CS 5785

Homework 0

Team
Jai Chaudhary
Paul Lewis

Problem 1

1. Teammate found
2. Kaggle Account made - http://www.kaggle.com/users/213523/jai-chaudhary#
3. Canopy and PyCharm are easy to install but not good editor. Using sublime for dev environment. Installed all package using pip in terminal

Problem 2

1. Iris Dataset
   a. How many features/attributes are there per sample?
      4
   b. How many different species are there
      3
   c. How many samples of each species did Anderson record?
      50
2. Figure out how to parse the dataset you downloaded. Load the samples into an array. Create a vector containing each sample's label (species).

Relevant Snippet

```python
for line in open('iris.data'):
    if line.split():
        data = line.split(',')
        data_features.append(map(float, data[:-1]))
        data_labels.append(data[-1].strip())

data_points_count = len(data_features)
attributes_count = len(data_features[0])
```

3. Create every possible scatterplot from all pairs of two attributes. Within each scatterplot, the color of each dot should correspond with the sample species.

Relevant Snippet

```python
colors = map(lambda data_label : labels.index(data_label), data_labels)

transpose_data_feature = [[row[i] for row in data_features] for i in range(attributes_count)]

figure, subplots = plt.subplots(attributes_count, attributes_count, sharex='row', sharey='col')
figure.suptitle('Iris Data, red=setosa, green=versicolor, blue=virginica')

label_names = ['Sepal length', 'Sepal width', 'Petal Length', 'Petal Width']

for x_axis in range(attributes_count):
    for y_axis in range(attributes_count):
        if x_axis != y_axis:
            subplot = subplots[x_axis][y_axis]
            subplot.scatter(transpose_data_feature[x_axis], transpose_data_feature[y_axis], c=colors)
            subplot.set_xlabel(label_names[x_axis])
            subplot.set_ylabel(label_names[y_axis])

plt.subplots_adjust(wspace=0.25, hspace=0.25,left=0.1,right=0.9, bottom=0.1)
plt.show()
```
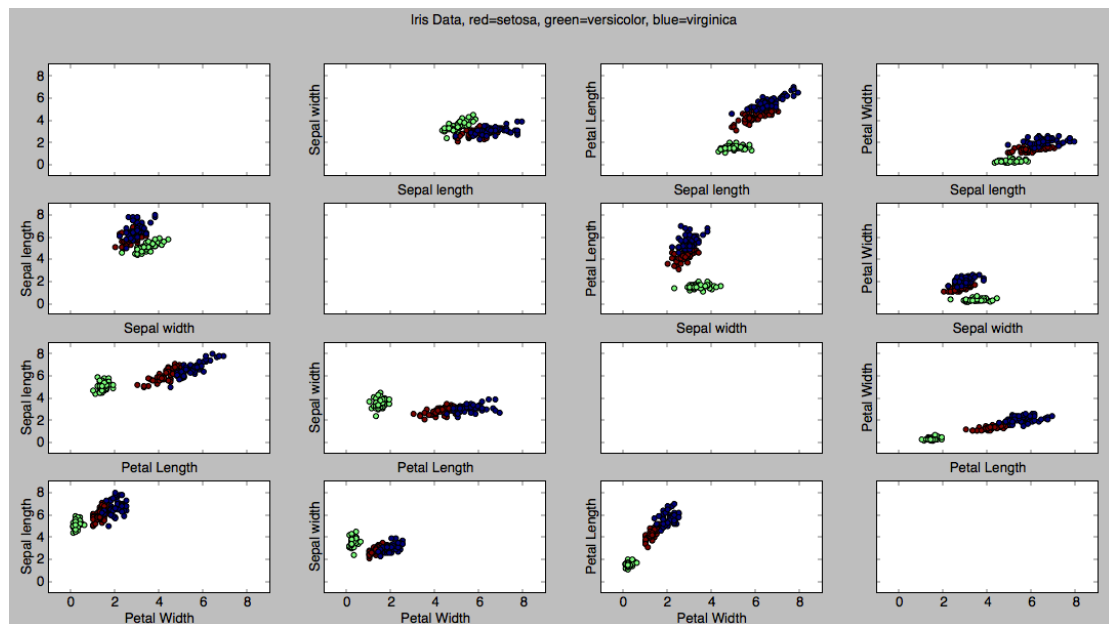
Iris Data, red=setosa, green=versicolor, blue=virginica

Homework0.py

```python
import matplotlib.pyplot as plt

def iris_flowers():
    data_features = []
    data_labels = []

    for line in open('iris.data'):
        if line.split():
            data = line.split(',')
            data_features.append(map(float, data[:-1]))
            data_labels.append(data[-1].strip())

    data_points_count = len(data_features)
    attributes_count = len(data_features[0])

    labels = list(set(data_labels))

    colors = map(lambda data_label : labels.index(data_label), data_labels)

    transpose_data_feature = [[row[i] for row in data_features] for i in range(attributes_count)]

    figure, subplots = plt.subplots(attributes_count, attributes_count, sharex='row', sharey='col')
    figure.suptitle('Iris Data, red=setosa, green=versicolor, blue=virginica')

    label_names = ['Sepal length', 'Sepal width', 'Petal Length', 'Petal Width']

    for x_axis in range(attributes_count):
        for y_axis in range(attributes_count):
            if x_axis != y_axis:
                subplot = subplots[x_axis][y_axis]
                subplot.scatter(transpose_data_feature[x_axis], transpose_data_feature[y_axis], c=colors)
                subplot.set_xlabel(label_names[x_axis])
                subplot.set_ylabel(label_names[y_axis])

    plt.subplots_adjust(wspace=0.25, hspace=0.25,left=0.1,right=0.9, bottom=0.1)
    plt.show()


iris_flowers()
```