Ling 572 HW7

Daniel Campos dacampos@uw.edu

02/27/2019

1 Q3

Table 1: Results of TBL

N	Training accuracy	Test accuracy
1	0.45296	0.41667
5	0.61481	0.63667
10	0.68407	0.69667
20	0.75296	0.73000
50	0.83668	0.75000
100	0.89556	0.78333
150	0.92815	0.78333
200	0.94667	0.78333
250	0.96407	0.78000

Looking at our results, we can see how this method quickly can learn the distribution of the training data set and this helps it perform reasoably well on the test set. Towards the tail end we see a drop in accuracy on test because we have begun to overfit to the training data. I would guess that if we had a larger set and thus had more transformations in our model(249 with minGain = 1) then the model would overfit fully to the training data. Despite this minor overfitting, despite the models simplicty it performs quite well on this task.

As far as my testing goes my model does everything per the HW specifications. Q1, the trainer, runs in about 4-5 minutes on condor and inference for Q2 is near immediate.