

Exercice 1

Considérons l'exemple très simple sur le football.

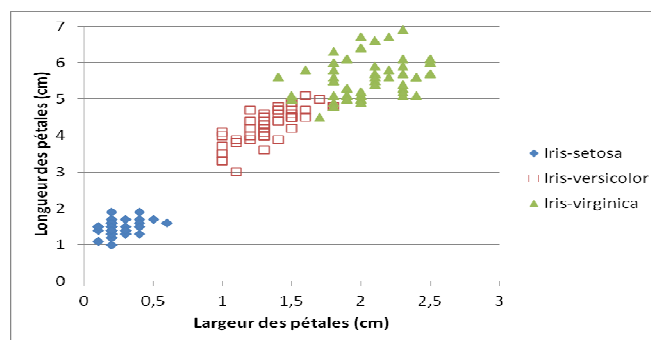
Match à domicile ?	Balance positive ?	Mauvaises conditions climatiques ?	Match précédent gagné ?	Match gagné
V	V	F	F	V
F	F	V	V	V
V	V	V	F	V
V	V	F	V	V
F	V	V	V	F
F	F	V	F	F
V	F	F	V	F
V	F	V	F	F

Quel est votre pronostic pour le match de samedi sachant qu'il est à domicile, que la balance est positive, les conditions climatiques seront bonnes et le match précédent a été gagné ?

- 1) Quelles sont les deux probabilités à calculer pour répondre à la question.
- 2) Transformer ces probabilités à l'aide de la formule de Bayes.
- 3) Appliquer l'hypothèse d'indépendance. Conclusion.

Exercice 2

Considérons le fameux jeu de données « Iris » : <http://archive.ics.uci.edu/ml/datasets/Iris>, représenté ci-dessous. Il s'agit de trois types d'iris caractérisés entre autres par la longueur et la largeur des pétales.



On ne considère que les deux variables concernant les pétales. Notons Y l'espèce d'iris, X_1 la longueur des pétales et X_2 la largeur.

- 1) Exprimer $P(Y=\text{setosa} | X_1, X_2)$ en fonction de $P(X_i | Y=\text{setosa})$. Quelle hypothèse faut-il faire ?
- 2) Calculez les moyennes et variances par classe (avec R).
- 3) En déduire la fonction de densité de chaque variable par classe.
- 4) Déterminer la forme du classifieur.
- 5) Déterminez la forme du classifieur et donnez la classe prédite pour les iris suivants :

	longueur des pétales	largeur des pétales
Iris 1	2	0,2
Iris 2	5	1,65
Iris 3	7	3

- 6) Vérifier graphiquement les hypothèses du modèle (indépendance + loi normale).

Exercice 3

Il s'agit maintenant d'utiliser le logiciel R pour construire le classifieur bayésien. Pour cela, nous allons utiliser les fonctions `naiveBayes` et `predict` du package `e1071`.

Installez et chargez le package `e1071`

1) Exemple guidé avec le jeu de données Iris.

L'instruction

```
naiveBayes(Y ~ X1+X2+...,data=Mydata)
```

permet de construire le classifieur bayésien où Y est la classe, X_1, X_2, \dots les variables explicatives du jeu de données `Mydata`.

```
modele=naiveBayes(Species ~ Petal.Length+Petal.Width,data=iris))
print(modele)
```

Le modèle est constitué des probabilités de chaque classe (A-priori probabilities) et des moyennes et écart-types des variables dans chaque classe (Conditional probabilities).

L'instruction

```
predict(model,newdata=Mynewdata)
```

permet de prédire la classe des nouveaux points de `Mynewdata` avec le modèle obtenu à l'aide de `naiveBayes`. Si on ajoute l'argument `type="raw"`, on obtient la probabilité de chacune des classes.

```
# Nouveaux iris (cf. exo 2)
newiris=matrix(0,3,2)
newiris[1,]=c(2,0.2)
newiris[2,]=c(5,1.65)
newiris[3,]=c(7,3)
newiris=as.data.frame(newiris)
names(newiris)=names(iris)[3:4] # "Petal.Length" et "Petal.Width"

predict(model,newdata=newiris,type="raw")
```

On retrouve bien les conclusions de l'exercice 2, à savoir une quasi-équiprobabilité pour l'iris 2 d'être dans deux classes.

2) Application sur un jeu de données simulé en 2D : `Test_Classif_Correl.txt`

```
tab=read.table("Test_Classif_Correl.txt",header=T)
summary(tab)
```

- Représentez le nuage de points. Pensez-vous que la méthode naïve Bayes est appropriée à ce jeu de données ?
- Construisez le modèle
- Calculez la matrice de confusion (sur la base d'apprentissage). Est-ce que les résultats étaient prévisibles ?

Exercice 4

Le jeu de données `Lenses.txt` constitué de 24 patients décrits par 5 variables qualitatives :

- Recommandation : (1) the patient should be fitted with hard contact lenses, (2) the patient should be fitted with soft contact lenses, (3) the patient should not be fitted with contact lenses.
- Age of the patient: (1) young, (2) pre-presbyopic.
- Spectacle prescription: (1) myope, (2) hypermetrope.

- Astigmatic: (1) no, (2) yes.
- Tear production rate: (1) reduced, (2) normal

Construire un modèle qui permet de prédire la recommandation en fonction des autres variables.

Que représentent les tableaux « Conditional probabilities » ?

Exercice 5 (autonomie)

Mettre en place la méthode Naïve Bayes sur le jeu de donnée Landsat ou Frogs.

- Vérifier la nature des variables et leur distribution
- Construire une base d'apprentissage et une base test
- Centrer et réduire les variables numériques sur la base d'apprentissage (pas indispensable pour Naive Bayes)
- Construire le modèle
- Calculer l'erreur sur la base d'apprentissage
- Transformer les variables de la base test et calculer l'erreur sur la base test.