Analyse discriminante (FDA – LDA –QLDA)

Master MLDS - 2017/2018

Université Paris Descartes

Lazhar.labiod@parisdescartes.fr

Analyse discriminante

- Introduction
- Approche géométrique
  - Analyse factorielle discriminante (AFD - FDA)
- Approche prédictive (probabiliste)
  - Analyse discriminante linéaire (LDA)
  - Analyse discriminante quadratique (QLDA)
- Exemple sous R

Deux approches

- **Méthode factorielle**: description "géométrique" de la séparation inter-classe (encore appelée analyse discriminante factorielle ou analyse discriminante linéaire de Fisher)

- M**éthode prédictive (interprétation bayesienne)** : classificateur bayesien (optimum au sens de la proba. de l'erreur) dans des conditions particulières pour les données! (encore appelée analyse discriminante décisionnelle, linéaire ou quadratique)

Analyse factorielle discriminante

- **Notations :**
- n observations, individus statistiques. $X_1, X_2, \ldots, X_n$ décrits par p variables quantitatives.
- Les variables descriptives $X^1, X^2, \ldots X^p$ doivent être ***quantitatives***!
- un critère qualitatif présentant K modalités, $G1, G2, \ldots, GK$
- ce critère divise la population en K sous populations (groupes)

Principes de l'AFD

- Objectif :
  - Discriminer (séparer, caractériser) K groupes d'individus *préalablement définis*, décrits par p variables quantitatives.
  - mettre en évidence des différences entre les classes c-à-d entre les observations appartenant à des classes différentes

- Moyen :
  - Rechercher des combinaisons linéaires des p variables initiales (axes discriminants) permettent de caractériser au mieux les groupes.
  - => description des liaisons entre la variable "classe" et les variables quantitatives: les *K* classes diffèrent-elles sur l'ensemble des variables numériques?

- D'un point de vue technique, l'AFD peut être vue comme l'ACP normée du nuage des centres de gravités des K groupes d'individus, munis du poids des groupes.

AFD: principes de base

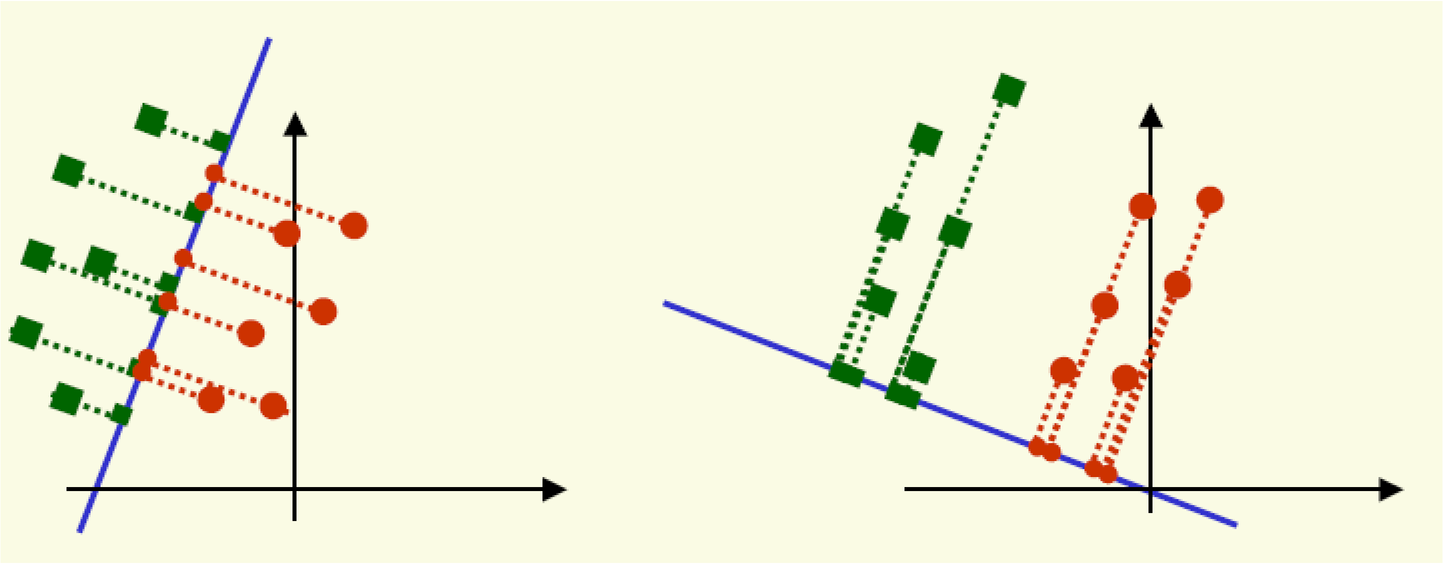
- ***Objectif :*** mettre en évidence des différences entre les classes c-à-d entre les observations appartenant à des classes différentes
- => description des liaisons entre la variable "classe" et les variables quantitatives: les K classes diffèrent-elles sur l'ensemble des variables numériques?

