Experiment 10

The aim of the experiment is to understand the types of classification metrics used and to understand the concepts of underfitting and overfitting in machine learning.

1. Read the file 'processed.cleveland.data' using read_csv() in pandas. Fill all NA values with column means. Create feature and label matrix. The first 13 columns are the features and 14th column is the label. Scale the features and split the dataset into 60% training set and 40% test set.

2. Create a six layer neural network for classification with the following number of neurons in the four hidden layers: (128,64,64,32). Vary the number of max_iter in the following range: [50,60,70,...,1000]. Evaluate test and training accuracy in each case.

3. Use step 2 results to create two plots: training accuracy vs max_iter and test_accuracy vs max_iter. The model is said to overfit if training accuracy keeps increasing with respect to a hyperparameter and test accuracy flattens or starts decreasing with respect to the same hyperparameter.

4. Print train accuracy, test accuracy, precision score and recall score for max_iter=4000.

5. Plot confusion matrix and Receiver Operating Characteristics (ROC) for the specific case in step4.