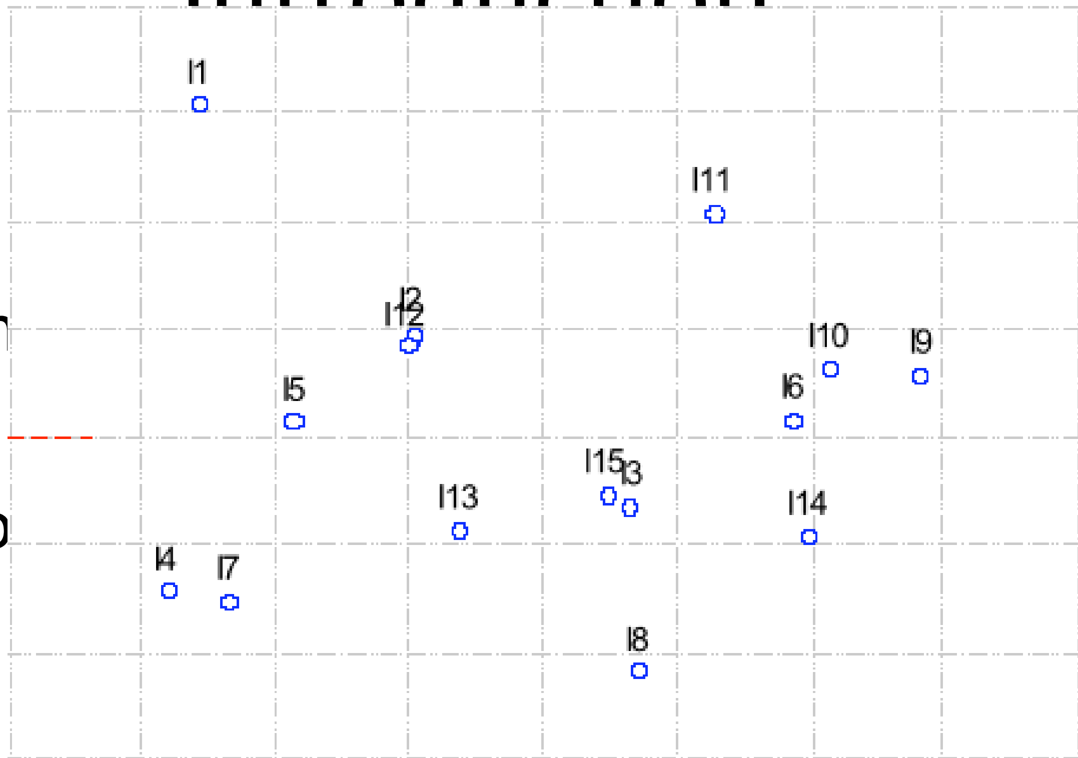
Chapitre I :
Analyse Descriptive/prédictive :
Apprentissage Non Supervisé/Supervisé :

ANALAYSE DISCRIMINANTE : AD

Introduction:



- ACP ana … nnées
  - Elle ch … es points individ
  - Par co … ent une bonne … ses des individ
- L'analyse discriminante cherche des projections qui sont efficaces pour discriminer(distinguer/séparer) les classes.

AD: Objectif:

- On dispose d'un tableau Individus/Variables **MAIS** les individus sont répartis en un ensemble de classes (groupes) définies selon des variables mesurées sur ses individus.
- L'objectif est de rechercher les combinaisons linéaires de variables :
- Qui permettent de séparer au mieux les différents groupes
  - Maximiser la variance inter classes
  - Minimiser la variance intra classes

AD: Définition

- L'AD permet d'étudier la différence entre 2 ou plusieurs groupes (pour donner le moyen d'une éventuelle classification) en tenant compte de multiples vars simultanément.
- C'est aussi : étudier les relations entre une variable qualitative et un ensemble de variable explicatives quantitatives

Approches de l’AD:

On distingue deux approches de l’AD

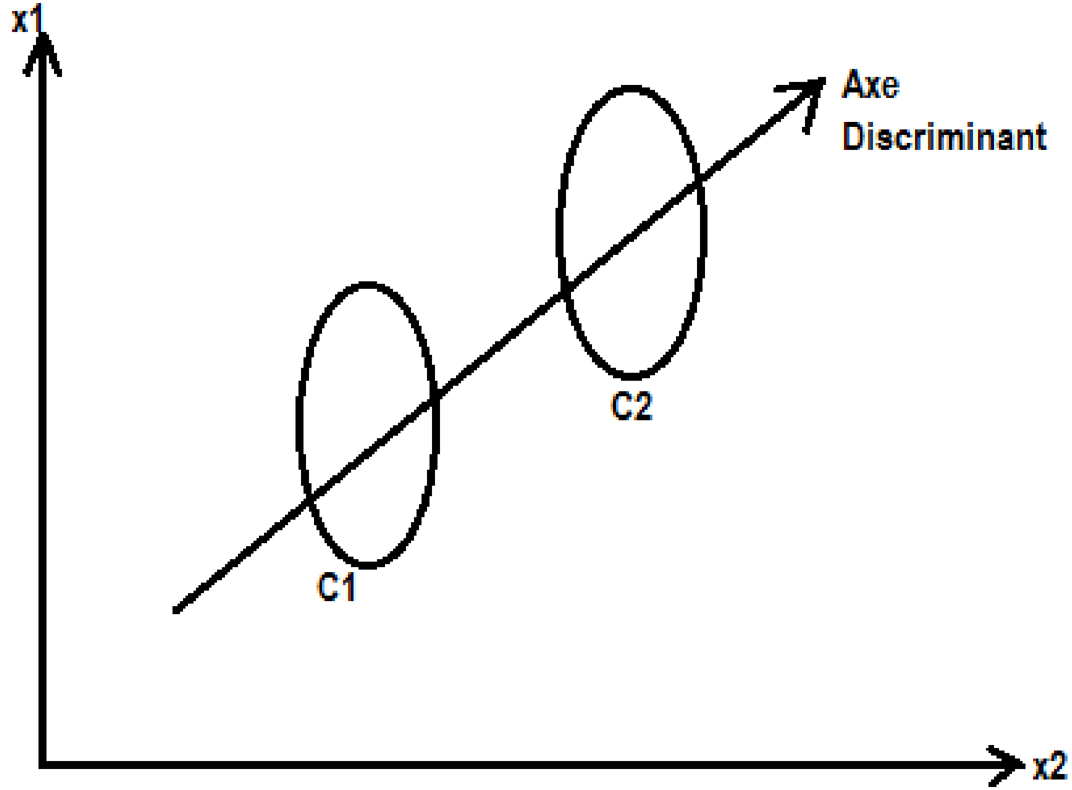
I. Approche descriptive appelée ADF :Analyse Discriminante de Fisher(Réduction de Dimension)

II. Approche Prédictive :classification:

1. Règle géométrique de classification;
2. Approche Probabiliste simplifiée de Classification.

I- AD de Fisher Descriptive:

- L'objectif est de créer de nouvelles variables qui sont particulièrement efficaces pour séparer les classe
- Ces vars appel Discriminants » combinaison linéa
- Il faut que la pro discriminant soit



I- ADF: Formulation du Problème:

- Y variable à expliquer qualitative à k groupes;
- X1,X2,…Xp P variables explicatives quantitatives
- N nombre d’observations Totale;
- Ni: nombre d’observations du groupe i i=1..k
- G est le centre de gravité des n observations
- Gi est le centre de gravité des observations du groupe i
- V = W +B
- W= Matrice de variance intragroupe(within)=

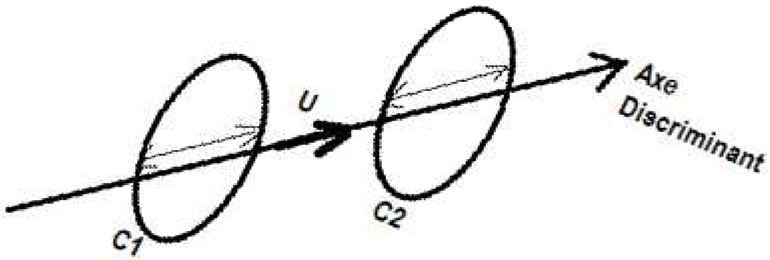
$$W_{(pxp)} = \frac{1}{n}\sum_{i}^{k} n_i V_i$$

- B= Matrice de variance intergroupe(Between)=

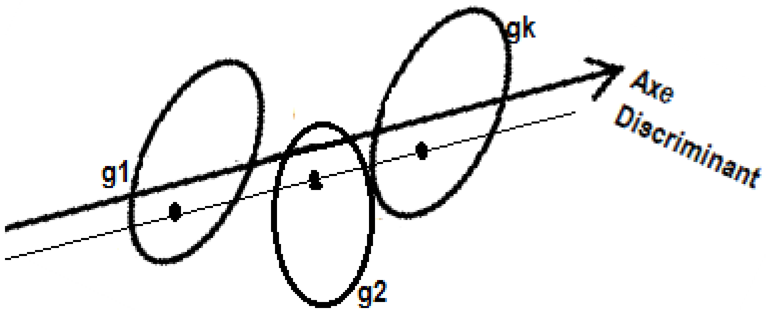
$$B_{(pxp)} = \frac{1}{n}\sum_{i}^{k} n_i (g_i - g)(g_i - g)^t$$

I- ADF: Résolution du Problème:

- Pour déterminer les axes discriminants il faut:
- Minimiser la Distance intra-classe : $\min\left(U^t W U\right)$
- U: Vecteur unité de l'axe discriminant:



- Maximiser la Distance inter-classe : $\max\left[\frac{U^t B U}{U^t V U}\right]$



λU

I- ADF: Résolution du Problème:

➔ il faut annuler la dérivée de la fonction pour trouver le Max.

