1. **SOLUZIONE ESERCIZIO RIASSUNTIVO**

n=1000

Z1 = rnorm(n); Z2 = rnorm(n); plot(Z1,Z2)

X1 = 3\*Z1; X2 = Z2; plot(X1,X2)

theta = pi/10

A = matrix(nrow=2,ncol=2)

A[1,1]= cos(theta)

A[1,2]= -sin(theta)

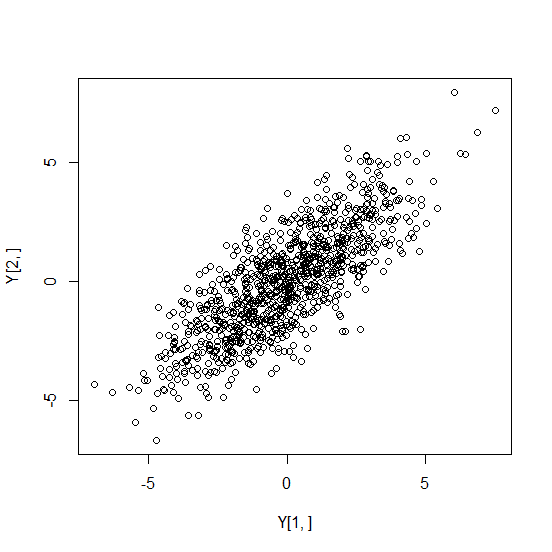
A[2,1]= sin(theta); A[2,2]= cos(theta)

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

X[1,]=X1; X[2,]=X2

Y=A%\*%X

plot(Y[1,], Y[2,])



e = eigen(Q)

U = e$vectors

B = U %\*% diag(sqrt(e$values)) %\*% t(U)

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

Z[1,]=Z1; Z[2,]=Z2

Y.sim = B %\*% Z

plot(Y.sim[1,], Y.sim[2,])

1. **VERSO PCA (l’osso di seppia)**

Questa parte di esercitazione è piuttosto avanzata rispetto al solito.

Richiede innanzi tutto l’installazione di pacchetti aggiuntivi di R. In fondo a questo file si trovano le istruzioni per installare pacchetti.

Supponiamo quindi di aver installato i pacchetti rgl, misc3d, plot3D, plot3Drgl.

I pacchetti, da R, vanno caricati (non vengono caricati tutti di default):

require(rgl)

require(misc3d)

require(plot3D)

require(plot3Drgl)

Generiamo (un osso di seppia, ovvero) dei punti gaussiani dello spazio coi comandi:

n=10000

Z1 = rnorm(n); Z2 = 10\*rnorm(n); Z3 = 5\*rnorm(n)

X01 = 0.707\*Z1-0.707\*Z2; X02 = 0.707\*Z1+0.707\*Z2; X03=Z3

X1=X01; X2 = 0.707\*X02 - 0.707\*X03; X3 = 0.707\*X02 + 0.707\*X03

plot3d(X1,X2,X3)

Diminuiamo drasticamente i punti, immaginando che siano individui (province, esperimenti, ecc.):

n=20

Z1 = rnorm(n); Z2 = 10\*rnorm(n); Z3 = 5\*rnorm(n)

X01 = 0.707\*Z1-0.707\*Z2; X02 = 0.707\*Z1+0.707\*Z2; X03=Z3

X1=X01; X2 = 0.707\*X02 - 0.707\*X03; X3 = 0.707\*X02 + 0.707\*X03

plot3d(X1,X2,X3)

Si ruoti per capire quale può essere la miglior visuale.

1. **PRIME PROVE SU PCA**

A <- read.table ('clipboard', header=TRUE)

PLIC SC SA.SC TD TMI

Piem 0.088 0.471 -0.707 -0.607 -0.3950

Vaos -1.545 0.348 -0.642 -0.813 1.5780

Lomb 0.202 1.397 -0.836 -0.790 -0.5380

TrAA 0.677 0.435 -1.269 -0.966 -0.0750

Vene 0.088 1.334 -1.210 -0.848 -0.4970

FrVG 0.639 -0.005 -1.028 -0.804 -1.3010

Ligu 1.190 -0.247 0.470 -0.429 -0.3540

EmRo 0.658 1.177 -1.315 -0.863 -0.3470

Tosc 0.126 1.092 -0.795 -0.644 -1.3550

Umbr -1.431 0.675 -0.140 -0.524 -1.2870

Marc 0.278 1.090 -0.265 -0.702 -0.0006

Lazi 2.329 0.546 -0.080 -0.113 -0.0140

Abru 0.335 -0.373 0.402 -0.456 0.0400

Moli 0.658 -1.289 0.065 0.451 -1.1510

Camp -1.811 -1.314 2.031 1.664 0.4140

Pugl -0.766 -0.926 1.038 0.648 1.1090

Basi -0.747 -1.154 0.661 0.844 2.0010

Cala -0.500 -1.727 1.571 2.153 0.6320

Sici -0.918 -1.130 1.332 1.517 1.7830

Sard 0.449 -0.403 0.717 1.285 -0.2380

cor(A)

PCA = princomp(A)

biplot(PCA)

plot(PCA)

summary(PCA)

Con questi comandi si trovano ad esempio le varianze spiegate.

