DSE 2242– Fundamentals of Machine Learning Lab
Week 3 – Date: 4th January 2025

EXER 1:

1. Use the "pima-indians-diabetes.csv" dataset and note down the meta information.
2. Compute mean & standard deviation, tabulate and visualize the age of the patients.
3. Analyze and tabulate the relationship of age, BMI of patients with respect to the class.
4. Tabulate the class label and comment on whether the classes are balanced.
5. Use the data set to build a logistic regression model (using sklearn) and predict the class label. Divide the dataset into training and test set (70,30) using train_test_split method in sklearn.
6. Use the test data set and evaluate the performance using a confusion matrix. Visualize the confusion matrix using a heat map.
7. Compute accuracy rate, true positive and true negative rate and comment on the performance.
8. Visualize the ROC curve, and comment on the performance of the classifier.

EXER 2:

1. For the IRIS data set write down the meta information.
2. Visualize the class label against the predictor variable using appropriate plots.
3. Use the IRIS data set to build a logistic regression model (using sklearn) and predict the class label 'Species'. Divide the dataset into training and test set (70,30) using train_test_split method in sklearn.
4. Analysis and visualize the performance of the classifier using metrics, confusion matrix.
5. Use the IRIS data and KNeighborsClassifier (using sklearn) and predict the class label 'Species' for k value between 2 and 20. Divide the dataset into training and test set (70,30) using train_test_split method in sklearn.
6. Identify the best k (for k between 2 and 20) for the model built.
7. Comment on the classifier (Logistic Regression or KNeighborsClassifier) that has a better performance for the IRIS dataset.