L'algorithme Naïve Bayes

(Apprentissage supervisé)

1. Introduction

Naïve Bayes Classifier est un algorithme populaire en Machine Learning. C'est un algorithme du *Supervised Learning* utilisé pour la classification. Il est particulièrement utilisé pour les problématiques de classification de texte (comme la classification des messages : SPAM).

Naïve Bayes permet de classifier un ensemble d'observations selon des règles déterminées par l'algorithme lui-même. Cet outil de classification doit dans un premier temps être entrainé sur un jeu de données d'apprentissage qui montre la classe attendue en fonction des entrées. Pendant la phase d'apprentissage, l'algorithme élabore ses règles de classification sur ce jeu de donnée, pour les appliquer dans un second temps à la classification d'un jeu de données de prédiction.

Le classificateur bayésien naïf implique que les classes du jeu de données d'apprentissage soient connues et fournit, d'où le caractère supervisé de l'outil.

2. Théorème de Bayes

Le classifier naïve Bayes se base sur le théorème de Bayes. Ce théorème est fondé sur les probabilités conditionnelles (Quelle est la probabilité qu'un événement se produise sachant qu'un autre événement s'est déjà produit).

Soit l'exemple suivant : supposons qu'on ait une classe de lycéens. Soit A et B les deux événements suivants :

- L'événement A : l'élève est une fille.
- L'événement B : l'élève pratique l'allemand.

Quelle est la probabilité qu'on choisisse au hasard une fille pratiquant l'allemand ?

Le théorème de Bayes permet de calculer ce genre de probabilité.

Notons P la probabilité d'un événement.

$$P(eleve\ est\ une\ fille\ ET\ eleve\ pratique\ allemand) = P(eleve\ est\ une\ fille) * P(eleve\ pratique\ allemand \mid eleve\ est\ une\ fille)$$

Ce qui est équivalent à :

$$P(eleve\ est\ une\ fille\ ET\ eleve\ pratique\ allemand) = P(eleve\ pratique\ allemand) * P(eleve\ est\ une \mid eleve\ pratique\ allemand)$$

Le terme P(A\B) : la probabilité que l'événement A se réalise sachant que l'événement B s'est déjà réalisé **(**A : **l'évidence**, B : **Outcome).**

Pour généraliser :

$$P(A \setminus B) = \frac{\mathrm{P(B\backslash A) * P(A)}}{\mathrm{P(B)}}$$

A, B : évènements
P(A\B) : probabilité de A sachant que B est vrai
P(B\A) : probabilité de B sachant que A est vrai
P(A), P(B): probabilités indépendantes de A et B

Nous allons maintenant reprendre l'exemple des élèves du lycée qui pratiquent l'allemand:

	Filles	**Garçons**	**Total**
Allemand	10	7	17
Autre langue	4	9	13
Total	14	16	30

Calculons la probabilité suivante : Quelle est la probabilité qu'on tire au hasard un élève parlant Allemand sachant qu'elle est une fille ?

Selon la formule de Bayes on a :

On considère Ω : l'ensemble des lycéens de notre exemple donc cardinal (Ω) = 30

Rappel : Cardinal d'un ensemble = nombre d'éléments de cet ensemble

$$P(Allemand \mid Fille) = P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B) * P(B)}{P(A)}$$

- $P(A)$ est la probabilité de prendre au hasard une fille de la population des élèves de la classe. On appelle $P(A)$ **la probabilité antérieure (*prior probability*)**.

$$P(A) = \frac{cardinal(A)}{cardinal(\Omega)} = \frac{14}{30} \approx 0.4666$$

$$P(B \cap A) = \frac{cardinal(B \cap A)}{cardinal(\Omega)} = \frac{10}{30} \approx 0.3333$$

Ce qui donne :

$$P(B \mid A) = \frac{\frac{10}{30}}{\frac{14}{30}} \approx \frac{0.3333}{0.4666} \approx 0.7143$$

Pour cet exemple nous avons appliqué le *théorème de Bayes* avec une seule variable prédictive (*Evidence) :* A savoir le sexe de l'élève (Fille). Dans les vraies applications du Naïve Bayes, on calcule le résultat (*Outcome)* en se basant sur **plusieurs variables et non sur une seule variable**. L'application du théorème de Bayes sur plusieurs variables rend le calcul plus complexe. Pour contourner cela, une approche consiste à **prendre en considération ces variables indépendamment les unes des autres**. Il s'agit d'une **hypothèse forte.** Généralement, les variables prédictives sont liées entre elles. Le terme "**naïve**" vient du fait qu'on suppose cette indépendance des variables.

