

1. Spam ou pas spam

On dispose des probabilités suivantes :

$$\mathbb{P}(S) = 0,4, \quad \mathbb{P}(N) = 0,6$$

où S signifie «spam» et N signifie «non spam». Pour un message contenant le mot *promo*, on sait que :

$$\mathbb{P}(\text{promo} | S) = 0,8, \quad \mathbb{P}(\text{promo} | N) = 0,2.$$

Classifier un message qui contient le mot *promo*.

2. Avis positif ou négatif

On veut classer un avis client comme positif + ou négatif -.

On suppose :

$$\mathbb{P}(+) = 0,5, \quad \mathbb{P}(-) = 0,5.$$

L'avis contient le mot *excellent*. On sait que :

$$\mathbb{P}(\text{excellent} | +) = 0,7, \quad \mathbb{P}(\text{excellent} | -) = 0,1.$$

Classifier l'avis.

3. Naïve Bayes en deux mots.

On veut classer un message comme spam S ou non spam N.

On suppose :

$$\mathbb{P}(S) = 0,5, \quad \mathbb{P}(N) = 0,5.$$

Le message contient les mots *gratuit* et *urgent*. On connaît :

$$\mathbb{P}(\text{gratuit} | S) = 0,8, \quad \mathbb{P}(\text{urgent} | S) = 0,6,$$

$$\mathbb{P}(\text{gratuit} | N) = 0,2, \quad \mathbb{P}(\text{urgent} | N) = 0,1.$$

Classifier le message avec l'hypothèse naïve d'indépendance conditionnelle.

4. Sport ou politique.

On veut classer un court texte comme parlant de sport S ou de politique P.

On suppose :

$$\mathbb{P}(S) = 0,6, \quad \mathbb{P}(P) = 0,4.$$

Le texte contient le mot *match*. On connaît :

$$\mathbb{P}(\text{match} | S) = 0,7, \quad \mathbb{P}(\text{match} | P) = 0,1.$$

Classifier le texte.

5. Classification de messages

On veut classer un court message selon une étiquette y pouvant prendre trois valeurs :

$$y \in \{A, B, C\}.$$

On dispose des probabilités a priori :

$$\mathbb{P}(A) = 0,5, \quad \mathbb{P}(B) = 0,3, \quad \mathbb{P}(C) = 0,2.$$

Le message contient les trois mots :

$$x_1 = \text{“urgent”}, \quad x_2 = \text{“réunion”}, \quad x_3 = \text{“gratuit”}.$$

On connaît les probabilités conditionnelles suivantes :

$$\mathbb{P}(x_1 | A) = 0,6, \quad \mathbb{P}(x_2 | A) = 0,7, \quad \mathbb{P}(x_3 | A) = 0,1,$$

$$\mathbb{P}(x_1 | B) = 0,2, \quad \mathbb{P}(x_2 | B) = 0,8, \quad \mathbb{P}(x_3 | B) = 0,2,$$

$$\mathbb{P}(x_1 | C) = 0,7, \quad \mathbb{P}(x_2 | C) = 0,1, \quad \mathbb{P}(x_3 | C) = 0,9.$$

On suppose que les mots sont conditionnellement indépendants sachant la classe.

1°. Classifier le message avec la méthode naïve de Bayes.

2°. Quelle est l'expression du classifieur naïf de Bayes ?

6. Diagnostic de l'état d'une machine.

Dans un atelier de production, on souhaite diagnostiquer l'état d'une machine à partir de deux mesures continues :

- x_1 : niveau vibratoire RMS en mm/s,
- x_2 : échauffement au-dessus de la température ambiante, en °C.

L'étiquette y peut prendre trois valeurs :

$$y \in \{N, D, R\},$$

où :

- N : fonctionnement nominal,
- D : désalignement mécanique,
- R : roulement dégradé.

On suppose les probabilités *a priori* suivantes :

$$\mathbb{P}(N) = 0,70, \quad \mathbb{P}(D) = 0,20, \quad \mathbb{P}(R) = 0,10.$$

On utilise un modèle de Bayes naïf, gaussien. Conditionnellement à la classe y , les variables x_1 et x_2 sont supposées indépendantes, et suivent des lois normales. Les paramètres sont les suivants :

$$x_1 | N \sim \mathcal{N}(1,2, 0,3^2), \quad x_2 | N \sim \mathcal{N}(8, 3^2),$$

$$x_1 | D \sim \mathcal{N}(2,2, 0,4^2), \quad x_2 | D \sim \mathcal{N}(12, 4^2),$$

$$x_1 | R \sim \mathcal{N}(3,0, 0,5^2), \quad x_2 | R \sim \mathcal{N}(22, 5^2).$$

Lors d'un contrôle, on observe :

$$x_1 = 2,6, \quad x_2 = 18.$$

1°. Préciser la forme du classifieur bayésien naïf.

2°. Classifier l'état de la machine.

7. Classification de commentaires.

Un magasin de vente en ligne souhaite mettre en place un système de traitement automatique des commentaires laissés par ses consommateurs. Chaque commentaire est résumé par trois notes allant de 0 à 3 étoiles :

Variable	Signification
S	qualité du service client
L	qualité de la livraison
C	conformité du produit à sa description

Le magasin souhaite prédire si un commentaire doit être traité en priorité par le service après-vente. La classe à prédire est donc :

Classe	Signification
Oui	commentaire prioritaire ou réclamation
Non	commentaire non prioritaire

On dispose des données suivantes :

Id	S	L	C	Prioritaire?
1	★★★	★★★	★★★	Non
2	★★★	★★☆	★★★	Non
3	★★☆	★★★	★★☆	Non
4	★★☆	★★☆	★★★	Non
5	★★★	★☆☆	★★☆	Non
6	★★☆	★★★	★★★	Non
7	★☆☆	★☆☆	★★☆	Oui
8	☆☆☆	★☆☆	★☆☆	Oui
9	★☆☆	☆☆☆	★☆☆	Oui
10	☆☆☆	☆☆☆	★★☆	Oui
11	★★☆	☆☆☆	★☆☆	Oui
12	★☆☆	★★☆	☆☆☆	Oui

On souhaite classifier le commentaire suivant :

S	L	C
★☆☆	☆☆☆	★★☆

On utilisera un classifieur naïf de Bayes avec lissage de Laplace.

1°. Déterminer les probabilités a priori des classes Oui et Non.

2°. Calculer les probabilités conditionnelles nécessaires pour classifier le commentaire :

3°. Utiliser la règle de Bayes naïve pour prédire si le commentaire est prioritaire.

4°. Vérifier le résultat à l'aide d'un petit script Python.

8.