Report : Assignment – 5

Training the Naïve Bayes Classifier Model

First, I built the Naïve Bayes Sentiment classifier, using an example which is described on pages 79-80 (Section 6.3, Worked Example) of the textbook. This involved calculation of the prior probabilities, and then calculation of the likelihoods and finally the class chosen by a Naïve Bayes classifier for a given sentence, the equation for which is given on page 78 of the textbook (Equation 6.9).

Next, I extended the program to be able to take as input both the training sets that were provided, and when I executed the classifier on the positive and negative training sets after training on both, I got accuracies of 62% and 80% respectively (on training sets).

In other words I built a unigram language model for each of the two classes (positive and negative), and assessed the probability of each test case/sentences with respect to each of the two models. The program then predicts the highest probability class as the correct choice.

I used a stop-list from nltk to filter out the most commonly used words.

Then I used a similar procedure build the deception classifier, using a Naïve Bayes classifier model and training on the datasets (True and False reviews) given for this assignment.

When I trained only on the datasets for this assignment, and tested on them, for

(A) The False review set, I got an accuracy of 98.14%
(B) The True review set, I got an accuracy of 77.10%

Hence the average accuracy I achieved on the training set was 87.62%.

Finally, I modified the program so that the

(1) Sentiment classifier returned both the probabilities for a review to be positive and negative
(2) Deception classifier returned both the probabilities for a review to be true and false

And then combined these probabilities for each review to get four probabilities for each review:

(a) True and Positive
(b) True and Negative
(c) False and Positive
(d) False and Negative

by adding the individual (respective) probabilities, since I was dealing with the logarithm of the probabilities.

Deception Classification/Detection

To classify a review at the end, I tagged it with the deception tag (True/False) of the highest probability from the four above-mentioned probabilities for the review in question.

With this method, after training on the training sets (Sentiment and Deception) and testing on:

(C) The False review set, I got an accuracy of 99.07%
(D) The True review set, I got an accuracy of 38.79%

Hence the average accuracy I achieved on the training set was 68.93%.

So, although the accuracy on the False reviews went up by a bit, the accuracy on the True review set dropped significantly.