Machine Learning **Matricola: ____________________**

12-Set-2023 (90 minuti) **Cognome: __________ Nome: ____________**

1) Cosa è possibile apprendere mediante tecniche di reinforcement learning? Fare un esempio.

Dispense “Deep Learning (parte 2)”

2) Nell’ambito di classificazione con SVM cosa si intende per pattern linearmente separabili e non linearmente separabili? Fare esempio grafico dei due casi.

Dispense “Classificazione (parte 2)”

3) Qual è l’idea di base dell’algoritmo di clustering EM con Gaussian mixture?

Dispense “Clustering”

4) Cosa si intende per iperparametri? Fornire esempi pratici di iperparametri. Come si ottimizzano?

Dispense “Fondamenti”

---

5) Data una rete neurale MLP a 3 livelli <u>senza bias</u> composta da:

- 16 neuroni per l’input layer
- 32 neuroni per l’hidden layer
- 4 neuroni di output

Quante somme e moltiplicazioni sono necessarie per il passo forward di un generico pattern trascurando le operazioni effettuate dalla funzione di attivazione? Motivare la risposta riportando anche il numero di operazioni per livello.

<u>Svolgimento</u>

Per ogni neurone del livello corrente si deve calcolare la seguente formula:

$$net_i = \sum_{j=1..d} w_{ji} \cdot in_j$$

che comprende una moltiplicazione e una somma per ogni neurone del livello precedente più la somma finale del bias. Pertanto:

Neuroni livello corrente | Una moltiplicazione per ogni neurone livello precedente | Una somma per ogni neurone livello precedente

Numero operazioni livello hidden: $32 \cdot (16 + 16) = 1024$

Numero operazioni livello di output: $4 \cdot (32 + 32) = 256$

Totale: **1280 (640 somme e 640 moltiplicazioni)**

> <u>NOTA:</u> Il calcolo a lato è eseguito trascurando il fatto che un’implementazione ottimizzata della sommatoria potrebbe evitare una somma per il primo elemento della sommatoria (somma con 0).

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):

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

<u>Svolgimento</u>

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:

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

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) Dato un livello di convoluzione in una CNN con un volume di Input pari a $128 \times 224 \times 16$ (nel formato $Width \times Height \times Depth$), e filtri di dimensioni $5 \times 5 \times 16$. Sì calcolino le dimensioni ($Width \times Height$) di ogni *feature map* nel volume di Output considerando un $Padding = 1$ e $Stride = 3$.

Svolgimento

Per calcolare la dimensione di ogni *feature map* del volume di Output si utilizza la formula seguente:

$$W_{out} = \frac{W_{in} - F + 2 \cdot Padding}{Stride} + 1$$

Sostituendo $F = 5$, $W_{in} = 128$ (oppure $W_{in} = 224$ a seconda della dimensione considerata), $Padding = 1$ e $Stride = 3$ otteniamo:

$$Width = \frac{128 - 5 + 2 \cdot 1}{3} + 1 = 42$$

$$Height = \frac{224 - 5 + 2 \cdot 1}{3} + 1 = 74$$

Pertanto la dimensione di ogni *feature map* del volume di Output sarà 42×74 ($Width \times Height$).

Si noti che la profondità del volume di Input che corrisponde alla profondità del filtro (in questo caso 16) costituisce un valore indipendente e dunque non utile per il calcolo della dimensione di ogni *feature map* del volume di Output.