CLASSIFICATEUR DE BAYES

INTRODUCTION

1.Définition :

- Le classificateur de Bayes est une approche probabiliste utilisée pour la **classification**.
- Il s'agit d'un **classificateur statistique** qui prédit la **probabilité d'appartenance** d'un élément à une classe donnée.

2.Fondement Théorique :

- Basé sur le **théorème de Bayes**, qui permet d’actualiser la probabilité d’une hypothèse en fonction des nouvelles données observées.

3.Hypothèses et Simplifications :

- Dans le cas du **classificateur bayésien naïf**, on suppose une **indépendance conditionnelle** entre les attributs, ce qui simplifie les calculs.

4.Avantages :

- **Précision élevée**, notamment sur de grandes bases de données.
- **Rapidité des calculs**, ce qui le rend efficace pour le traitement de grandes quantités de données.

CLASSIFICATEUR DE BAYES

Dans cette approche, on s'intéresse souvent à estimer la probabilité conditionnelle $p(\mathbf{x}|y)$, qui veut dire:

Étant donné y vérifié ou observé (ex. y peut être une classe), qu'elle est la probabilité d'observer la donnée $\mathbf{x}$?

RAPPELS SUR LE THÉORÈME DE BAYES

- Soient **A, B et C** trois événements, on a les relations suivantes :
- **Probabilité a posteriori** :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Probabilité a probabilité a posteriori **avec un contexte *C*** :

$$P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$$

- **P(A|B)** représente la vraisemblance de A sachant que B est observé.
- **P(A|B,C)** représente la vraisemblance de A sachant que B et C sont observés.
- **Note** : L'événement **C** représente le **contexte général** dans lequel les probabilités sont évaluées.

RAPPELS SUR LE THÉORÈME DE BAYES

- **Exemple concret : Diagnostic médical**
- Supposons qu'on veuille diagnostiquer une maladie H à partir d'un test E.
  - P(H) : Probabilité d'avoir la maladie avant de faire le test (ex. 1 % de la population).
  - P(E|H) : Probabilité que le test soit positif si la personne est malade (ex. 90 %).
  - P(E) : Probabilité globale d'un test positif dans toute la population
- Le théorème de Bayes permet de calculer la **vraie probabilité d'être malade si le test est positif**, et pas seulement la fiabilité du test.

CLASSIFICATION PAR LA RÈGLE DE BAYES

- Dans le cadre de la classification, nous supposons :
  - $y=C_k$ représente l'événement "l'observation appartient à la classe C_k".
  - $x=\mathbf{x}$ représente l'événement "j'observe la donnée $\mathbf{x}$".
- La probabilité conditionnelle suivante est utilisée :

$$P(y = C_k | x = \mathbf{x}, D)$$

- **Application de la règle de Bayes**
- En appliquant le **théorème de Bayes**, on obtient :

$$P(y = C_k | x = \mathbf{x}, D) = \frac{P(x = \mathbf{x} | y = C_k, D) P(y = C_k | D)}{P(x = \mathbf{x} | D)}$$

CLASSIFICATION PAR LA RÈGLE DE BAYES

1

2

$$P(y = C_k | x = \mathbf{x}, D) = \frac{P(x = \mathbf{x} | y = C_k, D) P(y = C_k | D)}{P(x = \mathbf{x} | D)}$$

3

1-La probabilité **a priori** d'une classe C_k est :$P(y=C_k|D)$

- Elle représente la **probabilité d'observer la classe C_k avant même d'analyser la donnée x**, compte tenu de l'ensemble des observations D.
- Elle est généralement estimée par la **fréquence** des exemples appartenant à la classe C_k dans la base de données D ,

- 2- $P(x = \mathbf{x} | y = C_k, D)$

  représente la **vraisemblance**, c'est-à-dire la probabilité **d'observer la donnée x sachant qu'elle appartient à la classe C_k**. Contrairement à la probabilité **a priori**, ce terme est souvent plus difficile à estimer, car il dépend de la distribution des données pour chaque classe.

- L'**hypothèse de Bayes naïve** repose sur une simplification qui facilite le calcul des probabilités conditionnelles. Elle suppose que la donnée x est composée d'un ensemble de caractéristiques indépendantes les unes des autres.

