

1) Quali sono le condizioni necessarie affinché le tecniche di deep learning siano più efficaci di altri approcci di machine learning?

Pag. 6 delle dispense “Deep Learning”

2) La formula di distanza di un pattern dall'iperpiano risultante dal training di un SVM dipende da tutti i pattern del training set o solo da una parte di questi? Motivare la risposta.

Pag. 5 delle dispense “Classificazione (2)”

3) Descrivere a grandi linee l'algoritmo di Clustering K-means.

Pag. 8 delle dispense "Clustering"

4) Qual'è la funzione obiettivo in formato matriciale della multiple linear regression?

Pag. 5-6 delle dispense "Regressione"

5) Data una rete neurale MLP a 3 livelli senza bias composta da:

- 5 neuroni per l'input layer
- 10 neuroni per l'hidden layer
- 4 neuroni di output

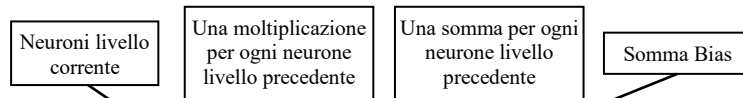
Quante somme e moltiplicazioni sono necessarie per il passo forward di un generico pattern trascurando le operazioni effettuate dalla funzione di attivazione? Motivare la risposta riportando anche il numero di operazioni per livello.

Svolgimento

Per ogni neurone del livello corrente si deve calcolare la seguente formula:

$$net_i = \sum_{j=1..d} w_{ji} \cdot in_j + w_{0i}$$

che comprende una moltiplicazione e una somma per ogni neurone del livello precedente più la somma finale del bias. Pertanto:



Numero operazioni livello hidden: $10 \cdot (5 + 5 + 0) = 100$

Numero operazioni livello di output: $4 \cdot (10 + 10 + 0) = 80$

NOTA: Il calcolo a lato è eseguito trascurando il fatto che un'implementazione ottimizzata della sommatoria potrebbe evitare una somma per il primo elemento della sommatoria (somma con 0).

Totale: **180**

6) Date tre distribuzioni multinormali identificate dai seguenti parametri:

$$\mu_1 = \begin{bmatrix} 9,40 \\ 2,18 \\ 6,79 \\ 3,52 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 5,22 \\ 8,15 \\ 6,82 \\ 2,94 \end{bmatrix}$$

$$\mu_3 = \begin{bmatrix} 5,66 \\ 0,80 \\ 7,51 \\ 8,10 \end{bmatrix}$$

$\Sigma_1 = \Sigma_2 = \Sigma_3 = I$ (matrice identità) e $P(w_1) = P(w_2) = P(w_3)$.

Indicare la classe assegnata ai seguenti pattern da un classificatore di Bayes multinormale (motivandone la risposta):

$$p_1 = \begin{bmatrix} 7,15 \\ 3,87 \\ 3,56 \\ 2,50 \end{bmatrix}, p_2 = \begin{bmatrix} 3,21 \\ 5,62 \\ 6,40 \\ 1,29 \end{bmatrix}, p_3 = \begin{bmatrix} 7,10 \\ 2,05 \\ 0,69 \\ 1,27 \end{bmatrix}$$

Svolgimento

La regola di classificazione di Bayes assegna un pattern \mathbf{x} alla classe w_i per cui è massima la probabilità a posteriori $P(w_i|\mathbf{x}) = \frac{p(\mathbf{x}|w_i) \cdot P(w_i)}{p(\mathbf{x})}$.

Dato che la probabilità a priori delle tre distribuzioni è la stessa, determinare la classe con probabilità a posteriori massima equivale a individuare la classe con densità di probabilità condizionale $p(\mathbf{x}|w_i)$ massima. Pertanto, sapendo che la densità di probabilità condizionale nella distribuzione multinormale è:

$$p(\mathbf{x}|w_i) = \frac{1}{(2\pi)^{\frac{d}{2}} \cdot |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \cdot \boldsymbol{\Sigma}_i^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu}_i)}$$

e che le tre matrici di covarianza sono uguali alla matrice identità, la classe restituita dal classificatore di Bayes sarà quella con il valore $(\mathbf{x} - \boldsymbol{\mu}_i)^t \cdot (\mathbf{x} - \boldsymbol{\mu}_i)$ minimo (che corrisponde alla distanza euclidea al quadrato D^2 tra il pattern \mathbf{x} e il vettore medio $\boldsymbol{\mu}_i$).

$$D^2(\mathbf{p}_1, \boldsymbol{\mu}_1) = \mathbf{19,39} \quad D^2(\mathbf{p}_1, \boldsymbol{\mu}_2) = 32,86 \quad D^2(\mathbf{p}_1, \boldsymbol{\mu}_3) = 58,61$$

$$D^2(\mathbf{p}_2, \boldsymbol{\mu}_1) = 55,27 \quad D^2(\mathbf{p}_2, \boldsymbol{\mu}_2) = \mathbf{13,34} \quad D^2(\mathbf{p}_2, \boldsymbol{\mu}_3) = 76,84$$

$$D^2(\mathbf{p}_3, \boldsymbol{\mu}_1) = \mathbf{47,58} \quad D^2(\mathbf{p}_3, \boldsymbol{\mu}_2) = 81,11 \quad D^2(\mathbf{p}_3, \boldsymbol{\mu}_3) = 96,80$$

\mathbf{p}_1 viene assegnato alla classe 1, \mathbf{p}_2 viene assegnato alla classe 2 e \mathbf{p}_3 viene assegnato alla classe 1.

7) Supponendo di utilizzare *K-fold Cross-Validation* con $K = 5$ per suddividere 8500 pattern in *training* e *validation set*, quanti diversi addestramenti (*run*) vengono effettuati? Ad ogni *run* quanti pattern vengono utilizzati per il training e quanti per la validazione?

Svolgimento

Con $K = 5$ gli 8500 pattern sono suddivisi in 5 partizioni da 1700 pattern l'una. Verranno eseguiti 5 addestramenti (*run*) utilizzando ogni volta una partizione diversa come *validation set* e le 4 restanti come *training set*. Pertanto ad ogni *run* verranno utilizzati 6800 pattern per il training e 1700 pattern per la validazione.