

CSL 603 - Machine Learning

Lab 1

Decision Trees and Forests

Aditya Gupta
2015csb1003

September 2, 2017

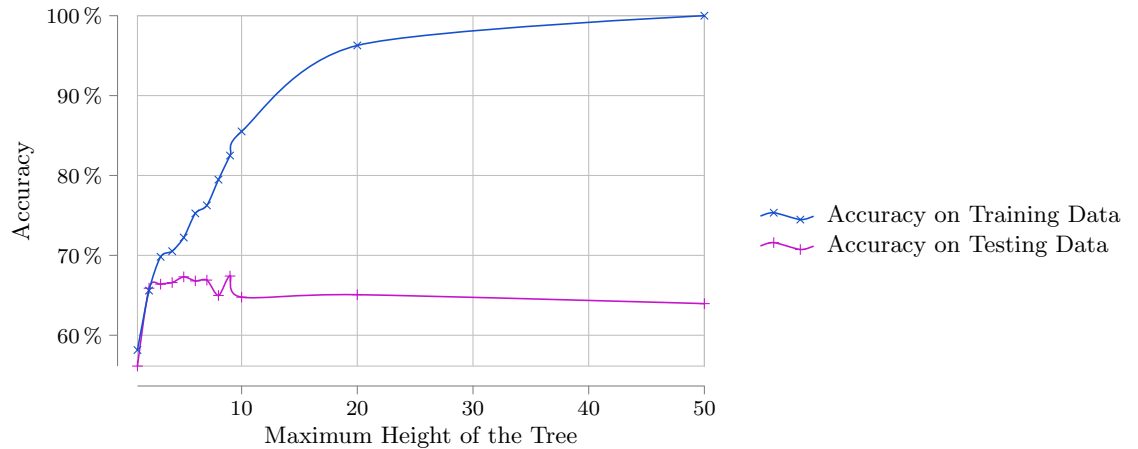
Decision Tree (Experiment 2)

A decision tree was learnt using 1000 reviews from the train set and tested on 1000 reviews from the test set. The learned tree showed the following statistics:

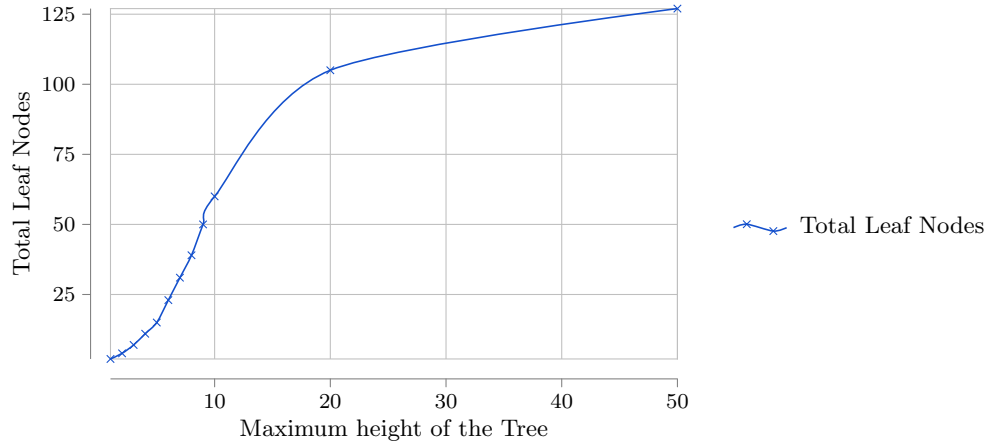
0.1 Effect of Early Stopping

The tree propagation was stopped when a critical height was reached and various properties of the tree such as total terminal nodes (leaf nodes) and the accuracy was noted. The following results were obtained:

- Increase in Training Accuracy as maximum height increased as tree more and more fitted to the training data. The Testing accuracy initially increased but then became almost constant and then slightly decreased due to overfitting.