CLASSIFICATION PAR LA RÈGLE DE BAYES

- Si l'on considère une donnée x décrite par D attributs $x_1, x_2, \ldots, x_D$, dont les valeurs observées sont $x_1, x_2, \ldots, x$, alors selon l'HBN :

$$P(\mathbf{x} = x | y = C_k, D) = P(x_1 = x_1 | y = C_k, D) \times P(x_2 = x_2 | y = C_k, D) \times \cdots \times P(x_D = x_D | y = C_k, D)$$

- **la vraisemblance d'une donnée appartenant à une classe est décomposée en un produit des probabilités de chacun de ses attributs**, conditionnées par la classe.

$P(x_d = x_d | y = C_k, D)$ correspond à la **probabilité d'observer la valeur x_d de l'attribut x_d lorsque la classe C_k est connue.**

CLASSIFICATION PAR LA RÈGLE DE BAYES

- 3- La probabilité $p(X=x/D)$ représente la probabilité d'observer la valeur $X=x=(x_1,\ldots,x_D)$ étant donné l'ensemble d'exemples D.
- Pour calculer cette probabilité, on applique la marginalisation suivante :

$$p(\mathbf{x}=x|D)=\sum_{j=1}^{K} p(\mathbf{x}=x, y=C_j|D)$$

En utilisant la règle du produit, cette expression peut être reformulée comme suit

$$p(\mathbf{x}=x|D)=\sum_{j=1}^{K} p(\mathbf{x}=x|y=C_j, D)p(y=C_j|D)$$

$$p(\mathbf{x}=x|y=C_k, D)=\sum_{j=1}^{K} p(\mathbf{x}=x|y=C_j, D)p(y=C_j|D)$$

Cette formulation met en évidence que la probabilité de x est obtenue en sommant les probabilités conditionnelles pondérées par les probabilités a priori des classes."

CLASSIFICATION PAR LA RÈGLE DE BAYES

- Une fois la probabilité d'appartenance $p(y=C_k|x=x,D)$ calculée pour chaque classe $k \in \{1,\ldots,K\}$ et pour la donnée $\mathbf{x}$=x, on peut prédire la classe de x en choisissant celle qui maximise la probabilité a posteriori :

$$y_{\text{MAP}} = \arg\max_{y=C_k} p(y = C_k | \mathbf{x} = x, D)$$

- En appliquant la règle de Bayes, cette expression peut être reformulée comme suit :

$$y_{\text{MAP}} = \arg\max_{y=C_k} p(\mathbf{x} = x | y = C_k, D) p(y = C_k | D)$$

- Si l'on ne tient pas compte de la probabilité a priori $p(y=C_k|D)$, on obtient alors :

$$y_{\text{ML}} = \arg\max_{y=C_k} p(\mathbf{x} = x | y = C_k, D)$$

- Ce critère est appelé le *maximum de vraisemblance* (ML).

Exemple

$\overbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxx}}^{D = 4}$ $\overbrace{\hphantom{xxxxxxxx}}^{K = 2}$

	x_1	x_2	x_3	x_4	$C_1 = Oui, C_2 = Non$
Jour	Ciel	Température	Humidité	Vent	Jouer au tennis ?
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

N = 14

$x = (Ciel{=}Couvert,\ Température{=}Chaud,\ Humidité{=}Normale,\ Vent{=}Faible)$
$y =$ Oui

CLASSIFICATION PAR LA RÈGLE DE BAYES

- **Exemple :** Prenons l'exemple de la décision de jouer au tennis.
Les caractéristiques sont représentées par le vecteur **x = (Ciel, Température, Humidité, Vent)**, et la variable cible **y** prend ses valeurs dans l'ensemble **{oui, non}**.
- Supposons que nous souhaitons prédire la classe de l'observation **x = (Ciel = ensoleillé, Température = fraîche, Humidité = élevée, Vent = fort)**. En utilisant la règle de Bayes, nous obtenons :

$$p(jouer = oui|x, \mathcal{D}) = \frac{p(x|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{p(x|\mathcal{D})}$$

$$= \frac{p(x|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{\sum_{a\in\{oui,non\}} p(x|jouer = a, \mathcal{D})p(jouer = a|\mathcal{D})}$$

