

1) Quali sono i parametri di una distribuzione multinormale? Per ciascun parametro specificare la dimensionalità.

Dispense “Classificazione (parte 1)”

2) Nell’ambito dei multi-classificatori che cosa è e come funziona il Borda count?

Dispense “Classificazione (parte 2)”

3) Che cosa si intende per clustering? Fare esempi di applicazioni.

Dispense “Clustering”

4) Cosa si intende per K-fold cross-validation? Quali sono i vantaggi rispetto a un semplice split a due dei dati di training?

Dispense “Fondamenti”

5) Un problema di multiple linear regression viene risolto ai minimi quadrati ottenendo su un training set i seguenti coefficienti $\beta = [-8.8 \quad 1.3 \quad -7.3]$ (dove l’ultimo elemento è il termine noto). Dato un test set costituito dai due pattern $\mathbf{x}_1$ e $\mathbf{x}_2$ (di cui y_1 e y_2 sono i valori veri della variabile dipendente):

$$\mathbf{x}_1 = [-2.3 \quad 0.6], y_1 = 2.3$$

$$\mathbf{x}_2 = [-0.1 \quad -2.9], y_2 = -5.5$$

determinare il RMSE (*Root Mean Squared Error*) sul test set, riportando i principali passaggi intermedi del calcolo.

Svolgimento

Un problema di multiple linear regression consiste nella generalizzazione a iperpiani di un problema di linear regression. Dato un pattern di test $\mathbf{x}$, è possibile predire il valore della variabile dipendente y a partire dai valori dei coefficienti β utilizzando la seguente formula:

$$y = \sum_{i=1 \dots d+1} x_i \cdot \beta_i$$

Con $x_{d+1} = 1$ (per considerare il termine noto β_{d+1}).

Eseguendo i calcoli per i pattern di test si ottengono le seguenti predizioni:

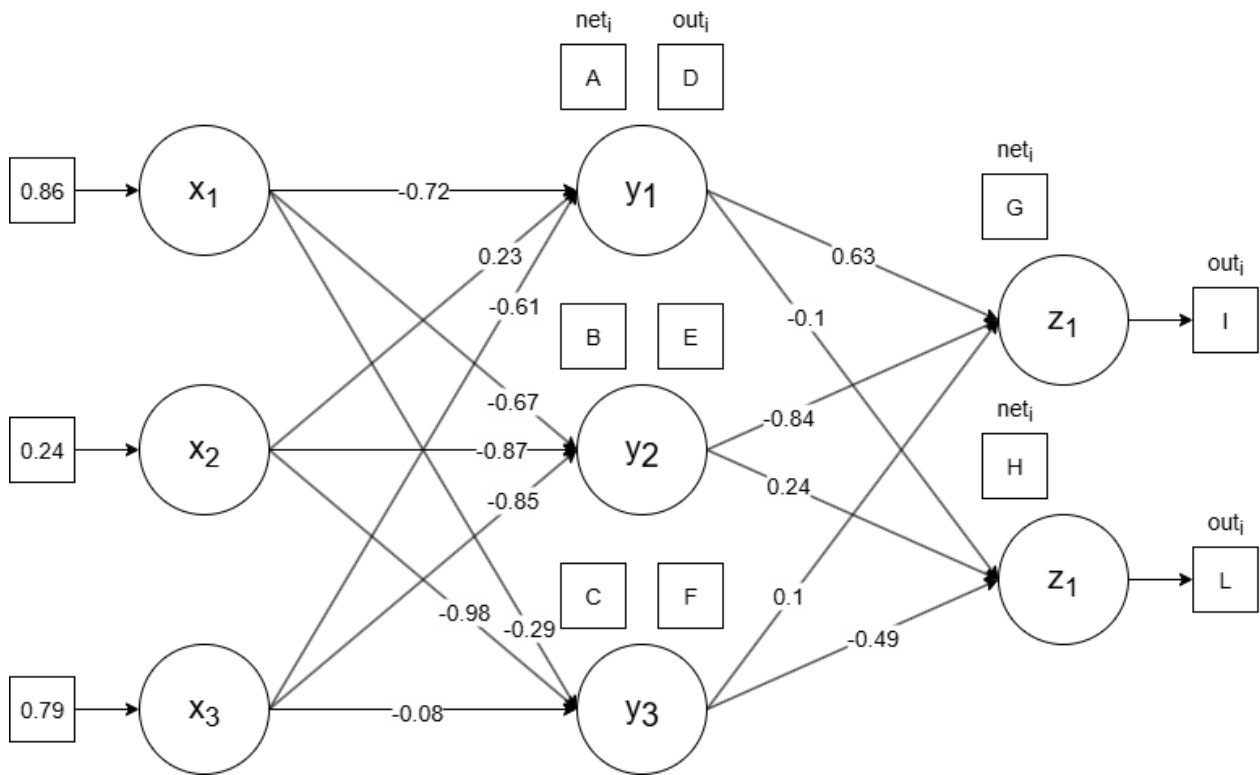
$$\hat{y}_1 = (-8.8) \cdot (-2.3) + 1.3 \cdot 0.6 + 1.0 \cdot (-7.3) = 13.7$$

$$\hat{y}_2 = (-8.8) \cdot (-0.1) + 1.3 \cdot (-2.9) + 1.0 \cdot (-7.3) = -10.2$$

