LING 570 – HW8

Q1(e)

Table 1: Classification results for Q1(e)

	Training accuracy	Test accuracy
NaiveBayes	0.9440740740740741	0.8833333333333333
MaxEnt	0.9714814814814815	0.87
DecisionTree	0.5981481481481481	0.6
Winnow	0.727037037037037	0.6833333333333333
BalancedWinnow	0.9714814814814815	0.8633333333333333

Q1(f)

The Naïve Bayes, Max Entropy and Balanced Winnow trainers produced classifiers with the highest training accuracy. These classifiers in turn produced the highest test accuracy. On the other hand, the Decision Tree trainer has the lowest classifier training accuracy along with the lowest test accuracy as compared to the rest.

Another observation was, for the Decision Tree trainer the accuracy difference between the classifier training accuracy versus the test accuracy (0.5981481481481481 versus 0.6) was the closest among the five trainers. In fact, in the case of DecisionTree trainer the test accuracy was slightly higher than the training accuracy.

Q1(g)

command 1:

> vectors2train - -training-file ~/dropbox/09-10/570/hw8/examples/train.vectors - -trainer MaxEnt - -output-classifier q1g_model - - report train:accuracy train:confusion > q1g.stdout 2>q1g.stderr

command 2:

> classify - -testing-file ~/dropbox/09-10/570/hw8/examples/test.vectors - -classifier q1g_model - - report test:accuracy test:confusion test:raw > q1g_res.stdout 2>q1g_res.stderr

Yes, we would get the same result as in (e) using the vectors2train and classify commands.

Q1(h)

command:

> classifier2info - -classifer q1g_model > q1g_model.txt

Q2

Table 2: Classification results for Q2(c)-(e)

	Training accuracy	Test accuracy
(3) three talk.politics.* groups	0.9685185185185186	0.8266666666666667
(4) four sci.* groups	0.9980555555555556	0.8625
(5) four rec.* groups	0.9977777777777778	0.95

End of HW8 – submitted by Wee Teck Tan

Student ID: 0937003

Course Name: LING 570