- ***Méthode :*** déterminer un/des facteur(s), combinaison(s) linéaire(s) des variables descriptives, qui prenne(nt) des valeurs les + proches possible pour des éléments de la même classe, et les + éloignées possible entre éléments de classes différentes. (= facteurs discriminants)

- **Formulation :**
- ***1) Décomposition de la matrice variance-covariance V***
- Ensemble des n observations $\mathbf{x}i$ = un nuage de points, de centre de gravité $\mathbf{m}$ et de matrice variance-covariance $\mathbf{V}$.
- Ce nuage est partagé en K ***sous-nuages*** par la variable "classe".
- Chaque sous-nuage (classe Gk) d'effectif nk est caractérisé par son centre de gravité (ou ***centroïde***) $\mathbf{m}k$ et sa matrice variance-covariance $\mathbf{V}k$.

AFD (K=2 classes)

- Idée : projeter les données sur un axe discriminant qui permet une meilleure séparation des classes

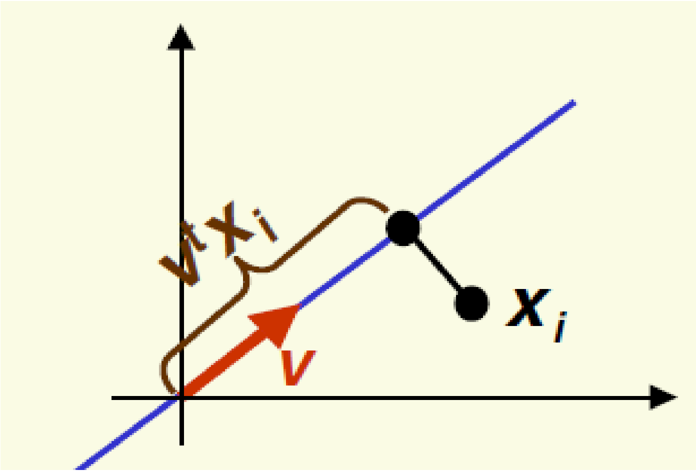


Mauvaise discrimination

Bonne discrimination

AFD..

- On suppose que nous avons deux classes et n observations X1,…, Xn définies dans un espace de dimension p. soit :
  - n1: le nombre d'observations dans la classe c1
  - n2: le nombre d'observations dans la classe c2
- On considère la projection du nuage de point sur l'axe définit par le vecteur unité V.
- Le scalaire $v^T x_i$ est la distance entre la projection de Xi est l'origine
- Alors $v^T x_i$ est la projection (la coordonnée) de Xi dans un sous espace de dimension 1



AFD..

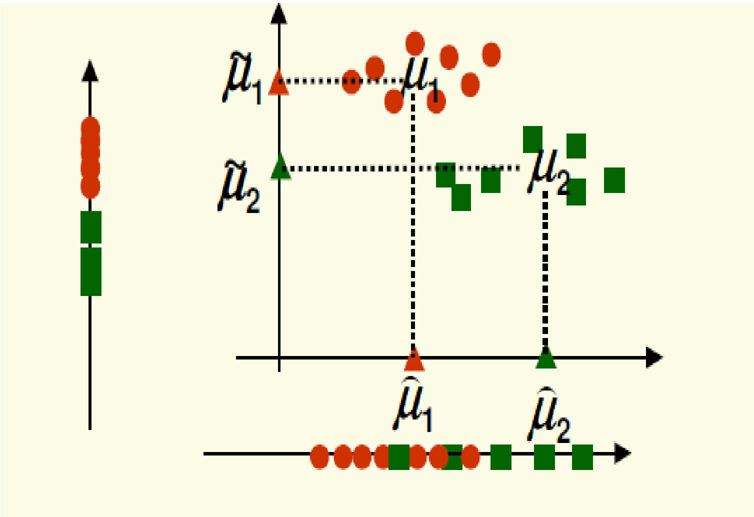
- La projection de xi sur l'axe v est donnée par $v^T x_i$
- **Comment mesurer la séparation entre les projections des différentes classes?**
- Soit $\tilde{m}_1$ et $\tilde{m}_2$ les centres des projections des classes 1 et 2
- Soit m1 and m2 les centres des classes 1 et 2

Centres des classes:

$$m_1 = \frac{1}{n_1} \sum_{x_i \in C_1} x_i, \quad m_2 = \frac{1}{n_2} \sum_{x_i \in C_2} x_i$$

Centres des projections des classes:

