31st Aug 2020

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

Assignment No. 2

i. Download Cancer Wisconsin (Diagnostic) Data Set (already in the needed format). The data-set is used to recognize 2 types of cancer to be predicted (benign or malignant).

ii. Implement Logistic regression using scikit-learn package in python after splitting the dataset 80:10:10 percent (use seed = 5 for splitting).

iii. Use 'newton-cg', 'lbfgs', 'liblinear' solver to train the Logistic regression model, and create a table for the coefficients of all the features along with accuracy.

iv. Use 'l1', 'l2', 'none' penality to train the Logistic regression model, and create a table for the coefficients of all the features along with accuracy.

v. Vary the l1 penalty over the range (0.1, 0.25, 0.75, 0.9) and compare the coefficients of the features.

vi. Estimate the average accuracy of the Naive Bayes algorithm using 5-fold cross-validation using a scikit-learn package in python. Plot the bar graph using matplotlib.

Submit a report with the result.