

# **Autres méthodes supervisées extrapolées du schéma bayésien**

Quelques approches pour rendre calculable  $P(Y/X)$

Ricco RAKOTOMALALA



# Théorème de Bayes

Probabilité conditionnelle

## Estimer la probabilité conditionnelle

$$\begin{aligned} P(Y = y_k / \mathfrak{X}) &= \frac{P(Y = y_k) \times P(\mathfrak{X} / Y = y_k)}{P(\mathfrak{X})} \\ &= \frac{P(Y = y_k) \times P(\mathfrak{X} / Y = y_k)}{\sum_{k=1}^K P(Y = y_k) \times P(\mathfrak{X} / Y = y_k)} \end{aligned}$$

## Déterminer la conclusion = déterminer le max.

$$\begin{aligned} y_{k^*} &= \arg \max_k P(Y = y_k / \mathfrak{X}) \\ \Leftrightarrow \\ y_{k^*} &= \arg \max_k P(Y = y_k) \times P(\mathfrak{X} / Y = y_k) \end{aligned}$$

Probabilité a priori  
Estimé facilement par  $n_k/n$

Comment estimer  $P(\mathfrak{X}/Y=y_k)$  ?

*Impossibilité à estimer avec des fréquences  
Le tableau croisé serait trop grand et rempli de zéros*



# Modèle bayésien naïf (I)

Modèle d'indépendance conditionnelle

Hypothèse d'indépendance conditionnelle

$$P(\mathbf{X} / Y = y_k) = \prod_{j=1}^J P(X_j / Y = y_k)$$

Les descripteurs sont deux à deux indépendants conditionnellement aux valeurs prises par Y



Pour un descripteur X discret quelconque, la probabilité conditionnelle pour qu'elle prenne la valeur  $x_l$  s'écrit

$$P(X = x_l / Y = y_k) = \frac{P(X = x_l \wedge Y = y_k)}{P(Y = y_k)}$$

Et son estimation (profil ligne)

$Y \setminus X$	$x_l$	$\Sigma$
$y_k$	$n_{kl}$	$n_k$
$\Sigma$		$n$

$$\hat{P}(X = x_l / Y = y_k) = \frac{\#\{\omega \in \Omega, X(\omega) = x_l \wedge Y(\omega) = y_k\}}{\#\{\omega \in \Omega, Y(\omega) = y_k\}}$$

$$= \frac{n_{kl}}{n_k}$$



# Modèle bayésien naïf (II)

## Exemple

Maladie	Marié	Etud.Sup
Présent	Non	Oui
Présent	Non	Oui
Absent	Non	Non
Absent	Oui	Oui
Présent	Non	Oui
Absent	Non	Non
Absent	Oui	Non
Présent	Non	Oui
Absent	Oui	Non
Présent	Oui	Non

### Estimation directe

$$\hat{P}(\text{Maladie} = \text{Absent} / \text{Marié} = \text{oui}, \text{Etu} = \text{oui}) = \frac{1}{1} = 1$$

$$\hat{P}(\text{Maladie} = \text{Présent} / \text{Marié} = \text{oui}, \text{Etu} = \text{oui}) = \frac{0}{1} = 0$$

→ Si Etu = oui et Marié = oui Alors Maladie = Absent !

(+) Calcul sans hypothèses restrictives, (-) effectifs indigents

NB Maladie			
Maladie	Somme		
<b>Absent</b>	<b>50.00%</b>		
Présent	50.00%		
Total	100.00%		

NB Maladie	Marié		
Maladie	Non	Oui	Total
<b>Absent</b>	<b>40.00%</b>	<b>60.00%</b>	100.00%
Présent	80.00%	20.00%	100.00%
Total	60.00%	40.00%	100.00%

NB Maladie	Etu.S		
Maladie	Non	Oui	Total
<b>Absent</b>	<b>80.00%</b>	<b>20.00%</b>	100.00%
Présent	20.00%	80.00%	100.00%
Total	50.00%	50.00%	100.00%