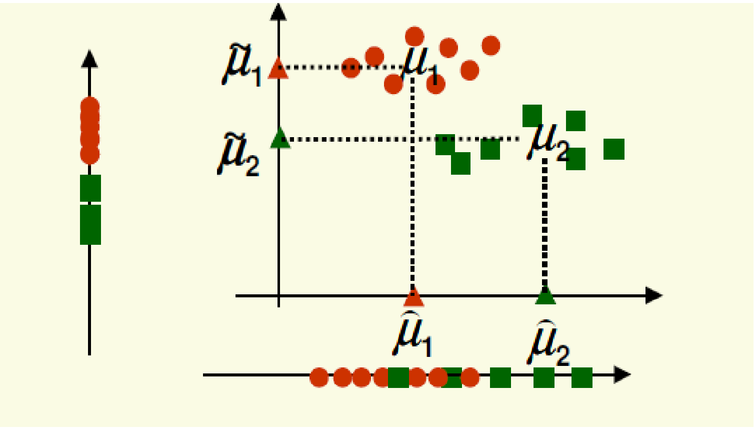
$$\tilde{m}_1 = \frac{1}{n_1} \sum_{x_i \in C_1} v^T x_i = v^T m_1, \qquad \tilde{m}_2 = \frac{1}{n_2} \sum_{x_i \in C_2} v^T x_i = v^T m_2$$



- $\left|\tilde{m}_2 - \tilde{m}\right| = \left|v^T (m_2 - m_1)\right|$ semble être une bonne mesure de séparation

AFD..

- $\left|\tilde{m}_2 - \tilde{m}_1\right| = \left|v^T(m_2 - m_1)\right|$ **Est elle une bonne mesure de séparation?**
- Plus $\left|\tilde{m}_2 - \tilde{m}_1\right|$ est grande, meilleure est la séparation attendue

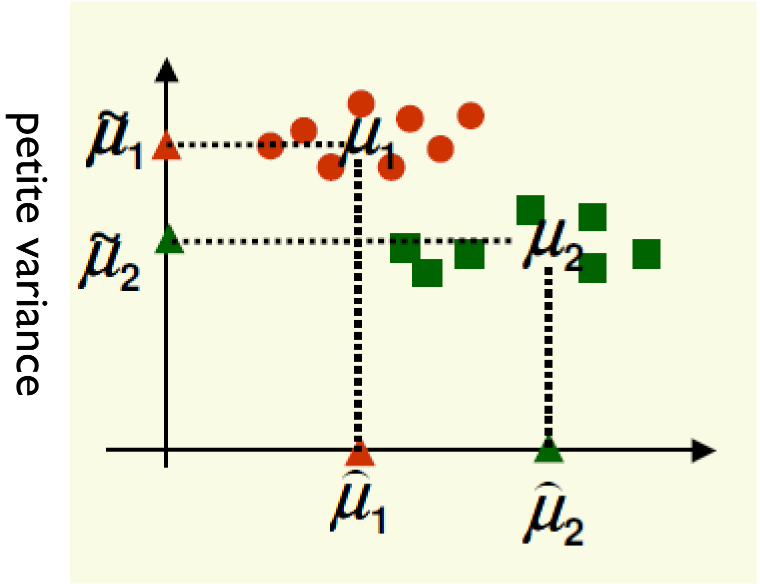


- Pour une meilleure séparation des classes, il est meilleur de projeter les classes sur l'axe vertical que sur l'axe horizontal

- Toutefois on a $\left|\tilde{m}_2 - \tilde{m}_1\right| \prec \left|\hat{m}_2 - \hat{m}_1\right|$

AFD..

- Le problème avec ce critère $\left|\tilde{m}_2 - \tilde{m}_1\right|$ est qu'il ne considère pas la variance des classes



Grande variance

- Il faut normaliser $\left|\tilde{m}_2 - \tilde{m}_1\right|$ par un facteur qui est proportionnel à la variance

AFD.. fonction objective

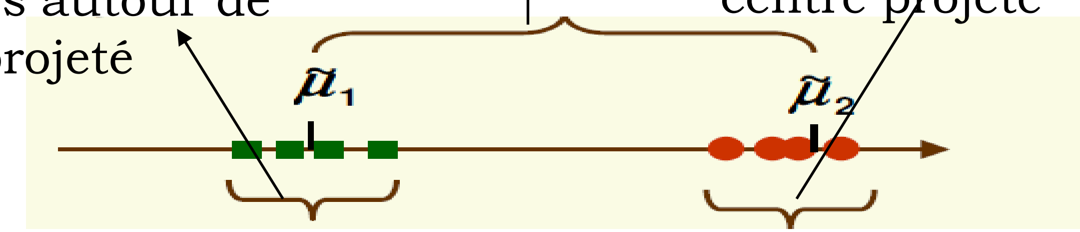
- Nous avons besoin de normaliser à la fois par l'inertie intra de la classe 1 et l'inertie intra de la classe 2
- Ainsi l'analyse factorielle discriminante de Fisher consiste à projeter le nuage de point sur l'axe de direction v qui maximise

Projeter les centres des classes loin les uns des autres

$$J(v) = \frac{(\tilde{m}_2 - \tilde{m}_1)^2}{\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}}$$

On veut que la dispersion intra la classe 1 soit la plus petit possible, les observations de la classe 1 sont centrées autour de son centre projeté

On veut que la dispersion intra la classe 2 soit la plus petit possible, les observations de la classe 2 sont centrées autour de son centre projeté



