

LING 572 – HW3

Q1

Commands:

```
info2vectors - -input train.vectors.txt - -output train.vectors
```

```
info2vectors - -input test.vectors.txt - -output test.vectors - -use-pipe-from train.vectors
```

```
vectors2classify - -training-file train.vectors - -testing-file test.vectors - -trainer  
NaiveBayes > q1.stdout 2>q1.stderr
```

Training accuracy = 0.94259

Test accuracy = 0.88333

Q2-3

Table 1: Results of **Bernoulli** NB model

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.93037037037037	0.88
0.5	0.91037037037037	0.863333333333333
1.0	0.897037037037037	0.84
2.0	0.87962962962963	0.823333333333333

Q4

Table 2: Results of **multinomial** NB model

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.957037037037037	0.913333333333333
0.5	0.95037037037037	0.906666666666667
1.0	0.944814814814815	0.9
2.0	0.94	0.896666666666667

Q5

Table 3: Results of **multinomial** NB model with **binary** features

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.95962962962963	0.91
0.5	0.956296296296296	0.903333333333333
1.0	0.952962962962963	0.896666666666667
2.0	0.947777777777778	0.896666666666667

Q6(a)

From the experiment above, we can see that the multinomial NB model consistently performs better than the Bernoulli NB model in terms of both training and test classification accuracies. Although the performance for both NB models have shown a declining trend as the conditional probability delta was increased, the multinomial NB model seems to be least affected by changes to the conditional probability delta value. The classification results also suggest that the performance of the multinomial NB model is more uniform across all conditional probability delta values as compared to the Bernoulli NB model.

Binarizing the features in the data improved the training accuracy of the multinomial model marginally and there were very little or no changes to the test accuracy overall.

Q6(b)

The $P(f|c)$ values produced by the Bernoulli NB model are generally higher as compared to the multinomial NB model. For a general feature with a high $P(f|c)$ value where the same high $P(f|c)$ value can also be found across different classes, it tells us nothing more than the fact that it can have the same likelihood of presence across different classes. Hence, such feature may not necessarily be an important feature for classification and an example of such feature is “a”.

For a subject matter specific feature, a high $P(f|c)$ value is often found in a given class along with lower $P(f|c)$ values of the same feature found in other classes. This would tell us that the feature is an important one for classification. Examples of subject matter specific features are “guns” to gun class and “israel” to mideast class.

Depending on the vocabulary size, the importance of a feature given its $P(f|c)$ value in the Bernoulli NB model may be reduced or emphasized during classification since the entire vocabulary would be considered in the process. In general, the Bernoulli NB model yields higher classification accuracy with smaller vocabulary size (McCallum and Nigam, 1998).

Q6(c)

Although the $P(f|c)$ values produced by the multinomial NB model are generally lower as compared to those produced from the Bernoulli NB model, we can identify important features from the $P(f|c)$ values. Similar to the Bernoulli NB model, we can have high $P(f|c)$ values occurring across different classes for a generic feature. We can also have high $P(f|c)$ value for subject matter specific feature that occurs only within a specific class and that would provide an indication on the importance of such feature.

Unlike the Bernoulli NB model, the importance of the $P(f|c)$ value is not affected by the size of the vocabulary. In fact, the multinomial NB model has been shown to outperform the Bernoulli NB model over larger vocabulary sizes (McCallum and Nigam, 1998). In addition, unlike the Bernoulli NB model, a feature with a high or low $P(f|c)$ value of significance in the multinomial NB model may be affected by the actual sum of occurrence of this feature in the documents since these information are considered in the classification process.

Q6(d)

From the experiment, we did not find any performance difference between the two models across all conditional probability delta scenarios.

End of HW3 – Joint submission by

Scott Mantei, Student ID:

Wee Teck Tan, Student ID: 0937003

Course Name: LING 572