Machine Learning
19-Gen-2022 (90 minuti)

Matricola: ____________________
Cognome: __________ Nome: ____________

1) Cosa si intende per future reward nell'ambito del reinforcement learning? Fare un esempio.

Dispense "Deep Learning (2)"

2) Cosa si intende per multi-classificatore? Quando un multi-classificatore è efficace?

Dispense "Classificazione (2)"

3) Che cos'è il learning rate nell'ambito dell'apprendimento di reti neurali? Cosa succede se viene scelto un learning rate troppo piccolo o troppo grande?

Dispense "Reti Neurali"

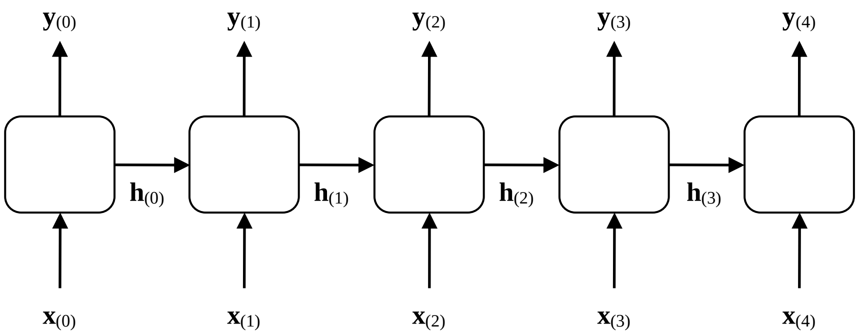
4) Cosa si intende per K-fold cross-validation? Quali sono i vantaggi rispetto a un semplice split a due dei dati di training?

Dispense "Fondamenti"

---

5) Data una cella di rete ricorrente (RNN) con 64 neuroni e 256 pesi addestrabili, disegnare lo schema grafico del suo *unfolding in time* su 5 stati. Quanti neuroni e pesi addestrabili ha la rete *unfolded*? Giustificare la risposta.

<u>Svolgimento</u>



Il numero totale di neuroni è pari a 320 (64 neuroni per ogni stato) mentre il numero totale di pesi addestrabili è pari a 256 in quanto i pesi della cella sono comuni a tutte le sue istanze.

6) Date due distribuzioni multinormali identificate dai seguenti parametri:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3{,}8 \\ -2{,}9 \\ -0{,}7 \\ 2{,}8 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} -2{,}4 \\ 0{,}5 \\ 2{,}5 \\ 3{,}8 \end{bmatrix}$$

$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \mathbf{I}$ (matrice identità) e $P(w_1) = P(w_2)$.
Indicare la classe assegnata ai seguenti pattern da un classificatore di Bayes multinormale (motivandone la risposta):

$$\mathbf{p_1} = \begin{bmatrix} -4{,}1 \\ 2{,}1 \\ 3{,}9 \\ 2{,}4 \end{bmatrix}, \mathbf{p_2} = \begin{bmatrix} 3{,}4 \\ -2{,}6 \\ 3{,}1 \\ 0{,}3 \end{bmatrix}, \mathbf{p_3} = \begin{bmatrix} 2{,}0 \\ 0{,}7 \\ 1{,}0 \\ 0{,}6 \end{bmatrix}$$

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:

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

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 $\boldsymbol{\mu}_i$).

$D^2(\mathbf{p}_1, \boldsymbol{\mu}_1) = 108{,}7$ $\qquad$ $D^2(\mathbf{p}_1, \boldsymbol{\mu}_2) = \mathbf{9{,}4}$

$D^2(\mathbf{p}_2, \boldsymbol{\mu}_1) = \mathbf{20{,}9}$ $\qquad$ $D^2(\mathbf{p}_2, \boldsymbol{\mu}_2) = 55{,}9$

$D^2(\mathbf{p}_3, \boldsymbol{\mu}_1) = \mathbf{23{,}9}$ $\qquad$ $D^2(\mathbf{p}_3, \boldsymbol{\mu}_2) = 31{,}9$

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

7) Un problema di multiple linear regression viene risolto ai minimi quadrati ottenendo su un training set i seguenti coefficienti $\boldsymbol{\beta} = [0{,}5 \quad 0{,}3 \quad -0{,}1 \quad 2{,}0]$ (dove l'ultimo elemento è il termine noto). Dato un test set costituito dai due soli pattern $\mathbf{x}_1$ e $\mathbf{x}_2$ (di cui y_1 e y_2 sono i valori veri della variabile dipendente):
$\mathbf{x}_1 = [3{,}2 \quad 8{,}4 \quad -3{,}1]$, $y_1 = 7{,}1$
$\mathbf{x}_2 = [2{,}2 \quad 2{,}4 \quad -4{,}1]$, $y_2 = 3{,}2$
determinare l'RMSE 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 $\boldsymbol{\beta}$ 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 $\mathbf{x}_1$ e $\mathbf{x}_2$ si ottengono le seguenti predizioni $\hat{y}_1$ e $\hat{y}_2$:

$\hat{y}_1 = 3{,}2 \cdot 0{,}5 + 8{,}4 \cdot 0{,}3 + (-3{,}1) \cdot (-0{,}1) + 1{,}0 \cdot 2{,}0 = 6{,}4$
$\hat{y}_2 = 2{,}2 \cdot 0{,}5 + 2{,}4 \cdot 0{,}3 + (-4{,}1) \cdot (-0{,}1) + 1{,}0 \cdot 2{,}0 = 4{,}2$

L'RMSE sul test set è dato dalla radice quadrata della media (su tutto il test set) degli errori al quadrato. In questo caso:

$$RMSE = \sqrt{\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2}{2}} = \sqrt{\frac{(7{,}1 - 6{,}4)^2 + (3{,}2 - 4{,}2)^2}{2}} = \sqrt{\frac{0{,}5 + 1{,}0}{2}} = 0{,}9$$