

Feuille d'Exercices : Analyse en Composantes Principales

Vincent Guillemot

14 février 2009

Exercice 1 : Calcul matriciel

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \text{ et } B = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \text{ et } C = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

1. Calculez les produits AB et BCA .
2. Montrez que la matrice $C - I$ est inversible.
3. Quelles sont les valeurs propres de C ? C est-elle inversible?

Exercice 2 : Plus grande valeur propre

Soit S une matrice carrée symétrique de taille $n \times n$. Soit la fonction f définie pour tout x de \mathbb{R}^n non nul :

$$f(x) = \frac{x^\top S x}{\|x\|}.$$

Montrez que f admet un maximum, et qu'il est égal à λ_{max} , la valeur propre maximale de S .

Exercice 3 : Directions principales

Soit un jeu de données \mathbf{X} de taille $n \times p$ (n individus et p variables). La matrice de variance covariance de \mathbf{X} est la matrice

$$S = \frac{1}{n} \mathbf{X}^\top \mathbf{X}.$$

La projection de \mathbf{X} sur un vecteur de \mathbb{R}^p de norme unitaire a s'exprime simplement par le produit scalaire

$$Y = \mathbf{X}a^\top.$$

La variabilité de ce nouveau vecteur est exprimé par le scalaire $I(a) = \|Y\|^2 = Y^\top Y$.

1. Exprimez $I(a)$ en fonction de S , a et n .
2. D'après l'exercice précédent, quand le scalaire $I(a)$ atteint-il sa valeur maximale?
3. En déduire les directions principales du jeu de données \mathbf{X} .

Exercice 4 :

Simulation de données

Pour simuler des données, téléchargez tout d'abord le package `rgl` :

```
source("http://www.bioconductor.org/biocLite.R")
biocLite(c("rgl","mvtnorm"))
require(rgl) ; require(mvtnorm)
```

Exécutez le petit exemple ci-dessous, interprétez chaque ligne !

```
sigma <- matrix(c(1,0.8,0.02,0.8,1,0,0.02,0,1),3,3)
eigen(sigma)$values
x <- rmvnorm(50,sigma=sigma);colnames(x) <- c("a","b","c")
plot3d(x,type = "s")
prcomp(x)
plot3d(prcomp(x)$x,type = "s")
v <- eigen(cov(x))$vectors
plot3d(x%*%v,type = "s")
biplot(prcomp(x))
```

Inspirez vous en pour construire un exemple à deux classes :

- (i) la première classe aura une moyenne nulle,
- (ii) la deuxième aura une moyenne égale à

$$\mu = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

```
(rmvnorm(n=10,mean=c(1,1,0),sigma=sigma))
```

Application à des données réelles

```
biocLite("multtest")
require(multtest)
data(golub)
plot(prcomp(t(golub))$x[,1:2],col=golub.c1+1,pch=20)
```