**Esercizio. Costruire artificialmente una tabella con 5 colonne e 30 righe, in cui le prime due colonne siano molto allineate e l’ultima colonna sia molto correlata ad esse. Applicare poi la regressione lineare multipla, in cui l’ultima colonna sia l’output del modello ed osservare che, nonostante il forte legame, i p-value sono scarsi.**

A = matrix(nrow=30 , ncol=5)

A[,1]=1:30; A[,2]=A[,1]+0.1\*rnorm(30)

(al posto di 1:30 nella definizione di A[,1] potevamo mettere più o meno qualsiasi cosa)

A[,3]=rnorm(30); A[,4]=rnorm(30)

A[,5]=A[,1]+rnorm(30)

Verifiche:

> round(cor(A),2)

[,1] [,2] [,3] [,4] [,5]

[1,] 1.00 1.00 -0.02 0.19 0.99

[2,] 1.00 1.00 -0.02 0.19 0.99

[3,] -0.02 -0.02 1.00 -0.14 -0.01

[4,] 0.19 0.19 -0.14 1.00 0.22

[5,] 0.99 0.99 -0.01 0.22 1.00

fit = lm(A[,5]~ A[,1]+ A[,2]+ A[,3]+ A[,4])

summary(fit)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.01045 0.37051 0.028 0.978

A[, 1] 0.20371 2.41236 0.084 0.933

A[, 2] 0.79689 2.41082 0.331 0.744

A[, 3] 0.15922 0.18411 0.865 0.395

A[, 4] 0.33524 0.19966 1.679 0.106

Residual standard error: 0.9711 on 25 degrees of freedom

Multiple R-squared: 0.9898, Adjusted R-squared: 0.9881

F-statistic: 605.4 on 4 and 25 DF, p-value: < 2.2e-16

Siete invitati a cambiare un po’ di parametri (il grado di allineamento e di legame con l’ultima colonna) per vedere come cambia il risultato.

1. **RADICE QUADRATA DI UNA MATRICE SIMMETRICA DEFINITA POSITIVA**

I comandi, data Q simmetrica definita positiva, sono

Qe = eigen(Q); U=Qe$vectors; lam=Qe$values

sqrtQ = U %\*% diag(sqrt(lam)) %\*% t(U)

ESEMPIO 1. Q diagonale

Q=diag(c(9,4))

> Q

[,1] [,2]

[1,] 9 0

[2,] 0 4

>

Qe = eigen(Q); U=Qe$vectors; lam=Qe$values

> U

[,1] [,2]

[1,] -1 0

[2,] 0 -1

> lam

[1] 9 4

sqrtQ = U %\*% diag(sqrt(lam)) %\*% t(U)

> sqrtQ

[,1] [,2]

[1,] 3 0

[2,] 0 2

Ha funzionato.

ESEMPIO 2. Radice quadrata di una matrice di covarianza empirica.

Intanto, se la matrice di covarianza empirica non è già disponibile, dobbiamo crearcela:

X=matrix(nrow=30,ncol=2)

X[,1]=3\*rnorm(30); X[,2]=rnorm(30)

plot(X,asp=1)

Q=cov(X)

> Q

[,1] [,2]

[1,] 6.1064841 -0.7835786

[2,] -0.7835786 1.6350809

Questa sappiamo che è simmetrica e definita positiva, quindi esiste la sua radice quadrata.

Qe = eigen(Q); U=Qe$vectors; lam=Qe$values

sqrtQ = U %\*% diag(sqrt(lam)) %\*% t(U)

> sqrtQ

[,1] [,2]

[1,] 2.462153 -0.210446

[2,] -0.210446 1.261267

Sarà vero?

> sqrtQ %\*% sqrtQ

[,1] [,2]

[1,] 6.1064841 -0.7835786

[2,] -0.7835786 1.6350809

È proprio Q.

1. **GENERAZIONE DI UN CAMPIONE GAUSSIANO DI COVERIANZA NOTA**

ESERCIZIO. Generare un campione gaussiano di numerosità 1000 a partire dalla matrice di covarianza precedente.

Ricordiamo che i punti generati da cui abbiamo estratto Q erano della forma 

Z=matrix(nrow=2,ncol=1000)

Z[1,]=rnorm(1000); Z[2,]=rnorm(1000)

X= sqrtQ %\*%Z

plot(X[1,],X[2,],asp=1)

plot(t(X),asp=1)



Funziona abbastanza.

1. **ESERCIZI DAL COMPITO DEL 3 settembre 2015**

**Esercizio 3**. Sia A una tabella con 30 righe e 5 colonne già caricata su R. Costruire una matrice B con

6 colonne, le cui prime 5 siano quelle di A, l’ultima contenga 15 zeri e 15 uni, nelle prime

ed ultime 15 righe rispettivamente. [Commento in vista della domanda successiva: gli zeri

corrispondono alla classe C0, gli uni alla classe C1. Abbiamo così classificato le 30 unità

corrispondenti alle 30 righe.]

Supponiamo che su R sia ancora caricata la tabella A dell’inizio di questa esercitazione.:

> round(A,2)

[,1] [,2] [,3] [,4] [,5]

[1,] 1 1.05 -0.04 -0.35 1.63

[2,] 2 2.08 0.05 -0.12 1.41

[3,] 3 2.97 0.13 0.20 3.81

ecc.

B=matrix(nrow=30,ncol=6)

B[,1:5]=A

> round(B,2)

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1 1.05 -0.04 -0.35 1.63 NA

[2,] 2 2.08 0.05 -0.12 1.41 NA

[3,] 3 2.97 0.13 0.20 3.81 NA

ecc.

B[1:15,6]=0; B[16:30,6]=1

> round(B,2)

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1 1.05 -0.04 -0.35 1.63 0

[2,] 2 2.08 0.05 -0.12 1.41 0

[3,] 3 2.97 0.13 0.20 3.81 0

…….

[16,] 16 16.00 -0.32 -0.05 17.29 1

[17,] 17 17.02 -1.04 1.62 16.77 1

[18,] 18 18.07 0.39 0.05 18.26 1

ecc.

**Esercizio 4**. Applicare la regressione lineare multipla alla matrice B della domanda precedente, usando

L’ultima colonna come variabile di output. Caricata su R una stringa di 5 numeri x=(x1,...,x5),

che corrisponde ad una nuova unità, con che comandi e ragionamenti sui risultati numerici

decidereste se essa sia di classe C0 o C1?

fit = lm(B[,6]~ B[,1]+ B[,2]+ B[,3]+ B[,4]+B[,5])

x=c(3,pi/2,1.2,0.5,10.56)

y=fit$coef%\*%c(1,x)

> y

[,1]

[1,] -0.480891

Il numero è più vicino a zero e quindi viene classificato come C0.

**Esercizio 6**. Generare 100 coppie di stringhe indipendenti lunghe 30, secondo una N(0;25), calcolare la covarianza di ciascuna coppia riempiendo così il vettore denominato COVA, tracciare un istogramma dei 100 valori ottenuti.

COVA=1:100

for (i in 1:100){

X=rnorm(30,0,5); Y=rnorm(30,0,5)

COVA[i]=cov(X,Y)

}

hist(COVA,20)