Jour	Ciel	Température	Humidité	Vent	Jouer
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

attribut	valeur	J = oui	J = non
Ciel	Ensoleillé	2	3
	Couvert	4	0
	Pluie	3	2
Température	Chaude		
	Tiède		
	Fraiche		
Humidité	Élevée		
	Normale		
Vent	Fort		
	Faible		
Jouer		9	5

CLASSIFICATION PAR LA RÈGLE DE BAYES

attribut	valeur	Jouer = oui	Jouer = non
Ciel	Ensoleillé	2	3
	Couvert	4	0
	Pluie	3	2
Température	Chaude	2	2
	Tiède	4	2
	Fraiche	3	1
Humidité	Élevée	3	4
	Normale	6	1
Vent	Fort	3	3
	Faible	6	2
Jouer		**9**	**5**

CLASSIFICATION PAR LA RÈGLE DE BAYES

x = (Ciel = ensoleillé, Température = fraîche, Humidité = élevée, Vent = fort).

$$p(jouer = oui|x, \mathcal{D}) = \frac{p(x|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{p(x|\mathcal{D})}$$

- $p(x|jouer = oui, \mathcal{D}) \overset{\text{HBN}}{=} p(\text{Ciel} = ensoleillé|jouer = oui, \mathcal{D})$
  $\times\, p(\text{Température} = fraiche|jouer = oui, \mathcal{D})$
  $\times\, p(\text{Humidité} = élevée|jouer = oui, \mathcal{D})$
  $\times\, p(\text{Vent} = fort|jouer = oui, \mathcal{D})$

  $= \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} = \frac{2}{243}$

- $p(jouer = oui|\mathcal{D}) = \frac{9}{14}$.
- $p(jouer = non|\mathcal{D}) = \frac{5}{14}$.

$$p(jouer = oui|x,\mathcal{D}) = \frac{\frac{2}{243} \times \frac{9}{14}}{\frac{2}{243} \times \frac{9}{14} + \boxed{?} \times \boxed{?}} \approx \boxed{?}$$

CLASSIFICATION PAR LA RÈGLE DE BAYES

x = (Ciel = ensoleillé, Température = fraîche, Humidité = élevée, Vent = fort).

$$p(x|jouer = non, \mathcal{D}) = p(\text{Ciel} = ensoleillé|jouer = non, \mathcal{D})$$

$$\times\, p(\text{Température} = fraiche|jouer = non, \mathcal{D})$$

$$\times\, p(\text{Humidité} = élevée|jouer = non, \mathcal{D})$$

$$\times\, p(\text{Vent} = fort|jouer = non, \mathcal{D})$$

$$= \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} = \frac{36}{625}$$

$$p(jouer = oui|x,\mathcal{D}) = \frac{\frac{2}{243} \times \frac{9}{14}}{\frac{2}{243} \times \frac{9}{14} + \frac{36}{625} \times \frac{5}{14}} \approx 0.205$$

$$p(jouer = non|x,\mathcal{D}) = 0.795$$

ATTRIBUTS NUMÉRIQUES ET LEUR TRAITEMENT DANS LA CLASSIFICATION

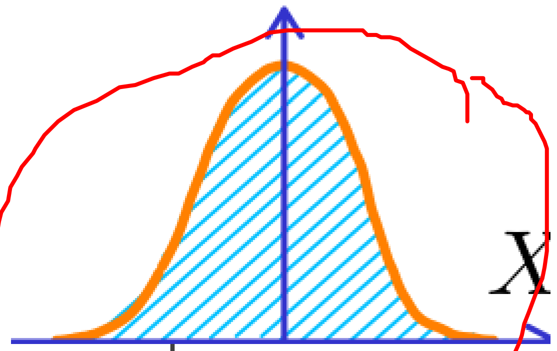
- Lorsque les attributs sont **numériques**, la détermination de la probabilité d'observer une valeur spécifique ne peut pas se faire de la même manière que pour les attributs catégoriels
- Pour modéliser la probabilité d'une valeur numérique, on suppose que la distribution des valeurs de l'attribut x_d suit une **loi normale** (ou distribution gaussienne). Cela s'écrit :

$$X \sim N(\mu; \sigma)$$

$$x_d \sim \mathcal{N}(\mu_d, \sigma_d^2)$$

- μ_d est la **moyenne** des valeurs de l'attribut x_d,
- σ_d est l'**écart-type**.

