$:

- $$w_8^{NEW} = w_8 - \varepsilon \frac{\partial E_{TOT}}{\partial w_8}$$ ε = LEARNING RATE

- $$w_7^{NEW} = w_7 - \varepsilon \frac{\partial E_{TOT}}{\partial w_7}$$

- $$w_6^{NEW} = w_6 - \varepsilon \frac{\partial E_{TOT}}{\partial w_6}$$

- $$w_5^{NEW} = w_5 - \varepsilon \frac{\partial E_{TOT}}{\partial w_5}$$

9) VOGLIAMO ADESSO MISURARE L'IMPATTO DI CIASCUN PESO DELL' HIDDEN LAYER SULL' ERRORE TOTALE. CALCOLIAMO QUINDI IL GRADIENTE DELL'ERRORE TOTALE RISPETTO A CIASCUN PESO DELL' HIDDEN LAYER COME FATTO PRIMA

- $\dfrac{\partial E_{TOT}}{\partial w_4}$ · $\dfrac{\partial E_{TOT}}{\partial w_3}$ · $\dfrac{\partial E_{TOT}}{\partial w_2}$ · $\dfrac{\partial E_{TOT}}{\partial w_1}$

(10) SI AGGIORNANO I NUOVI VALORI DEI PESI $w_4, \ldots, w_1$:

- $$w_4^{NEW} = w_4 - \varepsilon \frac{\partial E_{TOT}}{\partial w_4}$$

- $$w_3^{NEW} = w_3 - \varepsilon \frac{\partial E_{TOT}}{\partial w_3}$$

- $$w_2^{NEW} = w_2 - \varepsilon \frac{\partial E_{TOT}}{\partial w_2}$$

- $$w_1^{NEW} = w_1 - \varepsilon \frac{\partial E_{TOT}}{\partial w_1}$$

(11) SI RIPETE IL PROCESSO FINO A QUANDO L'ERRORE TOTALE NON È ABBASTANZA PICCOLO!