### Indépendance conditionnelle

$$\begin{aligned} &\hat{P}(\text{Maladie} = \text{Absent} / \text{Marié} = \text{oui}, \text{Etu} = \text{oui}) \\ &= \hat{P}(\text{Maladie} = \text{Absent}) \times \hat{P}(\text{Marié} = \text{oui} / M = \text{Abs.}) \times \hat{P}(\text{Etu} = \text{oui} / M = \text{Abs.}) \\ &= 0.5 \times 0.6 \times 0.2 \\ &= 0.06 \end{aligned}$$

$$\begin{aligned} &\hat{P}(\text{Maladie} = \text{présent} / \text{Marié} = \text{oui}, \text{Etu} = \text{oui}) \\ &= \hat{P}(\text{Maladie} = \text{présent}) \times \hat{P}(\text{Marié} = \text{oui} / M = \text{Abs.}) \times \hat{P}(\text{Etu} = \text{oui} / M = \text{Abs.}) \\ &= 0.5 \times 0.2 \times 0.8 \\ &= 0.08 \end{aligned}$$

→ Si Etu = oui et Marié = oui Alors Maladie = Présent !

(-) Hypothèse discutable, (+) estimations des probas (effectifs) plus fiables



# Modèle bayésien naïf (III)

## Avantages et inconvénients

- +**
- » Simplicité, rapidité de calcul (pas de risque de « plantage », cf. la régression logistique ou l'ADL)
- $K \times \sum_{j=1}^J L_j$  Probabilités à estimer contre  $K \times \prod_{j=1}^J L_j$  pour le modèle bayésien complet
- » Incrémentalité (table des probas conditionnelles à maintenir)
  - » Robustesse (performant même si hypothèse non-respectée)
  - » C'est un modèle linéaire (prouvé sur descripteurs binaires)
- 
- » Pas de sélection (mise en évidence) des variables pertinentes
  - » Nombre de règles égal au nombre de combinaisons de descripteurs  
(dans la pratique, les règles ne sont pas formées, nous conservons les probas conditionnelles que nous appliquons pour chaque individu à classer → pas d'interprétation des résultats)
- !**
- Traitement des variables continues
    - Discrétisation (supervisée)
    - Hypothèse de distribution, pour chaque descripteur  $f(X/Y) =$  loi normale

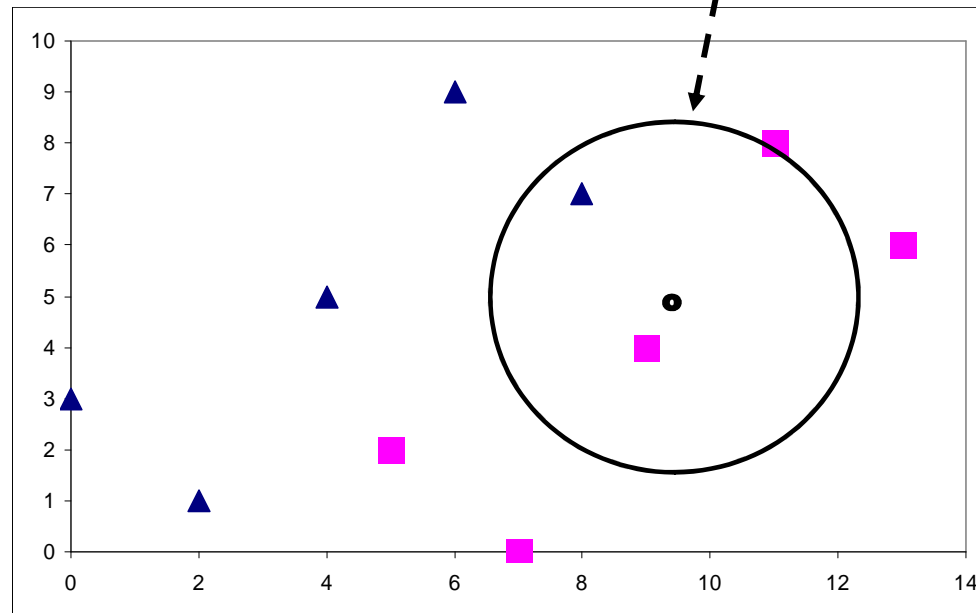


# Analyse discriminante non-paramétrique

Estimations locales des probabilités

Principe : Ne pas faire d'hypothèses sur les distributions !

Définir un voisinage autour du point « o » à classer et estimer localement les probabilités.

