Machine Learning
07-Dic-2016 (60 minuti)

Matricola: ____________________
Cognome: __________ Nome: ____________

1) Dare la definizione di Training, Validation e Test set e discutere una possibile suddivisione dei dati nei tre insiemi.

Pag. 20 delle dispense "Fondamenti"

2) Nella formulazione dell'SVM lineare la funzione obiettivo richiede di massimizzare il margine. L'ottimizzazione è però vincolata; in cosa consistono i vicoli? quanti sono?

Pag. 5 delle dispense "Classificazione (2)"

3) Descrivere le principali criticità e limitazioni dell'algoritmo di Clustering K-means.

Pag. 8-11 delle dispense "Clustering"

4) Nell'ambito di CNN, che cosa si intende per connessioni locali e condivisione di pesi?

Pag. 7 delle dispense "Deep Learning"

5) Date due distribuzioni multinormali identificate dai seguenti parametri:

$\boldsymbol{\mu}_0 = \begin{bmatrix} 1,03 \\ 1,03 \end{bmatrix}$

$\boldsymbol{\Sigma}_0^{-1} = \begin{bmatrix} 52,56 & -10,05 \\ -10,05 & 36,88 \end{bmatrix}$

$|\boldsymbol{\Sigma}_0| = 0,000544$

$P(w_0) = 0,6$

$\boldsymbol{\mu}_1 = \begin{bmatrix} 2,02 \\ 1,53 \end{bmatrix}$

$\boldsymbol{\Sigma}_1^{-1} = \begin{bmatrix} 100,58 & -22,34 \\ -22,34 & 37,85 \end{bmatrix}$

$|\boldsymbol{\Sigma}_1| = 0,000302$

$P(w_1) = 0,4$

Nell'ipotesi dell'impiego di un classificatore di Bayes multinormale, calcolare per il punto $\boldsymbol{x} = \begin{bmatrix} 1,60 \\ 1,25 \end{bmatrix}$:

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

Si ricorda che la densità di probabilità, nel caso della distribuzione multinormale è:

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

<u>Svolgimento</u>

$$\boldsymbol{x} - \boldsymbol{\mu}_0 = \begin{bmatrix} 1,60 \\ 1,25 \end{bmatrix} - \begin{bmatrix} 1,03 \\ 1,03 \end{bmatrix} = \begin{bmatrix} 0,57 \\ 0,22 \end{bmatrix}$$

$$p(\boldsymbol{x}|w_0) = \frac{1}{2\pi \cdot \sqrt{0,000544}} \cdot e^{-\frac{1}{2}\cdot[0,57 \quad 0,22]\cdot\begin{bmatrix} 52,56 & -10,05 \\ -10,05 & 36,99 \end{bmatrix}\cdot\begin{bmatrix} 0,57 \\ 0,22 \end{bmatrix}}$$

$$\begin{bmatrix} 52,56 & -10,05 \\ -10,05 & 36,88 \end{bmatrix} \cdot \begin{bmatrix} 0,57 \\ 0,22 \end{bmatrix} = \begin{bmatrix} 52,56 \cdot 0,57 + (-10,05) \cdot 0,22 \\ (-10,05) \cdot 0,57 + 36,88 \cdot 0,22 \end{bmatrix} = \begin{bmatrix} 27,7482 \\ 2,3851 \end{bmatrix}$$

$$[0,57 \quad 0,22] \cdot \begin{bmatrix} 27,7482 \\ 2,3851 \end{bmatrix} = 0,57 \cdot 27,7482 + 0,22 \cdot 2,3851 = 16,3412$$

$p(\boldsymbol{x}|w_0) = 6,8237 \cdot e^{-\frac{16,3412}{2}} = 0,00193$ (Densità di probabilità condizionale di $\boldsymbol{x}$ data w_0)

$$\boldsymbol{x} - \boldsymbol{\mu}_1 = \begin{bmatrix} 1,60 \\ 1,25 \end{bmatrix} - \begin{bmatrix} 2,02 \\ 1,53 \end{bmatrix} = \begin{bmatrix} -0,42 \\ -0,28 \end{bmatrix}$$

$$p(\boldsymbol{x}|w_1) = \frac{1}{2\pi \cdot \sqrt{0,000302}} \cdot e^{-\frac{1}{2}\cdot[-0,42 \quad -0,28]\cdot\begin{bmatrix} 100,58 & -22,34 \\ -22,34 & 37,85 \end{bmatrix}\cdot\begin{bmatrix} -0,42 \\ -0,28 \end{bmatrix}}$$

$$\begin{bmatrix} 100,58 & -22,34 \\ -22,34 & 37,85 \end{bmatrix} \cdot \begin{bmatrix} -0,42 \\ -0,28 \end{bmatrix} = \begin{bmatrix} 100,58 \cdot (-0,42) + (-22,34) \cdot (-0,28) \\ (-22,34) \cdot (-0,42) + 37,85 \cdot (-0,28) \end{bmatrix} = \begin{bmatrix} -35,9884 \\ -1,2152 \end{bmatrix}$$

$$[-0,42 \quad -0,28] \cdot \begin{bmatrix} -35,9884 \\ -1,2152 \end{bmatrix} = (-0,42) \cdot (-35,9884) + (-0,28) \cdot (-1,2152) = 15,4554$$

