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. Suppose you have a population of a thousand people. You want to find out if a person has any signs of weak knees or not. You have a test in hand that always returns True result for all potential True case, and False if the person is not likely to catch it. The caveat is that for 5 percent of the cases it return a False Positive.
The results have come out as True for a particular subject in the population. What is the probability that the subject actually has weak knees?

Question 4 : K-Nearest Neighbours

Write a code to perform KNN classification on Iris dataset provided. Use the stater 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.