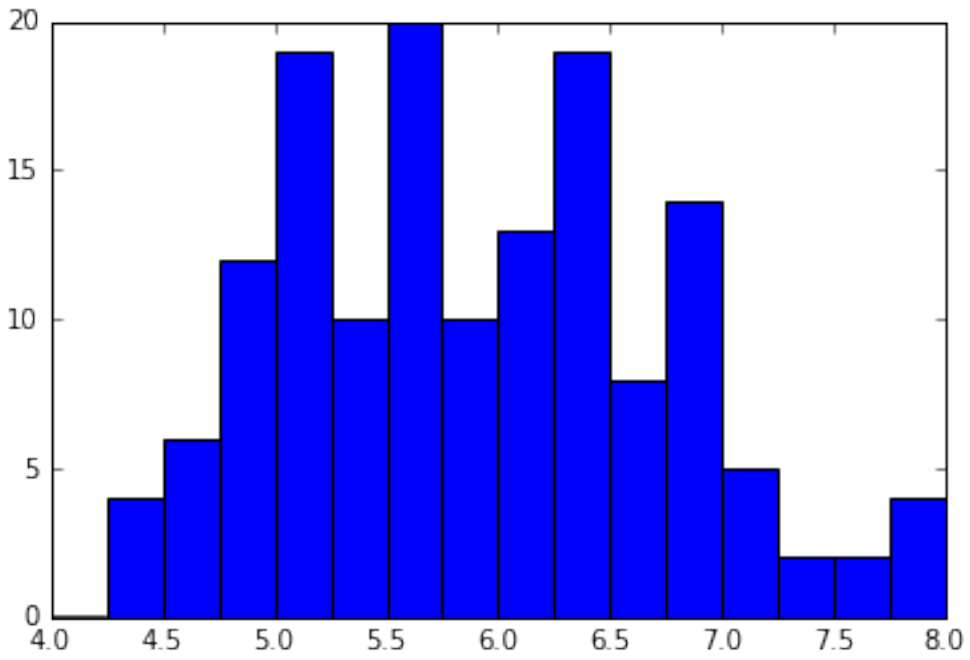
ML_HW1_Problem_1_Code

October 3, 2016

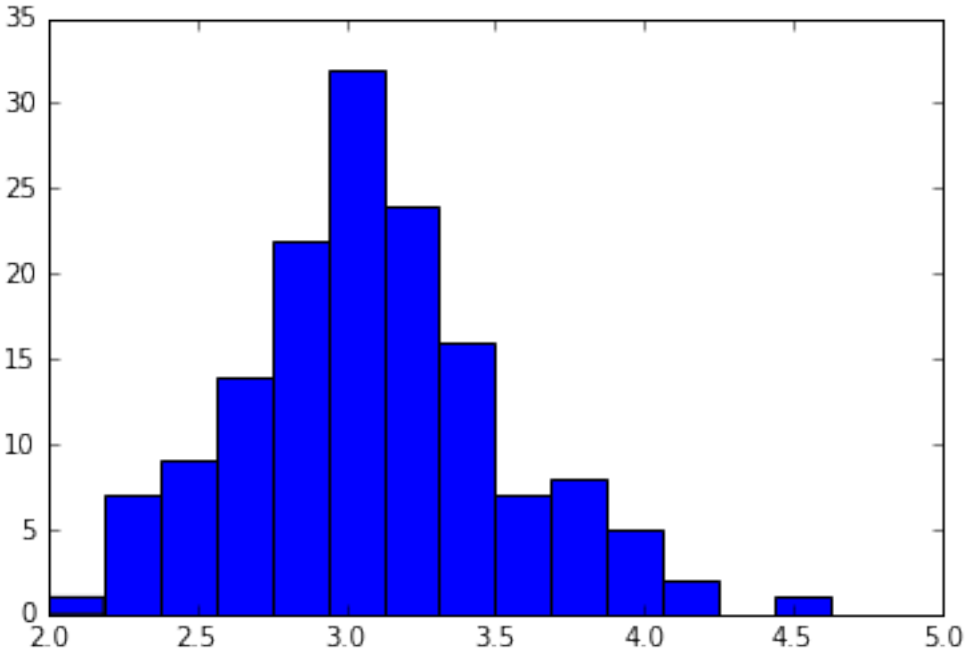
In [11]:
```
%matplotlib inline

import numpy as np
import matplotlib.pyplot as plt
import math
iris = np.genfromtxt("data/iris.txt",delimiter=None) # load the text file
Y = iris[:,-1] # target value is the last column
X = iris[:,0:-1] # features are the other columns

# Problem #1 Section A -- number of features and data points
print("Problem # 1 Section A -- number of features and data points -- : \n
featureCount      = X.shape[1]
dataPointsCount   = X.shape[0]

print ("Number of Features: "+ str(featureCount))
print ("Number of Data Points: " + str(dataPointsCount))



#section b , Histogram for each feature of data values
print("\nProblem # 1 Section B -- Histogram of each feature -- \n")


numberOfBins=17  # adjust it to data points if needed
for index in range(featureCount):
  print("Histogram for feature "+str(index+1)+" : \n")
  F = X[:,index] # extract each feature
  Bins = np.linspace(math.floor(np.min(F)),math.ceil(np.max(F)),numberOfBi
  plt.hist( F , bins=Bins ) #
  plt.show()


print("Section C -- Average of each feature --\n")
meanOfEachFeature = np.mean(X, axis=0)
for index in range(featureCount):
   print ("mean of feature "+ str(index+1) + " is : "+ str(meanOfEachFeatu
```

```
print("\nSection D -- Variane and Standard Deviation --\n")