AFD.. réécriture de la fonction objective

- Nous cherchons le vecteur v qui maximise le critère J(v) , afin d'assurer que les classes sont bien séparées
- Tout ce que nous devons faire maintenant est d'exprimer J explicitement en fonction de v et de le maximiser
- Définissons maintenant les inerties intra classe des classes 1 et 2, l'inertie inter classe et l'inertie totale (à partir de données originales avant projection)

$$S_1 = \sum_{x_i \in c_1} (x_i - m_1)(x_i - m_1)^T$$ inertie intra classe 1

$$S_2 = \sum_{x_i \in c_2} (x_i - m_2)(x_i - m_2)^T$$ inertie intra classe 2

$$S_W = S_1 + S_2$$ inertie intra classe

$$S_B = (m_2 - m_1)(m_2 - m_1)^T$$ inertie inter classe

$$S_T = \sum_i (x_i - m)(x - m)^T$$ inertie totale

$$S_T = S_W + S_B$$ Décomposition de l'inertie totale

AFD.. réécriture de la fonction objective

- Inertie intra classe des données projetées

$$y_i = v^T x_i$$

$$\tilde{s}_1^2 = \sum_{y_i : x_i \in c_1} (y_i - \tilde{m}_1)^2 = \sum_{x_i \in c_1} (v^T x_i - v^T m_1)^2 = v^T \left(\sum_{x_i \in c_1} (x_i - m_1)(x_i - m_1)^T \right) v = v^T S_1 v$$

$$\tilde{s}_2^2 = \sum_{y_i : x_i \in c_2} (y_i - \tilde{m}_2)^2 = \sum_{x_i \in c_2} (v^T x_i - v^T m_2)^2 = v^T \left(\sum_{x_i \in c_2} (x_i - m_2)(x_i - m_2)^T \right) v = v^T S_2 v$$

- Inertie inter classe des données projetées

$$S_B = (m_2 - m_1)(m_2 - m_1)^T$$

$$(\tilde{m}_2 - \tilde{m}_1)(\tilde{m}_2 - \tilde{m}_1)^T = (v^T m_2 - v^T m_1)(v^T m_2 - v^T m_1)^T$$

$$= v^T (m_2 - m_1)(m_2 - m_1)^T v = v^T S_B v$$

Ratio de la différence de moyennes projetées sur l'inertie intra classe

- Objective à maximiser

$$J(v) = \frac{(\tilde{m}_2 - \tilde{m}_1)^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{v^T S_B v}{v^T S_w v}$$

Rayleigh quotient

Problème d'optimisation

- Nous recherchons l'axe correspondant à la séparation maximale entre les classes

$$\max_v J(v) = \max_v \frac{v^T S_B v}{v^T S_w v}$$

- Et donc

$$\partial_v (J(v)) = \partial_v \left(\frac{v^T S_B v}{v^T S_w v} \right) = 0$$

- Ce qui revient à résoudre

$$v^T S_W v (S_B v) - v^T S_B v (S_W) v = 0 \Longrightarrow S_B v = \frac{v^T S_B v}{v^T S_W v} S_W v$$

- Posons $\lambda = \frac{v^T S_B v}{v^T S_W v}$
- On obtient le **problème aux valeurs/vecteurs propres généralisé**

$$S_B v = \lambda S_W v$$

Problème de valeurs/vecteurs propres

- Nous obtenons donc le problème aux valeurs/vecteurs propres

$$S_W^{-1} S_B v_1 = \lambda_1 v_1$$

- On peut montrer que

$$S_T^{-1} S_B v_1 = \delta_1 v_1$$

- Où $\lambda_1 = \dfrac{\delta_1}{1-\delta_1}$

- v_1 est la direction du premier axe discriminant
- $F_1 = X v_1$ est la première variable discriminante: vecteur constitué des coordonnées des n individus sur l'axe 1
- δ_1 est le pouvoir discriminant de l'axe 1

Généralisation au cas multi-classes (K>2)

- L'AFD du tableau X s'obtient en cherchant les vecteurs propres vk et les valeurs propres associées de : $S_W^{-1} S_B$ ou $S_T^{-1} S_B$
- Le kieme axe discriminant est le vecteur propre associé à la valeur propre de rang k de cette matrice.

- Le nombre maximum d'axes (= nombre de valeurs propres non nulles) que l'on puisse obtenir en effectuant l'AFD sur K groupes est (K-1).
- Recherche des facteurs discriminants
  - On travaille en variables centrées => m est ramené à l'origine!
  - Le 1er facteur discriminant (F1) est une nouvelle variable, combinaison linéaire des variables descriptives (centrées), dont la variance inter-classe est maximum (ou, de façon équivalente la variance intra-classe est minimum).
  - Géométriquement: le 1er facteur détermine un axe dans le nuage de points (passant par l'origine) tel que les projections des points sur cet axe aient une variance inter-classe (variance des moyennes de classe) maximale.
  - Le 2eme facteur (F2) est non corrélé (perpendiculaire) au 1er et de variance inter-classe max. Etc pour le 3eme …

Propriétés et représentation graphique

- **Propriétés**:
- les facteurs sont entièrement déterminés par la matrice définie par: $S_W^{-1} S_B$ ou $S_T^{-1} S_B$ (vecteurs propres).
- le nombre maximum de facteurs discriminants = $K - 1$
- la part de variance inter-classe expliquée = variance inter / variance totale est décroissante entre les facteurs successifs.
- Toutes ces propriétés s'expliquent par le fait que: une analyse **discriminante = ACP** sur le nuage des K centroïdes**,** pondérés par l'effectif des classes *nk*, dans un espace $\Re$^p avec S_W^{-1} comme métrique.

