

TP03 : Arbres de décision

- L'ensemble d'apprentissage est IRIS. Cet ensemble possède 3 classes $K = 3$ (3 espèces de la fleur d'Iris).



Iris setosa

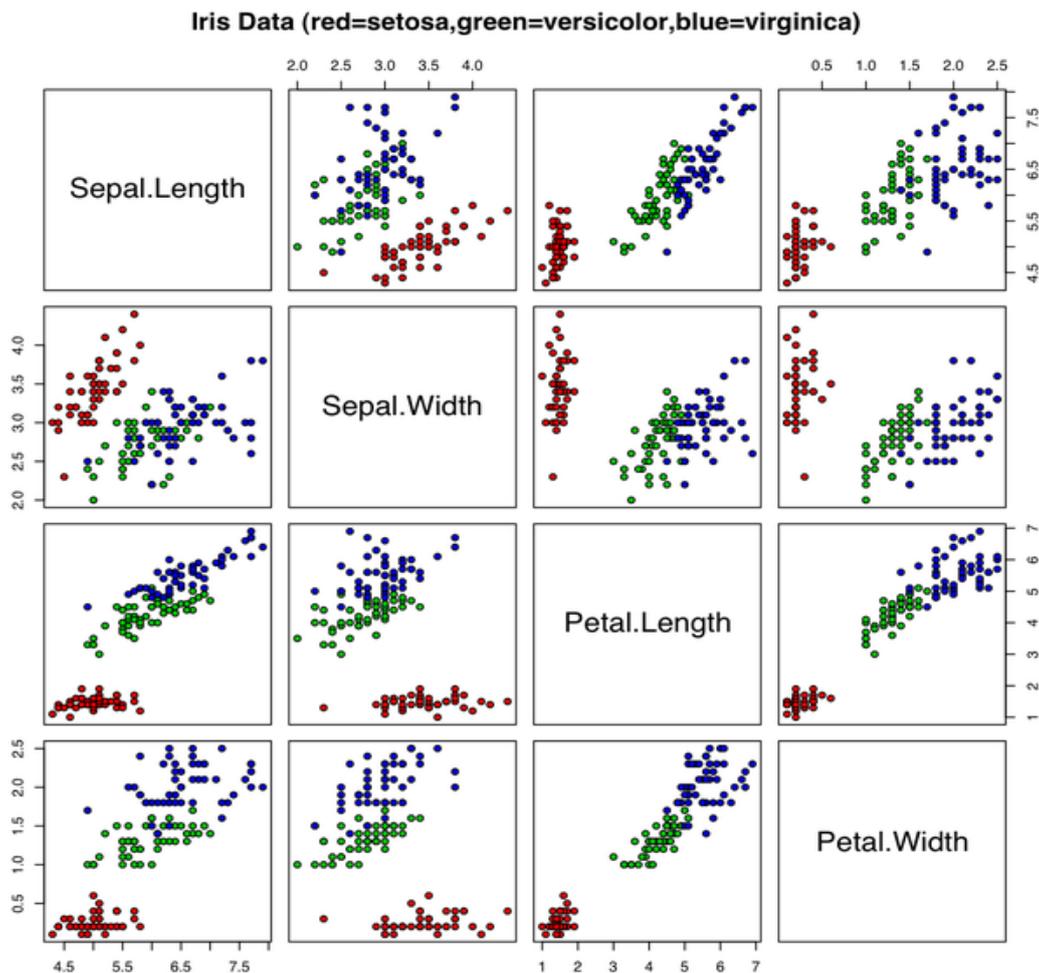


Iris versicolor



Iris virginica

- 50 échantillons de chacune des 3 espèces ont été collectés. ($N = 150$). On a mesuré la longueur et la largeur des sépales et pétales, respectivement.



1. Les données sont sauvegardées sous format Excel. Utiliser les fonctions Python pour lire le dataset à partir de `sklearn.datasets` ou à partir du fichier **Excel**.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

2. Séparer les données en: entrées (les attributs) et sorties (cibles). Dans ce fichier, les classes (qui sont le résultat attendu) sont dans la dernière colonne, et les autres caractéristiques (les entrées) sont dans les colonnes restantes.
3. Diviser l'ensemble de données en un ensemble d'apprentissage et un ensemble de test.
4. Générer un arbre de décision pour le problème cité en utilisant Python. Vous pouvez utiliser la classe **DecisionTreeClassifier** ou choisir une autre implémentation en citant la référence.
5. Valider le classificateur implémenté afin d'estimer la précision avec laquelle le classificateur ou le modèle peut prédire le type de fleurs.
6. Afficher et exporter l'arbre sous forme graphique et texte.

