

Apprentissage automatique

Lecture 2: Bayes naïf

Département Génie Informatique, FST de Tanger

M. AIT KBIR

Plan

- Introduction à la science des données
- Notions d'apprentissage statistique
- Bayes Naïf
- Régression logistique
- Arbres de décision
- K-means
- ...
- Labs (NoteBook Jupyter)
- 2 Devoir
- 1 CC

2023-2024

M. AIT KBIR (MST IASD-S1)

2

Loi de Bayes

Formules des probabilités conditionnelles:

$$p(A|B) = \frac{P(A \cap B)}{P(B)}, \quad p(B|A) = \frac{P(A \cap B)}{P(A)}$$

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

$$p(A, B|C) = p(A|B, C)P(B|C)$$

Formule des probabilités totales:

$w_i, i \in \{1, 2, \dots, C\}$, un système complet d'événements, avec $p(w_i) \neq 0, \forall i$

$$p(X) = \sum_{i=1}^C p(X \cap w_i) = \sum_{i=1}^C p(X|w_i)p(w_i)$$

2023-2024

M. AIT KBIR (MST IASD-S1)

3

Loi de Bayes: Exemple

Sachant que le pourcentage de trouver un restaurant classé dans une ville est de 60% et que le pourcentage des bons repas sont 85% et 50% respectivement pour un restaurant classé et un autre non classé.

Question:

Un touriste choisi au hasard un restaurant pour prendre un repas, quel est la probabilité qu'il prend un bon repas ?.

2023-2024

M. AIT KBIR (MST IASD-S1)

4

Loi de Bayes: Exemple

Une compagnie d'assurance constate que les jeunes conducteurs représentent 30% de sa clientèle et que les conducteurs qui ont eu un accident en 3 ans se présentent comme suite:

- Les jeunes avec un pourcentage d'accident de 40%
- les autres clients avec un pourcentage d'accident de 20%.

Question:

Quelle est la probabilité d'avoir un accident pour un client quelconque ?

2023-2024

M. AIT KBIR (MST IASD-S1)

5

Loi de Bayes : Exemple

Un patient va voir un médecin pour effectuer le test de dépistage d'une maladie. Le médecin effectue un test avec 99 % de fiabilité. Le médecin sait que seulement 1% de la population est malade.

Question: :

Si le test est positif pour un patient, quelles sont les chances que le patient soit malade ?

2023-2024

M. AIT KBIR (MST IASD-S1)

7

Loi de Bayes

Utiliser la loi de Bayes pour déduire les variables d'un modèle inconnu à partir des données observées.

Une observation représentée par le vector X :

$$(x_1, x_2, \dots, x_N)$$

qu'on veut assigner à une des C classes :

$$\{w_1, w_2, \dots, w_C\}$$

Vraisemblance

A priori

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

A posteriori

2023-2024

M. AIT KBIR (MST IASD-S1)

6

Loi de Bayes : Exemple

Un laboratoire commercialise un test médical le test est positif chez 90% des personnes malades, le test est négatif chez 99% des personnes saines, la maladie touche 0,5% de la population.

Question:

Une personne passe le test et le résultat est positif. Quelle est la probabilité qu'elle soit malade?

0,311

2023-2024

M. AIT KBIR (MST IASD-S1)

8

Apprentissage automatique: Naïf Bayes

Un exemple est affecté à la classe qui maximise:

$$p(w_j|X) > p(w_k|X), \quad k = \{1, 2, \dots, C\}; k \neq j$$

La probabilité a posteriori peut être exprimée en terme probabilité a priori et la probabilité conditionnelle de la loi de x connaissant sa classe $p(X/w_j)$ en utilisant le théorème de Bayes.

$$p(w_j|X) = \frac{p(X|w_j)p(w_j)}{p(X)} = \frac{p(X|w_j)p(w_j)}{\sum_{i=1}^C p(X|w_i)p(w_i)}$$

En supposant naïvement l'indépendance des variables:

$$p(X|w_j) = \prod_{i=1}^N p(x_i|w_j)$$

2023-2024

M. AIT KBIR (MST IASD-S1)

9

Apprentissage automatique, exemple: Données de la météo/Match de tennis

	Outlook		Temperature			Humidity		Windy		Play			
	yes	no	yes	no		yes	no	yes	no	yes	no		
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

Outlook	Temperature		Humidity		Windy		Play
sunny	cool		high		true		?

