

CS5785 Modern Analytics

Homework 0

September 5, 2014

Fanxing Meng

Part 1

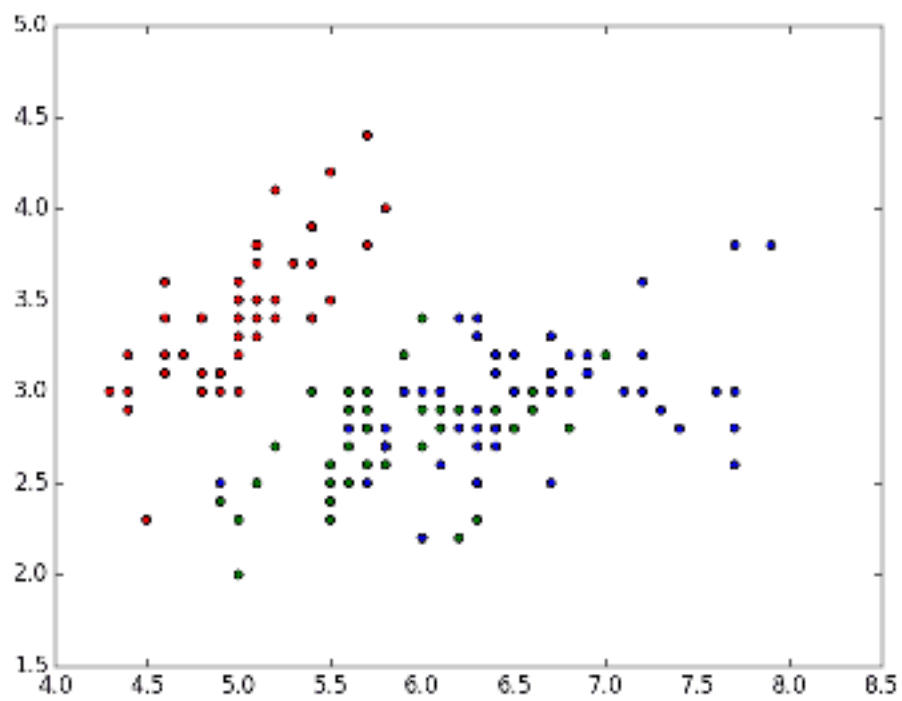
1. Did not have to find a teammate;
2. Signed up on Kaggle;
3. Installed Canopy using full installer;
4. Passed package import tests.

Part 2

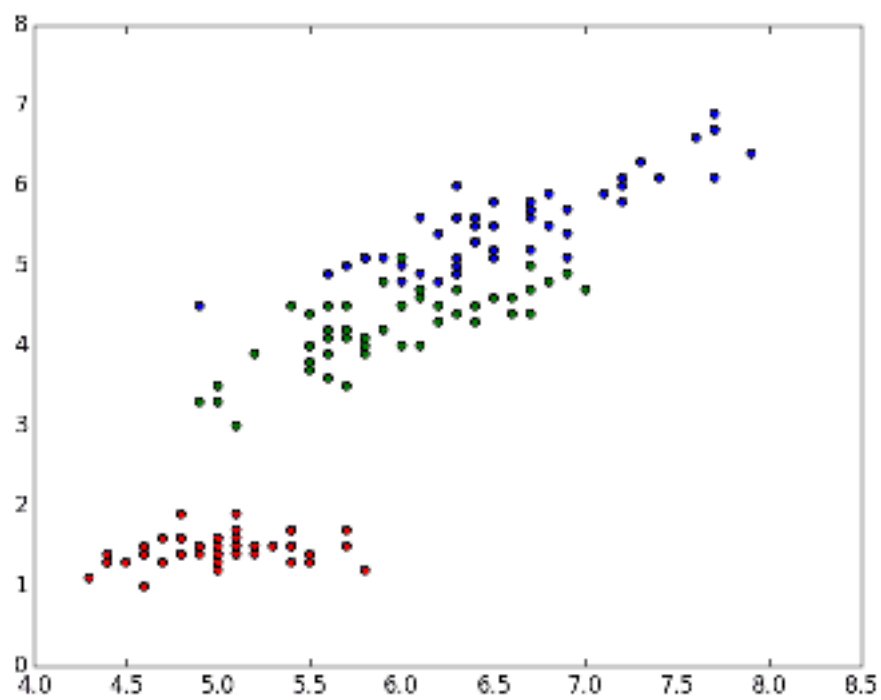
1. 4 numeric attributes and the class , 3 species, 50 samples in each species;
2. Lines 7 of the source code reads the file, line 8 parses it, line 9 and 10 appends line-by-line data to an array, and line 11 makes a transpose of the 2-D array to extract vectors row-by-row. To ensure accessing from the correct directory, I used system's path functions as shown in ref. 1;
3. Lines 13-19 creates the coloring array, with red representing setosa, green for versicolor, and blue for virginica. Lines 21 to 25 plots sepal length vs sepal width, 26-30 for sepal length vs petal length, 31-35 for sepal length vs petal width, 36-40 for sepal width vs petal length, 41-45 for sepal width vs petal width, and 46-50 for petal length vs petal width. To eliminate overlays, I adopted the approach from ref. 2.