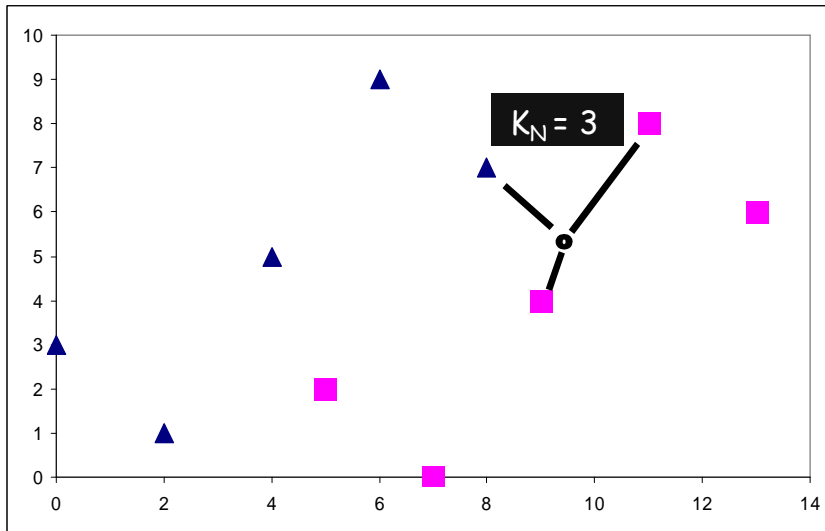


- » Principal problème à résoudre : comment définir le voisinage ?
- » La distance utilisée joue un rôle important !



# Méthode des plus proches voisins (I)

Paramètre :  $K_N$ , nombre d'observations autour du point à classer (voisinage)



$Y \setminus V$	$\in K_N$	$\notin K_N$	$\Sigma$
$y_k$	$K_N(y_k)$		$n_k$
$\Sigma$	$K_N$		$n$

« Rappel » des  $y_k$   
dans le voisinage

$$\frac{\hat{P}[\mathfrak{X}(\omega) / Y = y_k]}{\hat{P}[\mathfrak{X}(\omega)]} = \frac{\frac{K_N(y_k)}{n_k}}{\frac{K_N}{n}}$$

Taille (relative) du voisinage

Simplification de l'écriture (si échantillon extrait aléatoirement)

$$\begin{aligned} \hat{P}[Y = y_k / \mathfrak{X}(\omega)] &= \hat{P}[Y = y_k] \times \frac{\hat{P}[\mathfrak{X}(\omega) / Y = y_k]}{\hat{P}[\mathfrak{X}(\omega)]} \\ &= \frac{n_k}{n} \times \frac{\frac{K_N(y_k)}{n_k}}{\frac{K_N}{n}} \\ &= \frac{K_N(y_k)}{K_N} \end{aligned}$$

Proportion de  $Y_k$  dans le  
voisinage du point à classer



# Méthode des plus proches voisins (II)

## Avantages et inconvénients



- » Simplicité, pas d'apprentissage d'un modèle (lazy learning)
- » Incrémentalité (garder à disposition les individus de la base)
- » Bonnes performances en général
- »  $Err(1-ppv) < 2 \times Err(\text{Modèle bayésien idéal})$



- » Paramétrage difficile (choix de la taille du voisinage)
- » Impossibilité d'interprétation d'un classement proposé
- » Nécessité de garder sous la main la base de données
- » Lenteur en classement (passage en revue de tous les individus de la base)
- » Sensibilité à la dimensionnalité (et aux variables non pertinentes)



- » Traitement des variables discrètes (codage 0/1 ou axes factoriels)
- » Choix de la distance pèse sur les résultats
- » Attention aux problèmes d'échelle si distance euclidienne utilisée
- » Pondérer l'influence des observations selon leur éloignement dans le voisinage





# Bibliographie

E.DIDAY, L. LEMAIRE, J.POUGET, F. TESTU - « Éléments d'analyse de données », DUNOD, 1982.

L. LEBART, A. MORINEAU, M. PIRON - « Statistique exploratoire multidimensionnelle », DUNOD, 2000 (3ème édition).

G. CELEUX, J.P. NAKACHE, « Analyse discriminante sur variables qualitatives », POLYTECHNICA, 1994.

*Et les incontournables, FUKUNAGA, DUDA & HART, etc.*