Il RMSE del test set è dato dalla radice quadrata della media (su tutto il test set) del quadrato degli errori. In questo caso:

$$RMSE = \sqrt{\frac{(13.7-2.3)^2 + (-10.2-(-5.5))^2}{2}} = \sqrt{\frac{11.4^2 + (-4.7)^2}{2}} = \sqrt{\frac{152.05}{2}} = 8.7$$

6) Data la seguente rete neurale, calcolare net_i e out_i di ogni neurone (A..H) al seguito del passo forward del pattern di input, utilizzando la *tangente iperbolica* ($Tanh(net)$) come funzione di attivazione. Riportare il procedimento.



Svolgimento

$$\begin{aligned}
 A &= 0.86 \cdot (-0.72) + 0.24 \cdot 0.23 + 0.79 \cdot (-0.61) = -1.05 \\
 B &= 0.86 \cdot (-0.67) + 0.24 \cdot (-0.87) + 0.79 \cdot (-0.85) = -1.46 \\
 C &= 0.86 \cdot (-0.29) + 0.24 \cdot (-0.98) + 0.79 \cdot (-0.08) = -0.55 \\
 D &= \tanh(-1.05) = -0.78 \\
 E &= \tanh(-1.46) = -0.9 \\
 F &= \tanh(-0.55) = -0.5 \\
 G &= -0.78 \cdot 0.63 + (-0.9) \cdot (-0.84) + (-0.5) \cdot 0.1 = 0.21 \\
 H &= -0.78 \cdot (-0.1) + (-0.9) \cdot 0.24 + (-0.5) \cdot (-0.49) = 0.11 \\
 I &= \tanh(0.21) = 0.21 \\
 L &= \tanh(0.11) = 0.11
 \end{aligned}$$

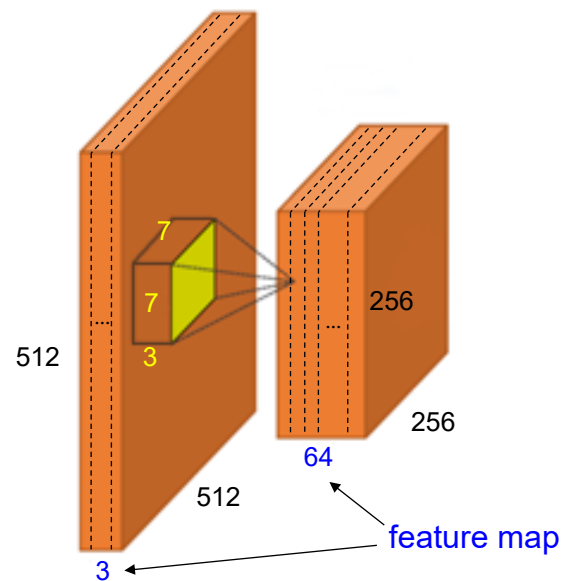
7) Dati un volume di input ed uno di output relativi a un livello di convoluzione in una CNN, aventi le seguenti dimensioni:

- *Volume Input:* $3 \times 512 \times 512$
- *Volume Output:* $64 \times 256 \times 256$

Considerando che ciascun filtro abbia dimensioni:

- *Dimensione Filtro:* $3 \times 7 \times 7$

Si calcoli il numero totale di connessioni e di pesi del livello (considerando i bias) motivando la risposta.



Svolgimento

Ogni neurone del livello di output ($64 \times 256 \times 256$) è connesso con tanti neuroni del livello di input pari alla dimensione del filtro ($3 \times 7 \times 7$).

Pertanto, **il numero totale di connessioni** è $(64 \times 256 \times 256) \cdot (3 \times 7 \times 7) = 616\,562\,688$.

Il **numero totale di pesi**, invece, risulta molto più piccolo giacché in una CNN i pesi di ciascun filtro sono condivisi da tutti i neuroni contenuti in una stessa feature map. Visto che il numero di feature map è uguale a 64, e il numero di input per ciascun filtro è pari a $(3 \times 7 \times 7)$, il numero totale di pesi (considerando il bias) è $(3 \times 7 \times 7) \times 64 + 64 = 9472$.