- **Représentation graphique:**
- – Si 2 groupes => 1 seul facteur = axe de projection où la séparation inter-classe est la mieux exprimée => coordonnées sur cet axe = scores discriminants.
- – Si + de 2 groupes => plan discriminant ($F1$ et $F2$) = plan de projection ou la variance inter-classe **SB** (=> dispersion des centroïdes dans le plan) sera la mieux représentée.

Interprétation

- Comme en ACP: corrélations facteurs aux variables initiales, + cercle des corrélations avec les 2 premiers facteurs (K> 2)

- **Analyse discriminante décisionnelle => méthode de classification:**

- ***1) règle géométrique (règle de Fisher)***:
- Les facteurs discriminants donnent la meilleure représentation de la séparation des K centroïdes de classe (dans un espace orthonormé).

- => pour un individu $\mathbf{x}$ projeté dans l'espace des facteurs: attribuer la classe dont le centroïde est le plus proche (au sens de la ***distance euclidienne***):

Interprétation

- ***Traduction dans l'espace de départ*** (variables descriptives): allocation au centroïde **m**k le plus proche au sens de la métrique S_W^{-1} (***distance de Mahalanobis***)

$$d_M{}^2(x, m_k) = (x - m_k)^T S_W{}^{-1}(x - m_k)$$

- ***Problèmes:***
- – La métrique S_W^{-1} est évaluée sur l'ensemble des données => problème si les classes ne sont pas de même "forme" (dispersion).
- – une classe est représentée par son centroïde => problème si le centroïde n'est pas représentatif d'une classe (cas des classes non ellipsoïdales ou composées de sous nuages différents => séparation fortement non linéaire).

Justification: lien avec la classification bayésienne

- On peut montrer que la règle de Fisher correspond à un classificateur bayésien (minimisation de la proba. de l'erreur) dans les conditions suivantes:
- chaque classe suit une distribution gaussienne (multivariée) de même matrice
- variance-covariance S_W (les nuages de points ont la même 'forme'),
- les classes sont équidistribuées: mêmes proba. *a priori*
- (très facilement généralisable si ce n'est pas le cas)
- En effet: Lorsque les distributions de classes sont gaussiennes de même matrice variance-covariance S_W, un classificateur bayésien définit les fonctions discriminantes $\delta_k(x)$ suivantes:
- (**x** alloué à G*k* si $\delta_k(x) \succ \delta_l(x)$
- Et $\delta_k(x) = -\frac{1}{2}\log(|S_W|) - \frac{1}{2}(x - m_k)^T S_W^{-1}(x - m_k) + \log(\pi_k)$
- <=> **x** alloué à G*k* si $d_M^{\ 2}(x, m_k) - \log \pi_k$ est ***minimum*** !!!
- ***règle de Fisher généralisée*** (dépendante des proba. *a priori*): favorise les classes fortement représentées !

Analyse discriminante linéaire (LDA)

Supposons que les observations de chaque classe sont générées par une Gaussienne:

- Où **x** est le vecteur des caractéristiques projeté sur les axes discriminants retenus
- = modèle paramétrique
- Dans ce cas-ci, nous supposons qu'il y a une matrice variance-covariance **Sw** égale pour toutes les classes, (posons dans la suite $Sw = \mathbf{\Sigma}$).

Probabilité a posteriori $\Pr(G = k \mid X = x) = \dfrac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}$

Application de la règle de Bayes

π_k est la probabilité a priori de la classe k

$f_k(x)$ est la densité conditionnelle de x sachant la classe k

$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(x - m_k)^T \mathbf{\Sigma}^{-1}(x - m_k)\right)$$

LDA…

$$\log \frac{\Pr(G=k \mid X=x)}{\Pr(G=l \mid X=x)} = \log \frac{\pi_k}{\pi_l} + \log \frac{f_k}{f_l}$$

$$= \underbrace{(\log \pi_k + x^T \boldsymbol{\Sigma}^{-1} m_k - \frac{1}{2} m_k^T \boldsymbol{\Sigma}^{-1} m_k)}_{\delta_k(x)} - \underbrace{(\log \pi_l + x^T \boldsymbol{\Sigma}^{-1} m_l - \frac{1}{2} m_l^T \boldsymbol{\Sigma}^{-1} m_l)}_{\delta_l(x)}$$

Règle de classification : $\hat{G}(x) = \arg\max_k \delta_k(x)$

Est equivalente à: $\hat{G}(x) = \arg\max_k \Pr(G=k \mid X=x)$

Qui n'est rien d'autre que le bon vieux classificateur de Bayes!