Si on prend l'exemple de classification de mails en (spam et non spam), Naïve Bayes se basera sur la **fréquence d'occurrence des mots** dans le mail pour en définir la catégorie. Lors de sa classification, l'algorithme supposera que les mots du mail “apparaissent” indépendamment les uns des autres. Évidemment, d’un point de vue linguistique et sémantique, cette supposition est fausse !

3. Fonctionnement

Exercice d'application :

Soit un ensemble d'individus ayant les caractéristiques : **cheveux, taille, poids, crème solaire**, on désire classer ces individus selon la classe **Classe.**

N°	Cheveux	Taille	Poids	Crème Solaire	Classe
1	Blond	Moyenne	Léger	Non	Coup de soleil
2	Blond	Grande	Moyen	Oui	Bronzé
3	Brun	Petite	Moyen	Oui	Bronzé
4	Blond	Petite	Moyen	Non	Coup de soleil
5	Roux	Moyenne	Lourd	Non	Coup de soleil
6	Brun	Grande	Lourd	Non	Bronzé
7	Brun	Moyenne	Lourd	Non	Bronzé
8	Blond	Petite	Léger	Oui	Bronzé

1. Donner le modèle de décision déduit de cette base en utilisant la classification bayésienne.
2. Trouver les classes des exemples suivants :

N°	Cheveux	Taille	Poids	Crème Solaire
1	?	Petite	?	Oui
2	?	Grande	Moyen	?
3	Brun	?	?	Non
4	?	?	Lourd	?

Solution

1. Commençons par remplir ce tableau qui résumé les probabilités des classes par rapport aux attributs :

Attributs	**Valeurs**	**P(classe =Oui = Coup de soleil) = 3/8**	**P(classe =Non = Bronzé) = 5/8**
Cheveux	Blond = 4	2/3	2/5
	Brun = 3	**0/3**	3/3
	Roux = 1	1/3	**0/5**
Taille	Petite = 3	1/3	2/5
	Moyenne = 3	2/3	1/5
	Grande = 2	**0/3**	2/5
Poids	Leger = 2	1/3	1/5
	Moyen = 3	1/3	2/5
	Lourd = 3	1/3	2/5
Crème solaire	Oui = 3	**0/3**	3/5
	Non = 5	3/3	2/5

On remarque que notre tableau comporte des valeurs nulles, dans ce cas, on utilisera l'estimateur de Laplace.

Important : Lorsqu'on a un effectif égal à 0 (pour une classe donnée, et pour un attribut *a* donné) : on ajoute une valeur (par exemple 1) à chaque décompte de la table des effectifs (pour la classe considérée). Il faudra ensuite considérer qu'il y a k exemples de plus (k : nb de valeurs possibles *a*)

L'idée générale est :

- D'ajouter une valeur μ à chaque dénominateur pour l'attribut considéré *a* et la classe considérée.
- D'ajouter μ/k à l'effectif associé à chaque valeur de l'attribut considéré et classe considérée. Cette quantité, μ/k de l'attribut considéré, peut être vue comme une probabilité *a priori* de l'observation de chacune des valeurs de l'attribut.

Avec l'estimateur de Laplace, on obtiendra les valeurs suivantes:

		P(Classe = Oui = Coup de soleil) = 3/8	P(Classe= Non = Bronzé)=5/8
Cheveux	Blond = 4	2/3 = > (2+1)/(3+3)=3/6	2/5 => (2+1)/(5+3)=3/8
	Brun = 3	0/3 => (0+1)/(3+3) = 1/6	3/5 => (3+1)/(5+3) = 4/8
	Roux = 1	1/3 => (1+1)/(3+3) = 2/6	0/5 => (0+1)/(5+3) = 1/8
Taille	Petite = 3	1/3 => (1+1)/(3+3) = 2/6	2/5
	Moyenne =3	2/3 => (2+1)/(3+3) = 3/6	1/5
	Grande = 2	0/3 => (0+1)/(3+3) = 1/6	2/5
Poids	Léger = 2	1/3	1/5
	Moyen =3	1/3	2/5
	Lourd = 3	1/3	2/5
Crème Solaire	Oui = 3	0/3 =>(0+1)/(3+2) =1/5	3/5
	Non =5	3/3 => (3+1)/(3+2) = 4/5	2/5

Nombre des valeurs d'un attribut

Le tableau ci-dessus présente le modèle bayésien que nous allons utiliser pour faire les prédictions.