$$p(x_d \mid \mu_d, \sigma_d) = \frac{1}{\sigma_d \sqrt{2\pi}} \exp\left(-\frac{(x_d - \mu_d)^2}{2\sigma_d^2}\right)$$



- Ces paramètres sont estimés à partir des données d'apprentissage *D*. Une fois ces valeurs calculées, on peut utiliser la formule de la densité de probabilité pour estimer la probabilité d'une nouvelle observation

ATTRIBUTS NUMÉRIQUES ET LEUR TRAITEMENT DANS LA CLASSIFICATION

Ciel	Température	Humidité	Vent	Joue
Ensoleillé	27, 5	85	Faible	Non
Ensoleillé	25	90	Fort	Non
Couvert	26, 5	86	Faible	Oui
Pluie	20	96	Faible	Oui
Pluie	19	80	Faible	Oui
Pluie	17, 5	70	Fort	Non
Couvert	17	65	Fort	Oui
Ensoleillé	21	95	Faible	Non
Ensoleillé	19, 5	70	Faible	Oui
Pluie	22, 5	80	Faible	Oui
Ensoleillé	22, 5	70	Fort	Oui
Couvert	21	90	Fort	Oui
Couvert	25, 5	75	Faible	Oui
Pluie	20, 5	91	Fort	Non

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

	Température		Humidité	
	oui	non	oui	non
	26, 5	27, 5	86	85
	20	25	96	90
	19	17, 5	80	70
	17	21	65	95
	19, 5	20, 5	70	91
	22, 5		80	
	22, 5		70	
	21		90	
	25, 5		75	
moyenne	21, 5	22, 3	79, 1	86, 2
écart-type	2, 91	3, 53	10, 2	9, 7

ATTRIBUTS NUMÉRIQUES ET LEUR TRAITEMENT DANS LA CLASSIFICATION

x = (Ciel= *ensoleillé*, Température = *18*, Humidité = *90*, Vent = *fort*), on aura:

- On peut alors calculer par exemple:

$$p(x_d \mid \mu_d, \sigma_d) = \frac{1}{\sigma_d\sqrt{2\pi}} \exp\left(-\frac{(x_d-\mu_d)^2}{2\sigma_d^2}\right)$$

$$p(\text{Température} = 18 | jouer = oui, \mathcal{D}) = \frac{1}{\sqrt{2\pi(2.91)^2}} e^{-\frac{(18-21.5)^2}{2(2.91)^2}}$$

$$\approx 0.0665$$

$$p(\text{Température} = 18 | jouer = non, \mathcal{D}) = \frac{1}{\sqrt{2\pi(3.53)^2}} e^{-\frac{(18-22.3)^2}{2(3.53)^2}}$$

$$\approx 0.0538$$

ATTRIBUTS NUMÉRIQUES ET LEUR TRAITEMENT DANS LA CLASSIFICATION

$$p(\text{Humidité} = 90 | jouer = oui, \mathcal{D}) = \frac{1}{\sqrt{2\pi(10.2)^2}} e^{-\frac{(90-79.1)^2}{2(10.2)^2}}$$

$$\approx 0.0221$$

$$p(\text{Humidité} = 90 | jouer = non, \mathcal{D}) = \frac{1}{\sqrt{2\pi(9.70)^2}} e^{-\frac{(90-86.2)^2}{2(9.70)^2}}$$

$$\approx 0.0381$$

$$\boxed{p(x_d \mid \mu_d, \sigma_d) = \frac{1}{\sigma_d\sqrt{2\pi}} \exp\left(-\frac{(x_d-\mu_d)^2}{2\sigma_d^2}\right)}$$