➔ On va avant remplacer la fonction à maximiser par une fonction avec les multiplicateurs de Lagrange λ

➔ Il faut donc trouver les λ et vecteurs propres :

$$\frac{\partial(U^t BU - \lambda(U^t VU^{-1}))}{\partial U} = 0$$

➔ 2BU -2 λVU = 0

➔ $V^{-1}BU = \lambda U$

I- ADF: Résolution du Problème:

- On recherche l'axe correspondant à la séparation maximale entre les classes➔
- $V^{-1}BU = \lambda$ U $0 < \lambda < 1$

λ = 0 aucune séparation linéaire entre les classes

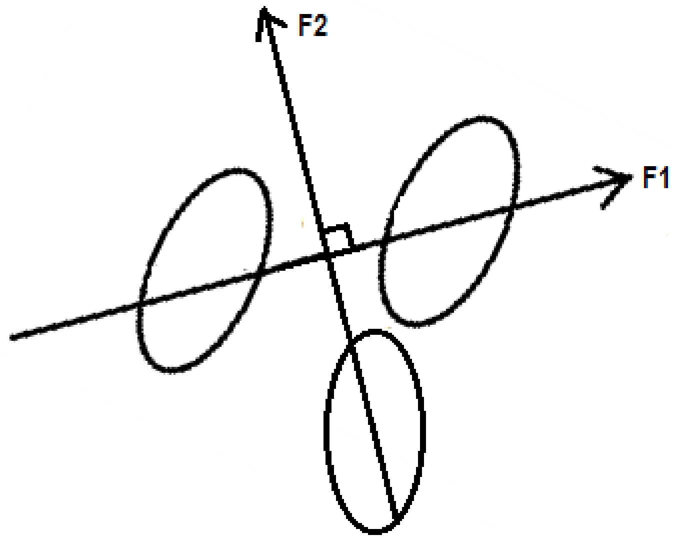
λ = 1 : Séparation parfaite entre les classes

λ : multiplificateur de lagrange;

U est donc vecteur propre de la matrice $V^{-1}B$ correspondant à la plus grande valeur propre.

ADF: Remarque:

- En général n> P >k ➔ il y a au plus (k-1) axes factorielles.
- Pour trouver les autres facteurs: Nous avons la projection des points sur le premier axe a une variance inter-classe maximale, le 2eme axe est perpenduculaire au 1er et de variance interclasse max….



II- AD PREDICTIVE :Classification:
1-Règle Géométrique

- Nous avons de nouvelles observations qu'on veut classer dans un des groupes:

A. Une idée est de calculer la distance entre la nouvelle observation et le centre de chacun des groupes, l'observation est classée dans le groupe pour lequel cette distance est minimale:

$$d^2(x, g_i) = (x - g_i)^t V^{-1} (x - g_i)$$

Ou g_i est le vecteur (p,1) des moyennes des p variables du groupe i

II- AD Classification:
1-Règle Géométrique

B. Une deuxième idée est de classer l’observation dans le groupe pour lequel fi est maximal avec :

$$f_i = \left(X^t V^{-1} g_i - \frac{1}{2} g_i^t V^{-1} g_i\right) = \left(X^t - \frac{1}{2} g_i^t\right) V^{-1} g_i$$

Les fi sont appelées fonctions de classification ou fonctions linéaires discriminantes, On en possède autant qu’il ya des groupes et on affecte la nouvelle obs au groupe pour lequel sa fct de classification est maximale.

II- AD Classification:
2-Approche probabiliste simplifiée:

- L'idée est de classer une observation dans le groupe pour lequel la probabilité conditionnelle d'appartenir à ce groupe est maximale sous l'hypothèse que les observations proviennent d'une loi multinormale (chaque groupe suit une loi normale)

II- AD Classification:

2-Approche probabiliste simplifiée:

- La probabilité qu'une observation appartienne à un groupe est :

$$P(groupei/x) = \frac{f_i(x)}{\sum_j^k f_j(x)}$$

- Les fi sont les fonctions de classification.
- Cette probabilité est maximale quand fi est maximale.

AD : Remarques:

1. Les approches géométriques et probabilistes sont équivalentes lorsque on a k populations multinormale avec <u>même matrice de covariance.</u>
2. Si les matrices varient d'un groupe à l'autre, on se trouve à effectuer une discrimination non Linéaire:

➔Les zones attachées à chaque groupe sont délimitées par des courbes ce qui est une discrimination quadratique (elle est rarement utilisée).

<u>Exemple</u> :

- Des ménages selon le nombre d’enfants(2,3,4,5) statut socio-professionnel(Cadre, Employé et Manuel) et la consommation en viandes volaille, legumes fruits et boissons….
- Le facteur de taille mis en évidence est lié à la taille du ménage et à son statut sociau-professionnel, ce qui implique la dépense globale qui augmente avec le nbre d’enfants et le statut social

Conclusion

- L'analyse discriminante permet de réduire la dimensionnalité des problèmes

➢ Pour l'apprentissage paramétrique, elle permet de réduire le nombre de paramètres à estimer(Réduction de Dimension)

➢ Et Beaucoup dans la reconnaissance des formes(Classification)