Connor Johnson

CS 1675

Assignment 7

1. The restricted tree appears to be better because it has a lower error on the test set. We should always try to backprune the tree because it is possible that the tree is overfit and will perform poorly on test data.

MinLeafSize = 2

|  |  |  |  |
| --- | --- | --- | --- |
| NumSplits/ParentSize | 5 | 10 | 20 |
| 5 | 0.2271 | 0.2271 | 0.2314 |
| 15 | 0.2358 | 0.2795 | 0.2314 |
| 25 | 0.2576 | 0.2576 | 0.2404 |
| 50 | 0.2664 | 0.2838 | 0.2620 |

MinLeafSize = 4

|  |  |  |  |
| --- | --- | --- | --- |
| NumSplits/ParentSize | 5 | 10 | 20 |
| 5 | 0.2314 | 0.2314 | 0.2314 |
| 15 | 0.2795 | 0.2795 | 0.2314 |
| 25 | 0.2445 | 0.2445 | 0.2489 |
| 50 | 0.2576 | 0.2489 | 0.2358 |

The tests show that the error is lowered when the number of splits and parent size is lower. Changing the minimum leaf size slightly raised the error, but also cause the error to vary less among changing the parameters. Changing the splitcriterion variable never had any effect on the error on the test set.

