**CSE-4211**

**Machine Learning**

**Assignment-1**

**Md. Asaf-Uddowla Golap**

**Roll: 1207005**

**Dept. of CSE**

**KHULNA UNIVERSITY OF ENGINEERING & TECHNOLOGY**

**Introduction:**

In machine learning several types of classifiers are used. Naive Bayes is a simple technique for constructing classifiers that is based on the popular Bayes probability theorem. models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle. All naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification.

**Data Set Description:**

I choose a movie review data set, where total 2000 review. A folder ‘neg’ contain 1000(.txt) negative comment and another folder ‘pos’ contain 1000(.txt) positive comment.

**Experiment setup and Result Analysis:**

According to Bayes theorem, P( A|B) =

To expand this using Bayes theorem,

we have P(C|d) = =

P(d) is same for all of the classes. Therefore, we can omit the this and we have following

equation: P(C| d) = P(d| C)P(C)

Now for our data we can compute the class probability using following formula:

P(comment/C=[neg or pos]) =

Lets split the comment in it’s constituent tokens such as comment = w1, w2, w3, w4, w5, … wn.

= argmax P(d = w1 w2w3w4w5..wn| C) P(C)

Assuming terms are independent, we have:

= argmax P(C)P(w1| C)P(w2| C)P(w3| C)..P(wn| C)

= argmax P(C) Ħwi ∈ V P(wi| C) , V is the Vocabulary

= argmax log P(C) + Ħwi ∈ V logP(wi| C) , V is the Vocabulary

From this we got

confusion Matrix: neg pos

neg [6571, 16]

pos [12, 6575]

Error: 0.212539851222 %

**References:**

[1] [*https://www.cs.cornell.edu/people/pabo/movie-review-data/*](https://www.cs.cornell.edu/people/pabo/movie-review-data/)

[2] [*https://en.wikipedia.org/wiki/Naive\_Bayes\_classifier*](https://en.wikipedia.org/wiki/Naive_Bayes_classifier)

[3] [*http://www.kuet.ac.bd/webportal/ppmv2/uploads/1477472586CSE4211%20Assignment%201.pdf*](http://www.kuet.ac.bd/webportal/ppmv2/uploads/1477472586CSE4211%20Assignment%201.pdf)