

**INF7370**

---

# **Classification naïve bayésienne**

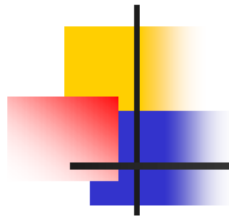
**Mohamed Bouguessa**



# Théorème de Bayes

---

- Soit  $X$  et  $Y$  deux variables aléatoires.
- $P(X, Y)$  : probabilité jointe.
- $P(X=x, Y=y)$  : probabilité que la variable  $X$  prend la valeur  $x$  et la variable  $Y$  prend la valeur  $y$ .
- $P(Y=y | X=x)$  : la probabilité que la variable  $Y=y$  sachant que  $X=x$ . (probabilité conditionnelle)
- $P(X, Y) = P(Y | X) * P(X) = P(X | Y) * P(Y)$



# Théorème de Bayes

---

Théorème de Bayes

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

$$P(X) = \sum_{i=1}^k P(X, Y_i) = \sum_{i=1}^k P(X|Y_i)P(Y_i)$$

$$X \in \{X_1, X_2, \dots, X_k\}$$

$$Y \in \{Y_1, Y_2, \dots, Y_k\}$$



# Théorème de Bayes pour la classification

---

- Soit  $X = \{X_1, X_2, \dots, X_d\}$  un ensemble de  $d$  attributs.
- Soit  $Y$  l'attribut des classes.
  - Le but est de prédire la classe  $Y$ .
  - Spécifiquement, le but est d'estimer  $P(Y | X)$ .  
cela revient à **estimer les probabilité a posteriori**  
 **$P(Y | X)$**  à partir des données d'apprentissage.
- En connaissant ces probabilités, un nouvel enregistrement  $X'$  peut être classé par l'identification de la classe  $Y'$  qui maximise la probabilité a posteriori  $P(Y' | X')$ .

# Exemple

- Soit l'ensemble d'apprentissage suivant :

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

classes

- Prédire la classe de l'enregistrement suivant:

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
11	No	Married	120K	?

$$\begin{aligned} &\longrightarrow P(Y=\text{Yes} \mid X_{11}) = ? \\ &P(Y=\text{No} \mid X_{11}) = ? \end{aligned}$$



## Exemple

---

- Pour prédire la classe de l'enregistrement précédent, on doit estimer les probabilité *a posteriori*  $P(\mathbf{Yes} | X)$  et  $P(\mathbf{No} | X)$  à partir des données d'apprentissage.

- Application du théorème de Bayes

$$P(Y = Yes | X) = \frac{P(X | Y = Yes)P(Y = Yes)}{P(X)}$$

- $P(Y=Yes)$  est la probabilité *a priori* qui peut être estimée directement à partir des données d'apprentissage.
- $P(X | Y=Yes)$  peut-être estimé en utilisant : un classificateur bayésien naïf.



# Classification bayésienne naïve

---

- Soit  $X = \{X_1, X_2, \dots, X_d\}$  un ensemble de  $d$  attributs.
- Soit  $y$  l'étiquette d'une classe.
- La classification bayésienne naïve suppose que les attributs sont indépendants. Formellement :

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y)$$

- Pour classifier un objet, le classifieur estime la probabilité *a posteriori* pour chaque classe  $Y$  comme :

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)}$$



# Classification bayésienne naïve

---

- Pour classer un objet, il suffit de choisir la classe qui maximise le terme

$$P(Y) \prod_{i=1}^d P(X_i | Y = y)$$

→ Comment estimer les probabilités  $P(Y)$  et  $P(X_i | Y=y)$  à partir des données d'apprentissage ?



# Estimation des probabilités

## □ Estimation des probabilités *a priori*

$$P(Y=y) = n_y / n$$

$$P(Y=Yes) = 3 / 10$$

$$P(Y=No) = 7 / 10$$

<i>Tid</i>	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

# Estimation des probabilités

## ❑ Estimation des probabilités conditionnelles

### Cas des attributs catégorique

Pour un attribut catégorique  $X_i$ , la probabilité conditionnelles  $P(X_i=x_i | Y=y)$  = le nombre des objets de la classe  $y$  tel que leurs valeurs dans l'attribut  $X_i = x_i$  divisé par le nombre total des objets de la classe  $y$

$$P(\text{Status}=\text{Married} | \text{No}) = 4/7$$

$$P(\text{Refund}=\text{Yes} | \text{Yes}) = 0$$

<i>Tid</i>	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



# Estimation des probabilités

---

## ❑ Estimation des probabilités conditionnelles

### Cas des attributs numérique

Deux façons possibles:

1. Discrétisation des attributs numérique continus pour les transformer en attributs catégorique (décomposition en plusieurs intervalles). Deux situations se présentent:
  - 1.1. Si le nombre d'intervalles est large, il y'aura peu de données d'apprentissage dans chaque intervalle pour fournir une estimation fiable de  $P(X_i | Y)$ .
  - 1.2. Si le nombre d'intervalle est petit, certains intervalles vont contenir des objets de classes différentes ce qui affecte la précision de la classification.



