INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY,
HYDERABAD

Assignment-2

Setters: *Shanthika and Abhinaba*

Course: *Statistical Methods in AI(CS7.403)* – Professor: *Dr. Anoop Namboodiri*
Due date: *11th November, 2021*

Question 1 : Eigen Values and Eigen Vectors

A. Among Eigen Value Decomposition and Singular Value Decomposition, which one is more generalizable to matrices and why?

B. Show the method and find the Singular Value Decomposition of the following matrix::

$$M = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \tag{1}$$

Question 2 : LDA and PCA

A. Suppose you want to apply PCA to your data X which is in 2D and you decompose X as UDV^T. Then, which of the following are correct:

(a) PCA can be useful if all elements of D are equal

(b) PCA can be useful if all elements of D are not equal

(c) D is not full-rank if all points in X lie on a straight line

(d) V is not full-rank if all points in X lie on a straight line

(e) D is not full-rank if all points in X lie on a circle

B. True/False
PCA will project the data points(multi-class) on a line which preserve information useful for data classification.

Question 3 : Bayes Theorem

The aim of this question is to understand Bayes Theorem. One very useful resource is Bayes Theorem video by 3b1b.

A. What is the difference between prior and posterior probabilities?

B. Let's say that you are at work one day and have just finished lunch. You suddenly feel horrible and find yourself lying down. Maybe it is because one of your friend was recently sick with flu.
You have a headache and sore throat, and you know that people with the flu have the same symptoms roughly 90% of the time. In other words, 90% of people with the flu have the same symptoms you currently have.
Wanting to gain a little more information you roll over, grab your phone and search Google. You find a reputable article that says that only 5% of the population will get the flu in a given year. Or, the probability of having the flu, in general, is only 5%.
You then spot one more statistic that says 20% of the population in a given year will have a headache and sore throat at any given time.
What is the probability of you having a flu given you have a sore throat and headache?

Question 4 : K-Nearest Neighbours

Write a code to perform KNN classification on Iris dataset provided. Use the statrter code for loading the train, test dataset. Report the accuracy obtained on test dataset. Do not use direct inbuilt functions. Numpy or other math libraries are allowed.

Question 5 : Logistic Regression

For the sample dataset provided, write a code to perform logistic regression. Plot a decision boundary between the two classes. Sample result image is provided. Do not use direct inbuilt functions. Numpy or other math libraries are allowed.