Reference

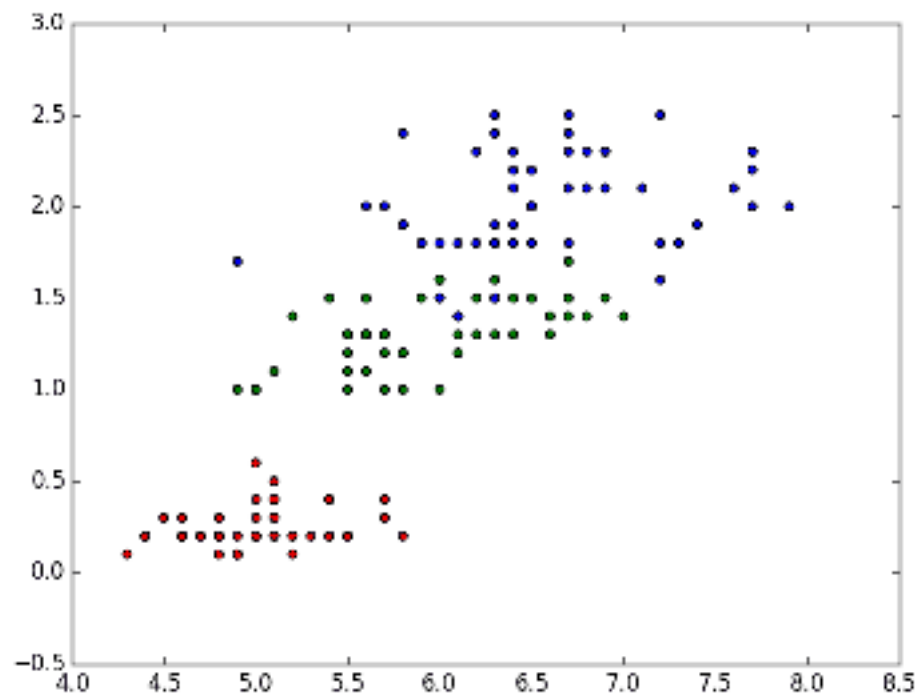
1. <http://stackoverflow.com/questions/4060221/how-to-reliably-open-a-file-in-the-same-directory-as-a-python-script>
2. <http://stackoverflow.com/questions/19176396/clear-overlay-scatter-on-matplotlib-image>



