Machine Learning
18-Gen-2023 (120 minuti)

Matricola: ____________________
Cognome: __________ Nome: ____________

1) Come classifica i pattern un classificatore K-NN? Discutere la complessità computazionale.

Dispense “Classificazione 1”

2) Nel classificatore SVM cosa sono i support vectors?

Dispense “Classificazione 2”

3) Cosa si intende per convergenza di un algoritmo di apprendimento iterativo? Accuratezza e loss come si comportano durante le iterazioni in caso di convergenza. Disegnare un semplice grafico.

Dispense “Fondamenti”

4) Nella regressione cosa si intende per variabile indipendente e variabile dipendente?

Dispense “Regressione”

---

5) In una rete CNN, data un’immagine di Input di dimensione $7 \times 7 \times 3$ (nel formato $Width \times Height \times Depth$) e un livello di convoluzione composto da 1 filtro di dimensioni $3 \times 3 \times 3$ con $padding = 0$ e $stride = 2$, si calcoli il valore dell’elemento del volume di output indicato con il ?.

Input

Depth 0

197	103	42	252	27	78	205
114	57	2	195	7	1	130
97	71	179	60	187	22	21
86	84	187	229	208	167	237
25	177	236	250	25	9	87
217	175	190	175	23	10	69
67	127	246	142	4	125	87

Depth 1

124	164	158	18	229	152	110
19	111	22	75	167	224	88
136	21	201	237	248	43	136
151	245	140	163	12	207	19
212	197	87	203	42	149	157
12	78	232	52	113	232	198
64	167	99	112	42	236	186

Depth 2

105	45	160	12	81	207	228
174	18	111	216	200	91	62
170	191	128	124	74	187	123
199	224	184	134	66	193	87
77	41	50	226	226	88	106
151	182	191	216	198	184	93
60	120	91	168	141	136	150

Filtro

Depth 0

0.89	0.87	0
0	0.30	0.52
0	0	0

Depth 1

0	0.24	0.90
0	0.07	0
0.64	0	0

Depth 2

0	0	0
0	0.71	0
0.40	0	0

Output

-	-	?
-	-	-
-	-	-

Svolgimento

Nella convoluzione 3D, per ogni elemento del volume di output, il filtro opera su una porzione diversa del volume di input. Tale posizione dipende dai parametri $padding$ e $stride$ oltre che dalla posizione dell'elemento che si vuole calcolare (nel volume di output). La regione bordata di nero nell'immagine di Input rappresenta la porzione da considerare per calcolare l'elemento di output evidenziato.
Il valore della cella di output viene calcolato come somma dei prodotti di ogni elemento della porzione di input per l'elemento corrispondente del filtro.
Nell'ambito delle CNN, la convoluzione non richiede nessuna operazione di "ribaltamento" del filtro e di normalizzazione del risultato.

$$\text{Depth } 0 = 27 \cdot 0.89 + 78 \cdot 0.87 + 1 \cdot 0.30 + 130 \cdot 0.52 = 159.79$$

$$\text{Depth } 1 = 152 \cdot 0.24 + 110 \cdot 0.90 + 224 \cdot 0.07 + 248 \cdot 0.64 = 309.88$$

$$\text{Depth } 2 = 91 \cdot 0.71 + 74 \cdot 0.40 = 94.21$$

$$\text{Risultato} = 159.79 + 309.88 + 94.21 = 563.88$$

Output

640.23	867.63	563.88
799.15	850.68	760.43
652.76	835.28	476.18

6) Date due distribuzioni multinormali identificate dai seguenti parametri:

$\boldsymbol{\mu}_0 = \begin{bmatrix} 3.21 \\ -0.57 \end{bmatrix}$

$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}$

$P(w_0) = 0.7$

$\boldsymbol{\mu}_1 = \begin{bmatrix} 1.27 \\ -2.14 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1/4 \end{bmatrix}$

$P(w_1) = 0.3$

Nell'ipotesi dell'impiego di un classificatore di Bayes multinormale, calcolare per il punto $\boldsymbol{x} = \begin{bmatrix} 3.80 \\ -2.40 \end{bmatrix}$:

- 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:

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

- 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

$$\boldsymbol{x} - \boldsymbol{\mu}_0 = \begin{bmatrix} 3.80 \\ -2.40 \end{bmatrix} - \begin{bmatrix} 3.21 \\ -0.57 \end{bmatrix} = \begin{bmatrix} 0.59 \\ -1.83 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_0^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

$$|\boldsymbol{\Sigma}_0| = 0.167$$

$$p(\boldsymbol{x}|w_0) = \frac{1}{2\pi \cdot \sqrt{0.167}} \cdot e^{-\frac{1}{2}\cdot[0.59 \quad -1.83]\cdot\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\cdot\begin{bmatrix} 0.59 \\ -1.83 \end{bmatrix}}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} 0.59 \\ -1.83 \end{bmatrix} = \begin{bmatrix} 2 \cdot 0.59 + 0 \cdot (-1.83) \\ 0 \cdot 0.59 + 3 \cdot (-1.83) \end{bmatrix} = \begin{bmatrix} 1.18 \\ -5.49 \end{bmatrix}$$

$$[0.59 \quad -1.83] \cdot \begin{bmatrix} 1.18 \\ -5.49 \end{bmatrix} = (0.59) \cdot (1.18) + (-1.83) \cdot (-5.49) = 10.74$$

$p(\boldsymbol{x}|w_0) = 0.389 \cdot e^{-\frac{10.74}{2}} = 0.0018$ (Densità di probabilità condizionale di $\boldsymbol{x}$ data w_0)

$$\boldsymbol{x} - \boldsymbol{\mu}_1 = \begin{bmatrix} 3.80 \\ -2.40 \end{bmatrix} - \begin{bmatrix} 1.27 \\ -2.14 \end{bmatrix} = \begin{bmatrix} 2.53 \\ -0.26 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

$$|\boldsymbol{\Sigma}_1| = 0.25$$

$$p(\boldsymbol{x}|w_1) = \frac{1}{2\pi \cdot \sqrt{0.25}} \cdot e^{-\frac{1}{2}\cdot[2.53 \quad -0.26]\cdot\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}\cdot\begin{bmatrix} 2.53 \\ -0.26 \end{bmatrix}}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2.53 \\ -0.26 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2.53 + 0 \cdot (-0.26) \\ 0 \cdot 2.53 + 4 \cdot (-0.26) \end{bmatrix} = \begin{bmatrix} 2.53 \\ -1.04 \end{bmatrix}$$

$$[2.53 \quad -0.26] \cdot \begin{bmatrix} 2.53 \\ -1.04 \end{bmatrix} = 2.53 \cdot 2.53 + (-0.26) \cdot (-1.04) = 6.67$$

$p(\boldsymbol{x}|w_1) = 0.318 \cdot e^{-\frac{6.67}{2}} = 0.0113$ (Densità di probabilità condizionale di $\boldsymbol{x}$ data w_1)

$p(\boldsymbol{x}) = 0.0018 \cdot 0.70 + 0.0113 \cdot 0.30 = 0.00465$ (Densità di probabilità assoluta dato $\boldsymbol{x}$)

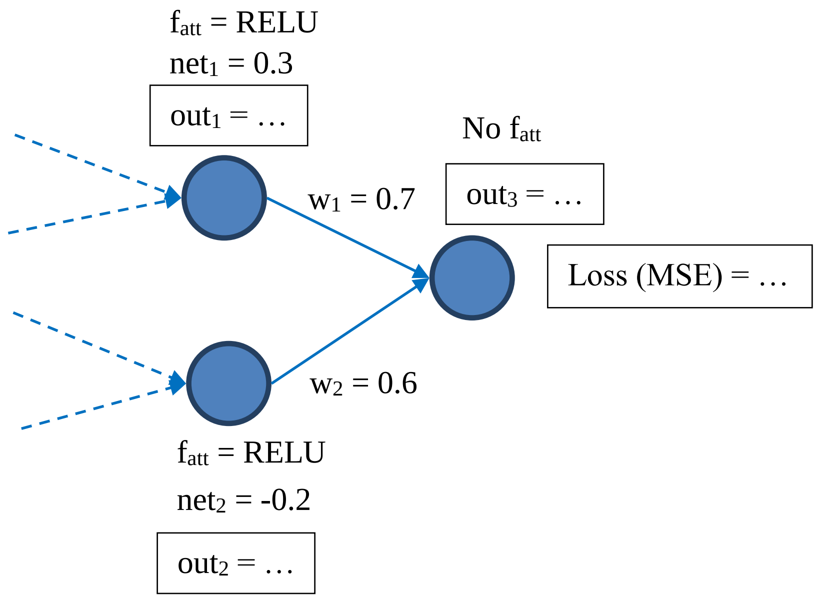
$p(w_0|\boldsymbol{x}) = \frac{0.0018 \cdot 0.70}{0.00465} = 0.271$ (Probabilità a posteriori di w_0 dato $\boldsymbol{x}$)

