

Continuous Evaluation: 70%, Viva: 30%

Assignment 2:

- i. Download [Cancer Wisconsin \(Diagnostic\) Data Set](#) (already in the needed format). The dataset is used to recognize 2 types of cancer to be predicted (benign or malignant).
- ii. Implement Logistic Regression using scikit-learn package.
 - a. Use 'newton-cg', 'lbfgs', 'liblinear' solver for the regression model.
 - b. For each solver use 'l1', 'l2', 'none' penalty to train the model.
 - c. Split the dataset in to 80:10:10 percent (train : validation : test) (use seed = 5 for splitting).
 - d. Train the model initially with 80% training data and create a table for the coefficients of all the features.
 - e. Fine-tune the model with the remaining validation partition of the dataset (consisting of 10% of the original dataset) and create a table for the updated coefficients of all the features.
- iii. For every solver vary the 'l1' penalty over the range (0.1, 0.25, 0.75, 0.9) and compare the coefficients of the features.
- iv. Using the test split (of 10%) to show the accuracy, precision, recall, F1-score of the regression model.
 - a. Show the output for every possible combination of solver-penalty (newton-cg-l1, newton-cg-l2, newton-cg-none, ...)
 - b. Show the output for both the situations: one before performing fine-tuning on the model (with validation data split) and one after performing the fine-tuning on the model (with validation split).
 - c. Comment on the improvement, if any.

Submit a report with the result.