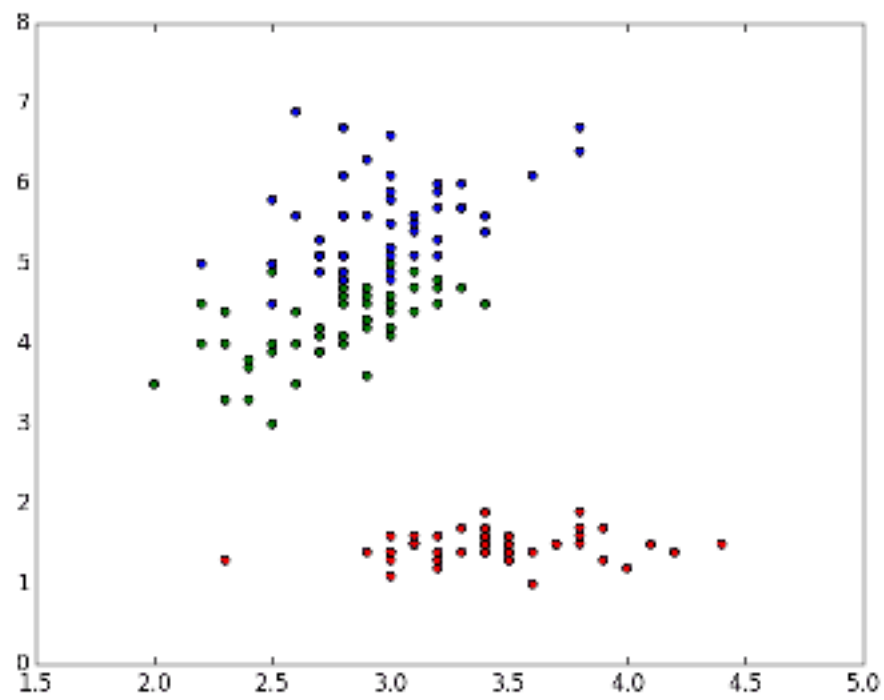
Plot 1. sepal length vs sepal width



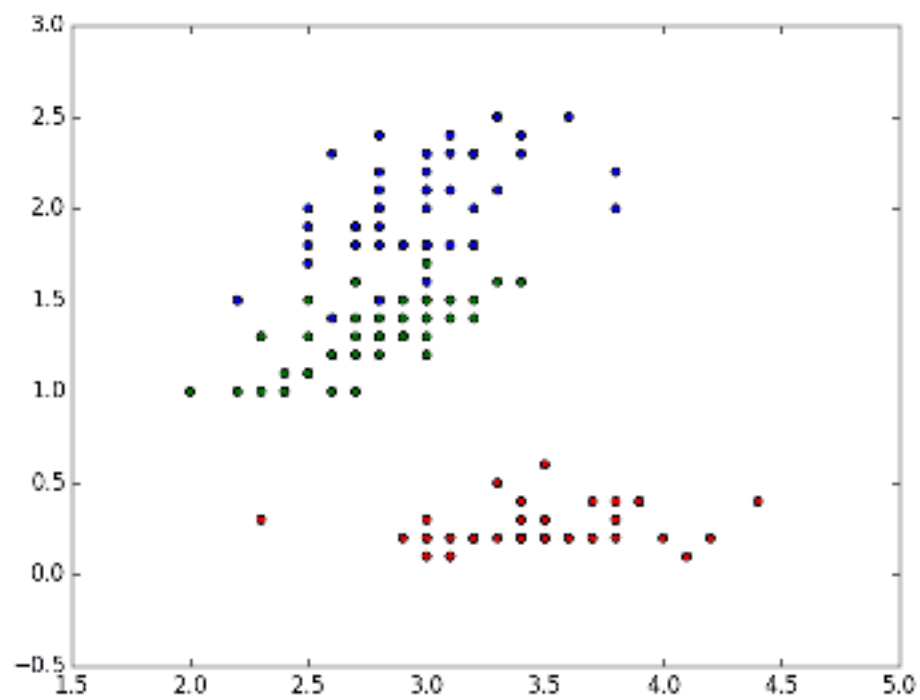
Plot 2. sepal length vs petal length



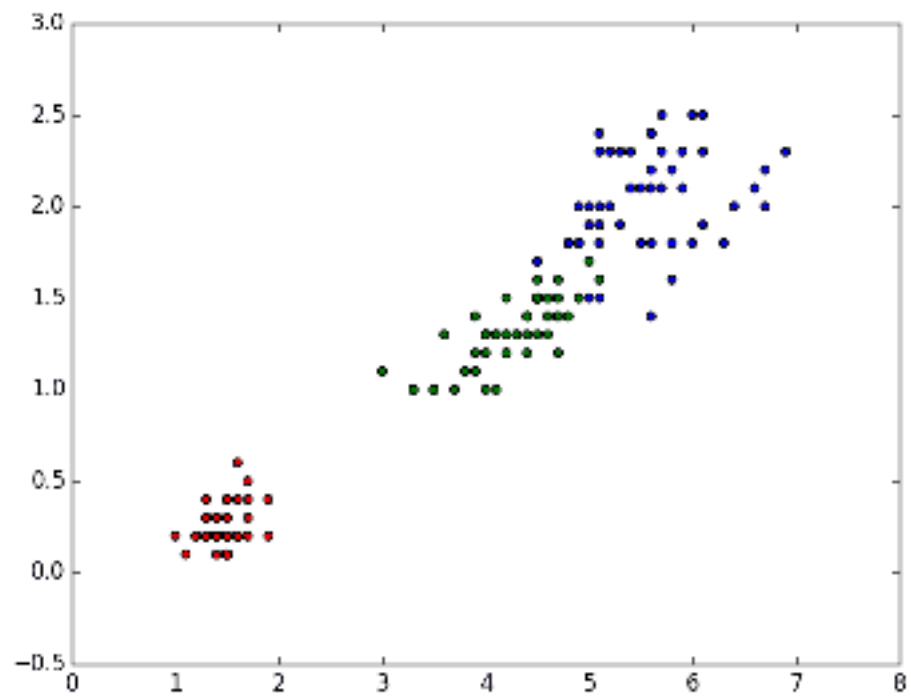
Plot 3. sepal length vs petal width



Plot 4. sepal width vs petal length



Plot 5. sepal width vs petal width



Plot 6. petal length vs petal width