Naive Bayes Classifier

BVS Ruthvik 2019A7PS0017H
R Vedang. 2019A7PS0150H
Rahul Mehndiratta 2017A7PS1479H

Introduction:

In statistics, **Naive Bayes classifiers** are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models, but coupled with kernel density estimation, they can achieve higher accuracy levels.

Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

Problem Definition and Implementation:

First, given data was loaded into a df - df_original by seperating the comments/messages from the corresponding labels. Then , two dataframes were created, df_raw which contained all the training data and df_test which contained all the testing data. df_positive contains all the data pertaining to the positive reviews and df_negative contains all the data of the negative reviews. The working of the functions used can be found in their docstrings. 2 dictionaries were created, dict_positive and dict_negative - each having the number of occurences of unique positive and unique negative words from their respective comments. ie. each key in, say dict_positive is a unique word contained in the positive comments and its value is the number of times it appeared in the positive comments. Then, laplace smoothening was used to find out the probability that a word can be observed, given that it belongs to a particular category. And to find out the probability that a list of words belong to a particular category,multiply the individual probabilities, assuming they are independent of each other. Then, the probability was calculated that a list of words belongs to a particular category, using the bayes theorem, then based on which probability is higher, categorize an example as either positive or negative.

Observations:

Accuracy in fold 1 is 82.44274809160305
Accuracy in fold 2 is 86.25954198473282
Accuracy in fold 3 is 75.57251908396947
Accuracy in fold 4 is 83.96946564885496
Accuracy in fold 5 is 77.09923664122137
Accuracy in fold 6 is 81.67938931297711
Accuracy in fold 7 is 80.1526717557252

Average accuracy across all folds is 0.8102508178844057

Major limitation of the Perceptron classifier :

1) Assumption that the feature parameters predict the probability independently of one another.
2) The order of the words is ignored while implementing the algorithm.
3) Without laplace smoothening, the presence of a word absent in the original dataset results in 0 probabilities for the whole sentence.