PR ASSIGNMENT - 1 (Deadline : 27/01/2020)

General Instruction:

1. All assignments need to be completed as a group of 3 members(maximum).
2. Kindly upload your group details in the excel sheet that will be attached to the next mail.
3. The assignments need to be submitted online in the drive link attached to the next mail.
4. Programming assignments can be done either in python3 or MATLAB.

Deliverables for this assignment:

1. Word document(not PDF) with typed solutions to Q1 and codes of Q2, Q3, Q4, Q5.
2. Code file and output screenshots for Q2, Q3, Q4, Q5.

Q1. Prove that following distances are metric:

I. Edit(Levenshtein distance)

II. Mahalanobis

Q2. Calculate the distance between the two normalized histograms H1 and H2 using each of the following methods:

I. KL Distance

II. Bhattacharyya Distance

H1 = [0.24, 0.2, 0.16, 0.12, 0.08, 0.04, 0.12, 0.04]

H2 = [0.22, 0.19, 0.16, 0.13, 0.11, 0.08, 0.05, 0.02]

Q3. Given $(hq-ht)^T$ = (0.5 0.5 -0.5 -0.25 -0.25) and

$$A = \begin{pmatrix} 1 & 0.135 & 0.195 & 0.137 & 0.157 \\ 0.135 & 1 & 0.2 & 0.309 & 0.143 \\ 0.195 & 0.2 & 1 & 0.157 & 0.122 \\ 0.137 & 0.309 & 0.157 & 1 & 0.195 \\ 0.157 & 0.143 & 0.122 & 0.195 & 1 \end{pmatrix}$$

Find the quadratic form distance.

Q4. Compare two text files doc1.txt and doc2.txt using cosine distance.

doc1.txt

MATLAB is a program for solving engineering and mathematical problems. The basic MATLAB objects are vectors and matrices, so you must be familiar with these before making extensive use of this program.

doc2.txt
MATLAB works with essentially one kind of object, a rectangular numerical matrix. Here is some basic information on using MATLAB matrix commands.

Q5. Classify flower 0, 50, and 100 from the Iris Dataset (.csv file) attached to this mail into one of the three classes as given in dataset specification:

Dataset Specifications:
Total number of samples = 150
Number of classes = 3 (Iris setosa, Iris virginica, and Iris versicolor)
The number of samples in each class = 50.

Directions to classify:

1. Use features PetalLengthCm and PetalWidthCm only for classification.
2. Consider flowers 0,50 and 100 as test cases.
3. Plot the distribution of rest 147 sample points along with their classes(differentiate classes with different colour). Consider PetalWidthCm along Y-axis and PetalLengthCm along X-axis.
4. Capture the properties of the distribution and use suitable distance metrics to classify the flowers 0,50 and 100 into one of the classes.
5. Print their class and plot the points on the previous plot with a marker differentiating the three points.
