**Assignment**

**(Numpy, Scipy, Matplotlib)**

1. Create random numbers and plot: Histogram, Line graph, scatter plot (use subplots for the same)
2. From given sequence.txt calculate A, G, T, C % and plot in the pie chart (eg. A%= Total number of A/Total length of the sequence \* 100)
3. From given a file data.dat plot the distribution of each data. N.B. Plot probability distribution (each column in the file depicts separate dataset.)
4. Write a python program to read 'random.txt' file. Count (“a”, “it”, “the”, “spot”, “among”, “income”, “several”) words in the text. Plot length of each word v/s count as histogram plot. Title of the plot should be “Word Count”, x-axis should be “Length” and y-axis should be “Count”.
5. Find values of <x, y, z, t> satisfying the following system of equations:

41\*x + 15\*y + 96\*z + 31\*t = 46

10\*x + 21\*y + 33\*z - 11\*t = 45

70\*x - 31\*y + 11\*z + 10\*t = 50

25\*x - 22\*y + 28\*z - 49\*t = 57

1. Integrate the following functions over the given limit:
2. y = x from (0, 10)
3. y = x\*\*2 from (-5, 5)
4. y = sin(x) from (0, pi/2)
5. The data in populations.txt describes the populations of hares and lynxes (and carrots) in northern Canada during 20 years. Answer the following questions:
6. The mean and std of the populations of each species for the years in the period.
7. Which year each species had the largest population.
8. Which species has the largest population for each year . (Hint: argsort & fancy indexing of np.array(['H', 'L', 'C']))
9. Which years any of the populations is above 50000. (Hint: comparisons and np.any)
10. The top 2 years for each species when they had the lowest populations. (Hint: argsort, fancy indexing)
11. Compare (plot) the change in hare population (see help(np.gradient)) and the number of lynxes.
12. Check correlation (see help(np.corrcoef)).

… all without for-loops.

1. The rainfall statistics have been downloaded from <http://www.mospi.gov.in/statistical-year-book-india/2016/203>

You need to analyze this time series data. Write a menu-driven program where the user selects two Indian states, followed by a month of interest. **Make a plot of the rainfall distribution across the years (2009-2014) for the given month for both the states as subplot.** You will then look to see if the data is correlated (are wet summers in State A also wet summers in State B) using the scipy.stats module. One way to test this is to calculate the correlation coefficient between two time series. The scipy.stats module has functions to do this. Use the scipy.stats.pearsonr function (which takes two numpy arrays as arguments) to calculate the Pearson correlation coefficient. Is the correlation positive? Is it significant?

**Note: You are free to re-arrange the excel data in a text file or multiple text files, depending on your choice. You can decide upon the rows and columns.**