MCA 5141 – Machine Learning Lab
Week – 6

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

Download fuel consumption dataset "FuelConsumption.csv", which contains model-specific fuel consumption ratings and estimated carbon dioxide emissions.

- Select the features 'ENGINESIZE', 'CYLINDERS', 'FUELCONSUMPTION_COMB', 'CO2EMISSIONS' to use for building the model. Plot Emission values with respect to Engine size.
- split the data into training and test sets (70:30) to create a model using training set, evaluate the model using test set, and use model to predict unknown value.
- Try to use a polynomial regression with the dataset of degree – 3, 4 & 5. Verify the accuracy by calculating Mean absolute error, Residual sum of squares, R2-score and comment on which model is the best.