Machine Learning	**Matricola: ____________________**
01-Set-2022 (90 minuti)	**Cognome: __________ Nome: ____________**

1) Nell'ambito dell'apprendimento automatico quali sono le principali cause di overfitting?

Dispense "Fondamenti"

2) Quanti sono i parametri indipendenti di una distribuzione multinormale nel caso 3-dimensionale? Motivare la risposta.

Dispense "Classificazione (1)"

3) Quali sono le più comuni funzioni di attivazione utilizzate per neuroni artificiali? Perché è necessario che siano non-lineari e differenziabili (esistenza derivata)?

Dispense "Reti Neurali"

4) Quali sono le più note tecniche di riduzione di dimensionalità? Quali i loro tipici utilizzi?

Dispense "Riduzione Dimensionalità"

---

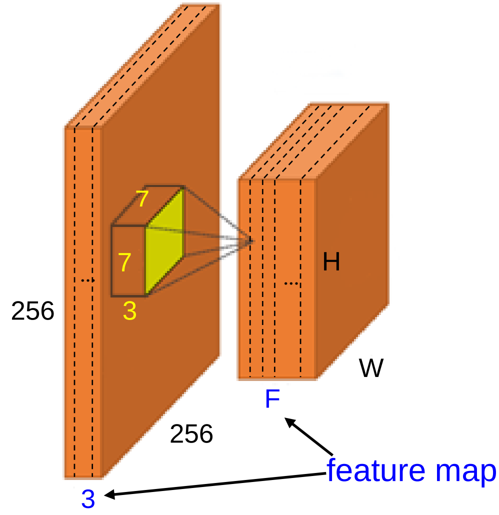
5) Un multiclassificatore, composto da 3 classificatori combinati a livello di decisione utilizzando Borda count come tecnica di fusione, viene utilizzato per riconoscere pattern appartenenti a 4 classi. Nella tabella seguente sono riportati i ranking restituiti dai singoli classificatori (C_i) dati in input 3 diversi pattern ($\boldsymbol{p}_j$). Completare la tabella nell'ipotesi che alla prima classe siano assegnati 10 punti, alla seconda 7, alla terza 5 e alla quarta 2.

	C_1				C_2				C_3			
$\boldsymbol{p}_1$	4	2	1	3	2	4	1	3	1	4	2	3
$\boldsymbol{p}_2$	1	2	3	4	1	3	2	4	2	1	3	4
$\boldsymbol{p}_3$	3	2	4	1	2	3	4	1	3	4	2	1

	Punteggi Classi				**Classe scelta**
	1	**2**	**3**	**4**	
$\boldsymbol{p}_1$	20	22	6	24	4
$\boldsymbol{p}_2$	27	22	17	6	1
$\boldsymbol{p}_3$	6	22	27	17	3

6) Dati un volume di input e un livello di convoluzione in una CNN, aventi le seguenti caratteristiche:

- *Volume Input*: $3 \times 256 \times 256$
- 64 filtri di dimensione $3 \times 7 \times 7$
- *Stride: 3*
- *Padding: 0*



Si calcoli motivando la risposta:
- La dimensione del volume di output: $F \times W \times H$;
- il numero totale di connessioni e di pesi del livello (senza considerare i bias).

Svolgimento

Il numero di feature map F del volume di output è pari al numero di filtri utilizzati: 64.
Essendo larghezza e altezza uguali tra loro sia nel volume di input che nei filtri, le dimensioni spaziali del volume di output (W e H) saranno anch'esse uguali tra loro.

$$W = H = \frac{D_{IN} - D_F + 2 \cdot Padding}{Stride} + 1 = \frac{256 - 7 + 2 \cdot 0}{3} + 1 = \frac{249}{3} + 1 = 84$$

Ogni neurone del livello di output ($64 \times 84 \times 84$) 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 $(64 \times 84 \times 84) \cdot (3 \times 7 \times 7) = 66\,382\,848$.

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 elementi di ciascun filtro è pari a $(3 \times 7 \times 7)$, il numero totale di pesi (non considerando il bias) è $(3 \times 7 \times 7) \times 64 = 9408$.

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:

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

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:

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