Machine Learning
17-Feb-2017 (90 minuti)

Matricola: ____________________
Cognome: __________ Nome: ____________

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

Pag. 14, 20 delle dispense “Fondamenti”

2) Come si imposta un problema di multiple linear regression? Come sono popolati X, y e β?

Pag. 5 delle dispense “Regressione”

3) Che cos’è il learning rate nell’ambito dell’apprendimento di reti neurali? Cosa succede se viene scelto un learning rate troppo piccolo o troppo grande?

Pag. 15 delle dispense “Reti Neurali”

4) Quali sono i più noti algoritmi di clustering?

Pag. 6 delle dispense “Clustering”

Machine Learning **17-Feb-2017 (90 minuti)**	**Matricola: ____________________** **Cognome: _________ Nome: ___________**

5) Un multiclassificatore, composto da 3 classificatori combinati a livello di confidenza, viene utilizzato per riconoscere pattern appartenenti a 3 classi (A, B, C). Nella tabella seguente sono riportate le confidenze restituite dai singoli classificatori (C_i) dati in input 3 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			Somma				Prodotto				Massimo				Minimo			
	A	B	C	A	B	C	A	B	C	A	B	C	Out	A	B	C	Out	A	B	C	Out	A	B	C	Out
$\boldsymbol{p}_1$	0,15	0,81	0,04	0,02	0,56	0,42	0,54	0,12	0,34	0,71	**1,49**	0,80	**B**	0,00	**0,05**	0,01	**B**	0,54	**0,81**	0,42	**B**	0,02	**0,12**	0,04	**B**
$\boldsymbol{p}_2$	0,31	0,24	0,45	0,54	0,41	0,05	0,02	0,03	0,95	0,87	0,68	**1,45**	**C**	0,00	0,00	**0,02**	**C**	0,54	0,41	**0,95**	**C**	0,02	0,03	**0,05**	**C**
$\boldsymbol{p}_3$	0,42	0,46	0,12	0,77	0,21	0,02	0,41	0,30	0,29	**1,6**	0,97	0,43	**A**	**0,13**	0,03	0,00	**A**	**0,77**	0,46	0,29	**A**	**0,41**	0,21	0,02	**A**

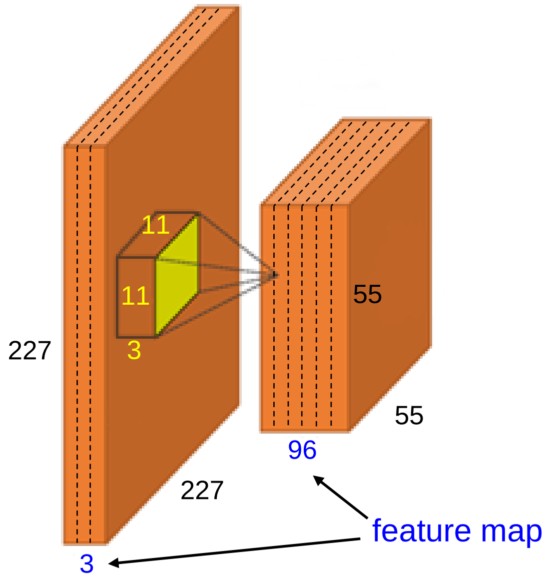
6) Dati un volume di input ed uno di output relativi a un livello di convoluzione in una CNN, aventi le seguenti dimensioni:

- *Volume Input*: $3 \times 227 \times 227$
- *Volume Output*: $96 \times 55 \times 55$

Considerando che ciascun filtro abbia dimensioni:

- *Dimensione Filtro*: $3 \times 11 \times 11$

Si calcoli il numero totale di connessioni e di pesi del livello (senza considerare i bias) motivando la risposta.



Svolgimento

Ogni neurone del livello di output $(96 \times 55 \times 55)$ è connesso con tanti neuroni del livello di input pari alla dimensione del filtro $(3 \times 11 \times 11)$. Pertanto **il numero totale di connessioni** è $(96 \times 55 \times 55) \cdot (3 \times 11 \times 11) = 105\,415\,200$.

Il **numero totale di pesi**, invece, risulta molto più piccolo giacché in una CNN i pesi di ciascun filtro sono condivisi da tutti i neuroni contenuti in una stessa feature map. Visto che il numero di feature map è uguale a 96, ed il numero di input per ciascun filtro è pari a $(3 \times 11 \times 11)$, il numero totale di pesi (senza considerare il bias) è $(3 \times 11 \times 11) \times 96 = 34\,848$.

7) Date tre distribuzioni multinormali identificate dai seguenti parametri:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0{,}75 \\ 4{,}69 \\ 9{,}57 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 1{,}74 \\ 0{,}80 \\ 9{,}59 \end{bmatrix} \qquad \boldsymbol{\mu}_3 = \begin{bmatrix} 6{,}32 \\ 7{,}94 \\ 1{,}82 \end{bmatrix}$$

$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 =$ I (matrice identità) e $P(w_1) = P(w_2) = P(w_3)$.
Indicare la classe assegnata ai seguenti pattern da un classificatore di Bayes multinormale (motivandone la risposta):

$$\mathbf{p}_1 = \begin{bmatrix} 0{,}85 \\ 5{,}12 \\ 9{,}52 \end{bmatrix}, \mathbf{p}_2 = \begin{bmatrix} 9{,}73 \\ 4{,}30 \\ 5{,}41 \end{bmatrix}$$

Svolgimento

La regola di classificazione di Bayes assegna un pattern **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 **x** e il vettore medio $\boldsymbol{\mu}_i$).

$D^2(\mathbf{p}_1, \boldsymbol{\mu}_1) = \mathbf{0,20}$ $\qquad$ $D^2(\mathbf{p}_1, \boldsymbol{\mu}_2) = 19{,}46$ $\qquad$ $D^2(\mathbf{p}_1, \boldsymbol{\mu}_3) = 97{,}16$

$D^2(\mathbf{p}_2, \boldsymbol{\mu}_1) = 98{,}10$ $\qquad$ $D^2(\mathbf{p}_2, \boldsymbol{\mu}_2) = 93{,}56$ $\qquad$ $D^2(\mathbf{p}_2, \boldsymbol{\mu}_3) = \mathbf{37,77}$

$\mathbf{p}_1$ viene assegnato alla classe 1 mentre $\mathbf{p}_2$ viene assegnato alla classe 3.