

# Analyse Discriminate Linéaire

HUYLENBROECK Florent, DELFOSSE Charly, JOSSE Thomas

6 juin 2019

# Table des matières

- 1 Introduction
- 2 Fonctionnement
  - Principe géométrique
  - Principe statistique
- 3 Exemple

# Qu'est-ce que l'ADL ?

Cette technique fait partie des techniques d'analyse discriminante prédictive. Le but est de pouvoir expliquer et prédire l'appartenance d'un individu à un groupe prédéfini à partir de caractéristiques qui ont été mesurées au préalable à l'aide de variables prédictives. On peut la comparer à la régression logistique.

# La règle bayesienne

Cette règle consiste à produire une estimation de la probabilité après notre affectation. Cela veut dire que nous devons réaliser une estimation pour une probabilité conditionnelle :

$$P(Y = y_k|X) = \frac{P(Y = y_k) \times P(X|Y = y_k)}{\sum_{i=1}^K P(Y = y_i) \times P(X|Y = y_i)}$$

Où  $X$  contient les  $j$  variables prédictives  $[X = X_1, \dots, X_j]$ ,  $n$  est le nombre d'observations réparties dans  $K$  groupe d'effectifs  $n_k$ ,  $Y$  est notre variable à prédire qui prend des valeurs dans l'ensemble  $\{y_1, \dots, y_K\}$ . Nous avons aussi  $P(Y = y_k)$  qui est la probabilité d'appartenance à la classe  $y_k$  et  $P(X|Y = y_k)$  qui est la fonction de densité des  $x$  si par rapport à l'appartenance à la classe  $y_k$ . Et donc, ce qu'on recherche avec l'analyse discriminante (en général car il existe d'autres type d'analyse discriminante) est de trouver l'estimation de  $P(X|Y = y_k)$ .

# Ce qu'on veut

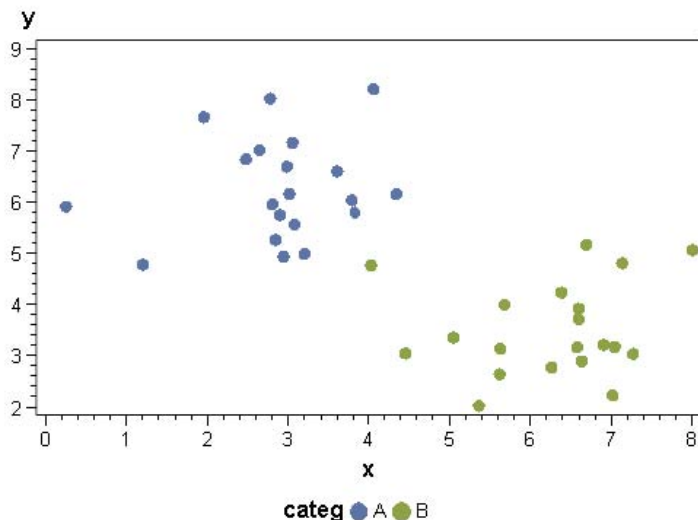
# Développement

- Chaque groupe d'observations remplacé par barycentre
- On trace une droite perpendiculaire au segment qui relie les deux barycentres
- L'intersection entre la droite tracée et le segment est appelé le point "zéro"
- Cette droite est placée en fonction des échantillons (s'ils sont équilibrés la droite est équidistante des barycentres, sinon celle-ci est plus proche du barycentre dont l'échantillon est le plus représenté).

# Ce qu'on veut

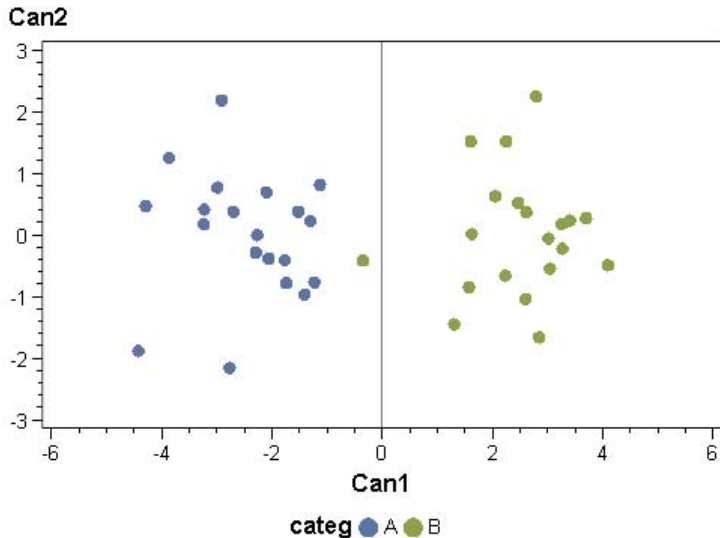
On veut replacer nos données dans un nouveau repère tel que les points sont le plus possible au sein d'un même groupe, mais ceux-ci doivent être aussi distants que possible entre les 2 groupes.  
⇒ Exemple aux slides suivantes.

# Données de base





# Nouveau repère



Mettre package utilisé avec la méthode discrimin + explications

Montrer exemple