print("Here is the Variance of Each Feature of the data points\n")
varianceOfEachFeature = np.var(X,axis=0)
for index in range(featureCount):
   print ("variance of feature "+ str(index+1) + " is : "+ str(varianceOfE

print("\nHere is the Standard Deviation of Each Feature of the data points
stdOfEachFeature = np.std(X,axis=0)
for index in range(featureCount):
   print ("standard deviation of feature "+ str(index+1) + " is : "+ str(s


print("\nSection E -- Normalizing Data --\n")
normalizedData=np.empty([dataPointsCount,featureCount])
for index in range(featureCount):
   normalizedData[:,index]=(X[:,index]-meanOfEachFeature[index])/stdOfEach

np.savetxt('Problem1_SectionE_NormalaizedData.txt', normalizedData, delimi
print("Saved Normalized Datapoints into Problem1_SectionE_NormalaizedData.



print("\nSection F -- Scatterplot  --\n")

colors = ['b','g','r']
numberOfClasses = np.unique(Y)
for c in numberOfClasses:
      plt.figure(1)
      plt.plot( X[Y==c,0], X[Y==c,1], 'o', color=colors[int(c)])
      plt.xlabel("feature 1")
      plt.ylabel("feature 2")


#
for c in numberOfClasses:
    plt.figure(2)
    plt.plot( X[Y==c,0], X[Y==c,2], 'o',color=colors[int(c)])
    plt.xlabel("feature 1")
    plt.ylabel("feature 3")

#
for c in numberOfClasses:
      plt.figure(3)
      plt.plot( X[Y==c,0], X[Y==c,3], 'o', color=colors[int(c)] )
      plt.xlabel("feature 1")
      plt.ylabel("feature 4")
```