ATTRIBUTS NUMÉRIQUES ET LEUR TRAITEMENT DANS LA CLASSIFICATION

x = (Ciel= *ensoleillé*, Température = *18*, Humidité = *90*, Vent = *fort*), on aura:

$$p(jouer = oui|x, \mathcal{D}) = \frac{2/9 \times \boxed{0.0665} \times 0.0221 \times 3/9 \times 9/14}{p(x|\mathcal{D})}$$

$$\approx \frac{2.1 \times 10^{-4}}{p(x|\mathcal{D})}$$

$$p(jouer = non|x, \mathcal{D}) = \frac{3/5 \times \boxed{0.0538} \times \boxed{0.0381} \times 3/5 \times 5/14}{p(x|\mathcal{D})}$$

$$\approx \frac{2.635 \times 10^{-4}}{p(x|\mathcal{D})}$$

ATTRIBUTS NUMÉRIQUES ET LEUR TRAITEMENT DANS LA CLASSIFICATION

En ayant:

$$p(x|\mathcal{D}) = 2.1 \times 10^{-4} + 2.635 \times 10^{-4}$$

On obtient:

$$p(jouer = oui|x, \mathcal{D}) = \frac{2.1 \times 10^{-4}}{2.1 \times 10^{-4} + 2.635 \times 10^{-4}} \approx 0.44$$

$$p(jouer = non|x, \mathcal{D}) = \frac{2.635 \times 10^{-4}}{2.1 \times 10^{-4} + 2.635 \times 10^{-4}} \approx 0.56$$

GESTION DES VALEURS D'ATTRIBUTS MANQUANTES DANS LA CLASSIFICATION BAYÉSIENNE

- Lors de l'application de la méthode de Bayes, un problème survient lorsque certaines **valeurs d'attributs** ne sont pas présentes dans l'ensemble d'entraînement *D*. Cela conduit à des estimations de probabilités nulles, ce qui pose à la fois des problèmes conceptuels et pratiques.

GESTION DES VALEURS D'ATTRIBUTS MANQUANTES DANS LA CLASSIFICATION BAYÉSIENNE

- Le problème vient du **manque de données** : l'absence d'exemples pour une valeur spécifique dans l'ensemble d'entraînement conduit à une estimation erronée. En réalité, une probabilité nulle cache souvent une **valeur très petite**, mais non nulle. Il est donc nécessaire de corriger cette estimation pour éviter des résultats absurdes.
- Pour résoudre ce problème, on utilise une technique appelée **lissage de Laplace** (ou estimateur de Laplace). Cette méthode consiste à **ajouter une petite valeur** (par exemple 1) à chaque décompte, ce qui évite les probabilités nulles tout en respectant les règles des probabilités.

APPLICATION DE L'ESTIMATEUR DE LAPLACE EXEMPLE CONCRET :

- Supposons que pour l'attribut "Ciel", les probabilités estimées soient initialement :3/5, 0/5 , 2/5
- Avec l'estimateur de Laplace (en ajoutant 1 à chaque décompte), **Ensoleillé** : 3+1
  - •**Nuageux** : 0+1=1
  - •**Couvert** : 2+1=3 aucune probabilité n'est nulle, et les estimations sont plus robustes.
- les probabilités deviennent :4/8 , 1/8 , 3/8
- **Formalisation générale** : Pour un attribut donné, on ajoute une valeur μ (par exemple 1) à chaque décompte au numérateur, et on ajoute µ-arité de l'attribut au dénominateur.
- µ-arité d'un attribut correspond au nombre de valeurs possibles qu'il peut prendre. Cela revient à introduire une **probabilité a priori** pour chaque valeur de l'attribut.

$$\mu = 3$$

$$p(jouer = oui|\text{Ciel} = \text{couvert}, \mathcal{D})$$
$$= \frac{p(\text{Ciel} = \text{couvert}|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{p(\text{Ciel} = \text{couvert}|\mathcal{D})}$$

$$= \frac{4/9 \times 9/14}{p(\text{Ciel} = \text{couvert}|\mathcal{D})}$$

La probabilité 0 est remplacé par une petite valeur

$$p(jouer = non|\text{Ciel} = \text{couvert}, \mathcal{D}) = \frac{1/8 \times 5/14}{p(\text{Ciel} = \text{couvert}|\mathcal{D})}$$

$$p(jouer = oui|\text{Ciel} = \text{couvert}, \mathcal{D}) = \frac{\frac{4}{9} \times \frac{9}{14}}{\frac{4}{9} \times \frac{9}{14} + \frac{1}{8} \times \frac{5}{14}} \approx 0.8649$$