1. **ARRICCHIMENTO GRAFICO**

> summary(PCA)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5

Standard deviation 1.7779900 0.9103965 0.7114299 0.40172430 0.29690293

Proportion of Variance 0.6661239 0.1746455 0.1066499 0.03400577 0.01857485

Cumulative Proportion 0.6661239 0.8407694 0.9474194 0.98142515 1.00000000

varianza.cumulativa=c(0, 0.666, 0.840, 0.947, 0.981, 1.000)

plot(varianza.cumulativa,type="b")

1. **ESPLORAZIONE DI TUTTI I PIANI**

C<-predict(PCA)

i=2; j=3

plot(C[,c(i,j)],type="n",asp=1)

text(C[,c(i,j)],labels=as.character(row.names(A)))

Mettiamo le esplorazioni insieme al biplot

par(mfrow=c(1,2))

biplot(PCA)

i=2; j=3

plot(C[,c(i,j)],type="n",asp=1)

text(C[,c(i,j)],labels=as.character(row.names(A)))

può servire a riconoscere se certi clusters sono davvero tali.

1. **ESPLORAZIONE TRIDIMENSIONALE**

require(rgl)

require(misc3d)

require(plot3D)

require(plot3Drgl)

par(mfrow=c(1,1))

biplot(PCA)

plot3d(PCA$scores[,1:3])

text3d(PCA$scores[,1:3],texts=rownames(A))

text3d(PCA$loadings[,1:3], texts=rownames(PCA$loadings), col="red")

coords <- NULL

for (i in 1:nrow(PCA$loadings)) {coords <- rbind(coords, rbind(c(0,0,0),PCA$loadings[i,1:3]))}

lines3d(coords, col="red", lwd=4)

Si suggerisce di ruotare la figura fino a farla coincidere col biplot.

Spettacolare osservare il posizionamento reale delle frecce e delle regioni.

1. **PUNTEGGI E CLASSIFICHE**

predict(PCA)

benessere = -predict(PCA)[,1]

benessere

sort(benessere)

1. **LOADINGS**

PCA$loadings

Essi possono essere usati o qualitativamente per “dare un nome” alle componenti principali (in questo senso il SW pone =0 i valori piccoli, per aiutare le associazioni), oppure quantitativamente per calcolare delle relazioni lineari, come detto nelle slide.

1. **APPENDICE: SCARICAMENTO PACCHETTI AGGIUNTIVI**

Primo modo (ad es. per scatterplot3D): dopo averlo trovato alla pag.

<http://cran.r-project.org/web/packages/scatterplot3d/index.html>

si esegue il download. Ad es. chi ha Windows può scaricare

|  |  |
| --- | --- |
| Windows binaries: | r-release: [scatterplot3d\_0.3-35.zip](http://cran.r-project.org/bin/windows/contrib/3.1/scatterplot3d_0.3-35.zip) |

Poi, si inserisce l’intera cartella (decompressa, ma forse questo non è necessario) in library. Chiudere R e ripartire (può darsi che questa operazione non sia necessaria). A questo punto è apribile.

Secondo modo: usare la tendina “Pacchetti” dalla pagina di R, da cui si può o installare da zip locali (saltando quindi metà delle operazioni precedenti) oppure usare da subito “installa pacchetti” che reca ad un indirizzo a scelta, e. Padova, dove poi si trovano tutti i pacchetti.

---- NOTA PER UTENTI LINUX ----

Relativamente a rgl e misc3d le funzioni associate sono presenti nelle versioni di sistema di R.

E’ sufficiente installare i pacchetti: r-cran-rgl r-cran-misc3d

ad esempio in distribuzioni derivate debian/ubuntu con il comando

sudo apt-get install r-cran-rgl r-cran-misc3d

Per quanto riguarda plot3D e plot3Drgl si possono installare all'interno del software R con i comandi

install.packages("plot3D", repos="<http://R-Forge.R-project.org>")

install.packages("plot3Drgl", repos="<http://R-Forge.R-project.org>")

(si può scegliere volendo un altro source per i pacchetti ad esempio padova)

Su linux inoltre si può avere l'accortezza di avviare R con i privilegi di amministratore in modo da poter installare le librerie di cui sopra nella cartella di sistema, altrimenti R propone di installarle in una cartella locale.

--- fine nota ----