

## Question 1:

150 entries

variable cible : Species

3 classes : setosa, versicolor, virginica

5 attributs :

- id : identifiant de la fleur ( pas utile )
- SepalLengthCm : la longueur de la sépale ( utile )
- SepalWidthCm : la largeur de la sépale ( utile )
- PetalLengthCm : la longueur de la pétale ( utile )
- PetalWidthCm : la largeur de la pétale ( utile )

répartition :

- Iris-setosa 50
- Iris-versicolor 50
- Iris-virginica 50

## Question 2 :

La première est à ignorer car on classe en fonction de l'id.

sepalwidth en fonction de sepallength : tout est mélangé, du coup pas utile

sinon les autres sont bien séparés

personnellement le petalwidth en fonction du petallength me paraît être le mieux.

## Question 3

a)

$\frac{2}{3}$  pour l'apprentissage, du coup 100 (150 échantillons en tout)

b)

'Iris-setosa' : 50

'Iris-versicolor' : 50

'Iris-virginica' : 0

on va s'entraîner sur les setosa et les versicolor et tester sur les virginica, ce qui est pas pertinent du tout

c)

'Iris-setosa' : 33

'Iris-versicolor' : 31

'Iris-virginica' : 36

c'est plus pertinent car c'est mieux réparti entre les 3 espèces

## Question 4

le classifieur se trompe 4 fois

```
for i in range(len(y_pred)):
    if y_pred[i] != y_test[i]:
        print(y_pred[i], y_test[i])
```

qui me rend

Iris-setosa Iris-versicolor  
Iris-virginica Iris-versicolor  
Iris-virginica Iris-versicolor  
Iris-versicolor Iris-virginica

## Question 5

Accuracy:  $0.92 = \frac{TP + TN}{TP + TN + FP + FN}$

## Question 6

erreur réelle 0.07999999999999996 (erreur entre la valeur de test et sa prédiction)  
erreur empirique 0.06999999999999995 (erreur entre la valeur d'entraînement et sa prédiction)

savoir si notre entraînement donne des résultats corrects. Si les 2 erreurs sont similaires et petites, notre modèle est bien.

## Question 7

[17 0 0]  
[ 1 16 2]  
[ 0 1 13]

	Prédites setosa	Prédites versicolor	Prédites virginica
Réelles setosa	setosa	setosa prédite en versicolor	setosa prédite en virginica
Réelles versicolor	versicolor prédite en setosa	versicolor	versicolor prédite en virginica
Réelles virginica	virginica prédite en setosa	virginica prédite en versicolor	virginica

## Question 8

- Retrouvez la valeur de l'accuracy à partir de cette matrice

$$\text{sommeDiagonale} / \text{total} = (16+17+13)/(17+16+1+2+1+13) = 0.92$$

- Quelle est la précision pour la classe 'Iris-virginica' ?

$$\text{TP}/(\text{TP}+\text{FP}) = 13/(13+2+0) = 0,8666667$$

- Quel est le rappel pour la classe 'Iris-virginica' ?

$$\text{TP}/(\text{TP}+\text{FN}) = 13/(13+1+0) = 0,9285714$$

- Quelle est la classe qui a le meilleur rappel ? Qu'est-ce que cela signifie ?

$$\text{Rappel}(\text{Virginica}) = 0,93$$

$$\text{Rappel}(\text{versicolor}) = 16/(16+2+1) = 0,84$$

$$\text{Rappel}(\text{setosa}) = 17/(17+0+0) = 1$$

La setosa a le meilleur rappel, cela veut dire que la plus grande proportion de setosa a été prédit en tant que setosa, ici elles le sont toutes car égale à 1

- Quelle est la classe qui a la plus faible précision ? Qu'est-ce que cela signifie ?

$$\text{Précision}(\text{Virginica}) = 0,87$$

$$\text{Précision}(\text{versicolor}) = 16/(16+1+0) = 0,94$$

$$\text{Précision}(\text{setosa}) = 17/(17+1+0) = 0,94$$

La virginica est la moins précise, elle a beaucoup de versicolor et setosa considéré comme virginica.

## Question 9