

Homework 1

COMP 7150/8150

FALL 2016

Instructor: Deepak Venugopal

Due Date: September 26, 2016

1. (25 points) Numpy Exercise: Create a numpy matrix with random numbers (use the `numpy.random.randn` function) with 5 rows and 5 columns. Compute the following functions for this matrix.
 - a. Rescale each element such that it lies in the range between (-1,1). That is the values in the i-th column needs to be rescaled with respect to the min and max values in the i-th column.
 - b. Standardize the elements of the matrix. That is, for each element at position (i,j), divide the element by the mean of the i-th column and divide by the standard deviation of the i-th column.

Display the original and the changed matrices

2. Pandas Exercise (25 points) Load the CSV file named "311-service-requests.csv" as a pandas dataframe and perform the following
 - a. We want to find compare complaints for different cases for each "Borough" (one of the feature-names in the csv). Specifically, Draw a bar chart that compares for each "Borough" the total number of complaints where the "Complaint Type" includes the phrase "Noise", and the total number of complaints where the "Complaint Type" includes the phrase "Parking". (Hints: Look at function `value_counts()` for this problem)
3. Pandas Exercise (50 points): Load the CSV file named "weather-2012.csv"
 - a. What month of the year has the ideal temperature? To answer this question, draw the average temperature ("Temp (C)" column) for each month in the year as a line graph (25 points)
 - b. Distribution Fitting: What is the natural distribution for the feature "Wind Spd (km/h)"? To answer this, fit three distributions, normal, lognormal and beta for this feature. Show the histogram for this feature assuming 10 bins and all the fitted distributions on the histogram. (25 points)

Please write your code using ipython and submit your ipynb file to dropbox on ecourseware.