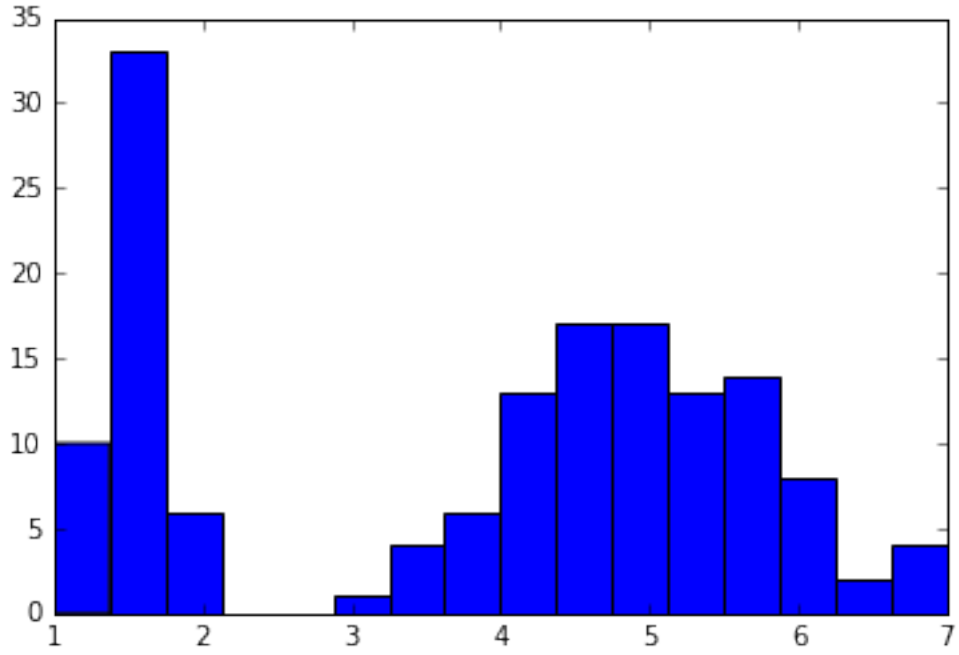
```
Problem # 1 Section A -- number of features and data points -- :

Number of Features: 4
Number of Data Points: 148

Problem # 1 Section B -- Histogram of each feature --

Histogram for feature 1 :
```



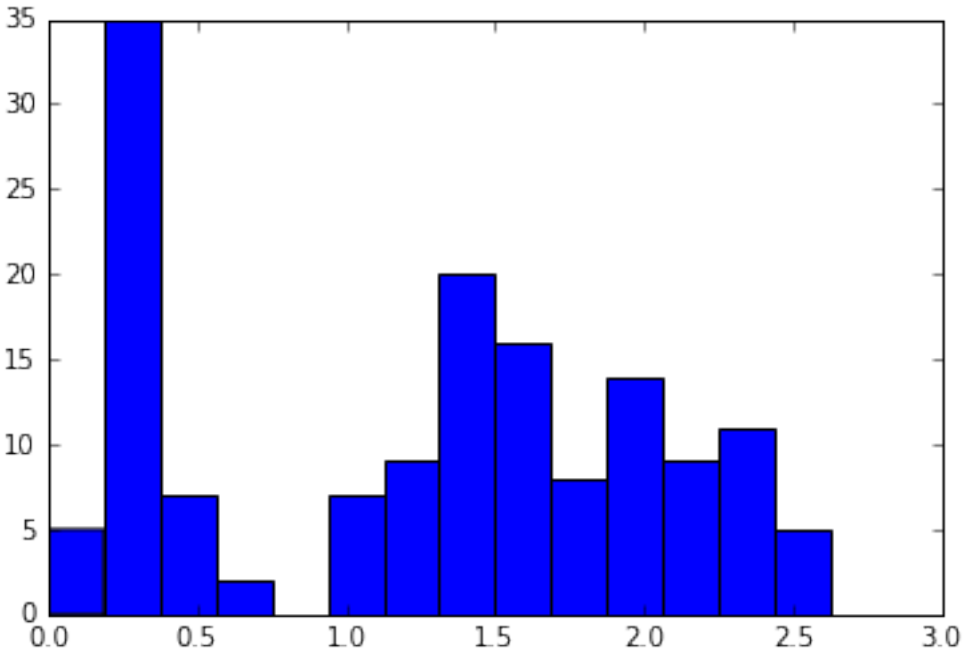

```
Histogram for feature 2 :
```



Histogram for feature 3 :



Histogram for feature 4 :



