Hello World of Machine Learning

The best small project to start with on a new tool is the classification of iris flowers (e.g. the iris dataset).

- Attributes are numeric so you have to figure out how to load and handle data.
- It is a classification problem, allowing you to practice with perhaps an easier type of supervised learning algorithm.
- It is a multi-class classification problem (multi-nominal) that may require some specialized handling.
- It only has 4 attributes and 150 rows, meaning it is small and easily fits into memory (and a screen or A4 page).
- All of the numeric attributes are in the same units and the same scale, not requiring any special scaling or transforms to get started.

To do

1. Installing the Python and SciPy platform.
2. Loading the dataset.
3. Summarizing the dataset.
    - Dimensions of the dataset.
    - Peek at the data itself.
    - Statistical summary of all attributes.
    - Breakdown of the data by the class variable.

4. Visualizing the dataset.
    - Univariate plots to better understand each attribute.
    - Multivariate plots to better understand the relationships between attributes.

5. Evaluating some algorithms.
    - Separate out a validation dataset.
    - Set-up the test harness to use 10-fold cross validation.
    - Build multiple different models to predict species from flower measurements
    - Select the best model.

        test 6 different algorithms:

        - Logistic Regression (LR)
        - Linear Discriminant Analysis (LDA)
        - K-Nearest Neighbors (KNN).
        - Classification and Regression Trees (CART).
        - Gaussian Naive Bayes (NB).
        - Support Vector Machines (SVM).

6. Making some predictions.