

Homework #0 (due 08/30/17)

Name: Weisi Zhang, Dering Xu
Email: wz337@cornell.edu, dx49@cornell.edu

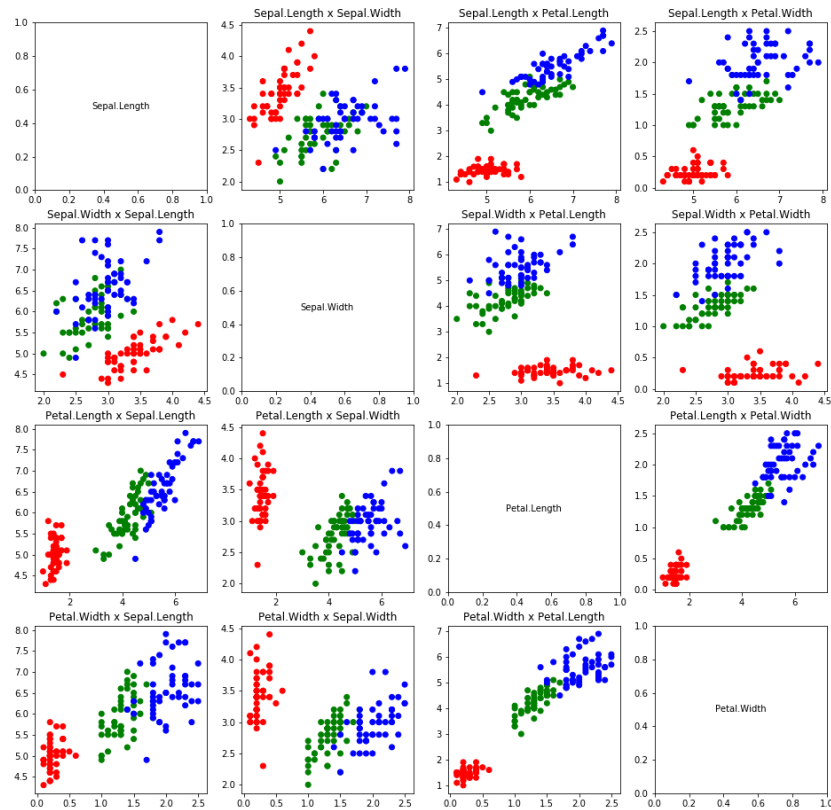
Introduction

The goal of this assignment is to help us setting up the development environment. After the python environment is set up, we are expected to load and parse the Iris Flowers Dataset[1] and create every possible scatter-plot from all pairs of two attributes.

Data Observation

In the dataset, there are four features per sample: Sepal Length, Sepal Width, Petal Length, and Petal Width. There are three different species, and Edgar Anderson recorded 50 samples for each species recorded, thereby 150 samples in total.

Iris Data(red=setosa,green=versicolor,bule=virginica)



Implementation Details

See details in *.ipynb file.

```

1 import matplotlib.pyplot as plt
2 import numpy as np
3 % matplotlib inline
4
5 result = [] # N * p array, N is the number of samples and p is the number of attributes per sample
6 colors = [] # N dimensional vector for labels
7
8 for line in open("iris.csv"):
9     line = line.strip().split(",")
10    result.append(line[:-1])
11    if (line[4] == 'Iris-setosa'):
12        colors.append('r')
13    if (line[4] == "Iris-versicolor"):
14        colors.append('g')
15    if (line[4] == "Iris-virginica"):
16        colors.append('b')
17
18 plt.figure(figsize=(16, 16))
19 plt.suptitle(' Iris Data(red=setosa, green=versicolor, blue=virginica)', fontsize=24)
20 name = ['Sepal.Length', 'Sepal.Width', 'Petal.Length', 'Petal.Width']
21
22 index = 1
23 for i in range(4):
24     for j in range(4):
25         plt.subplot(4,4,index)
26         index += 1
27
28         if (i != j):
29             x = np.array(result)[: , i]
30             y = np.array(result)[: , j]
31             plt.title(name[i]+'_x_' + name[j])
32             plt.scatter(x, y, c=colors)
33         else:
34             plt.annotate(name[i], (0.5, 0.5), xycoords='axes_fraction', ha='center', va='center')
35 plt.savefig("plot.png")

```

Listing 1: Homework 0 Implementation

References

- [1] Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.