LDA

Quand allons-nous utiliser les données d'apprentissage?

n le nombre total d'observations
n_k nombre d'observations dans la class k $(g_i, x_i),\ i = 1:n$
Nombre de classes: K

Les données d'apprentissage
sont utilisées pour estimer

les probabilités a priori: $\hat{\pi}_k = n_k / n$

Moyennes: $\hat{m}_k = \sum_{g_i = k} x_i / n_k$

Matrice de covariance: $\hat{\Sigma} = \sum_{k=1}^{K} \sum_{g_i} (x_i - \hat{m}_k)(x_i - \hat{m}_k)^T / n$

Analyse discriminante quadratique

- ***Généralisation*** au cas où les matrices variance-covariance S_{Wk} des classes ne sont pas égales: les fcts discriminantes du classif. bayesien deviennent:

$$\delta_k(x) = -\frac{1}{2}\log|S_{Wk}| - \frac{1}{2}(x-m_k)^T S_{Wk}^{-1}(x-m_k) + \log \pi_k$$

- $|S_{Wk}|$ déterminant de la matrice
- Dans ce cas, les ***surfaces de séparation*** entre 2 classes (définies par $\delta_k(x) = \delta_l(x)$

  ne sont ***plus linéaires*** => ***analyse discriminante quadratique (QDA)***
- ***En pratique:***
- La matrice S_W, ou les matrices S_{Wk} doi(ven)t être estimée(s) à partir des exemples disponibles pour chaque classe, ainsi que les π_k
- Lorsqu'on fait l'hypothèse d'égalité des matrices S_{Wk}, la matrice S_W est obtenue par estimation 'poolée':

$$S_W = (n_1 S_{W1} + n_2 S_{W2} + \ldots + n_K S_{WK}) / n$$

(n = effectif total)

- L'usage et l'estimation de matrice particulière S_{Wk} demande que les effectifs de classe soient suffisamment importants ! Pour des faibles effectifs l'existence de S_{Wk}^{-1} n'est pas tjrs assurée, de même $|S_{Wk}|$ peut être nul !

Analyse discriminante quadratique

- On relaxe l'hypothèse de variance égale dans les K classes, on suppose que chaque classe ait sa propre matrice de variance-covariance.
- Dans ce cas, la surface de séparation entre 2 classes est quadratique, elle est définie par :

$$\log \frac{\Pr(G=k \mid X=x)}{\Pr(G=l \mid X=x)} = \log \frac{\pi_k}{\pi_l} + \log \frac{f_k}{f_l} =$$

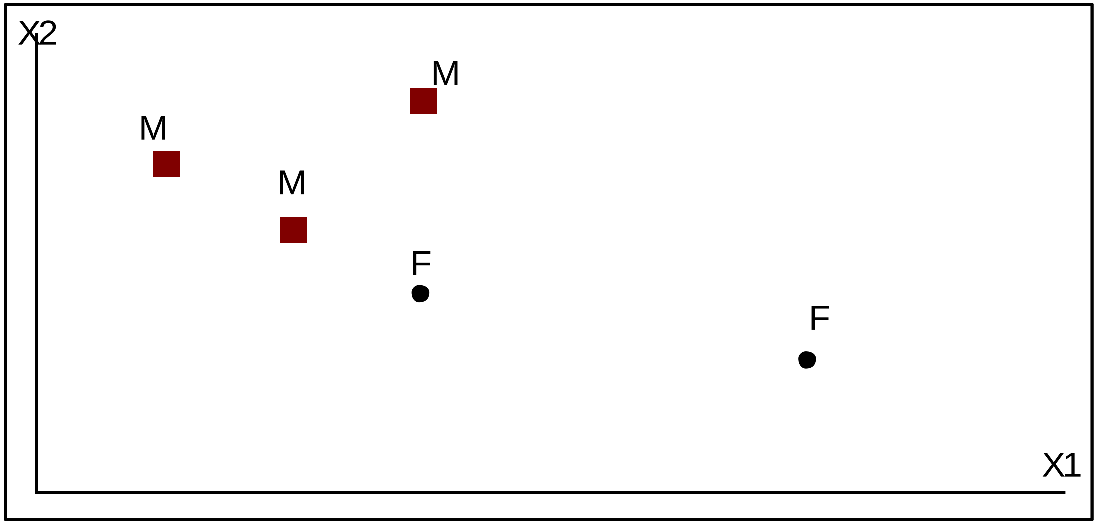
$$\underbrace{(\log \pi_k - \frac{1}{2}(x-m_k)^T \mathbf{\Sigma}_k^{-1}(x-m_k) - \frac{1}{2}\log|\mathbf{\Sigma}_k|)}_{\delta_k(x)} - \underbrace{(\log \pi_l - \frac{1}{2}(x-m_l)^T \mathbf{\Sigma}_l^{-1}(x-m_l) - \frac{1}{2}\log|\mathbf{\Sigma}_l|)}_{\delta_l(x)}$$

Exemple AFD à la main

On observe deux variables quantitatives X_1 et X_2 sur un ensemble de n=5 individus de même poids, supposés répartis en deux groupes (M : masculin et F : féminin) :