Vraisemblance de yes = $2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$

Vraisemblance de no = $3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$

Probabilité(yes) = $0.0053 / (0.0053 + 0.0206) = 0.205$

Probabilité(no) = $0.0206 / (0.0053 + 0.0206) = 0.795$

2023-2024

M. AIT KBIR (MST IASD-S1)

11

Apprentissage automatique, exemple: Données de la météo/Match de Tennis

Outlook: sunny, overcast, or rainy;

Temperature : hot, mild, or cool;

Humidity : high or normal;

Windy can be true or false.

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

If outlook = sunny and humidity = high then play = no
 If outlook = rainy and windy = true then play = no
 If outlook = overcast then play = yes
 If humidity = normal then play = yes
 If none of the above then play = yes

2023-2024

M. AIT KBIR (MST IASD-S1)

10

Apprentissage automatique, exemple: Données de la météo/Match de Tennis

Outlook	Temperature	Humidity	Windy	Play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

Avec des attributs à valeurs numériques continues, la première règle pourrait prendre la forme suivante:

If outlook = sunny and humidity > 83 then play = no

2023-2024

M. AIT KBIR (MST IASD-S1)

12

Apprentissage automatique:

Les attributs numériques sont traités en considérant qu'ils suivent une distribution gaussienne:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

	Outlook		Temperature		Humidity		Windy		Play	
	yes	no	yes	no	yes	no	yes	no	yes	no
sunny	2	3	83	85	86	85	false	6	2	9
overcast	4	0	70	80	96	90	true	3	3	5
rainy	3	2	68	65	80	70				
			64	72	65	95				
			69	71	70	91				
			75	80	70					
			75	70	70					
			72	90						
			81	75						

Outlook	Temperature		Humidity		Windy	Play
sunny	66		90		true	?

$$f(\text{temperature} = 66 | \text{yes}) = \frac{1}{\sqrt{2\pi} \cdot 6.2} e^{-\frac{(66-73)^2}{2 \cdot 6.2^2}} = 0.0340$$

$$f(\text{humidity} = 90 | \text{yes}) = 0.0221$$

$$\text{likelihood of yes} = 2/9 \times 0.0340 \times 0.0221 \times 3/9 \times 9/14 = 0.000036$$

$$\text{Probability of yes} = \frac{0.000036}{0.000036 + 0.000108} = 0.25$$

$$\text{likelihood of no} = 3/5 \times 0.0221 \times 0.0381 \times 3/5 \times 5/14 = 0.000108$$

$$\text{Probability of no} = \frac{0.000108}{0.000036 + 0.000108} = 0.75$$

2023-2024

M. AIT KBIR (MST IASD-S1)

13

Bayes naïf : modèle de Bernoulli

Utilisé avec des attributs binaires. C'est un modèle très utilisé pour la classification des documents, soit avec un problème à deux classes (classer les messages texte en tant que spam ou non spam) ou un problème avec plusieurs classes (classer un document par domaine : Informatique, Economie, Philosophie ...).

Chaque attribut est associé à un mot de l'ensemble du vocabulaire pris en considération. Pour un document, à l'attribut x_i est associé la valeur 1 (si le mot apparaît dans le document) ou 0 (sinon).

$$p(X|w_j) = \prod_{i=1}^N p(x_i|w_j) = \prod_{i=1}^N \theta_{ij}^{x_i} (1 - \theta_{ij})^{(1-x_i)}$$

$$\theta_{ij} = \frac{n_{ij}}{n_j} = \frac{\text{Nombre des documents de la classe } w_j \text{ qui contiennent le } i\text{ème mot}}{\text{Nombre des documents de la classe } w_j}$$

2023-2024

M. AIT KBIR (MST IASD-S1)

15

Bayes naïf: modèle gaussien

Utilisé avec des attributs(caractéristiques) continus.

