

LES IRIS (150 DONNÉES)

Classes :



Iris setosa



Iris versicolor



Iris virginica

Attributs :

1. Longueur des sépales
2. Largeur des sépales
3. Longueur des pétales
4. Largeur des pétales

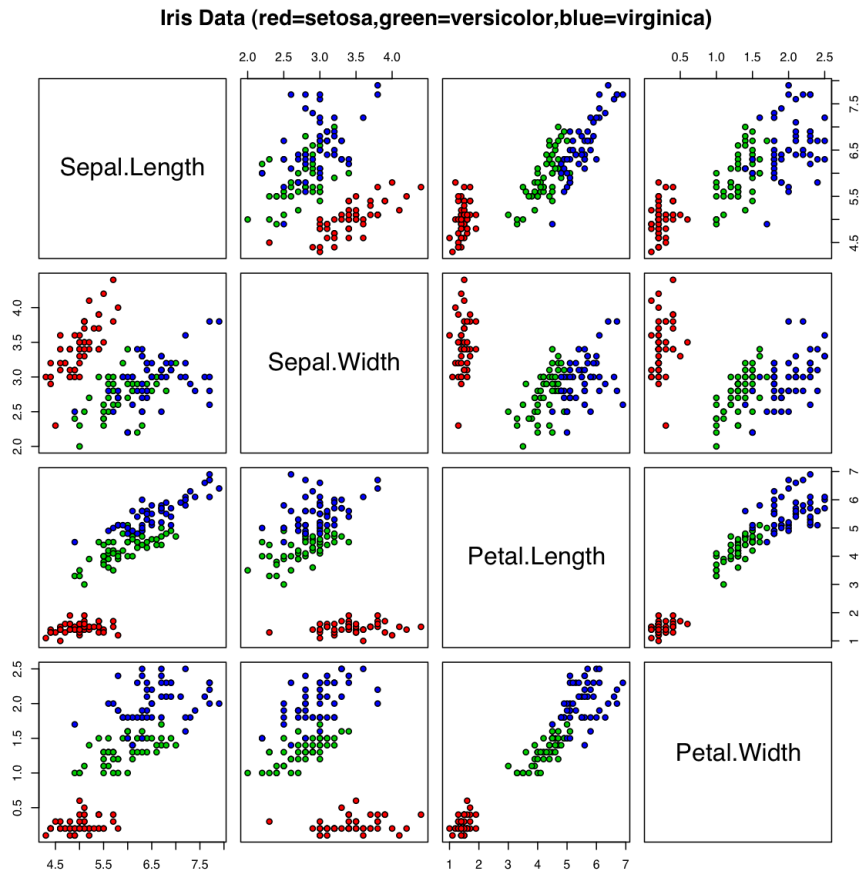
SÉPALE, subst. masc.

BOT. Chacun des éléments foliacés, généralement verts, dont la réunion compose le calice et supporte la corolle de la fleur.

PÉTALE, subst. masc.

A. – BOT. Chacun des éléments foliacés, généralement colorés, qui composent la corolle d'une fleur.

Répartition :



APPRENTISSAGE DU MEILLEUR K POUR LA CLASSIFICATION DES IRIS, À L'AIDE DE L'ALGORITHME DES K-PLUS PROCHES VOISINS (K-PPV).

Le programme devra :

1 - Prendre en entrée quatre fichiers :

- iris_learn_data
- iris_learn_label
- iris_test_data
- iris_test_label

Les deux fichiers data auront le même format, une donnée par ligne sous la forme :

```
5.0;3.3;1.4;0.2;  
7.0;3.2;4.7;1.4;  
6.5;3.2;5.1;2.0;
```

Les deux fichiers label auront le même nombre de lignes que le fichier data correspondant, chaque ligne indiquant la classe de la ligne correspondante dans le fichier data. Avec :

- 0 désigne la classe des iris setosa
- 1 désigne la classe des iris versicolor
- 2 désigne la classe des iris virginica

2 - Calculer la classe de chaque entrée du fichier iris_test_data, en appliquant les k-ppv.

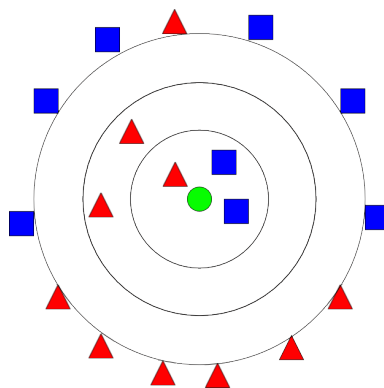
3 - Calculer le taux de bonne classification, grâce à iris_test_label

4 - Faire varier k, de 2 à 10 et trouver le meilleur taux de classification.

5 - Stocker le taux de bonne classification pour chaque k.

Rappels de l'algorithme des k-plus proches voisins :

Pour attribuer une classe à un nouvel élément x (ici, à une donnée de iris_test_data) on va calculer sa distance à chacun des points de l'ensemble d'apprentissage (ici, iris_learn_data) et affecter la classe majoritaire parmi les k points les plus proches de x (la classe d'un point est la valeur de la ligne correspondante dans iris_learn_label).



Les 2-ppv et 3-ppv, appliqués à la figure ci-dessus, affecteront la classe carré bleu au point vert. Les 5-ppv lui affecteront la classe triangle rouge.

Cas de points équidistants : si tous les points équidistants font partie des k -plus proches voisins, on les traitera comme des points distincts. Si leur nombre, additionné aux éventuels voisins encore plus proches, dépasse k , on en conservera suffisamment pour arriver à k voisins et on intégrera l'ensemble au calcul de la classe majoritaire. Dans la figure ci-dessus, les 6-ppv attribueront la classe triangle rouge au point vert, puisqu'il y a dix triangles rouges, pour huit dix, dans la zone considérée.