

INTELLIGENCE ARTIFICIELLE

MODELE BAYESIEN NAIF

L'intelligence artificielle consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines de pouvoir choisir et réaliser la meilleure solution selon la situation. L'une des techniques est l'utilisation des probabilités dont le modèle Bayésien naïf fait partie.

■ Un algorithme d'apprentissage supervisé

Initialement, il faut une base de donnée afin d'apprendre et de pouvoir ensuite prédire des futures données.

Ce classificateur est basé sur le théorème de Bayes :

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

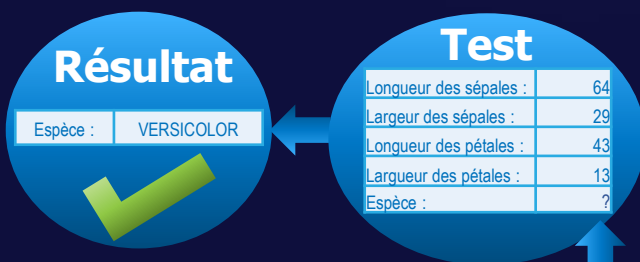
La proportion de chaque classe ainsi que les écart-types et moyennes de chaque caractéristique en fonction des classes permet de réaliser la prédiction.

La probabilité qu'une donnée appartienne à une classe déjà existante est :

$$P(\text{Classe}|\text{Critère}) = P(\text{Classe}) * \prod P(\text{Critère } i|\text{Classe})$$

Néanmoins, notre classificateur est dit naïf pour plusieurs raisons :

- Il dépend du nombre de données, plus il y en a, plus notre classificateur sera juste et précis.
- On fait l'hypothèse que les critères suivent une loi normale et sont indépendants. Ainsi il dépend de la loi normale et de l'indépendance des variables
- De plus, si une valeur n'a jamais été rentrée apprise initialement, alors la prédiction sera toujours nulle.



■ L'implantation de la base IRIS

Nous avons effectué l'implantation du classificateur Bayésien naïf en langage Python. La base de données utilisée fut la base IRIS répertoriant trois types de plantes en fonction de la taille de pétale, et celle des sépales.

On apprend 80% de la base pour pouvoir en tester 20% de celle-ci.

ESPECE	Longueur des sépales	Largeur des sépales	Longueur des pétales	Largeur des pétales
SETOSA	50	33	14	2
VIRGINICA	64	28	56	22
VERSICOLOR	65	28	46	15

■ Conclusion

Notre classificateur Bayésien naïf a un ratio de 86.7%.

Pour la même base de données, on voit que notre classificateur est en dessous de celui avec la méthode des K plus proches voisins KNN qui a un ratio de 97% ou encore du classificateur par régression linéaire qui a un ratio de 92%. Par conséquent, notre classificateur est loin d'être le meilleur pour cette base de données.