En supposant que les densités de probabilité des attributs suivent une distribution gaussienne, le modèle de Bayes naïf gaussien peut être écrit :

$$p(X|w_j) = \prod_{i=1}^N p(x_i|w_j) = \prod_{i=1}^N \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}}$$

2023-2024

M. AIT KBIR (MST IASD-S1)

14

Bayes naïf : modèle multinomial

Utilisé avec les attributs qui portent sur le comptage des données. Dans le cas de la classification des documents, on peut caractériser un document texte par la fréquence d'apparition d'un mot au lieu d'utiliser une valeur binaire. En pratique, la fréquence des termes est souvent normalisée en divisant la fréquence brute des mots par le nombre de mots du document.

$$p(X|w_j) = \prod_{i=1}^N p(x_i|w_j) \propto \prod_{i=1}^N \theta_{ij}^{x_i}$$

$$\theta_{ij} = \frac{n_{ij}}{n_j} = \frac{\text{Somme des fréquences du } i\text{ème mot de tous les documents de la classe } w_j}{\text{Somme de toutes les fréquences des documents de la classe } w_j}$$

2023-2024

M. AIT KBIR (MST IASD-S1)

16

Bayes naïf

Des probabilités estimées nulles de $p(x_i|w_j)$ auraient pour conséquence rendre obsolète le calcul de la probabilité conditionnelle $p(X|w_j)$.

En pratique, on procède souvent par le lissage des estimations sur les petits effectifs.

$$\theta_{ij} = \frac{n_{ij} + \alpha}{n_j + \alpha N}$$

α : paramètre N: nombre d'attributs

Estimateur Laplacien des probabilités: $\theta_{ij} = \frac{n_{ij} + 1}{n_j + 2}$

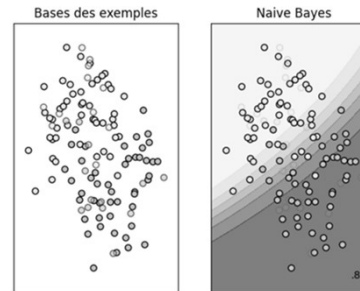
2023-2024

M. AIT KBIR (MST IASD-S1)

17

Bayes naïf

Illustrer la nature des bordures de décision réalisées par le classificateur NB. Les graphiques montrent les points d'entraînement en couleurs unies et les points de test semi-transparents (Voir le notebook NB Borders (2D).pdf).



2023-2024

M. AIT KBIR (MST IASD-S1)

19

Apprentissage automatique: Bayes naïf

- Avec un grand nombre de caractéristiques lors de la mise en œuvre de Bayes naïf, le produit des probabilités conditionnelles estimées peut provoquer des débordements de capacités, d'où la nécessité du passage par le logarithme des probabilités.
- Il est facile de travailler dans le domaine logarithmique en maximisant:

$$\log(p(w|X)) = \log(p(X|w)) + \log(p(w)) - \log(p(X))$$

- En supposant l'indépendance des variables et en éliminant le terme de normalisation $p(X)$, indépendant de w , cela revient à maximiser:

$$\log \left(\prod_{i=1}^N p(x_i|w) \right) + \log(p(w)) = \sum_{i=1}^N \log(p(x_i|w)) + \log(p(w))$$

2023-2024

M. AIT KBIR (MST IASD-S1)

18

Bayes naïf

Voir le chapitre 4 "Classifying with probability theory : naïve Bayes " du Livre "Machine learning in action".

Peter Harrington.

2023-2024

M. AIT KBIR (MST IASD-S1)

20