Machine Learning **13-Giu-2022 (90 minuti)**	**Matricola: ____________________** **Cognome: _________ Nome: ___________**

1) Come opera un livello di convoluzione di una CNN?

Dispense "Deep Learning (2)"

2) La densità locale di pattern con quali metodi (parametrici e non-parametrici) può essere stimata? Fare un esempio.

Dispense "Classificazione (1)"

3) Definire i problemi di Classificazione e Regressione evidenziandone le differenze e fornendo per ciascuno esempi reali della loro applicazione.

Dispense "Fondamenti"

4) Come può essere scelto nella pratica il numero di cluster in un algoritmo di clustering come K-means?

Dispense "Clustering"

---

5) Un multiclassificatore, composto da 3 classificatori combinati a livello di confidenza, viene utilizzato per riconoscere pattern appartenenti a 4 classi (A, B, C, D). Nella tabella seguente sono riportate le confidenze restituite dai singoli classificatori (C_i) dati in input 2 diversi pattern ($\boldsymbol{p}_j$). Completare la tabella riportando, per ogni metodo di fusione (Somma, Prodotto, Massimo e Minimo), le confidenze ottenute e la classe di output restituita dal multiclassificatore.

	C_1				C_2				C_3			
	A	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
$\boldsymbol{p}_1$	0.18	0.16	0.15	0.51	0.16	0.23	0.15	0.46	0.35	0.21	0.22	0.22
$\boldsymbol{p}_2$	0.51	0.15	0.15	0.19	0.17	0.42	0.21	0.20	0.16	0.39	0.25	0.20

<u>Svolgimento</u>

	Somma					**Prodotto**					**Massimo**					**Minimo**				
	A	**B**	**C**	**D**	**Out**	**A**	**B**	**C**	**D**	**Out**	**A**	**B**	**C**	**D**	**Out**	**A**	**B**	**C**	**D**	**Out**
$\boldsymbol{p}_1$	0.69	0.60	0.52	1.19	D	0.01	0.01	0.00	0.05	D	0.35	0.23	0.22	0.51	D	0.16	0.16	0.15	0.22	D
$\boldsymbol{p}_2$	0.84	0.96	0.61	0.59	B	0.01	0.02	0.01	0.01	B	0.51	0.42	0.25	0.20	A	0.16	0.15	0.15	0.19	D

6) Data una rete neurale MLP e un training set di 25000 pattern, si decide di eseguire il training con SGD e mini-batch di 500 pattern. Si eseguono 50 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.

<u>Svolgimento</u>

Il numero di volte in cui viene calcolato il vettore gradiente e aggiornati i pesi, corrisponde al numero di iterazioni eseguite.
Avendo 25000 pattern suddivisi in mini-batch di 500 pattern l'uno, ad ogni epoca vengono eseguite $\frac{25000}{500} = 50$ iterazioni. Dovendo eseguire 50 epoche, il numero totale di iterazioni può essere calcolato come $50 \cdot 50 = 2500$.

7) Dato un training set composto dai seguenti pattern:

$$\mathbf{x}_1 = \begin{bmatrix} 3.4 \\ 6.0 \\ 9.2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 6.5 \\ -1.6 \\ 6.3 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 4.9 \\ -7.5 \\ -0.7 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} -0.6 \\ 1.9 \\ 1.3 \end{bmatrix}, \quad \mathbf{x}_5 = \begin{bmatrix} 1.3 \\ -6.5 \\ -4.4 \end{bmatrix}$$

a cui sono associate le seguenti osservazioni (variabile dipendente):

$$y_1 = -7.9, \quad y_2 = 3.6, \quad y_3 = 1.5, \quad y_4 = 2.6, \quad y_5 = 0.4$$

formulare il problema di *multiple linear regression* definendo la matrice **X** e il vettore **y**.

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

La matrice **X** di dimensione $n \times (d+1)$ si ottiene disponendo su ciascuna riga gli n pattern del training set e inserendo a destra una colonna con tutti valori 1:

$$\mathbf{X} = \begin{bmatrix} 3.4 & 6.0 & 9.2 & 1.0 \\ 6.5 & -1.6 & 6.3 & 1.0 \\ 4.9 & -7.5 & -0.7 & 1.0 \\ -0.6 & 1.9 & 1.3 & 1.0 \\ 1.3 & -6.5 & -4.4 & 1.0 \end{bmatrix}$$

Il vettore **y** di dimensione n si ottiene disponendo in colonna i valori y_i:

$$\mathbf{y} = \begin{bmatrix} -7.9 \\ 3.6 \\ 1.5 \\ 2.6 \\ 0.4 \end{bmatrix}$$

Anche se non richiesto dall'esercizio, definita la matrice **X** e il vettore **y** le relazioni tra variabili indipendenti e dipendente possono essere scritte in forma matriciale come:

$$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta}$$

Dove $\boldsymbol{\beta}$ rappresenta il vettore di dimensione $d+1$ dei parametri da determinare:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

In formato matriciale la funzione obiettivo da minimizzare (*Least Square*) può essere scritta come:

$$||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2$$