Il est à noter que l'étape de calcul de l'estimateur de Laplace intervient que si on a des probabilités nulles et concerne la plage de l'attribut en question.

2. Trouver les classes des exemples suivants (ces données n'appartiennent pas au dataset) :

N°	Cheveux	Taille	Poids	Crème Solaire
1	?	Petite	?	Oui
2	?	Grande	Moyen	?
3	Brun	?	?	Non
4	?	?	Lourd	?

Le théorème de bayes :

$$P(X\backslash Y) = \frac{\mathrm{P(Y\backslash X) * P(X)}}{\mathrm{P(Y)}}$$

X(? , petite, ?, Oui)

Pour faire la classification de X dans une classe ou une autre, on va calculer On va calculer :

P(Classe = Oui\X) = $\frac{P(X\backslash Classe=Oui)*P(Classe=Oui)}{P(X)}$

P(Classe = Non\X) = $\frac{P(X\backslash Classe=Non)*P(Classe=Non)}{P(X)}$

Et on choisira la plus grande probabilité.

- **Classe coup de soleil :**

(1) P(Classe = Oui\X) = $\frac{P(X\backslash Classe=Oui)*P(Classe=Oui)}{P(X)}$

(2) P(Classe = Non\X) = $\frac{P(X\backslash Classe=Non)*P(Classe=Non)}{P(X)}$

- **Classe bronzé :**

(1) P(Classe = Oui\X) = $\frac{P(X\backslash Classe=Oui)*P(Classe=Oui)}{P(X)}$

(2) P(Classe = Non\X) = $\frac{P(X\backslash Classe=Non)*P(Classe=Non)}{P(X)}$

(1) P(Classe = Oui\X) = $\frac{P(X\backslash Classe=Oui)*P(Classe=Oui)}{P(X)} =$

$$\frac{P(?\backslash classe=Oui)*P(petite\backslash classe=Oui)*P(?\backslash classe=Oui)*P(Oui\backslash classe=Oui)*P(Classe=Oui)}{P(?)*P(petite)*P(?)*P(Oui)}$$

On calcule toutes les probabilités et on considère la probabilité la plus grande.

N°	Cheveux	Taille	Poids	Crème Solaire	
1	?	Petite	?	Oui	← Bronzé
2	?	Grande	Moyen	?	← Bronzé
3	Brun	?	?	Non	← Bronzé
4	?	?	Lourd	?	← Bronzé

4. Application en python

Voir les notebooks associés (à vérifier, corriger et améliorer)

5. Exercice :

Une banque dispose des informations suivantes sur un ensemble de clients :

client	M	A	R	E	I
01	moyen	moyen	village	oui	oui
02	élevé	moyen	bourg	non	non
03	faible	âgé	bourg	non	non
04	faible	moyen	bourg	oui	oui
05	moyen	jeune	ville	oui	oui
06	élevé	âgé	ville	oui	non
07	moyen	âgé	ville	oui	non
08	faible	moyen	village	non	non

L'attribut client indique le numéro du client ; l'attribut M indique la moyenne des crédits sur le compte du client ; l'attribut A donne la tranche d'âge ; l'attribut R décrit la localité du client ; l'attribut E possède la valeur oui si le client possède un niveau d'études supérieur au bac ; l'attribut I (la classe) indique si le client exécute ses opérations de gestion de compte via Internet.

1. Donner le modèle de décision déduit de cette base en utilisant la classification bayésienne naïve.
2. Trouver les classes des exemples suivants :

client	M	A	R	E
01	?	âgé	?	oui
02	élevé	?	ville	?
03	faible	?	?	?
04	?	moyen	bourg	?

6. Conclusion

Aujourd'hui Naïve Bayes, est un algorithme renommé dont les applications peuvent être rencontrées dans de nombreux domaines. La classification naïve bayésienne obtient des résultats remarquables dans de nombreuses applications de la vie courante ce qui en fait un algorithme de choix parmi les outils du Machine Learning. Parmi ces atouts, son apprentissage rapide qui ne nécessite pas un gros volume de données, les calculs de probabilités ne sont pas très coûteux, aussi, son extrême rapidité d'exécution comparé à d'autres méthodes plus complexes.

Cependant, l'algorithme Naïve Bayes Classifier suppose l'indépendance des variables : C'est une hypothèse forte et qui est violée dans la majorité des cas réels.