$p(w_1|\boldsymbol{x}) = \frac{0.0113 \cdot 0.30}{0.00465} = 0.729$ (Probabilità a posteriori di w_1 dato $\boldsymbol{x}$)

Indice della classe restituita: **1**

7) Data la seguente porzione di rete neurale (per un problema di regressione), durante il passo forward un pattern ha prodotto le attivazioni net_1 e net_2 indicate in figura. Motivando le risposte si richiede di:

1. Completare il passo forward calcolando out_1, out_2, out_3 (i neuroni 1 e 2 hanno funzione di attivazione RELU, mentre il neurone 3 non ha funzione di attivazione).
2. Calcolare la loss MSE considerando che il valore atteso per il pattern è $t = 2$.
3. Calcolare i nuovi valori che i pesi w_1 e w_2 assumeranno a seguito del passo backward con un learning rate (η) di 0.1. *Si ricorda che il* δ_3 *(errore) sul neurone 3 in caso di loss MSE senza funzione di attivazione è uguale alla differenza tra valore atteso e attivazione del neurone* $(\mathrm{t} - \mathrm{out}_3)$ *e pertanto il gradiente rispetto al peso* w_i *corrisponde a* $-(\delta_3 \cdot \mathrm{out}_i)$.



<u>Svolgimento</u>

Gli output dei neuroni 1 e 2, data la funzione di attivazione RELU $f(net) = \max(0, net)$, hanno rispettivamente un valore pari a $\mathrm{out}_1 = 0.3$ e $\mathrm{out}_2 = 0$.
Per poter calcolare l'output del neurone 3 è prima necessario calcolare $\mathrm{net}_3 = \mathrm{w}_1 \cdot \mathrm{out}_1 + \mathrm{w}_2 \cdot \mathrm{out}_2 = 0.7 \cdot 0.3 + 0.6 \cdot 0 = 0.21$. Visto che il neurone 3 non ha funzione di attivazione $\mathrm{out}_3 = \mathrm{net}_3 = 0.21$.
Avendo un solo neurone di output, il valore della loss si calcola come $\mathrm{MSE} = (\mathrm{t} - \mathrm{out}_3)^2 = (2 - 0.21)^2 = 3.20$.
Dato $\delta_3 = \mathrm{t} - \mathrm{out}_3 = 1.79$ possiamo calcolare il gradiente rispetto ai pesi w_1 e w_2 come $\frac{\partial J}{\partial \mathrm{w_i}} = -(\delta_3 \cdot \mathrm{out}_i)$:

$$\frac{\partial J}{\partial \mathrm{w}_1} = -(1.79 \cdot 0.3) = -0.54$$

$$\frac{\partial J}{\partial \mathrm{w}_2} = -(1.79 \cdot 0) = 0$$

I nuovi valori dei pesi w_1 e w_2 dopo il passo di backward sono calcolati utilizzando la formula $\mathrm{w}_i' = \mathrm{w}_i - \eta \cdot \frac{\partial J}{\partial \mathrm{w_i}}$. Pertanto:

$$\mathrm{w}_1' = 0.7 - 0.1 \cdot (-0.54) = 0.75$$

$$\mathrm{w}_2' = 0.6 - 0.1 \cdot 0 = 0.6$$