26th Aug,2022

Attendance: 10%, Continuous evaluation: 70%, Viva-20%

Assignment No. 3

i. Download Titanic Dataset (https://www.kaggle.com/heptapod/titanic/version/1#) and do initial pre-processing and train a Logistic Regression for the classifier.

ii. Analyze and control the overfitting by varying the inverse of regularization strength parameter (0.1, 0.25,0.5, 0.75, 0.9) and plot the accuracy graph for the test set.

iii. Using the same dataset train a Decision Tree classifier and vary the maximum depth of the tree to train at least 5 classifiers to analyze the effectiveness.

iv. Download the Forest Cover Type dataset (https://www.kaggle.com/uciml/forest-cover-type-dataset) and preprocess the dummy variables to create training, test, and development set. Reduce the train data size if the system unable to process the whole dataset.

v. Apply multiclass classification in SVM using Forest Cover Type
vi. dataset.
vii. vii. Plot and Analyze the Confusion matrix for the above applied SVM
viii. method.
ix.

x. Consider only two features and three classes and train Logistic Regression 3-class Classifier (Any three-class) to show the training and test area in a 2-Dplane, using matplotlib.

Submit a report with results.