Machine Learning
15-Feb-2018 (90 minuti)

Matricola: ____________________
Cognome: __________ Nome: ____________

1) Cosa si intende con SVM lineari? Cosa sono le superfici di separazione nel caso d=2 e d=3?

Pag. 4-10 delle dispense "Classificazione (2)"

2) Rispetto a K-means l'approccio di clustering EM con Gaussian mixture quali maggiori flessibilità consente?

Pag. 10, 20 delle dispense "Clustering"

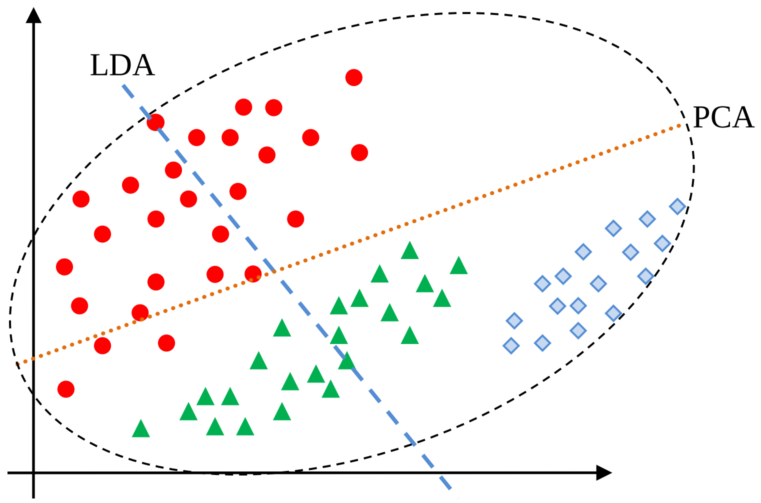
3) Cosa si intende per K-fold cross-validation?

Pag. 20 delle dispense “Fondamenti”

4) Indicare le differenze tra reti neurali feedforward e le reti neurali ricorrenti, disegnando un esempio di entrambe.

Pag. 7 delle dispense “Reti Neurali”

5) Date le distribuzioni riportate nel grafico sottostante, indicare graficamente le soluzioni ottenute (iperpiani) con gli algoritmi PCA e LDA per ridurre la dimensionalità dei pattern (da $d = 2$ a $k = 1$). Motivare la risposta.



Svolgimento

La tecnica PCA identifica l'iperpiano sul quale proiettando i pattern (indipendentemente dalla loro classe) si conserva al massimo l'informazione (linea puntinata arancione). In questo caso corrisponde all'asse principale di un ellisse che comprende tutti i pattern.
La tecnica LDA identifica l'iperpiano sul quale proiettando i pattern si distinguono al meglio le tre classi (linea tratteggiata blu). Nella linea disegnata in figura le proiezioni dei pattern delle classi sono completamente separate.

6) Dato un insieme di pattern tri-dimensionali composto da 4 elementi:

$$\left\{\begin{bmatrix}4{,}8\\1{,}3\\6{,}9\end{bmatrix},\begin{bmatrix}-0{,}7\\1{,}8\\3{,}2\end{bmatrix},\begin{bmatrix}-5{,}8\\9{,}6\\6{,}5\end{bmatrix},\begin{bmatrix}5{,}1\\-5{,}3\\-1{,}5\end{bmatrix}\right\}$$

Calcolare il vettore medio ($\boldsymbol{\mu}$) e la matrice di covarianza ($\boldsymbol{\Sigma} = [\sigma_{ij}]$).

Si ricorda che ogni elemento della matrice di covarianza può essere calcolato come
$\sigma_{ij} = \sigma_{ji} = \frac{1}{n}\sum_{k=1}^{n}(\mathrm{x}_{ki} - \mu_i)\cdot(\mathrm{x}_{kj} - \mu_j)$
dove x_{km} è l'm-esimo elemento del k-esimo pattern, e n il numero di pattern.

