Apprentissage automatique supervisé

Amira Barhoumi

amira.barhoumi@univ-grenoble-alpes.fr

Année universitaire : 2025-2026

Naïve Bayes

- Méthode de Bayes naïf (*Naïve Bayes*)
- Méthode d'apprentissage automatique supervisé
- Comment classer un nouvel exemple en fonction d'un ensemble d'exemples pour lesquels nous connaissons la classe?

	var 1	…	var k	classe
obj				oui
…				…
obj				non

	var1	…	var k	classe
new				?

Naïve Bayes

- Méthode de classification probabiliste
- Détermination l'appartenance la plus probable d'une observation X à une classe connue c en se basant sur les différents descripteurs
- Attribution de la probabilité d'attribution à une classe
- Méthode générative :
    - Modéliser chaque classe
    - Déterminer la probabilité d'appartenance d'une observation aux classes
- Application du théorème de Bayes

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B|A) \times p(A)}{p(B)}$$

Naïve Bayes

- $X = (x_1, x_2, \dots, x_n)$ est une observation
- $D=\{(X_i,Y_i);\ X_i$ et une observation et Y_i est sa classe correspondante$\}$
- Probabilités à priori $p(Y_i)$
- Détermination des probabilités à posteriori $p(Y|X)$

Vraisemblance

$$p(Y|X) = \frac{p(Y \cap X)}{p(X)}$$
$$= \frac{p(X|Y) \times p(Y)}{p(X)}$$

Probabilité à priori de Y

Probabilité à posteriori

Probabilité à priori de X

Naïve Bayes

- $X = (x_1, x_2, \ldots, x_n)$ est une observation
- $D=\{(X_i,Y_i); X_i$ et une observation et Y_i est sa classe correspondante$\}$
- Maximisation de la probabilité à posteriori

$$\begin{aligned}
\hat{Y} &= \underset{Y_i}{argmax}\ p(Y_i|X) \\
&= \underset{Y_i}{argmax}\ \frac{p(X|Y_i) \times p(Y_i)}{p(X))} \\
&\approx \underset{Y_i}{argmax}\ p(X|Y_i) \times p(Y_i) \\
&\approx \underset{Y_i}{argmax}\ p((x_1, x_2, ..., x_n)|Y_i) \times p(Y_i)
\end{aligned}$$

Naïve Bayes

- Estimation des deux termes à partir de l'équation en se basant sur le corpus d'apprentissage

$$\hat{Y} \approx \underset{Y_i}{argmax}\ p((x_1, x_2, ..., x_n)|Y_i) \times p(Y_i)$$

- $p(Y_i)$ est facile à calculer => compter la fréquence de chaque classe dans le corpus d'apprentissage
- Estimer $p(x_1,x_2,...,x_n|Y_i)$ est plus difficile
- Le nombre de ces termes est égal au nombre d'instances possibles multiplié par le nombre de classes
- Besoin de considérer chaque instance dans le corpus d'apprentissage pour obtenir une estimation fiable
- Simplification du calcul en utilisant la règle de chaînage

- Règle de chaînage (*chain rule*) :

$$
\begin{aligned}
P(X_1, \ldots, X_n) &= P(X_1,\ldots,X_{n-1})\, P(X_n \mid X_1,\ldots,X_{n-1}) \\
&= P(X_1,\ldots,X_{n-2})\, P(X_{n-1} \mid X_1,\ldots,X_{n-2})\, P(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \prod_{i=1..n} P(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}
$$

- Par application de la règle de chaînage, on obtient :

$$
\begin{aligned}
p((x_1, x_2, \ldots, x_n)|Y_i) &= p(x_1, x_2, \ldots, x_n|Y_i) \\
&= p(x_1|Yi)p(x_2|x_1, Yi)p(x_3|x_1, x_2, Yi)\ldots p(x_n|x_1, \ldots, x_{n-1}, Yi)
\end{aligned}
$$

Naïve Bayes

- Hypothèse de Naïve Bayes : les descripteurs sont indépendants étant donné la classe

$$p(x, y|c) = p(x|c)p(y|c) \quad ; \qquad \text{c : une classe}$$
$$p(x|y, c) = p(x|c) \quad ; \qquad \forall x, y, c$$

$$\begin{aligned} p(x_1, x_2, ..., x_n|Y_i) &= p(x_1|Yi)\ p(x_2|x_1, Yi)\ p(x_3|x_1, x_2, Yi) ...\ p(x_n|x_1, ..., x_{n-1}, Yi) \\ &= p(x_1|Yi)\ p(x_2|Yi)\ p(x_3|Yi)\ ...\ p(x_n|Yi) \\ &= \prod_{j=1}^{n} p(x_j|Y_i) \end{aligned}$$

facile à estimer

Naïve Bayes

$$\hat{Y} \approx \underset{Y_i}{argmax}\ p(x_1, x_2, ..., x_n|Y_i) \times p(Y_i)$$
$$\approx \underset{Y_i}{argmax}\ \prod_{j=1}^{n} p(x_j|Yi) \times p(Y_i)$$

- Calcul de $p(x_j|Y_i)$ est beaucoup plus facile que calculer $p(x_1,x_2,...,x_n|Y_i)$
- Fréquence des occurrences dans le corpus d'apprentissage

Naïve Bayes : Exemple

	TEMPS	HUMIDITE	VENT	TENNIS
Ex1	Soleil	Haute	Oui	Oui
Ex2	Soleil	Basse	Non	Non
Ex3	nuageux	Basse	Oui	Oui
Ex4	pluvieux	Haute	Oui	Non
Ex5	pluvieux	Basse	Oui	Non
Ex6	Soleil	Basse	Oui	Oui
Ex7	pluvieux	Basse	Non	Non
	Soleil	***haute***	***Non***	***?***

Va-t-on jouer s'il y a du soleil, beaucoup d'humidité et pas de vent ?

