Department of Computer Engineering
University of Peradeniya

CO 544 Machine Learning and Data Mining

Lab 03

19^{th} July 2021

1. Objective

Provide students a hands on experience to *Matplotlib* Python module.

2. Introduction

Making plots and visualizations are one of the most important tasks in data mining and machine learning. It may be a part of the exploratory process; for example, identify outliers, data transformations, or coming up with ideas for models. Matplotlib is a Python package which provides a wide variety of plot types such as lines, bars, pie charts and histograms.

3. Matplotlib

(a) Importing Matplotlib

```
import matpoltlib    #importing Matplotlib module
import matplotlib.pyplot as plt   #pyplot is a collection of command style functions
from mpl_toolkits import mplot3d  #importing modules for 3D plotting
```

(b) Fundamendals of plotting

```
plt.plot([1,2,3,4,5],[1,4,9,16,25])     # to plot x versus y,
plt.show()     #display the plot
```

(c) Figures and Subplots

```
t = np.arange(0., 5., 0.2)

fig = plt.figure(figsize =(10 , 10))    #creating a figure
fig.subplots_adjust(hspace =1.0)

axes_1 = plt.subplot (4,1, 1)    #first axes in the figure

plt.plot(t, t,'r^',markersize=8,label='line1')   #plotting with red marker '^'

legend = plt.legend(loc='upper right', shadow=True,fontsize='x-large') #adding the legend
plt.title('First Plot')    #adding the title
plt.xlabel('t')      #labeling x axis
plt.ylabel('t')      #labeling y axis
plt.xlim([0,10])     #limits of x axis

axes_2 = plt.subplot (4,1,2)     #second axes in the figure
plt.plot(t, t**2, 'b*',markersize=8)    #plotting
axes_2.set_title('Second Plot')      #adding the title
```

```
axes_2.set_xlabel('t')     #labeling x axis
axes_2.set_ylabel('t squred')     #labeling y axis
axes_2.set_ylim([0 ,40])    #limits of y axis
```

(d) Saving plots to file

```
plt.savefig('plot1.pdf')       #saving the plot as a pdf
plt.savefig('plot1.png',dpi =400,bbox_inches ='tight')
plt.savefig('plot1.jpg')       #saving the plot as a jpg file
plt.savefig('plot11.jpg', dpi=100, quality=50, optimize=True, progressive=True) #jpg options
```

(e) 3D plots

```
fig = plt.figure()      #creating a figure
ax = fig.add_subplot(311, projection='3d')      #creating 3D subplot
xs=([29, 24, 25, 23, 30 ,31, 26, 26, 30, 28])
ys=([ 7, 53 , 33 , 66,  1 ,11, 91, 51, 83,  6])
zs=([-25, -25, -19, -23,-6,  -9, -11 , -11,-5, 14])

ax.scatter(xs, ys, zs, c='r', marker='o')

ax.set_xlabel('X Label')
ax.set_ylabel('Y Label')
ax.set_zlabel('Z Label')
```

TODO 1: Explain the reason to set,

```
fig.subplots_adjust(hspace =1.0)
```

TODO 2: Visualize the above 3D plot in different angle.

TODO 3:

i. Load the **winequality_red.csv** file. (This is a dataset from UCI Machine Learning Repository open datasets)

```
df=pd.read_csv('winequality_red.csv', sep=';')
```

ii. Explore the correlation between each attribute and *quality* of the wine graphically. (Take the direction and magnitude of correlation)

iii. Visualize the relationship between each pair of attribute by a scatter matrix.

iv. Visualize the distribution of data for each attribute using a box plot for each class.

Use different colours to denote each class and your figures must contain titles, axis labels and appropriate legends.

4. Submission

(a) In Lab Submission
Please submit whatever you have done within the lab. Rename your file as 16xxxlab03.txt where xxx is your registration number.

(b) Complete Submission
Submit three different text files three TODO sections. Your files should contain all your commands you have tried out in the lab with the answers for TODO sections (as comments in the text). Rename the files as 16xxxlab03_TD1.txt , 16xxxlab03_TD2.txt and 16xxxlab03_TD3.txt where xxx is your registration number.

5. Deadline

The deadline: July 26, 2021, by 11.59 p.m
If you do not understand any concepts, make sure you get some help from instructors.