# Estimation des probabilités

---

## □ Cas des attributs numérique

2. On suppose que les valeurs d'un attribut suivent une certaine distribution statistique.

- En général, la distribution **Gaussienne** est souvent utilisée.
- Les paramètres de cette distribution sont : la moyenne  $\mu$  et la variance  $\sigma$ .



# Estimation des probabilités

---

## □ Cas des attributs numérique

- Pour chaque classe  $y_j$ , la probabilité conditionnelle pour l'attribut  $X_i$  est définie par

$$P(X_i = x_{ij} | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_{ij} - \mu_j)^2}{2\sigma_j^2}\right)$$

$$\mu_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

$$\sigma_j^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_{ij} - \mu_j)^2$$

# Estimation des probabilités

## □ Cas des attributs numérique

<i>Tid</i>	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

$$P(\text{Income} = 120 \mid \text{No}) = \frac{1}{\sqrt{2\pi \times 2975}} \exp\left(-\frac{(120 - 110)^2}{2(2975)}\right) = 0.0072$$

# Un exemple de la classification bayésienne naïve

Ensemble d'apprentissage

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Un objet X à classer

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
11	No	Married	120K	?

➤ On doit estimer  $P(\text{Yes} | X)$  et  $P(\text{No} | X)$

$$P(\text{Yes} | X) = \frac{P(\text{Yes}) \prod_{i=1}^d P(X_i | \text{Yes})}{P(X)}$$

- Estimation des probabilités *a priori*

$$P(\text{Yes}) = 0.3 \text{ et } P(\text{No}) = 0.7$$

# Un exemple de la classification bayésienne naïve

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
11	No	Married	120K	?

- Estimation des probabilités cond. :  $P(X_i | \text{Yes})$  et  $P(X_i | \text{No})$

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- pour les attributs catégoriques

$$P(\text{Owner}=\text{Yes}|\text{No}) = 3/7$$

$$P(\text{Owner}=\text{No}|\text{No}) = 4/7$$

$$P(\text{Owner}=\text{Yes}|\text{Yes}) = 0$$

$$P(\text{Owner}=\text{No}|\text{Yes}) = 1$$

$$P(\text{Marital Status}=\text{Single}|\text{No}) = 2/7$$

$$P(\text{Marital Status}=\text{Divorced}|\text{No}) = 1/7$$

$$P(\text{Marital Status}=\text{Married}|\text{No}) = 4/7$$

$$P(\text{Marital Status}=\text{Single}|\text{Yes}) = 2/3$$

$$P(\text{Marital Status}=\text{Divorced}|\text{Yes}) = 1/3$$

$$P(\text{Marital Status}=\text{Married}|\text{Yes}) = 0$$

- pour l'attribut numérique « Annual Income »

$$P(\text{Income} = 120K | \text{No}) = \frac{1}{\sqrt{2\pi \times 2975}} \exp\left(-\frac{(120 - 110)^2}{2(2975)}\right) = 0.0072$$

$$P(\text{Income} = 120K | \text{Yes}) = \frac{1}{\sqrt{2\pi \times 25}} \exp\left(-\frac{(120 - 90)^2}{2(25)}\right) = 1.2 \times 10^{-9}$$



# Un exemple de la classification bayésienne naïve

- $$\begin{aligned} P(X | \text{No}) &= P(\text{Owner}=\text{No} | \text{No}) \\ &\quad \times P(\text{Marital Status} = \text{Married} | \text{No}) \\ &\quad \times P(\text{Income}=120\text{K} | \text{No}) \\ &= 4/7 \times 4/7 \times 0.0072 = 0.0024 \end{aligned}$$

- $$\begin{aligned} P(X | \text{Yes}) &= P(\text{Owner}=\text{No} | \text{Yes}) \\ &\quad \times P(\text{Married} | \text{Yes}) \\ &\quad \times P(\text{Income}=120\text{K} | \text{Yes}) \\ &= 1 \times 0 \times 1.2 \times 10^{-9} = 0 \end{aligned}$$

- Puisque  $P(X | \text{No})P(\text{No}) > P(X | \text{Yes})P(\text{Yes})$

doc  $P(\text{No} | X) > P(\text{Yes} | X)$

→ Conclusion : l'enregistrement sera placé dans la classe No

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
11	No	Married	120K	

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
11	No	Married	120K	No

# Exemple

Temps	Temperature	Humidite	Vent	Class
soleil	chaud	élevé	faux	N
soleil	chaud	élevé	VRAI	N
couvert	chaud	élevé	faux	P
pluie	tiede	élevé	faux	P
pluie	froid	normal	faux	P
pluie	froid	normal		N
couvert	froid	normal	VRAI	P
soleil	tiede	élevé	faux	N
soleil	froid	normal	faux	P
pluie	tiede	normal	faux	P
soleil	tiede	normal	VRAI	P
couvert	tiede	élevé	VRAI	P
couvert	chaud	normal	faux	P
pluie	tiede	élevé	VRAI	N

