Aman Patel

April 8, 2021

CSCI-B 365

Homework 5

1. 2-class tree classifier
   1. See “HW5 Written Problems” PDF
   2. The error rate of the tree is high, but the tree does not have overfitting. If the tree had more splits and a lower error rate, the accuracy in practice would likely be smaller, as the tree may not be general enough.
      1. prune
      2. prune
      3. split
      4. split
      5. split
   3. Prune at 4 and 9 (pruning at 4 will remove node 9)
   4. If is increased by , the tree will be pruned at node 3.
2. Tree traversal in R
   1. See attached R file.
   2. See attached R file.
   3. See “HW5 Written Problems” PDF
   4. See “HW5 Written Problems” PDF
3. Cross validation
   1. When the tree is split 52 times, all terminal nodes have 0 error. The “rel error” column is the error of each node relative to the root node error
   2. Error for non-tree data is found in the “xerror” column. For row 23, the cross-validated error is .
   3. Rel error at 5 splits is
   4. A value of is best for because it has the lowest cross-validated error (
4. See attached R file.