$p(\boldsymbol{x}|w_1) = 9,1583 \cdot e^{-\frac{15,4554}{2}} = 0,00403$ (Densità di probabilità condizionale di $\boldsymbol{x}$ data w_1)

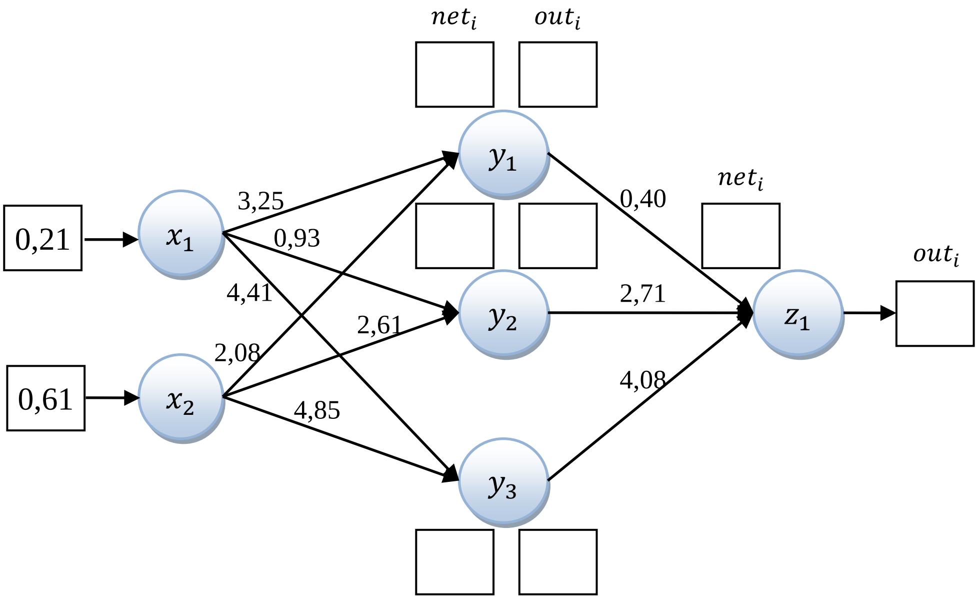
$p(\boldsymbol{x}) = 0,00193 \cdot 0,6 + 0,00403 \cdot 0,4 = 0,00277$ (Densità di probabilità assoluta dato $\boldsymbol{x}$)

$p(w_0|\boldsymbol{x}) = \frac{0,00193\cdot 0,6}{0,00277} = 0,418$ (Probabilità a posteriori di w_0 dato $\boldsymbol{x}$)

$p(w_1|\boldsymbol{x}) = \frac{0,00403\cdot 0,4}{0,00277} = 0,582$ (Probabilità a posteriori di w_1 dato $\boldsymbol{x}$)

Indice della classe restituita: **1**

6) Data la seguente rete neurale, calcolare net_i e out_i di ogni neurone al seguito del passo forward propagation del pattern di input, utilizzando la *standard logistic function* $\left(\frac{1}{1+e^{-net}}\right)$ come funzione di attivazione.



<u>Svolgimento</u>

$net(y_1) = 3{,}25 \cdot 0{,}21 + 2{,}08 \cdot 0{,}61 = 1{,}95$ $\qquad out(y_1) = 0{,}88$
$net(y_2) = 0{,}93 \cdot 0{,}21 + 2{,}61 \cdot 0{,}61 = 1{,}79$ $\qquad out(y_2) = 0{,}86$
$net(y_2) = 4{,}41 \cdot 0{,}21 + 4{,}85 \cdot 0{,}61 = 3{,}88$ $\qquad out(y_3) = 0{,}98$

$net(z_1) = 0{,}40 \cdot 0{,}88 + 2{,}71 \cdot 0{,}86 + 4{,}08 \cdot 0{,}98 = 6{,}68$ $\qquad out(z_1) = 1{,}00$

7) Un multiclassificatore, composto da 5 classificatori combinati a livello di decisione utilizzando la Majority vote rule, viene utilizzato per riconoscere pattern appartenenti a 4 classi. Nella tabella seguente sono riportati gli output restituiti dai singoli classificatori (C_i) dati in input 3 diversi pattern ($\boldsymbol{p}_j$). Riportare nell'ultima colonna la classe di output restituita dal multiclassificatore.

	C_1	C_2	C_3	C_4	C_5	MV
$\boldsymbol{p}_1$	1	2	1	1	3	**1**
$\boldsymbol{p}_2$	1	2	2	3	4	**2**
$\boldsymbol{p}_3$	3	2	1	3	3	**3**