4181788, 4178520
aabraham2@seattleu.edu, knalla@seattleu.edu

TW7: Supervised Learning

Names: Alfin Abraham, Karthik Nalla

GitHub: (Alfin) TW7 GitHub link, (Karthik) TW7 GitHub link

Summary:

Decision tree is a classification method which in general is a tree structure that simulates a flowchart, in which each internal node represents a test on an attribute, each branch a test result, and each leaf node (terminal node) a class label.

We have gained an understanding of decision tree classifier in supervised learning by applying it on breast cancer dataset from the sklearn.datasets. After importing the dataset, we split the dataset into training dataset and test dataset. Next, we used Entropy method with max_depth 2 followed by K fold validation where k value is 3. Then we obtained the summary statistics of each fold. Next, we used Gini method with max_depth 2 followed by K fold validation where k value is 3. Then we obtained the summary statistics of each fold.

Next, for both trees we have changed the parameters, Max_depth = 3,4,5 and also changed k value for cross validation such as k = 5,7,10.