

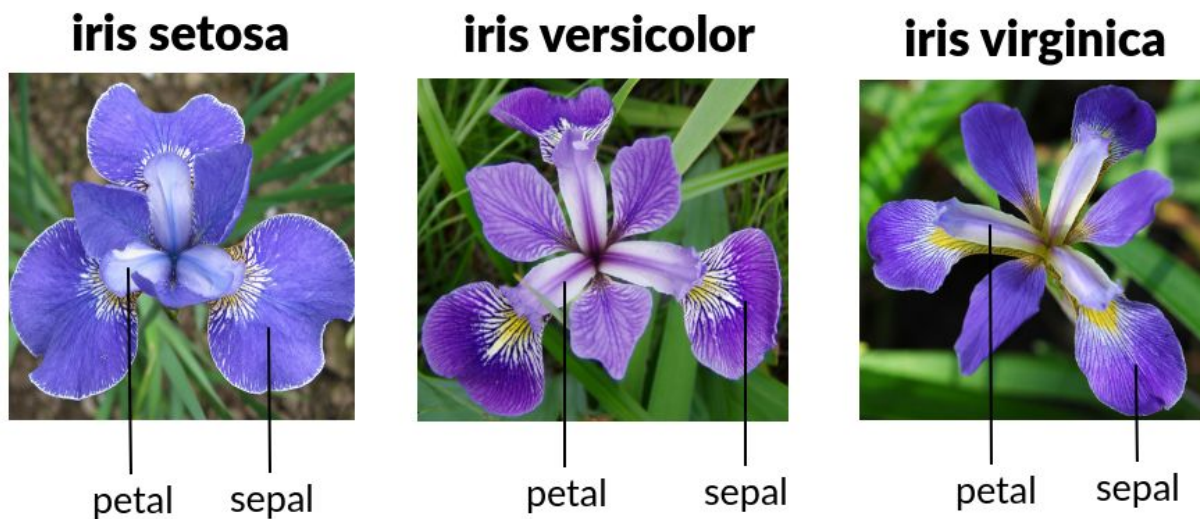
CSCU9YE - Artificial Intelligence

Lab 4: Supervised Machine Learning

For machine learning we will be using the Python library Scikit-learn, which implements a wide range of algorithms, and has wonderful documentation, examples and tutorials.

The dataset we will be using in the lab is provided by the Scikit-learn library. It is a well-known dataset named IRIS flowers. The data set consists of 50 samples from each of three species of Iris flowers: setosa, virginica and versicolor. For each flower, 4 features are stored: the length and the width of the sepals and petals, in centimeters.

Examples of Iris flowers, showing their petal and sepal.



Activities

Your tasks in this lab are the following:

- Browse the Scikit-learn documentation at: <http://scikit-learn.org/stable/documentation.html>
- Browse the structure and content of the IRIS dataset at: <https://gist.github.com/netj/8836201>
- Run the python script provided in Canvas/Units/Lab sheets: supervised.py. In this script the K Nearest Neighbour (KNN) classifier is used to model and classify the Iris dataset.
- Analyse and understand each line of code and the outputs/printouts of running the script

- The scatter plot created by running the script, visualises the first two attributes (columns 0 and 1). Create another plot that visualise another pair of attributes, for example columns 1 and 3.
- In the script the following example vector of attributes: [5, 2.9, 1, 0.2] is used for prediction, try with other example vectors.
- Using the Scikit-learn documentation for decision trees implement a decision tree to classify the Iris dataset:
 - Use the same training and testing set created in the supervised.py script for the KNN classifier
 - Compute Accuracy obtained with the decision tree
 - Compare the score obtained by the decision tree with the score obtained by the KNN classifier
 - How the accuracies of the two classification methods compare?
 - Optional: Try to visualise the obtained decision tree, using some available methods provided by Scikit-learn or other related libraries and examples.