LING 572 Homework 8
David McHugh and Chris Curtis
16 Mar 2013

Question 3

Results on the binary classification task:

N	Training accuracy	Test accuracy
1	0.452963	0.416667
5	0.614444	0.613333
10	0.684444	0.710000
20	0.753704	0.733333
50	0.837037	0.773333
100	0.897407	0.753333
150	0.926667	0.776667
200	0.945556	0.776667
250	0.962963	0.763333

From these results, we can see a rapid rise in both training and test accuracy with modest increases in number of transformations. However, while training accuracy continues to rise rapidly, test accuracy plateaus and begins to decline slightly.

This shows that the transformations are overfitting the training data. With transformation-based learning, it is conceivable that we could eventually train a model that perfectly fits the training set—essentially, a 2700-leaf-node decision tree. As these results show, that tree would generalize poorly to test data, as it is unlikely a test instance would precisely match any of those leaf nodes.