Scoring Appliqué à la Détection du Risque – Chapitre 1 – Exercices

Exercice 1.

Montrez que l'Analyse Factorielle Discriminante à plusieurs facteurs conduit à des variables discriminantes non corrélées.

Exercice 2.

Des observations de $\mathbb{R}^d$ sont réparties en deux groupes de centres $\bar{\boldsymbol{x}}_1$ et $\bar{\boldsymbol{x}}_2$. Montrer que l'axe factoriel dans l'Analyse Factorielle Discriminante basée sur la métrique de Mahalanobis est dirigé par $\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}}_1$.

Exercice 3.

Le fichier `client`[a] donne pour cent clients bancaires : la liquidité (`cash`), le flux (`flow`), l'épargne (`saving`), le niveau de consommation (`consume`) et la classe de risque (`risk`).

A quelle classe de risque l'Analyse Factorielle Discriminante basée sur la métrique de Mahalanobis affecte-t-elle un nouveau client avec les caractéristiques suivantes : liquidité $= 6,2$; flux $= 2,9$; épargne $= 4,9$; consommation $= 1,7$?

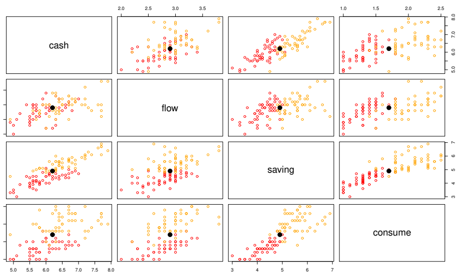

Fig. 1: *Cent clients bancaires décrits par quatre variables continues (cash, flow, saving, consume) et répartis en deux classes de risque (0 en rouge, 1 en orange) ; à quelle classe de risque doit-on affecter le nouveau client (point noir) ?*

Exercice 4.

On considère les données `banknote{mclust}` de R. A quelle classe (genuine/counterfeit) l'Analyse Factorielle Discriminante basée sur la métrique euclidienne affecte-t-elle un billet présentant les caractéristiques suivantes : `Length:214.9`, `Left:130.1`, `Right:129.8`, `Bottom:9.4`, `Top:10.6`, `Diagonal:140.3` ?

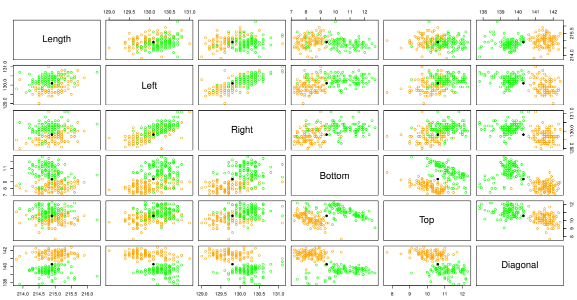

Fig. 2: *Deux cents billets de banque décrits par six variables continues (Length, Left, Right, Bottom, Top, Diagonal) et répartis en deux classes : genuine, counterfeit ; à quelle classe doit-on affecter le nouveau billet (point noir) ?*

Exercice 5.

Le fichier `invest`[b] donne pour cent cinquante investissements la valeur de quatre ratios : `R1`, `R2`, `R3`, `R4`, ainsi que la classe de risque : `low`, `medium`, `high`. A quelle classe de risque l'Analyse Factorielle Discriminante à deux facteurs basée sur la métrique de Mahalanobis affecte-t-elle le dernier investissement dont le risque est inconnu ?

A. Lourme, Faculté d'économie, gestion & AES, Université de Bordeaux `http://alexandrelourme.free.fr`

[a] disponible sous : `http://alexandrelourme.free.fr/M2IREF/SCORING/client.csv`
[b] disponible sous : `http://alexandrelourme.free.fr/M2IREF/SCORING/invest.csv`

Exercice 3.

Analyse Factorielle Discriminante à un facteur

```
client <- read.table(file='http://alexandrelourme.free.fr/M2IREF/SCORING/client.csv'
,sep=',',dec='.',header=TRUE)
```

Les données

```
X=as.matrix(client[,1:4]) # matrice des descripteurs
n=nrow(X) # taille de l'echantillon
d=ncol(X) # nombre de descripteurs/dimension de l'espace
G=length(unique(client[,5])) # nombre de groupes
Z=matrix(0,nrow=n,ncol=G) # tableau disjonctif complet/matrice des appartenances
for (i in 1:n){
Z[i,1]=ifelse (client[i,5]==0,1,0)
Z[i,2]=ifelse (client[i,5]==1,1,0)
}
```

Conditionnement des données

```
one=matrix(rep(1,n),nrow=n)
barx=t(colMeans(X)) # centre du nuage
X <- X-one%*%barx # centrage du nuage
new=as.vector(c(6.2,2.9,4.9,1.7)) # nouveau client
new <- new - barx # recentrage du nouveau client
```

Effectifs des groupes

```
ng <- NULL
for (g in 1:G){
ng[[g]]=sum(Z[,g]==1) # effectif du groupe g
}
```

Centres des groupes

```
barxg <- NULL
for (g in 1:G){
barxg[[g]]=colMeans(X[Z[,g]==1,]) # centres du groupe g
}
```

Matrices de variance des groupes

```
Vg <- NULL
for (g in 1:G){
Vg[[g]]=var(X[Z[,g]==1,])*(ng[[g]]-1)/ng[[g]] # matrice de variance du groupe g
}
```

Matrice de variance intra groupes

```
W <- matrix(0,nrow=d,ncol=d) # matrice de variance intra groupes
for (g in 1:G){W=W+ng[[g]]/n*Vg[[g]]}
```

Matrice de variance inter groupes

```
B <- matrix(0,nrow=d,ncol=d) # matrice de variance inter groupes
for (g in 1:G){B=B+ng[[g]]/n*barxg[[g]]%*%t(barxg[[g]])}
```

Matrice de variance

```
V=B+W # théorème de Konig-Huygens
V=var(X)*(n-1)/n # calcul direct
```

Analyse spectrale de $V^{-1}B$

```
# facteur discriminant
# M=diag(rep(1,d)) # choix de la métrique de Mahalanobis
M=solve(V) # choix de la métrique de Mahalanobis
EIG <- eigen(solve(V)%*%B)
lambda=EIG$values[1] # rapport de corrélation maximal d'une série obtenue par projection
u=as.vector(EIG$vectors[,1]) # facteur discriminant
u <- u/c(sqrt(t(u)%*%solve(M)%*%u)) # normalisation de u
```

Vecteur directeur M-unitaire de l'axe factoriel

```
# axe discriminant
a=as.vector(solve(M)%*%u)
a=a/c(sqrt(t(a)%*%M%*%a))
```

Allocation du nouveau client

```
s=X%*%u # variable discriminante
barvg <- NULL
for (g in 1:G){# centres des groupes sur la variable discriminantes
barvg[[g]]=mean(s[Z[,g]==1,])
}
snew=new%*%u #abscisse du nouveau client sur l'axe factoriel

dist2group1=abs(snew-barvg[[1]]) # distance au groupe 1
dist2group2=abs(snew-barvg[[2]]) # distance au groupe 2
print(dist2group1)

##            [,1]
## [1,] 0.9251935

print(dist2group2)

##            [,1]
## [1,] 0.8455574
```

Le nouveau client est affecté au groupe 2 (codé 1 dans les données).