Name: Wei Tang
Student ID: 907 176 1275

Part 2: Plotting Learning Curves

Plot learning curves that characterize the predictive accuracy of your learned trees as a function of the training set size.

Here, I plot the figures for Heart dataset and Diabetes dataset respectively. In each plot, the data points represent training set sizes that represent 5%, 10%, 20%, 50% and 100% of the instances, and the average test-set accuracy, the minimum and maximum test-set accuracy are plotted for each training set size.
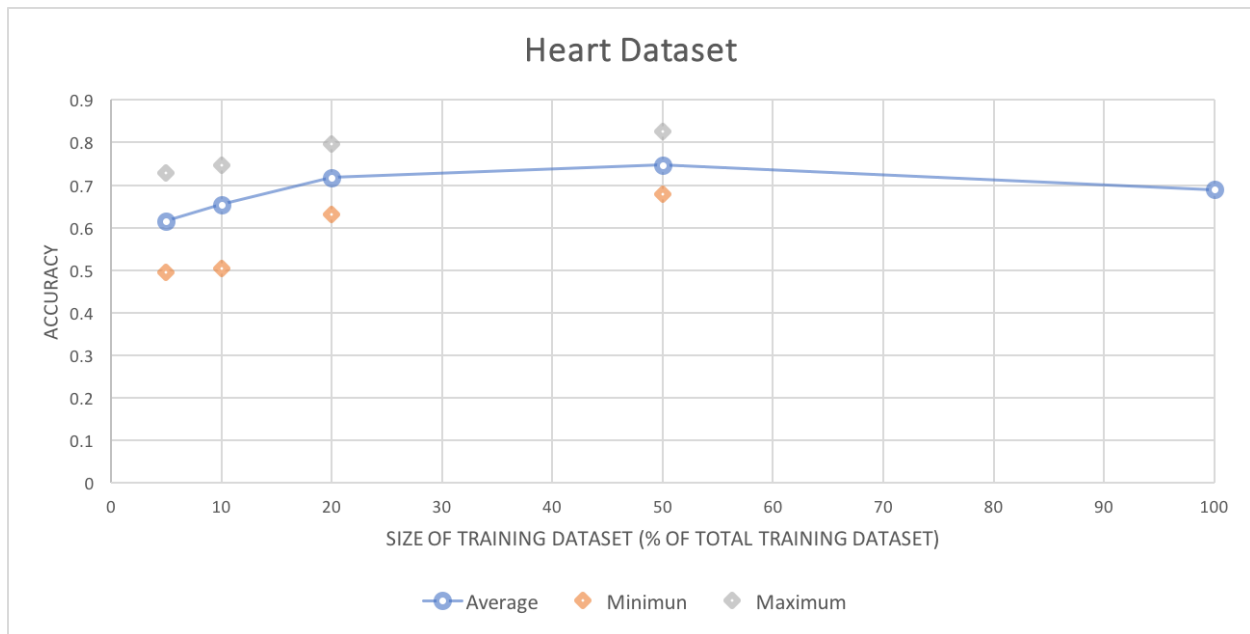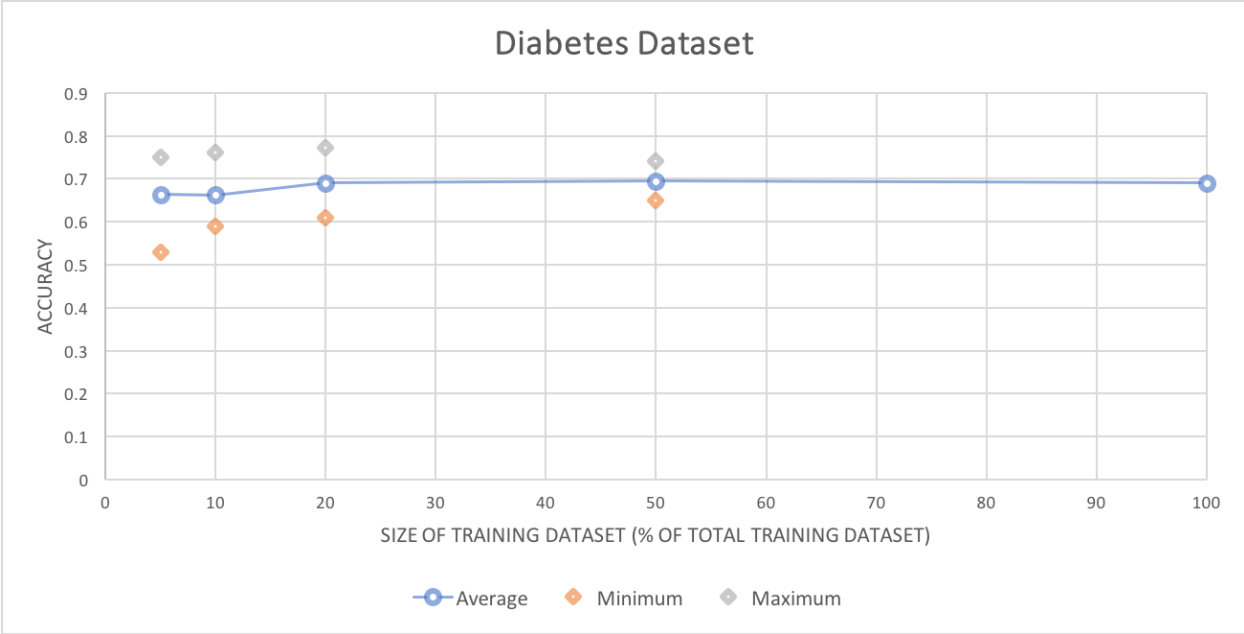