Groupe	X1	X2
M	1	5
M	3	6
M	2	4
F	3	3
F	6	2

AFD à la main



AFD à la main

Grandeurs d'intérêt :

n1=3, n2=2, n=5

$$X = \begin{pmatrix} -2 & 1 \\ 0 & 2 \\ -1 & 0 \\ 0 & -1 \\ 3 & -2 \end{pmatrix} \qquad G = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\tilde{X}_1 = \begin{pmatrix} -2 & 1 \\ 0 & 2 \\ -1 & 0 \end{pmatrix} \quad G_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad \tilde{X}_2 = \begin{pmatrix} 0 & -1 \\ 3 & -2 \end{pmatrix} \quad G_2 = \begin{pmatrix} 1.5 \\ -1.5 \end{pmatrix}$$

AFD à la main

Recherche de l'axe discriminant :

- ✓ Matrice variance totale: V = X'X/n
- ✓ Matrice de variance inter-classes : $$V = \frac{1}{5}\begin{pmatrix} 14 & -8 \\ -8 & 10 \end{pmatrix}$$
- ✓ Matrice de variance intra-classes : $$B = \frac{1}{5}\begin{pmatrix} 7.5 & -7.5 \\ -7.5 & 7.5 \end{pmatrix}$$

$$W = \frac{1}{n}\sum n_k V_k$$

$$V_1 = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad V_2 = \frac{1}{2}\begin{pmatrix} 4.5 & -1.5 \\ -1.5 & 0.5 \end{pmatrix} \quad W = \frac{1}{5}\begin{pmatrix} 6.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix}$$

AFD à la main

Matrice à diagonaliser :

$$V^{-1}B = \frac{7.5}{76}\begin{pmatrix} 2 & -2 \\ -6 & 6 \end{pmatrix}$$

AFD à la main

- La valeur propre <u>non nulle</u> de $V^{-1}B$ est = 0.79, qui est le pouvoir discriminant de l'axe (rappelons que plus cette valeur est proche de 1 meilleure est la discrimination)

- Le vecteur propre unitaire associé à cette valeur propre est donné par :

$$u = \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix}$$

- Les coordonnées sur cet axe F=Xu sont :

C=

M	5
M	6
M	1
F	-3
F	-9

AFD à la main

```
>c=read.table("cours.txt",header=T)
 Groupe X1 X2
1     M  1  5
2     M  3  6
3     M  2  4
4     F  3  3
5     F  6  2
>mc=matrix(apply(c[,2:3],2,mean),5,2,byrow=T
 )
>X=as.matrix(c[,2:3]-mc)
> X1=X[X$Groupe=="M",]
> X2=X[X$Groupe=="F",]
>G1=apply(X1[,2:3],2,mean)
> G2=apply(X2[,2:3],2,mean)
>V=(t(X)%*%X)/5
>P=diag(c(3/5,2/5)) ## poids des classes
>C=rbind(G1,G2)
>B=t(C)%*%P%*%C
 >V1=(t(X1)%*%X1)/3-G1%*%t(G1)
> V2=(t(X2)%*%X2)/2-G2%*%t(G2)
>W=(3*V1+2*V2)/5
>I=solve(V)%*%B
> v=eigen(I)$vector
>lambda= eigen(I)$values
>F=X%*%v
```

AFD sous R

1) On effectue l'ACP sur le nuage de point des centres de gravités du tableau centré.

2) On utilise la fonction lda() de la library MASS

3) On utilise la fonction discrimin() de la library ade4

AFD sous R

Library(MASS)

lda(formula, data, ...,)

Formula : A formula of the form 'groups ~ x1 + x2 + ...' That is, the response is the grouping factor and the right hand side specifies the (non-factor) discriminators.

data: Data frame from which variables specified in 'formula' are preferentially to be taken.

prior: the prior probabilities of class membership. If unspecified, the class proportions for the training set are used. If present, the probabilities should be specified in the order of the factor levels.

AFD sous R

age	revenu	patrimoine	emprunt	groupe
45	250	1300	600	3
47	160	1150	450	3
38	165	850	370	1
36	175	770	250	1
29	99	450	400	1
39	170	1400	120	3
27	120	1400	160	2
51	160	1300	320	3
32	155	1500	350	2
35	170	1400	180	2

AFD sous R

```
>a=lda(groupe~age+revenu+patrimoine+emprunt,d) #d=Données

Call:
lda(groupe ~ age + revenu + patrimoine + emprunt, data = d)

Prior probabilities of groups: probabilités à priori des classes
  1   2   3
0.3 0.3 0.4

Group means: moyennes par groupe des variables du tableau d
       age   revenu patrimoine emprunt
1 34.33333 146.3333    690.000   340.0
2 31.33333 148.3333   1433.333   230.0
3 45.50000 185.0000   1287.500   372.5
```

AFD sous R

Coefficients of linear discriminants: Coordonnées des vecteurs v1 et v2 renormalisées

	LD1	LD2
age	0.048261265	-2.169801e-01
revenu	0.025594479	5.976213e-04
Patrimoine	-0.011352863	1.724238e-04
emprunt	-0.005286007	-5.039816e-05

