Machine Learning	**Matricola: ____________________**
31-Gen-2025 (90 minuti)	**Cognome: __________ Nome: ____________**

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

Dispense "Deep Learning (parte 2)"

2) Nell'ambito dei multi-classificatori come si può ottenere indipendenza tra i singoli classificatori utilizzati?

Dispense "Classificazione (parte 2)"

3) Che cosa si intende con tokenizzazione negli LLM? Perchè si preferisce usare token invece di singoli caratteri o intere parole?

Dispense "Deep Learning (parte 2)"

4) Indicare le principali "stagioni" nello sviluppo dell'intelligenza artificiale e machine learning.

Dispense "Introduzione"

---

5) Data una rete neurale MLP e un training set di 1000000 pattern, si decide di eseguire il training con SGD e mini-batch di 250 pattern. Si eseguono 180 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 1000000 pattern suddivisi in mini-batch di 250 pattern l'uno, ad ogni epoca vengono eseguite $\frac{1000000}{250} = 4000$ iterazioni. Dovendo eseguire 180 epoche, il numero totale di iterazioni può essere calcolato come $4000 \cdot 180 = 720000$.

6) Supponendo di utilizzare *K-fold Cross-Validation* con $K = 8$ per suddividere 32000 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?

<u>Svolgimento</u>

Con $K = 8$ i 32000 pattern sono suddivisi in 8 partizioni da 4000 pattern l'una. Verranno eseguiti 8 addestramenti (*run*) utilizzando ogni volta una partizione diversa come *validation set* e le 7 restanti come *training set*. Pertanto, ad ogni *run* verranno utilizzati 28000 pattern per il training e 4000 pattern per la validazione.

7) Date due distribuzioni multinormali identificate dai seguenti parametri:

$\boldsymbol{\mu}_0 = \begin{bmatrix} 5.40 \\ -2.48 \end{bmatrix}$ $\boldsymbol{\mu}_1 = \begin{bmatrix} 3.01 \\ -0.93 \end{bmatrix}$

$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1/10 & 0 \\ 0 & 3/5 \end{bmatrix}$ $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}$

$P(w_0) = 0.4$ $P(w_1) = 0.6$

Nell'ipotesi dell'impiego di un classificatore di Bayes multinormale, calcolare per il punto $\boldsymbol{x} = \begin{bmatrix} 4.30 \\ -1.80 \end{bmatrix}$:

- le densità di probabilità condizionali;
- l'indice della classe restituita in output.

Si ricorda che:

- la densità di probabilità, nel caso della distribuzione multinormale è:

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

- l'inversa di una matrice diagonale si ottiene invertendo i singoli elementi;
- il determinante di una matrice diagonale si ottiene moltiplicando gli elementi della diagonale.

Svolgimento

$$\boldsymbol{x} - \boldsymbol{\mu}_0 = \begin{bmatrix} 4.30 \\ -1.80 \end{bmatrix} - \begin{bmatrix} 5.40 \\ -2.48 \end{bmatrix} = \begin{bmatrix} -1.10 \\ 0.68 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_0^{-1} = \begin{bmatrix} 10 & 0 \\ 0 & 1.67 \end{bmatrix}$$

$$|\boldsymbol{\Sigma}_0| = 0.06$$

$$p(\boldsymbol{x}|w_0) = \frac{1}{2\pi \cdot \sqrt{0.06}} \cdot e^{-\frac{1}{2}\cdot[-1.10 \quad 0.68]\cdot\begin{bmatrix} 10 & 0 \\ 0 & 1.67 \end{bmatrix}\cdot\begin{bmatrix} -1.10 \\ 0.68 \end{bmatrix}}$$

$$\begin{bmatrix} 10 & 0 \\ 0 & 1.67 \end{bmatrix} \cdot \begin{bmatrix} -1.10 \\ 0.68 \end{bmatrix} = \begin{bmatrix} -11.00 \\ 1.13 \end{bmatrix}$$

$$[-1.10 \quad 0.68] \cdot \begin{bmatrix} -11.00 \\ 1.13 \end{bmatrix} = 12.87$$

$p(\boldsymbol{x}|w_0) = 0.65 \cdot e^{-\frac{12.87}{2}} = 0.001042$ (Densità di probabilità condizionale di $\boldsymbol{x}$ data w_0)

$$\boldsymbol{x} - \boldsymbol{\mu}_1 = \begin{bmatrix} 4.30 \\ -1.80 \end{bmatrix} - \begin{bmatrix} 3.01 \\ -0.93 \end{bmatrix} = \begin{bmatrix} 1.29 \\ -0.87 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1^{-1} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

$$|\boldsymbol{\Sigma}_1| = 0.02$$

$$p(\boldsymbol{x}|w_1) = \frac{1}{2\pi \cdot \sqrt{0.02}} \cdot e^{-\frac{1}{2}\cdot[1.29 \quad -0.87]\cdot\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}\cdot\begin{bmatrix} 1.29 \\ -0.87 \end{bmatrix}}$$

$$\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} \cdot \begin{bmatrix} 1.29 \\ -0.87 \end{bmatrix} = \begin{bmatrix} 6.45 \\ -8.70 \end{bmatrix}$$

$[1.29 \quad -0.87] \cdot \begin{bmatrix} 6.45 \\ -8.70 \end{bmatrix} = 15.89$

$p(\boldsymbol{x}|w_1) = 1.12 \cdot e^{-\frac{15.89}{2}} = 0.000399$ (Densità di probabilità condizionale di $\boldsymbol{x}$ data w_1)

$p(\boldsymbol{x}) = 0.001042 \cdot 0.40 + 0.000399 \cdot 0.60 = 0.000656$ (Densità di probabilità assoluta dato $\boldsymbol{x}$)

$p(w_0|\boldsymbol{x}) = \frac{0.001042 \cdot 0.40}{0.000656} = 0.635$ (Probabilità a posteriori di w_0 dato $\boldsymbol{x}$)

$p(w_1|\boldsymbol{x}) = \frac{0.000399 \cdot 0.60}{0.000656} = 0.365$ (Probabilità a posteriori di w_1 dato $\boldsymbol{x}$)

Indice della classe restituita: **0**