Project 3: Decision Trees

The aim of this project is to practice tree structured classifiers. You are strongly encouraged to use datasets related to your research. You only need to choose one dataset to work on. Six datasets are available at `http://www.stat.psu.edu/˜jiali/stat557/material.html`.
Requirements for the project are listed below.

1. Implement a tree structured classifier using the splitting method in CART and a chosen split stopping criterion. Example stopping criteria: 1. the goodness of split is lower than a threshold, 2. the number of leaf nodes reaches a limit, 3. the number of samples in every leaf is below a threshold.

2. Optional: implement CART. If you finish this requirement, you don't need to do the first one.

3. Apply your tree structured classifier to a data set and evaluate it.

4. Write a report. Provide pseudo-code of your algorithm. Describe the decision tree obtained for your dataset: what are the candidate splits used, what are the splits in the tree, what is the goodness of every split, how many training and testing samples fall into each node, what is the resubstitution error rate of the tree. As usual, you are required to explain the contribution of each individual group member.