Naïve Bayes : Exemple

Classe Oui			Classe Non		
Temps	Soleil		Temps	Soleil	
	Nuageux			Nuageux	
	pluvieux			pluvieux	
Humidité	Haute		Humidité	Haute	
	Basse			Basse	
Vent	Oui		Vent	Oui	
	Non			Non	

Naïve Bayes : Exemple

Classe Oui			Classe Non		
Temps	Soleil	2/3	Temps	Soleil	1/4
	Nuageux	1/3		Nuageux	0/4
	pluvieux	0/3		pluvieux	3/4
Humidité	Haute	1/3	Humidité	Haute	1/4
	Basse	2/3		Basse	3/4
Vent	Oui	3/3	Vent	Oui	2/4
	Non	0/3		Non	2/4

P (Oui) = 3/7

P (Non) = 4/7

Test : classifier (Temps = soleil, humidité = haute, Vent = non)

Max (3/7 * 2/3 * 1/3 * 0/3, 4/7 * ¼ * ¼ * 2/4) = max (0, 0.017) = 0.017

Naïve Bayes : Exemple

Classe Oui			Classe Non		
Temps	Soleil	2/3	Temps	Soleil	1/4
	Nuageux	1/3		Nuageux	0/4
	pluvieux	0/3		pluvieux	3/4
Humidité	Haute	1/3	Humidité	Haute	1/4
	Basse	2/3		Basse	3/4
Vent	Oui	3/3	Vent	Oui	2/4
	Non	0/3		Non	2/4

P (Oui) = 3/7

P (Non) = 4/7

Test : classifier (Temps = soleil, humidité = haute, Vent = non)

Max (3/7 * 2/3 * 1/3 * 0/3, 4/7 * ¼ * ¼ * 2/4) = max (0, 0.017) = 0.017

→ Pas de tennis aujourd'hui !!

Naïve Bayes : Exemple

- 0 et 0.017 ne sont pas des probabilités
- Normalisation avec la somme des valeurs obtenues pour des différentes classes
- Valeur petite ⇒ *underflow problem*
- Utiliser la logarithmique (dans la pratique)

Naïve Bayes : Exemple

P (Oui) = 3/7

P (Non) = 4/7

Test : classifier (Temps = soleil, humidité = haute, Vent = non)

Max (3/7 * 2/3 * 1/3 * 0/3, 4/7 * ¼ * ¼ * 2/4) = max (0, 0.017) = 0.017

- p(vent = non | classe = oui) = 0
- La classe “*non*” est attribuée à chaque nouvelle instance avec vent = non
- Corpus d’apprentissage est petit ⇒ *smoothing*

Smoothing

- Introduction des instances irréelles (fake instances)
- Distribution uniforme
- Une possibilité : *m-estimate of probability*

$$P = \frac{n_c + mp}{n + m}$$

- p est une estimation de la probabilité à priori
- m est une constante appelée *equivalent sample size* (nombre d'instances irréelles)
- Par exemple : si l'attribut admet 2 valeurs possibles

  alors on peut fixer **p** = 1/2 = 0.5 (probabilité à priori uniforme)
- Étant cette probabilité à priori, on introduit **m** *fake instances*
- Par exemple, si l'attribut admet 2 valeurs possibles

  alors on peut fixer **m** = 2

$$P = \frac{n_c + 1}{n + 2}$$

Naïve Bayes : Exemple

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D3	overcast	hot	high	weak	yes
D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D6	rain	cool	normal	strong	no
D7	overcast	cool	normal	strong	yes
D8	sunny	mild	high	weak	no
D9	sunny	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes
D11	sunny	mild	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes
D14	rain	mild	high	strong	no

- Classer la nouvelle instance (sunny, cool, high, strong)

Naïve Bayes : Exemple

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D3	overcast	hot	high	weak	yes
D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D6	rain	cool	normal	strong	no
D7	overcast	cool	normal	strong	yes
D8	sunny	mild	high	weak	no
D9	sunny	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes
D11	sunny	mild	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes
D14	rain	mild	high	strong	no

- Classer la nouvelle instance (sunny, cool, high, strong)

$$\begin{aligned} P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes) &= 0.00529 \\ P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no) &= 0.02057 \end{aligned}$$

→ Pas de tennis aujourd'hui !!

Naïve Bayes

- Gaussian Naïve Bayes : il suppose que les données de chaque classe sont tirées d’une distribution gaussienne simple
- Multinomial Naïve Bayes : il suppose que les caractéristiques sont tirées d’une distribution multinomiale simple
- Bernoulli Naïve Bayes : il suppose que les caractéristiques sont de nature binaire (0 et 1)

Naïve Bayes

- Avantages :
  - Très répandu
  - Facile à mettre en place

- Inconvénients
  - Hypothèse d’indépendance des attributs naïve