Politecnico di Milano
Facoltà di Ingegneria dei Sistemi
IV Appello di Statistica Applicata
28 Settembre 2006

Nome e cognome:
Numero di matricola:

Problema 1

Si consideri un problema di classificazione relativo a due distribuzioni Normali su $\Re^n$ con uguale matrice di covarianza Σ non singolare e vettori delle medie rispettivamente uguali a μ_1 e μ_2. Si assuma che i costi di misclassificazione siano uguali e che siano parimenti uguali le probabilità a priori di appartenenza alle due distribuzioni.

(a) Si verifichi che, qualora $\Sigma = \sigma^2 I$ con $\sigma^2 > 0$, allora le due regioni di classificazione coincidono con i luoghi di punti più vicini rispettivamente a μ_1 e a μ_2 secondo la metrica euclidea.

(b) Si fornisca una descrizione geometrica, e non algebrica, della procedura per ottenere la superficie di separazione.

(c) Nel caso in cui $\Sigma \neq \sigma^2 I$ il risultato enunciato in (a) non è più vero. Tuttavia, rappresentiamo con $\Sigma = P\Lambda P'$ la decomposizione spettrale di Σ e consideriamo la trasformazione da $\Re^n$ in $\Re^n$ definita da:

$$x^* = \Lambda^{-\frac{1}{2}} P' x, \quad x \in \Re^n.$$

Si verifichi che nelle nuove coordinate le due regioni di classificazione corrispondono ai luoghi di punti più vicini rispettivamente a $\mu_1^* = \Lambda^{-\frac{1}{2}} P' \mu_1$ e a $\mu_2^* = \Lambda^{-\frac{1}{2}} P' \mu_2$ secondo la metrica euclidea.

Problema 2

Siano X_1 e X_2 due variabili congiuntamente normali tali che: $E(X_1) = 1$, $E(X_2) = 2$, $var(X_1) = 1$, $var(X_2) = 2$ e $cor(X_1, X_2) = 1/\sqrt{2}$. Sia $Y = (X_1 + X_2 - 3)^2$:

(a) Si individui la distribuzione di Y e si calcolino media e varianza.

(b) Si individui la distribuzione condizionata di $Y|(X_1 + 2X_2 = 5)$ e si calcolino media e varianza condizionate.

Problema 3
In seguito è riportato uno script di R utilizzato per generare una serie di numeri pseudocasuali:

```
> library(mvtnorm)
> A <- 10
> B <- c(0,0,0)
> C <- rbind(c(3,1,1),c(1,2,1),c(1,1,1))
> D <- rmvnorm(A,B,C)
> E <- mean(D)
> F <- cov(D)
> E
[1]  0.5  1.0  0.0
> F
     [,1] [,2] [,3]
[1,]  1.5  0.5  0.5
[2,]  0.5  1.0  0.5
[3,]  0.5  0.5  0.5
```

(a) Si riporti il significato delle variabili `A,B,C,D,E,F` riportate nello script.

(b) Utilizzando i dati generati nel precedente script, si effettui un test per la media:

$$H_0 : \mu = (1\ 0\ 0)' \ vs \ H_1 : \mu \neq (1\ 0\ 0)'$$

supponendo la matrice di covarianza ignota e si calcoli il p-value relativo.

(c) Utilizzando i dati generati nel precedente script, si effettui un test per la media:

$$H_0 : \mu = (1\ 0\ 0)' \ vs \ H_1 : \mu \neq (1\ 0\ 0)'$$

supponendo la matrice di covarianza nota e uguale a quella utilizzata nello script per generare i dati e si calcoli il p-value relativo.

(d) Si commentino i risulati dei test e la loro validità.

Problema 4

Un campione *iid* di numerosità 13 è estratto casualmente da una popolazione normale bivariata. Le usuali statistiche sono:

$$\bar{X} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad S = \begin{pmatrix} 2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 \end{pmatrix}$$

Fissato il livello nominale di confidenza $\gamma = 0.9$:

(a) Si individuino e si riportino sullo stesso grafico:

- La regione di confidenza ellissoidale.
- La regione di confidenza individuata dagli intervalli simultanei.
- La regione di confidenza individuata dagli intervalli di Bonferroni.
- Una regione di confidenza rettangolare ottenibile assumendo verificata l'ipotesi di indipendenza tra le due covariate.

(b) Assumendo verificata l'ipotesi di indipendenza tra le due covariate si ordinino i livelli di confidenza reali.

Problema 5

In un ospedale è stato organizzato un esperimento clinico per indagare riguardo ad eventuali differenze tra l'efficacia di 3 farmaci utilizzati per abbassare la febbre ai pazienti febbricitanti. Il farmaco A è stato somministrato a 9 pazienti, il farmaco B a 6 ed il farmaco C a 3. Per ogni paziente è stato misurato, dopo un'ora dalla somministrazione, il decremento in gradi Celsius della temperatura corporea del paziente; i dati raccolti sono stati poi analizzati mediante regressione lineare. Affinché la stima dell'intercetta coincidesse con una stima del decremento medio della temperatura sono state utilizzate come regressori due opportune variabili dummy:

$$X_1 = \begin{cases} 1 & farmaco\ A \\ 0 & farmaco\ B \\ \dots & farmaco\ C \end{cases} \qquad X_2 = \begin{cases} 0 & farmaco\ A \\ 1 & farmaco\ B \\ \dots & farmaco\ C \end{cases}$$

Sono stati ottenuti i seguenti valori:

$$R^2 = 0.1109 \quad \hat{\beta}_0 = 0.87325 \quad \hat{\beta}_1 = -0.20899 \quad \hat{\beta}_2 = -0.05107 \quad \hat{\sigma} = 1.035$$

(a) Si individuino i valori assunti dalle due variabili dummy in corrispondenza della somministrazione del farmaco C.

(b) Si riporti la matrice disegno utilizzata.

(c) Si riportino media e varianza campionaria delle 18 misure.

(d) Si riportino le tre medie campionarie relative ai tre gruppi di pazienti.

(e) Si esegua un test per vedere se almeno uno dei tre farmaci ha un'efficacia diversa dagli altri due.