

## TD K-NN

- 1) Combien de lignes comporte le fichier iris.csv ?
- 2) Combien de lignes comportent des longueurs de pétales ?
- 3) Le fichier iris.csv comporte t-il des doublons ? Est-ce une erreur ? Expliquer votre réponse.
- 4) Un point vert correspond-il toujours à un unique iris ?
- 5) Indiquer quels sont les codages retenus dans le programme pour identifier les iris setosa, virginica et versicola.
- 6) Pourquoi fait-on commencer la boucle à 1 et non à 0 à la ligne 15 ?
- 7) Si un nouvel iris a un pétale de 0,8 cm de longueur et 0,5 de large et que l'on choisit  $k=3$ , quelles seront les couleurs de ses 3 plus proches voisins ? Quelle sera la variété qui lui sera attribuée ?
- 8) Si un nouvel iris a un pétale de 5cm de longueur et 1.6 de large et que l'on choisit  $k=7$ , quelles seront les couleurs de ses 7 plus proches voisins ? Quelle sera la variété qui lui sera attribuée ?
- 9) Compléter la fonction `espece_majoritaire(Tab,k)` dans le fichier python joint, afin qu'elle détermine le numéro de l'espèce majoritaire dans les  $k$  éléments dans Tab.

Le paramètre Tab est un tableau de n-uplets contenant des réels (les distances à la nouvelle fleur) classés par ordre **croissant** ainsi qu'un entier (indiquant l'espèce associée). Elle retournera le numéro de l'espèce majoritaire trouvée.

- 10) Si Tab =[(0.22360679774997877, 2), (0.22360679774997877, 2), (0.28284271247461906, 1), (0.28284271247461906, 1), (0.28284271247461906, 1), (0.30000000000000004, 1), (0.31622776601683766, 2)...]
  - a) Avec la valeur de Tab ci-dessus et  $k=3$ , quel sera le numéro renvoyé, par cette fonction ?
  - b) Avec la valeur de Tab ci-dessus et  $k=5$ , quel sera le numéro renvoyé, par cette fonction ?