Section C -- Average of each feature --

mean of feature 1 is : 5.90010376419
mean of feature 2 is : 3.09893091689
mean of feature 3 is : 3.81955484054
mean of feature 4 is : 1.25255548459

Section D -- Variane and Standard Deviation --

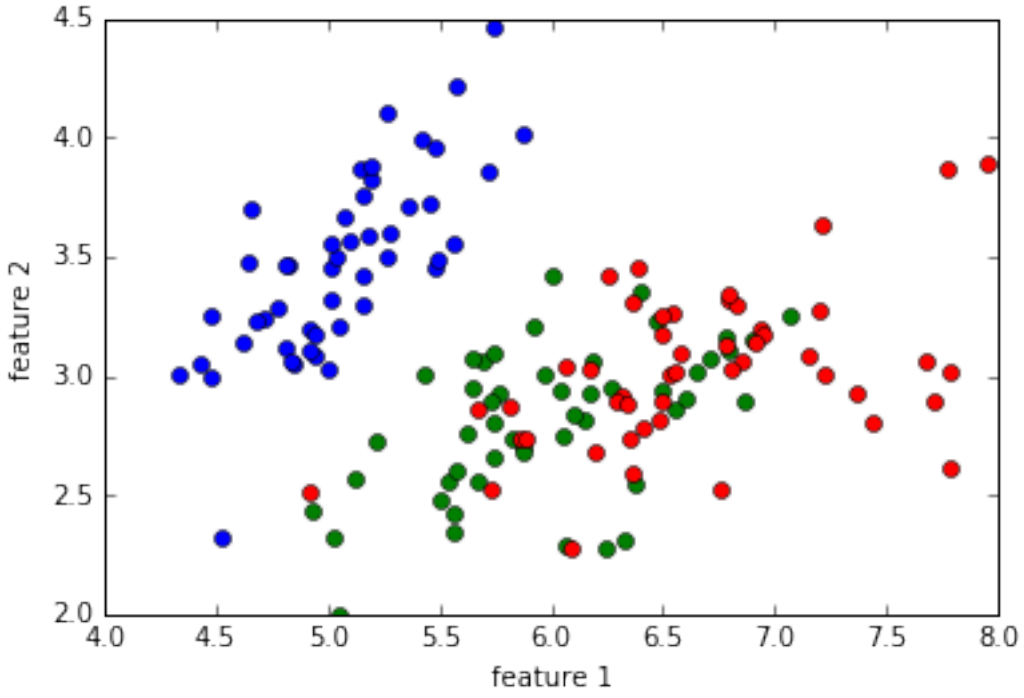
Here is the Variance of Each Feature of the data points

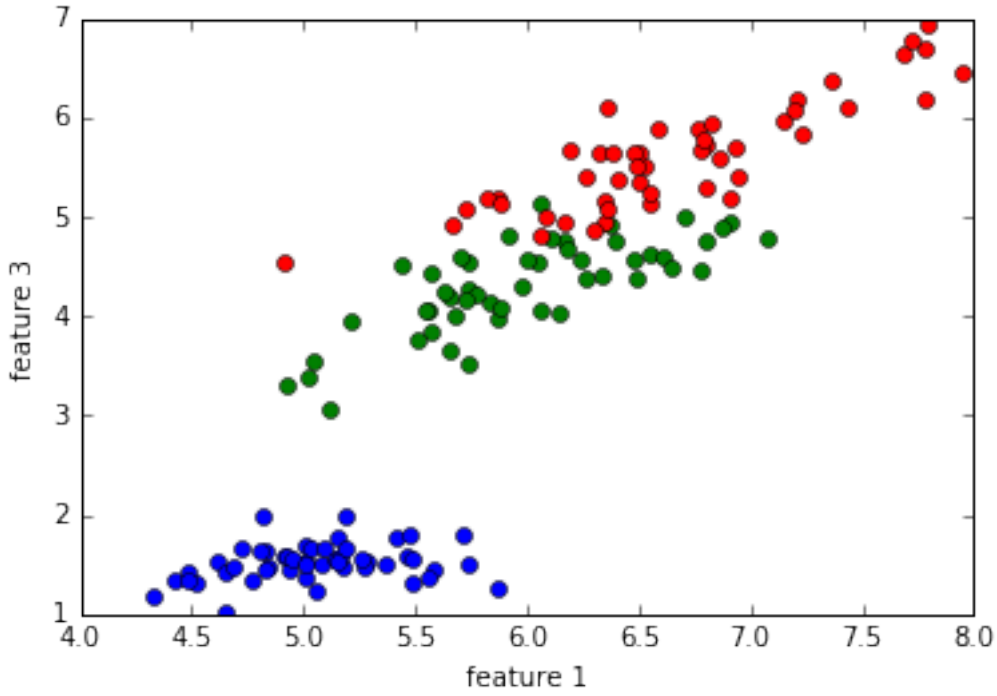
variance of feature 1 is : 0.694559004905
variance of feature 2 is : 0.190350567906
variance of feature 3 is : 3.07671634284
variance of feature 4 is : 0.575735641542

