

Compréhension du dataset

- 3 types d'Iris :
 - Setosa (classe = 0)
 - Versicolor (classe = 1)
 - Virginica (classe = 2)

- 4 features :
 - Longueur sépale
 - Largeur sépale
 - Longueur pétale
 - Largeur pétale

```
[[5.1 3.5 1.4 0.2]
[4.9 3. 1.4 0.2]
[4.7 3.2 1.3 0.2]
[4.6 3.1 1.5 0.2]
[5. 3.6 1.4 0.2]]
```

Fonctionnement Machine Learning

Scikit-learn (sklearn): Librairie ML de python

Division en deux de la BDD : entraînement (train) - test

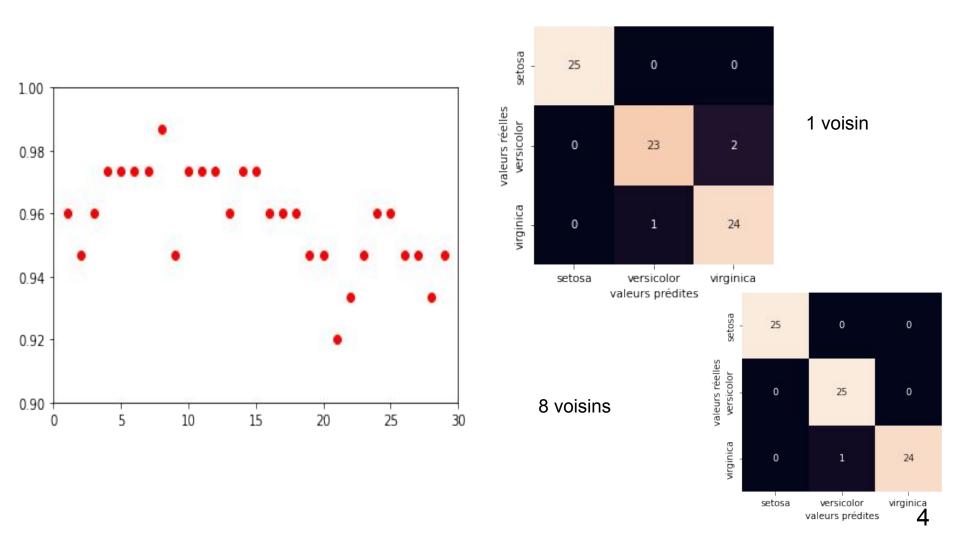
- => entrainement du modèle puis test (prédiction)
- => résultat:
 - Accuracy
 - Matrice de confusion

Modèle KNN

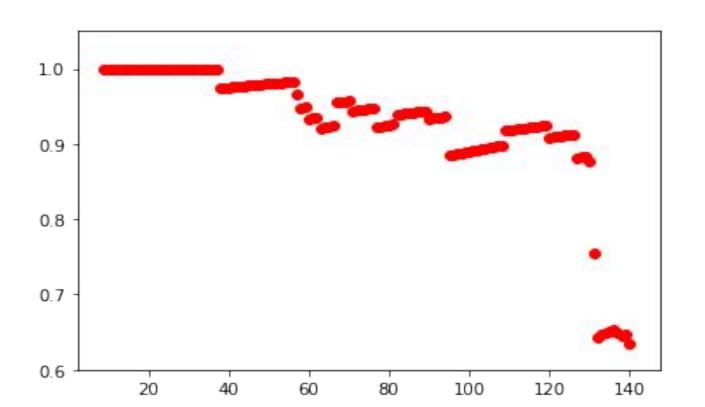
Supervised Machine Learning

- Modèle de classification
- K plus proches voisins:
 - Valeurs a "à placer"
 - Calcul de la distance entre a et chaque K voisins
 - Classification par comparaison





Train-Test ratio

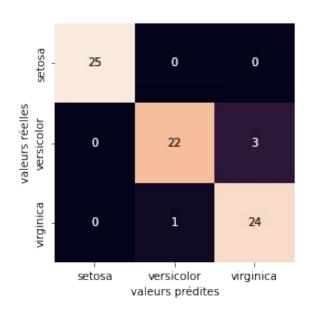


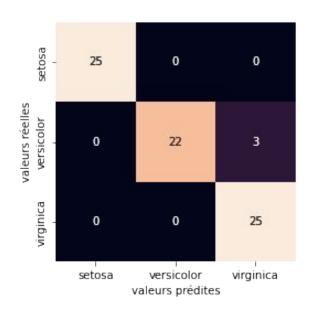
Modèle Naive Bayes

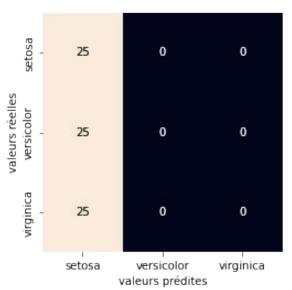
- Supervised Machine Learning
- Modèle de classification
- Utilisation du théorème de Bayes:

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

Gaussian Multinomial Bernoulli







Conclusion

- KNN: meilleur à 8 voisins (1 seule erreur)
- Naive Bayes : Meilleur résultat Multinomial
- Split train-test => impact sur les résultats.
- KNN mieux que Naive Bayes dans ce des de figure :
 - Plus de paramètres modifiables
 - Naive Bayes approximatifs
 - Naive Bayes peu gourmand en données et en temps

