IC152: Assignment 5

Mean, Median and Correlation

From this assignment and onwards, the assignments will be evaluated via viva TAs will take.

- Save a python file with the name “problem<n>.py” where <n> should be replaced with a problem number.

- Put all the python files in “<RollNo>assignment5.zip” and submit the zip file on LMS, <RollNo> should be replaced with the roll number of the student submitting the assignment. Only one student per group has to submit the assignment on LMS.

Problem 1: Write a python program to take a list of 10 numbers from the user and find the mean of the numbers in the input list. Print the output on the console/shell. You **cannot** use any inbuilt function for finding the mean, feel free to use other inbuilt functions.

Problem 2: Write a python program to take a list of 10 numbers from the user and find the median of the numbers in the input list. Print the output on the console/shell. In case of multiple medians you should print the range. You **cannot** use any inbuilt function for finding the median, feel free to use other inbuilt functions.

Problem 3: Show through the code that the mean is the minimizer of the sum S: $\sum_{i=1}^{N} (x - x_i)^2$ for the example in problem 1, where x_i represents each array in the input list/array. In order to show that S is minimum for mean, find S for different values of x. Print a list containing x and the corresponding S, including one of the x as mean.

Problem 4: Show through code that the median is the minimizer of the sum S: $\sum_{i=1}^{N} |x - x_i|$ for the example in problem 2, where x_i represents each array in the input list/array. In order to show that S is minimum for median, find S for different values of x. Print a list containing x and the corresponding S, including one of the x as the median.

Problem 5: Write a python program to take a 2D array from the user in the form of a list and find the mean of each **row**. Print the output in the form of a list.

Algorithm:
Input: 2 D Array or Matrix of dimension m X n
Output: 1D Array or Vector of dimension m x 1

Algo:
Create an empty output array.
Repeat m times following:
    For ith row in input array, find its mean
    Append the mean to the output array.

Algo find mean of an array/row:
    Add all elements in the row/array
    Divider the above sum by number of elements in the row/array

Code:

```
#Code for finding mean of each row in a matrix/2d array:
def checkIntegersInList(myRowList):
    #homework

def averageOfRow(myRowList):
    if len(myRowList) > 0:
        if checkIntegersInList(myRowList) == True:
                return sum(myRowList)/len(myRowList)
    return []

inputList = [[ 1,2,3], [], ["5634","25","23"]]
m = len(inputList)
outputList = []
for i in range(m): #0 to m-1, m will not be included
        mthRow = inputList[i]
        avgi = averageOfRow(mthRow)
        outputList.append(avgi)
print(outputList)
```

Problem 6: Correlation

Consider two vectors: $X = [x_1-\mu_x, x_2-\mu_x, \ldots, x_n-\mu_x]$ and $Y = [y_1-\mu_y, y_2-\mu_y, \ldots, y_n-\mu_y]$. μ_x is the mean of x_i's, and μ_y of y_i's. The cosine of the angle between these two vectors can be given by ratio of dot product of these two vectors with their magnitude, i.e. ratio of X.Y and |X||Y|.

Interestingly, correlation is a statistical method that measures the similarity of the variation between two random vectors. The correlation coefficient (value in between -1 to +1 similar to cosine) in between two vectors can be calculated with the help of the given formula:

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]^{\frac{1}{2}} \left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]^{\frac{1}{2}}}$$

Where, n = sample size, x_i and y_i are the sample points with index i.

a. Write a python code that takes 5 values of x and 5 values of y from the user in the form of a list and find the correlation using the formulae given above.
b. Prove that the ratio of X.Y and |X||Y| is equal to r given in above equation (X and Y are vectors defined in the starting of problem 6) . Show your proof to the lab TAs for evaluation. Reason why correlation will always lie between -1 and + 1.
c. With the same values given in part a by the user, find the correlation using method given in part b, and verify that you are getting the same correlation as in part a.

Create a single python file for problem 6.