Svolgimento

$$\boldsymbol{\mu} = \begin{bmatrix}\dfrac{4{,}8-0{,}7-5{,}8+5{,}1}{4}\\ \dfrac{1{,}3+1{,}8+9{,}6-5{,}3}{4}\\ \dfrac{6{,}9+3{,}2+6{,}5-1{,}5}{4}\end{bmatrix} = \begin{bmatrix}0{,}85\\1{,}85\\3{,}78\end{bmatrix}$$

$$\sigma_{11} = \frac{(4{,}8 - 0{,}85)^2 + (-0{,}7 - 0{,}85)^2 + (-5{,}8 - 0{,}85)^2 + (5{,}1 - 0{,}85)^2}{4} = 20{,}07$$

$$\sigma_{22} = \frac{(1{,}3 - 1{,}85)^2 + (1{,}8 - 1{,}85)^2 + (9{,}6 - 1{,}85)^2 + (-5{,}3 - 1{,}85)^2}{4} = 27{,}87$$

$$\sigma_{33} = \frac{(6{,}9 - 3{,}78)^2 + (3{,}2 - 3{,}78)^2 + (6{,}5 - 3{,}78)^2 + (-1{,}5 - 3{,}78)^2}{4} = 11{,}34$$

$$\sigma_{12} = \sigma_{21} = \frac{(4{,}8 - 0{,}85)(1{,}3 - 1{,}85) + (-0{,}7 - 0{,}85)(1{,}8 - 1{,}85) + (-5{,}8 - 0{,}85)(9{,}6 - 1{,}85) + (5{,}1 - 0{,}85)(-5{,}3 - 1{,}85)}{4} = -21{,}01$$

$$\sigma_{13} = \sigma_{31} = \frac{(4{,}8 - 0{,}85)(6{,}9 - 3{,}78) + (-0{,}7 - 0{,}85)(3{,}2 - 3{,}78) + (-5{,}8 - 0{,}85)(6{,}5 - 3{,}78) + (5{,}1 - 0{,}85)(-1{,}5 - 3{,}78)}{4} = -6{,}83$$

$$\sigma_{23} = \sigma_{32} = \frac{(1{,}3 - 1{,}85)(6{,}9 - 3{,}78) + (1{,}8 - 1{,}85)(3{,}2 - 3{,}78) + (9{,}6 - 1{,}85)(6{,}5 - 3{,}78) + (-5{,}3 - 1{,}85)(-1{,}5 - 3{,}78)}{4} = 14{,}29$$

$$\mathbf{\Sigma} = \begin{bmatrix} 20{,}07 & -21{,}01 & -6{,}83 \\ -21{,}01 & 27{,}87 & 14{,}29 \\ -6{,}83 & 14{,}29 & 11{,}34 \end{bmatrix}$$

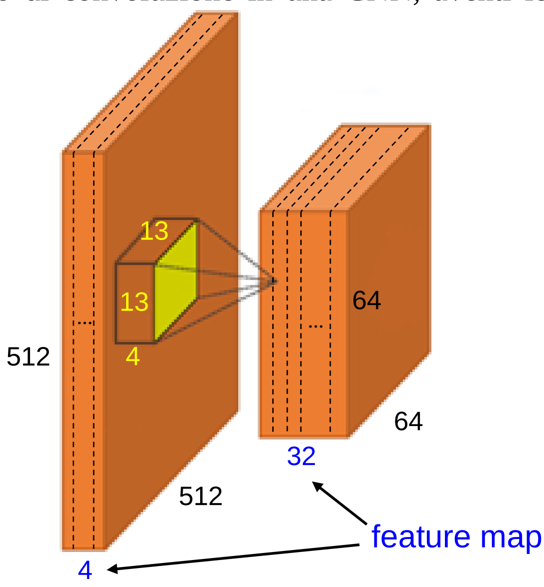
7) Dati un volume di input ed uno di output relativi a un livello di convoluzione in una CNN, aventi le seguenti dimensioni:

- *Volume Input*: $4 \times 512 \times 512$
- *Volume Output*: $32 \times 64 \times 64$

Considerando che ciascun filtro abbia dimensioni:

- *Dimensione Filtro*: $4 \times 13 \times 13$

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



Svolgimento

Ogni neurone del livello di output $(32 \times 64 \times 64)$ è connesso con tanti neuroni del livello di input pari alla dimensione del filtro $(4 \times 13 \times 13)$. Pertanto **il numero totale di connessioni** è $(32 \times 64 \times 64) \cdot (4 \times 13 \times 13) = 88604672$.
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 32, ed il numero di input per ciascun filtro è pari a $(4 \times 13 \times 13)$, il numero totale di pesi (considerando il bias) è $\left((4 \times 13 \times 13) + 1\right) \times 32 = 21664$.