Proportion of trace: mu/somme(mu): % d'inertie conservé par chaque axe

LD1	LD2
0.8451	0.1549

AFD sous R

```
>names(a)
[1] "prior"   "counts"  "means"   "scaling" "lev"     "svd"     "N"
[8] "call"    "terms"   "xlevels"
```

a$prior: poids des groupes

a$counts : nombre d'individus dans les groupes

a$means : moyenne des variables dans les groupes

a$scaling: coordonnées des axes discriminants

a$lev : nombre de niveaux du facteur groupe

a$svd: valeurs singulières

AFD sous R

>a$means

	age	revenu	patrimoine	emprunt
1	34.33333	146.3333	690.000	340.0
2	31.33333	148.3333	1433.333	230.0
3	45.50000	185.0000	1287.500	372.5

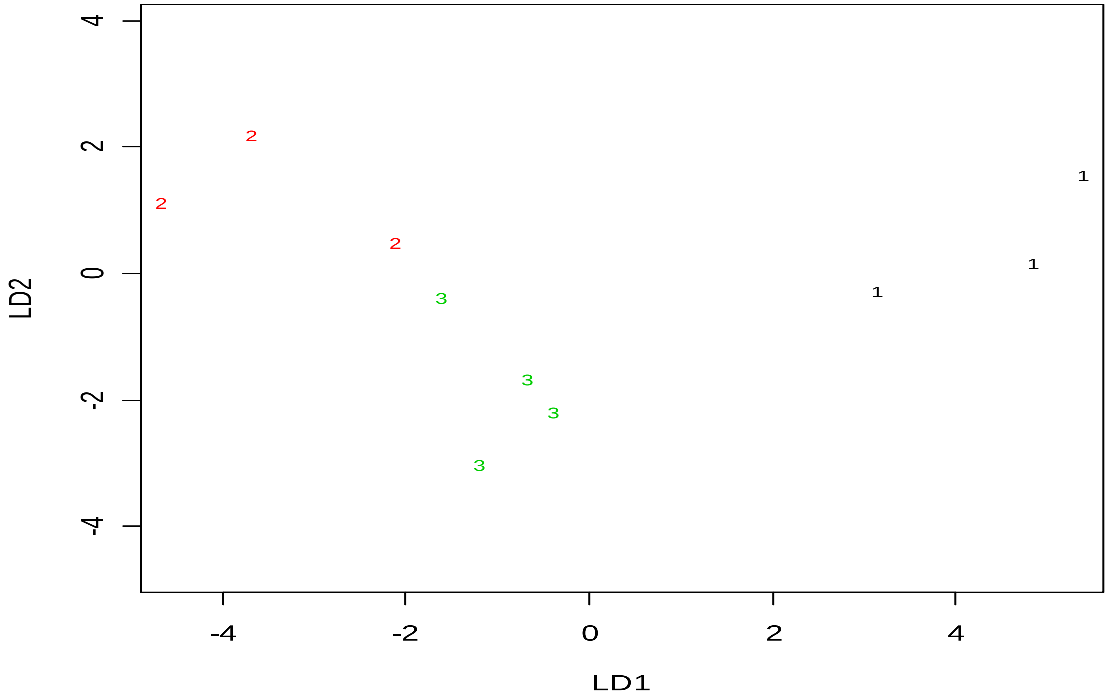
Le groupe 1 est un groupe de gens assez jeunes à revenus plus faibles que la moyenne dont le patrimoine est nettement plus faible que dans les autres classes et le taux d'emprunt plus élevé que la moyenne

Le groupe 2 est caractérisé par des gens jeunes de revenus moyens, mais dont le patrimoine est très important et le taux d'emprunt très faible

Le groupe 3 est caractérisé par des gens plus âgés de revenus confortables et de patrimoine assez important, ayant un taux d'emprunt plus élevé que dans les autres classes

AFD sous R

>plot(a, col = as.numeric(d[,5])) Graphe de X%*%a$scaling



AFD sous R

- ✓ Sur le graphique, on voit que l'axe 1 sépare bien les 3 groupes, en particulier le groupe 1 des deux autres groupes. Le pouvoir discriminant de l'axe 2 est moindre:

- ✓ L'interprétation des facteurs discriminants peut se faire comme en ACP en calculant les coordonnées des variables sur les axes (corrélations r(Xj,Dk))

AFD sous R

```
>D=d%*%a$scaling
>cor(d,D)
                  LD1         LD2
age         -0.03673802 -0.9991592
revenu      -0.13640214 -0.5627526
patrimoine  -0.96159782 -0.1540505
emprunt     -0.23596443 -0.4321157
```

L'axe 1 est un effet taille et isole les individus ayant des valeurs importantes des variables, en particulier à gros patrimoine. Ils s'opposent aux individus du groupe 1.

L'axe 2 est aussi un effet taille et isole les individus plus âgés que les autres: on y trouve les individus du groupe 3, qui s'opposent à ceux des deux autres groupes.