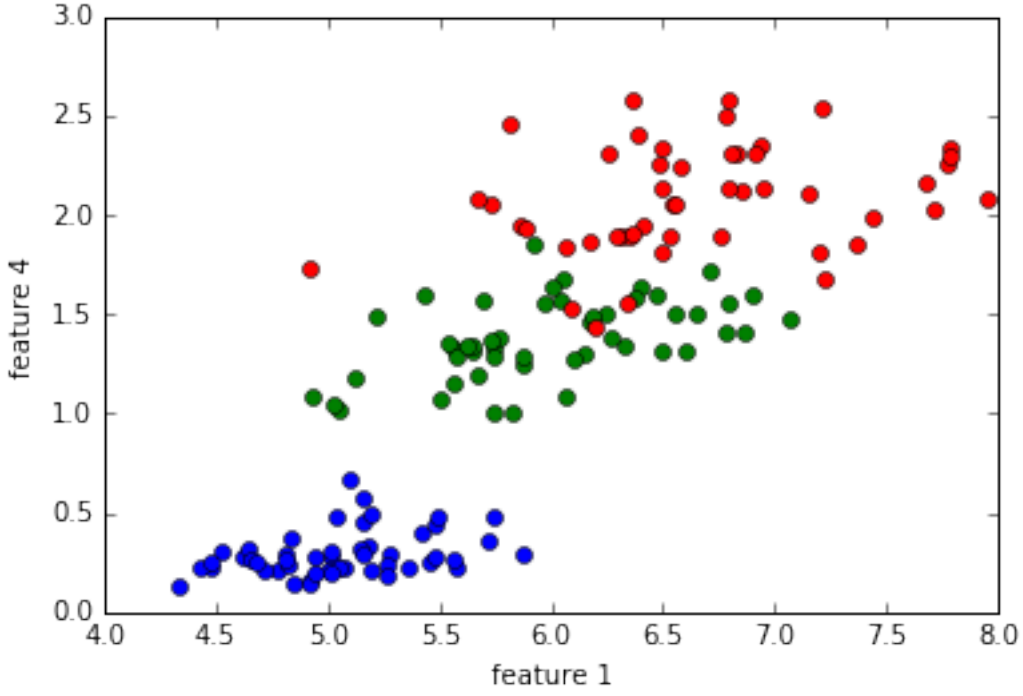
Here is the Standard Deviation of Each Feature of the data points

standard deviation of feature 1 is : 0.833402066775
standard deviation of feature 2 is : 0.436291838001
standard deviation of feature 3 is : 1.75405710934

```
standard deviation of feature 4 is : 0.758772457026

Section E -- Normalizing Data --

Saved Normalized Datapoints into Problem1_SectionE_NormalaizedData.txt file

Section F -- Scatterplot  --
```




feature 3
feature 1
7
6
5
4
3
2
1
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0


feature 4
feature 1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0