# 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
   1. How many features/attributes are there per sample?

4

* 1. How many different species are there

3

* 1. How many samples of each species did Anderson record?

50

1. 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

**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))

1. 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

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]

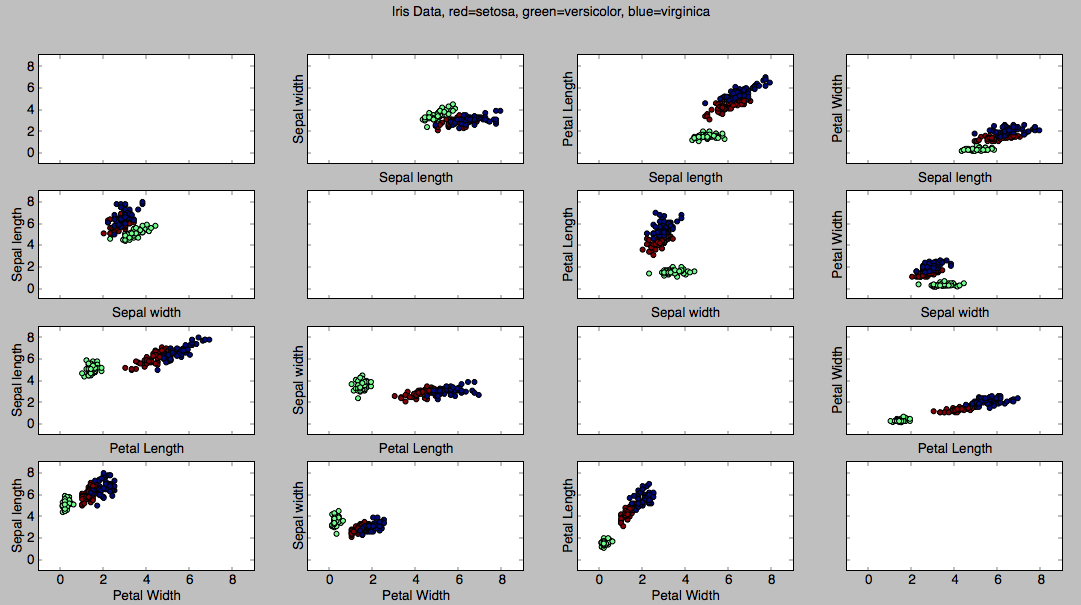
subplot.scatter(transpose\_data\_feature[x\_axis], transpose\_data\_feature[y\_axis], c=colors)

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()



**Homework0.py**

**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()