Machine Learning	**Matricola: ____________________**
21-Lug-2023 (90 minuti)	**Cognome: __________ Nome: ____________**

1) Perché le recenti reti neurali deep sono più efficaci delle MLP a tre livelli?

Dispense “Deep Learning (parte 1)”

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.

Dispense “Classificazione (parte 2)”

3) Nel classificatore di Bayes cosa si intende per densità di probabilità condizionale e probabilità a priori.

Dispense “Classificazione (parte 1)”

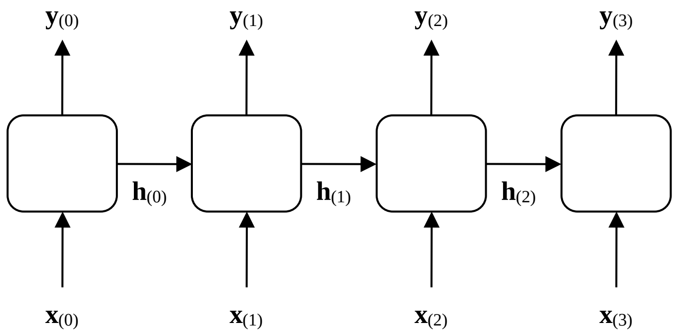
4) Dare la definizione di Training, Validation e Test set e discutere una possibile suddivisione dei dati nei tre insiemi.

Dispense “Fondamenti”

---

5) Data una cella di rete ricorrente (RNN) con 32 neuroni e 64 pesi addestrabili, disegnare lo schema grafico del suo *unfolding in time* su 4 stati. Quanti neuroni e pesi addestrabili ha la rete *unfolded*? Giustificare la risposta.

<u>**Svolgimento**</u>



Il numero totale di neuroni è pari a 128 (32 neuroni per ogni stato) mentre il numero totale di pesi addestrabili è pari a 64 in quanto i pesi della cella sono comuni a tutte le sue istanze.

6) Data una rete neurale MLP e un training set di 100000 pattern, si decide di eseguire il training con SGD e mini-batch di 1000 pattern. Si eseguono 150 epoche di addestramento. Calcolare il numero di volte in cui viene calcolato il (vettore) gradiente ed aggiornati i pesi durante l'apprendimento, motivando il calcolo.

Svolgimento

Il numero di volte in cui viene calcolato il vettore gradiente e aggiornati i pesi, corrisponde al numero di iterazioni eseguite.
Avendo 100000 pattern suddivisi in mini-batch di 1000 pattern l'uno, ad ogni epoca vengono eseguite $\frac{100000}{1000} = 100$ iterazioni. Dovendo eseguire 150 epoche, il numero totale di iterazioni può essere calcolato come $100 \cdot 150 = 15000$.

7) Un problema di multiple linear regression viene risolto ai minimi quadrati ottenendo su un training set i seguenti coefficienti $\boldsymbol{\beta} = [1{,}2 \quad -0{,}7 \quad -1{,}9 \quad 3{,}5]$ (dove l'ultimo elemento è il termine noto). Dato un test set costituito dai tre pattern $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$ (di cui y_1, y_2 e y_3 sono i valori veri della variabile dipendente):
$\mathbf{x}_1 = [-2{,}6 \quad 7{,}9 \quad 5{,}1]$, $y_1 = 5{,}8$
$\mathbf{x}_2 = [3{,}8 \quad -0{,}6 \quad -9{,}1]$, $y_2 = 1{,}3$
$\mathbf{x}_3 = [0 \quad 10{,}8 \quad -7{,}4]$, $y_3 = -3{,}4$

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 si ottengono le seguenti predizioni:

$$\hat{y}_1 = (-2{,}6) \cdot 1{,}2 + 7{,}9 \cdot (-0{,}7) + 5{,}1 \cdot (-1{,}9) + 1{,}0 \cdot 3{,}5 = -14{,}8$$
$$\hat{y}_2 = 3{,}8 \cdot 1{,}2 + (-0{,}6) \cdot (-0{,}7) + (-9{,}1) \cdot (-1{,}9) + 1{,}0 \cdot 3{,}5 = 25{,}8$$
$$\hat{y}_3 = 0 \cdot 1{,}2 + 10{,}8 \cdot (-0{,}7) + (-7{,}4) \cdot (-1{,}9) + 1{,}0 \cdot 3{,}5 = 10{,}0$$

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

$$RMSE = \sqrt{\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2}{3}} =$$
$$= \sqrt{\frac{(-14{,}8 - 5{,}8)^2 + (25{,}8 - 1{,}3)^2 + (10{,}0 - (-3{,}4))^2}{3}} = \sqrt{\frac{426{,}0 + 598{,}8 + 179{,}6}{3}} = 20{,}0$$