CS5785 Applied Machine Learning
HW0
Due Date: 08/31/2016

Siyadong Xiong
sx225@cornell.edu
Teammates: N/A

Problem Statement

Set up Python environment. Visualize *Iris Flowers* dataset[1]. Create every possible scatterplot from all pairs of two attributes among four features.

Python Code

See details in `*.ipynb` file.

```
import numpy as np
import matplotlib.pyplot as plt
import itertools


data = np.loadtxt("iris.data", delimiter=',',dtype=str)
label_dict = {l:i for i, l in enumerate(set(data[:, -1]))}

colors = ['r', 'g', 'b']
NUM = 4

plt.figure(figsize=(8, 5))
for i, idx in enumerate(list(itertools.combinations(range(NUM), 2))):
    plt.subplot(2, 3, i+1)
    plt.scatter(data[:, idx[0]], data[:, idx[1]],
                c=[colors[label_dict[x]] for x in data[:, -1]])
plt.tight_layout()
```

Attachments

`hw0_sidxiong.ipynb` in `hw0_sidxiong.zip`

References

[1] *http://archive.ics.uci.edu/ml/datasets.html*