

Intelligence Artificielle et Données

Apprentissage supervisé : Théorie de décision Bayésienne

Akka Zemmari

LaBRI, Université de Bordeaux

2024 - 2025

Rappels et notations

Notations

- ▶ $\mathbb{P}r(A)$: probabilité de l'évènement A ,
- ▶ $\mathbb{P}r_X(.)$: distribution de la v.a. discrète X ,
- ▶ $f_X(.)$: densité de probabilité de la v.a. continue X ,
- ▶ $f_{X,Y}(.)$: densité jointe de probabilité de X et Y ,
- ▶ $\mathbb{P}r_{X|Y}(.)$: distribution conditionnelle de la v.a. discrète X , sachant Y ,
- ▶ $f_{X|Y}(.)$: densité conditionnelle de X sachant Y .

Rappels et notations

Rappels

- Probabilité conditionnelle :

$$\mathbb{P}r(A \mid B) = \frac{\mathbb{P}r(A \cap B)}{\mathbb{P}r(B)},$$

- Loi des probabilités totales :

$$\mathbb{P}r(A) = \sum_{i=1}^n \mathbb{P}r(A \mid B_i) \mathbb{P}r(B_i)$$

- Théorème de Bayes :

$$\mathbb{P}r(B_i \mid A) = \frac{\mathbb{P}r(A \mid B_i) \mathbb{P}r(B_i)}{\sum_{j=1}^n \mathbb{P}r(A \mid B_j) \mathbb{P}r(B_j)}$$

Des chiens et des chats



- Supposons que l'on a les probabilités conditionnelles :

$$\mathbb{P}r(\text{petites oreilles} \mid \text{chien}) = 0.1, \quad \mathbb{P}r(\text{grandes oreilles} \mid \text{chien}) = 0.9$$

$$\mathbb{P}r(\text{petites oreilles} \mid \text{chat}) = 0.8, \quad \mathbb{P}r(\text{grandes oreilles} \mid \text{chat}) = 0.2$$

- On considère un animal avec de grandes oreilles. Peut-on dire que c'est un chat ou c'est un chien ?

Des chiens et des chats



- Supposons que l'on a les probabilités conditionnelles :

$$\mathbb{P}r(\text{petites oreilles} \mid \text{chien}) = 0.1, \quad \mathbb{P}r(\text{grandes oreilles} \mid \text{chien}) = 0.9$$

$$\mathbb{P}r(\text{petites oreilles} \mid \text{chat}) = 0.8, \quad \mathbb{P}r(\text{grandes oreilles} \mid \text{chat}) = 0.2$$

- On considère un animal avec de grandes oreilles. Peut-on dire que c'est un chat ou c'est un chien ?
Une réponse : chien car

$$\mathbb{P}r(\text{grandes oreilles} \mid \text{chien}) = 0.9 > 0.2 = \mathbb{P}r(\text{grandes oreilles} \mid \text{chat})$$

- $\Rightarrow$ Nous choisissons l'événement de plus grande probabilité, c'est-à-dire l'événement de **vraisemblance maximale**.

Fonction de vraisemblance

Un exemple : classification des poissons

- ▶ Supposons que :
 - ▶ la longueur l des saumons suit une distribution $\mathcal{N}(5; 1)$,
 - ▶ la longueur l des bars suit une distribution $\mathcal{N}(10; 4)$,
- ▶ Les densités conditionnelles des classes sont alors :

$$f(l \mid \text{saumon}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(l-5)^2}{2}} \text{ et } f(l \mid \text{bar}) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(l-10)^2}{2 \cdot 4}}$$

Fonction de vraisemblance

Un exemple : classification des poissons

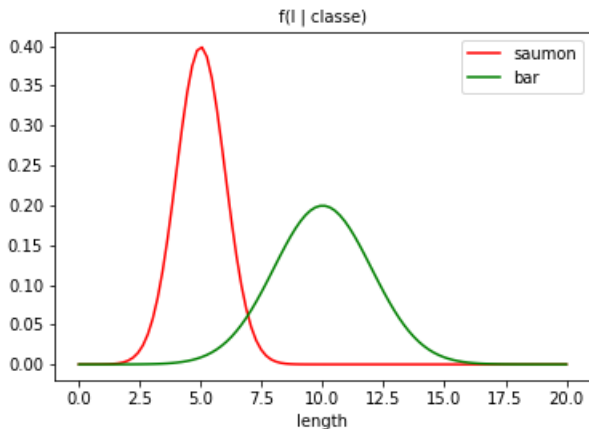
- Les densités conditionnelles des classes sont alors :

$$f(l \mid \text{saumon}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(l-5)^2}{2}} \text{ et } f(l \mid \text{bar}) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(l-10)^2}{2 \cdot 4}}$$

- Si on fixe la longueur l et qu'on fait varier la classe, on obtient la fonction de vraisemblance :

$$f(l \mid \text{classe}) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{(l-5)^2}{2}} & \text{si classe} = \text{saumon} \\ \frac{1}{2\sqrt{2\pi}} e^{-\frac{(l-10)^2}{8}} & \text{si classe} = \text{bar} \end{cases}$$

Vraisemblance vs densité conditionnelle des classes



Question : On a un poisson de longueur 6, où le classer ?