$$P(p) = 9/14$$

$$P(n) = 5/14$$

## Temps

$$P(\text{soleil}|P) = 2/9$$

$$P(\text{soleil}|N) = 3/5$$

$$P(\text{couvert}|P) = 4/9$$

$$P(\text{couvert}|N) = 0$$

$$P(\text{pluie}|P) = 3/9$$

$$P(\text{pluie}|N) = 2/5$$

## Température

$$P(\text{chaud}|P) = 2/9$$

$$P(\text{chaud}|N) = 2/5$$

$$P(\text{tiède}|P) = 4/9$$

$$P(\text{tiède}|N) = 2/5$$

$$P(\text{froid}|P) = 3/9$$

$$P(\text{froid}|N) = 1/5$$

## Humidité

$$P(\text{élevée}|P) = 3/9$$

$$P(\text{élevée}|N) = 4/5$$

$$P(\text{normale}|P) = 6/9$$

$$P(\text{normale}|N) = 2/5$$

## Vent

$$P(\text{Vrai}|P) = 3/9$$

$$P(\text{vrai}|N) = 3/5$$

$$P(\text{faux}|P) = 6/9$$

$$P(\text{faux}|N) = 2/5$$



## Exemple

---

- $X = \langle \text{pluie, chaud, élevée, faux, ?} \rangle$
- $P(X|P) \cdot P(p) =$   
 $P(\text{pluie}|p) \cdot P(\text{chaud}|p) \cdot P(\text{élevée}|p) \cdot P(\text{faux}|p) \cdot P(p)$   
 $= 3/9 \cdot 2/9 \cdot 3/9 \cdot 6/9 \cdot 9/14 = 0.010582$
- $P(X|n) \cdot P(n) =$   
 $P(\text{pluie}|n) \cdot P(\text{chaud}|n) \cdot P(\text{élevée}|n) \cdot P(\text{faux}|n) \cdot P(n)$   
 $= 2/5 \cdot 2/5 \cdot 4/5 \cdot 2/5 \cdot 5/14 = 0.018286$
- $X$  est classifié en  $N$  (ne pas jouer au tennis)

# Classification du texte

- Classer les documents comme intéressent ou non :

Document  $\rightarrow \{+, -\}$

- Phase d'apprentissage : estimation de

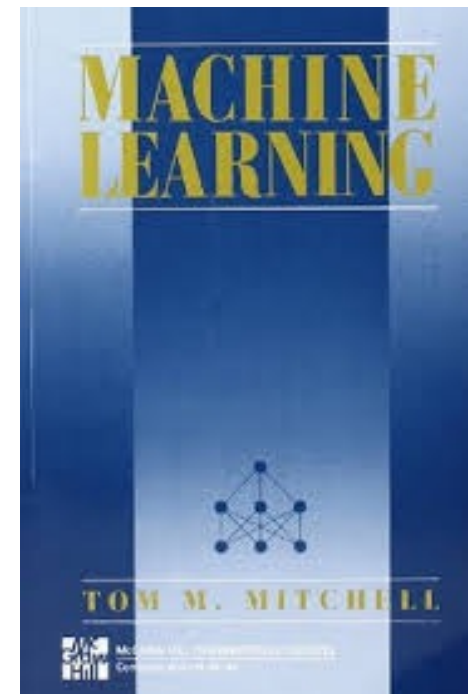
➤  $P(+)$

➤  $P(-)$

➤  $P(\text{Doc} | +)$

➤  $P(\text{Doc} | -)$

Algorithme tiré du livre de Tom Mitchell



## Learn\_Naive\_Bayes\_Text(TrainData, Y)

//Collecter tous les mots dans TrainData

- $vocabulaire \leftarrow$  l'ensemble de mots TrainData //ne pas considérer les mots vides. Considérer qu'une seule occurrence par mot.

//Calcule de la probabilité à priori de chaque classe  $P(Y=y_j)$  ainsi que les probabilités conditionnels associés au mots  $P(w_i | y_j)$

- **pour** chaque classe  $y_j$  dans Y **faire**

- $docs_j \leftarrow$  l'ensemble de mots appartenant à la classe  $y_j$ ;
- Calcule de la probabilité à priori de la classe  $y_j$   $P(y_j) \leftarrow \frac{|docs_j|}{|TrainData|}$
- $\mathcal{W} \leftarrow$  l'ensemble de mots qui sont à la fois dans  $docs_j$  ET  $vocabulaire$

- **pour** chaque mot  $w_i$  dans  $\mathcal{W}$  **faire**

- Calcule des probabilités  $P(w_i | y_j) \leftarrow \frac{n_i + 1}{n + |vocabulaire|}$ 
  - $n_i \leftarrow$  la fréquence du mot  $w_i$  dans  $docs_j$
  - $n \leftarrow$  le nombre de mots distinct dans  $docs_j$

- Return *Décision* tel que  $Décision = \max_{y_j \in Y} \left[ P(y_j) \prod_{w_i} P(w_i | y_j) \right]$