Figure 1. Heart Dataset – Part 2

Figure 2. Diabetes Dataset – Part 2

Part 3: Relation Between Predictive Accuracy and Tree Size

For both the Heart dataset and the Diabetes dataset, learn trees using the entire training set. Plotting curves showing how test-set accuracy varies with the value m used in the stopping criteria. Here, we can observe a trend that accuracy is increasing with value of m, which means, decreasing with tree size.
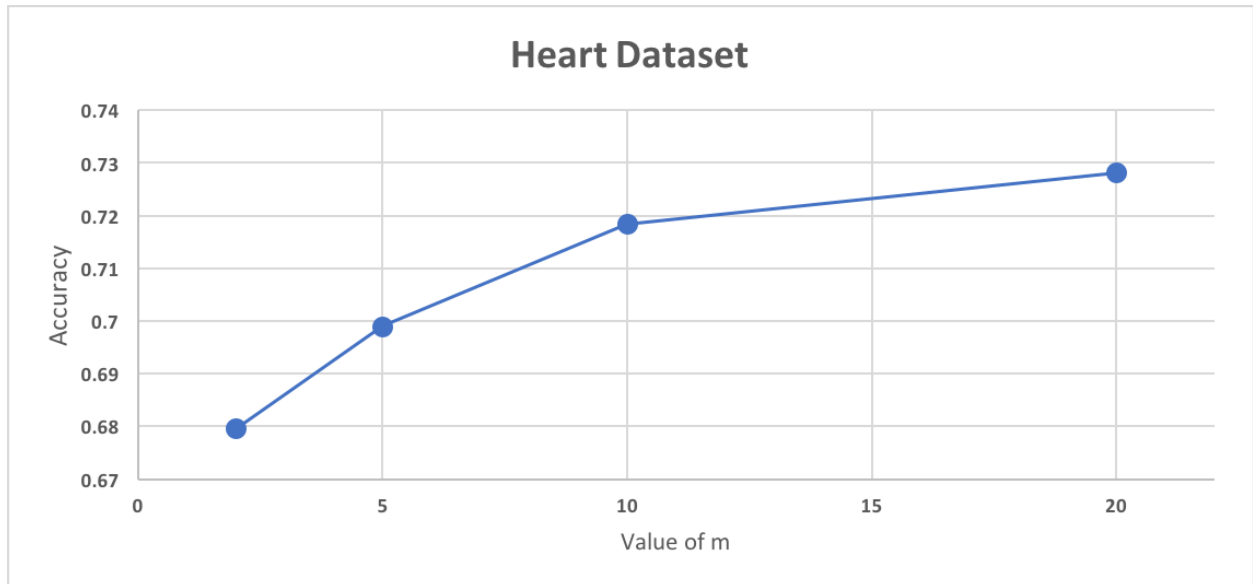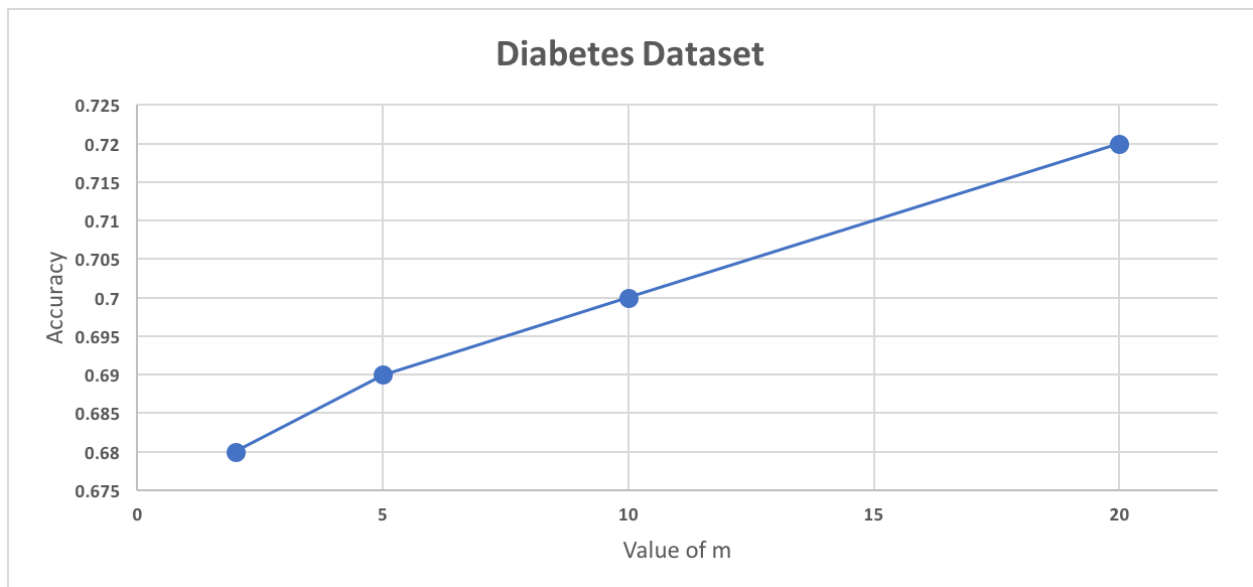

Figure 3. Heart Dataset - Part3

Figure 4. Diabetes Dataset – Part 3