

Corso di Riconoscimento e Recupero dell'Informazione per Bioinformatica - A.A. 2011/2012

Laboratorio - Esercitazione 5

Dr. Marco Cristani

23 aprile 2012

1 Distribuzioni Gaussiane

Considerate due distribuzioni Gaussiane monodimensionali $\mathcal{N}(\mu_1, \sigma_1)$ e $\mathcal{N}(\mu_2, \sigma_2)$, dove

$$\mathcal{N}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

1. Fissati $\mu_1, \mu_2, \sigma_1, \sigma_2$, provate a generare numerosi campioni da queste distribuzioni, visualizzateli e verificate la media e la deviazione standard;
2.
 - Generate due campionature significative (un migliaio di campioni) $\{x_1\}$ e $\{x_2\}$ $x_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ e $x_2 \sim \mathcal{N}(\mu_2, \sigma_2)$.
 - Valutate likelihood e log-likelihood di singoli campioni e di gruppi di campioni.
 - Calcolate la variabile $\{x_3\}$, con $x_3 = x_1 + x_2$.
 - Visualizzate la distribuzione ottenuta.
 - Verificate che sia Gaussiana calcolando e visualizzando l'istogramma.
 - Verificate la relazione tra la media e la varianza di x_3 e le medie e le varianze di x_1 e x_2 .
3. Provate con gaussiane multivariate, per esempio bidimensionali e tridimensionali, prestando particolare attenzione alle matrici di covarianza.

2 Classificatore Bayesiano: caso semplice

Date due distribuzioni Gaussiani monodimensionali di parametri $\mu_A = 3$, $\mu_B = 0$, $\sigma_A = \sigma_B = 1.5$

1. Calcolare la formula del classificatore Bayesiano lineare, nel caso

(a) $P(\omega_A) = P(\omega_B) = 0.5$

(b) $P(\omega_A) = \frac{|\omega_A|}{|\omega_A|+|\omega_B|}$ e $P(\omega_B) = \frac{|\omega_B|}{|\omega_A|+|\omega_B|}$, dove $|\omega_i|$ è la cardinalità della classe i .

2. Visualizzare il confine di decisione.
3.
 - Generare un insieme di dati dalle due distribuzioni gaussiane
 - Stimare il classificatore Bayesiano lineare a partire dai dati e calcolare l'errore di classificazione che si commette
 - Visualizzare il confine di decisione stimato
4. Ripetere il tutto al variare della cardinalità degli insiemi delle due classi. Verificare in particolare che, al crescere della cardinalità delle classi, il confine di decisione stimato approssima il confine di decisione reale.