Classifieur ML (Maximum likelihood)

- ▶ On voudrait choisir "saumon" si

$$\mathbb{P}r(I = 7 \mid \text{saumon}) > \mathbb{P}r(I = 7 \mid \text{bar})$$

- ▶ Or I est une v.a. continue, donc

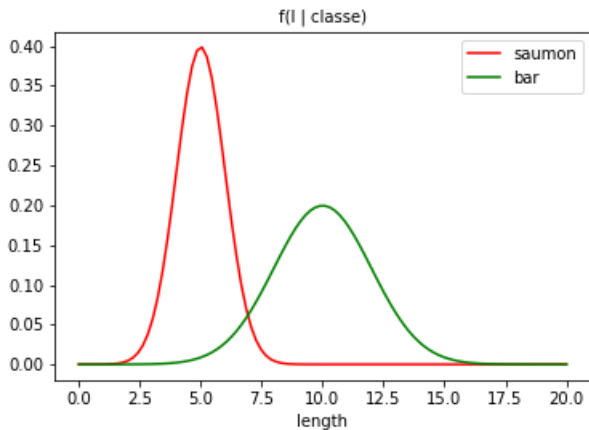
$$\mathbb{P}r(I = 7 \mid \text{saumon}) = \mathbb{P}r(I = 7 \mid \text{bar}) = 0$$

- ▶ Au lieu de ça, on choisit la classe qui maximise la vraisemblance :

$$f(I \mid \text{saumon}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(I-5)^2}{2}}; f(I \mid \text{bar}) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(I-10)^2}{2 \cdot 4}}$$

- ▶ **Classifieur ML** : pour une observation I ,
Si $f(I \mid \text{saumon}) > f(I \mid \text{bar})$ alors, on classe l'instance en "saumon", sinon, on la classe en "bar".

Frontière de décision



Classification apriori

- ▶ On ne dispose pas des données mais d'une connaissance **apriori**,
- ▶ On va exploiter une connaissance comme : " dans la mer, il y a deux fois plus de saumons que de bars"
- ▶ On en déduit une distribution apriori :

$$\mathbb{P}r(Saumon) = \frac{2}{3} \text{ et } \mathbb{P}r(Bar) = \frac{1}{3}.$$

- ▶ Retour à la question de toute à l'heure : *Comment classer un poisson de longueur $l = 6$?*
Intuitivement, on a envie de prendre en compte et le ML et les probabilités apriori ... Comment faire ?

Classifieur MAP (Maximum a posteriori)

Règle de décision de Bayes

1. on dispose de la fonction de vraisemblance :

$$f(I \mid \text{classe}) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{(I-5)^2}{2}} & \text{si classe} = \text{saumon} \\ \frac{1}{2\sqrt{2\pi}} e^{-\frac{(I-10)^2}{8}} & \text{si classe} = \text{bar} \end{cases}$$

2. on a les probabilités apriori :

$$\mathbb{Pr}(\text{Saumon}) = \frac{2}{3} \text{ et } \mathbb{Pr}(\text{Bar}) = \frac{1}{3}.$$

3. Règle de décision :

$$\text{classe} = \begin{cases} \text{Saumon} & \text{si } \mathbb{Pr}(\text{Saumon} \mid I) > \mathbb{Pr}(\text{Bar} \mid I) \\ \text{Bar} & \text{sinon} \end{cases}$$

Classifieur MAP (Maximum a posteriori)

Probabilités a posteriori

- ▶ $\mathbb{P}r(Saumon \mid I)$ et $\mathbb{P}r(Bar \mid I)$ sont les distributions a posteriori (Pourquoi ?)
- ▶ Comment les calculer ?

Classifieur MAP (Maximum a posteriori)

Probabilités a posteriori

- ▶ $\mathbb{P}r(Saumon \mid I)$ et $\mathbb{P}r(Bar \mid I)$ sont les distributions a posteriori (Pourquoi ?)
- ▶ Comment les calculer ?
- ▶ Théorème de Bayes (voir le tableau)

Et pour une dimension supérieure ?

Quid de la classification avec plusieurs "features" ?

Un Exemple

S(ex)	H(eight) (m)	W(eight) (kg)	F(oot) size (cm)
M	1.82	82	30
M	1.80	86	28
M	1.70	77	30
M	1.80	75	25
F	1.52	45	15
F	1.65	68	20
F	1.68	59	18
F	1.75	68	23

Un Exemple

S(ex)	H(eight) (m)	W(eight) (kg)	F(oot) size (cm)
M	1.82	82	30
M	1.80	86	28
M	1.70	77	30
M	1.80	75	25
F	1.52	45	15
F	1.65	68	20
F	1.68	59	18
F	1.75	68	23

- ▶ Question : Est-ce que (1.81, 59, 21) est M ou F ?
- ▶ $\Rightarrow$ Voir le tableau.

$$\mathbb{P}r(Y | X_1, \dots, X_n) = \frac{1}{Z} \times \mathbb{P}r(Y) \times \prod_{i=1}^n \mathbb{P}r(X_i | Y)$$

- ▶ X_i est continue $\rightarrow X_i | Y = y \sim \mathcal{N}(\mu_{iy}, \sigma_{iy})$
- ▶ X_i est binaire, $\rightarrow X_i | Y = y \sim \mathcal{B}(p_{iy})$

L'algorithme

- ▶ Entraînement :
Pour toutes les valeurs possibles de Y et de X_i , calculer $\hat{\mathbb{P}}_r(Y = y)$ et $\hat{\mathbb{P}}_r(X_i = x_i \mid Y = y)$
- ▶ Prédiction : étant donné $(x_1, x_2, \dots, x_n)$, retourner y qui maximise $\hat{\mathbb{P}}_r(Y = y) \hat{\mathbb{P}}_r(X_i = x_i \mid Y = y)$

Et en pratique

Voir le notebook