

Projet Scoring 1_Version_fin_2

Youming & Quentin

2023-09-21

Exercice 1

$$B = \sum_{g=1}^G (\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})' \frac{n_g}{n}$$

$$\bar{x} = \sum_{i=1}^n \frac{n_i}{n} \bar{x}_i = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2$$

étape 1

$$B = \sum_{g=1}^G \left\{ (\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})' \times \frac{n_1}{n} + (\bar{x}_2 - \bar{x})(\bar{x}_2 - \bar{x})' \times \frac{n_2}{n} \right\}$$

$$= \frac{n_1}{n} \left[\bar{x}_1 - \left(\frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2 \right) \right] \left[\right]' + \frac{n_2}{n} \left[\bar{x}_2 - \left(\frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2 \right) \right] \left[\right]'$$

$$= \frac{n_1}{n} \left[\left(1 - \frac{n_1}{n} \right) \bar{x}_1 - \frac{n_2}{n} \bar{x}_2 \right] \left[\right]' + \frac{n_2}{n} \left[\left(1 - \frac{n_2}{n} \right) \bar{x}_2 - \frac{n_1}{n} \bar{x}_1 \right] \left[\right]'$$

$$= \frac{n_1}{n} \left[\frac{n_2}{n} (\bar{x}_1 - \bar{x}_2) \right] \left[\right]' + \frac{n_2}{n} \left[\frac{n_1}{n} (\bar{x}_2 - \bar{x}_1) \right] \left[\right]'$$

$$= \frac{n_1}{n} \times \frac{n_2^2}{n^2} (\bar{x}_1 - \bar{x}_2) (\bar{x}_1 - \bar{x}_2)' + \frac{n_2}{n} \times \frac{n_1^2}{n^2} (\bar{x}_2 - \bar{x}_1) (\bar{x}_2 - \bar{x}_1)'$$

$$= \left(\frac{n_1}{n} \times \frac{n_2^2}{n^2} + \frac{n_2}{n} \times \frac{n_1^2}{n^2} \right) (\bar{x}_1 - \bar{x}_2) (\bar{x}_1 - \bar{x}_2)'$$

$$= \frac{n_1 n_2}{n^2} \left(\frac{n_2}{n} + \frac{n_1}{n} \right) (\bar{x}_1 - \bar{x}_2) (\bar{x}_1 - \bar{x}_2)'$$

$$= \frac{n_1 n_2}{n^2} (\bar{x}_1 - \bar{x}_2) (\bar{x}_1 - \bar{x}_2)'$$

On est dans le cas particulière. a est colinéaire à $(\bar{x}_1 - \bar{x}_2)$
et u est colinéaire à $V(\bar{x}_1 - \bar{x}_2)$

Donc a est colinéaire à $V(\bar{x}_1 - \bar{x}_2)$

Exercice 2

Bonjour, ceci est notre version finale__2 , veuillez oublier les versions précédentes, merci pour votre compréhension.

Bonjour, ceci est notre version finale__2 , veuillez oublier les versions précédentes, merci pour votre compréhension.

Bonjour, ceci est notre version finale__2 , veuillez oublier les versions précédentes, merci pour votre compréhension.

Merci encore!!!

```
I <- read.table(file = "http://alexandre.lourme.free.fr/M2IREF/SCORING/invest.csv" , sep=',', dec='.
```

Les données

```
X=as.matrix(I[1:150,1:4]) # matrice des descripteurs
n=nrow(X) # taille de l'échantillon
d=ncol(X) # nombre de descripteurs/dimension de l'espace
G=length(unique(I[1:150,5])) # 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 (I[i,5]=="low",1,0)
  Z[i,2]=ifelse (I[i,5]=="medium",1,0)
  Z[i,3]=ifelse (I[i,5]=="high",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 = I[151,1:4] # nouveau investissement
new <- new - barx # recentrage du nouveau investissement
```

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$values[1:2] # rapport de corrélation maximal d'une série obtenue par projection
u=as.matrix(EIG$vectors[,1:2]) # facteur discriminant
for (i in 1:2) {# normalisation de u
```

```

u[,i] <- u[,i]/c(sqrt(t(u[,i])%*%solve(M)%*%u[,i]))
}

```

Vecteur directeur M-unitaire de l'axe factoriel

```

# axe discriminant
a <- NULL
a= solve(M)%*%u
for (i in 1:2) {
  a[,i]=a[,i]/c(sqrt(t(a[,i])%*%M%*%a[,i]))
}

```

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]]=colMeans(s[Z[,g]==1,])
}

snew=as.matrix(new) %*% u #abscisse du nouveau investissement sur l'axe factoriel
dist2group1=abs(snew-barvg[[1]]) # distance au groupe 1
dist2group2=abs(snew-barvg[[2]]) # distance au groupe 2
dist2group3=abs(snew-barvg[[3]]) # distance au groupe 3
print(dist2group1)

```

```

##           [,1]           [,2]
## 151  1.382833  0.0003685579

```

```

##[,1]      [,2]
## 1.382833  0.0003685579
print(dist2group2)

```

```

##           [,1]           [,2]
## 151  0.2710489  0.8398573

```

```

##[,1]      [,2]
## 0.2710489  0.8398573
print(dist2group3)

```

```

##           [,1]           [,2]
## 151  0.9649406  0.265555

```

```
##[,1]      [,2]
## 0.9649406 0.265555

for (i in 1:3) {
  dist = abs(snew - barvg[[i]])
  print(paste("La distance de group", i, "=", sqrt(dist[1]^2 + dist[2]^2)))
}

## [1] "La distance de group 1 = 1.38283275002058"
## [1] "La distance de group 2 = 0.882512210862827"
## [1] "La distance de group 3 = 1.00081461639357"
```

Le nouveau investissement est affecté au groupe 2 (codé “medium” dans les données).