UECS3213 / UECS3453 Data Mining

SESSION: January 2019

Lab 5: Introduction to Data Visualization in Python using matplotlib

Introduction

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc.

- **Matplotlib**: the whole Python data visualization package
- **pyplot:** a module in the matplotlib package. The module provides an interface that allows you to implicitly and automatically create figures and axes to achieve the desired plot.

The first step that you have to take in order to start plotting in Python is to consider revising NumPy, the Python library for scientific computing. This Matplotlib tutorial takes you through the basics Python data visualization: the anatomy of a plot, pyplot and pylab, and more.

Two ways to plot

- Functional Method
- Object-Oriented Method

Object-oriented:

Create figure objects and call methods of it

```
fig = plt.figure()
axes = fig.add_axes([0.1,0.1,0.8,0.8])

axes.plot(x,y)

axes.set_xlabel("X Label")
axes.set_ylabel("Y Label")

axes.set_title("X vs Y")
```

Functional:

Use matplotlib command

```
x = np.linspace(0,10,100)
y = x**2
y = np.sqrt(x)

plt.plot(x,y,'r')

plt.xlabel("X Label")
plt.ylabel("Y Label")

plt.title("X vs Y")
plt.show()
```

General steps:

1. Manipulate / Massage your data to the desired format (use general Python data structure of numpy arrays) and feed into the plot function
2. Format the chart: title, axis labels etc. etc.
3. Plot (show) the chart

Summary about Matplotlib:

- Matplotlib is the most popular plotting library for python.
- It gives you control over every aspect of a figure.
- It was designed to have a similar feel to Matlab's graphical plotting

Objectives

At the end of this lab, you are expected to acquire the following:

a) The anatomy of a Matplotlib plot: what is a subplot? What are the Axes? What exactly is a figure?
b) Plot creation, how you exactly should go about initializing the figure and the Axes of your plot, how to use matplotlib in Jupyter notebooks, etc.
c) Plotting routines, plotting and visualizing your data.
d) Basic plot customizations, with a focus on plot legends and text, titles, axes labels and plot layout.
e) Saving, showing, clearing, … your plots: show the plot, save one or more figures to, for example, pdf files, clear the axes, clear the figure or close the plot, etc.
f) Two ways in which you can customize Matplotlib: with style sheets and the rc settings.

Instruction

1. Visit the “Matplotlib Tutorial: Python Plotting” at the following link: and https://matplotlib.org/tutorials/introductory/pyplot.html#sphx-glr-tutorials-introductory-pyplot-py and https://medium.com/incedge/data-visualization-using-matplotlib-50ffc12f6af2

2. Follow the step-by-step instructions in the tutorial.

3. Also, explore the different examples of plots that matplotlib library provides at https://matplotlib.org/gallery/index.html Focus on the following five types of plots:
   a) Line Plot
   b) Bar Chart
   c) Histogram Plot
   d) Box and Whisker Plot
   e) Scatter Plot

Other Related References

- https://matplotlib.org/
- https://matplotlib.org/tutorials/index.html
- https://www.datacamp.com/community/tutorials/matplotlib-tutorial-python
- https://machinelearningmastery.com/data-visualization-methods-in-python/

The End