Trevor Slaton (tms45)
Joseph Vokt (jpv52)

1 Model Selection and Validation

(a) S_{train} accuracy: 22/24. S_{test1} accuracy: 8/10. S_{test2} accuracy: 16/20.

(b) S_{train} accuracy: 24/24. S_{test1} accuracy: 8/10. S_{test2} accuracy: 15/20.

(c) Let $m = \frac{(n_{01}-n_{10})^2}{n_{01}+n_{10}}$ where $n_{01} = 1$ corresponds to the number of cases which were missclassified by the decision tree hypothesis and not the linear hypothesis, and $n_{10} = 1$ corresponds to the number of cases which were missclassified by the linear hypothesis and not the decision tree hypothesis. For S_{test1}, $m = 0$. Thus we can't say that one model generalizes any better than the other because m is much less than the critical value 3.84.

(d) For S_{test2}, $n_{01} = 2$ and $n_{10} = 3$, so $m = \frac{1}{5} = .2$. Similarly, we can't say that one model generalizes any better than the other because m is much less than the critical value 3.84.

(e) The accuracy for S_{test1} stays the same for both the decision tree hypothesis and linear hypothesis. The accuracy for S_{test2} is different for the decision tree hypothesis and linear hypothesis. The χ^2 statistic for the McNemar test the same for both S_{test1} and S_{test2}.

2 Model Averaging with Decision Trees

(a)

(b)

(c)

(d)

3 Text Categorization with Decision Trees

(a)

(b)

(c)

(d)

(e)

(f)