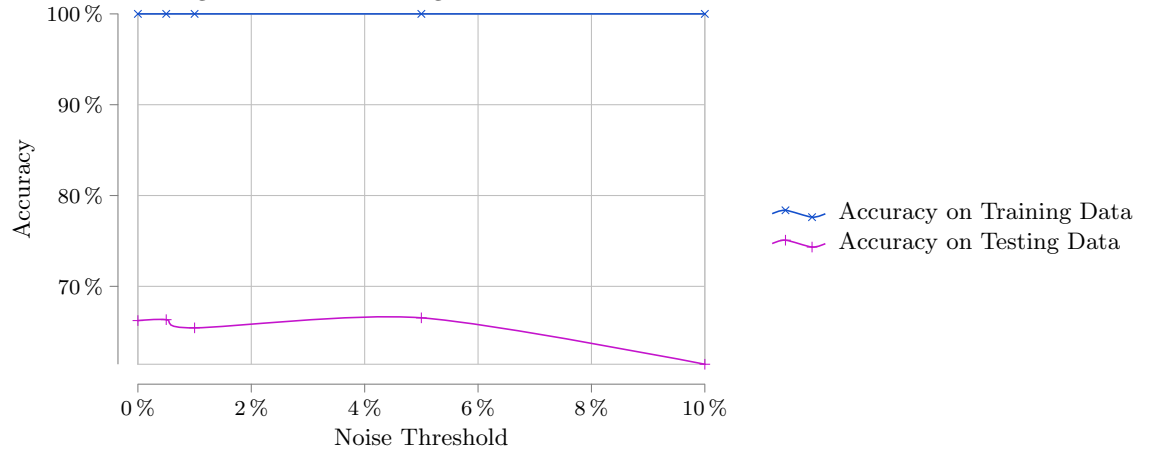
- The total number of leaf nodes increased slowly as height increased as more and more splits occurred and thus resulting in more leaf nodes.



- Some of the commonly used words to decide were “this”, “where”, “town”, “moments” and “he”.

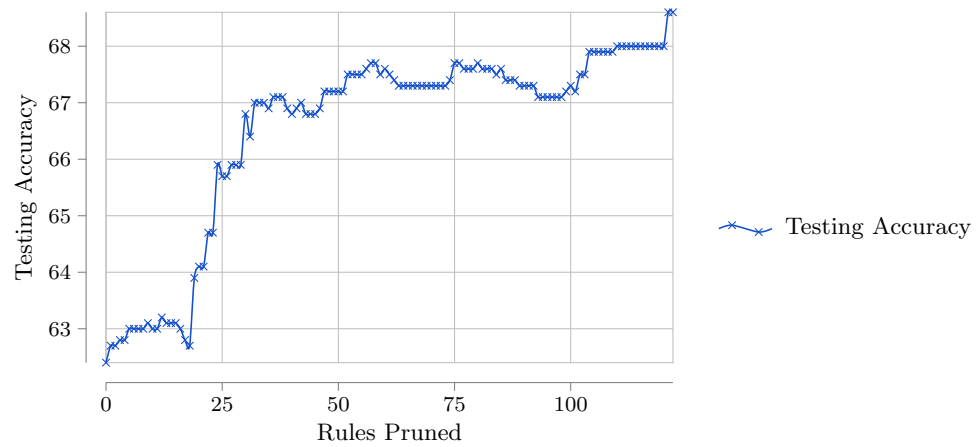
Noise (Experiment 3)

The Noise was changed and the following results were obtained:



Pruning (Experiment 4)

Each rule was pruned by removing antecedents greedily increasing the rule's accuracy and finally the sorted rules were used to calculate the accuracy of the rules.



Decision Forests (Experiment 5)

A decision forest was constructed using 5, 10, ...50 trees with \sqrt{N} where N are the attributes occurring in the 1000 trees and the following results were obtained:

