# Methods in classifier framework

## classification.NaiveBayes

*buildDataModel ()* - Constructs the data structure such that, the probabilities can be calculated easily while constructing the model

*buildTrainData()* - Coverts data in train file to array

*buildTestData()* - Converts data in test file to array

*buildClassifier()* - Builds the classifier model

*classifyTestData()* - Uses the classifier model to classify the test data

*measureClassifierQuality()* - Measures the quality of classifier by counting four parameters

*countXGivenY()* - Calculates count of A given condition B. Used in classifier model to calculate the conditional probability.

## classification.AdaBoost

*AdaBoost()* - Gets an instance of naiveBayes to construct the weak classifiers

*boost()* - Forms 5 classifierModel based on the error and updatedWeights

*calculateError()* - Calculates the error of the new classifier model. The train data is made the test data to calculate the error

*sampleDataBasedOnWeight()* - Implements roulette-wheel selection to sample based on weight and updates the weights array accordingly

*ensemble()* -Used the 5 classifierModels and its weight to classify the test data

*measureClassifierQuality()* - Measures the quality of the ensemble classifier by counting four parameters

# Metrics Measured

## Data Set Chess:

| | Naïve Bayes | AdaBoost |
|---|---|---|
| Accuracy | 86.34590377 | 86.60598179 |
| Error rate | 13.65409623 | 13.39401821 |
| Sensitivity | 89.52618454 | 86.78304239 |
| Specificity | 89.52618454 | 86.78304239 |
| Precision | 85.07109005 | 87.43718593 |
| F-1 Score | 87.2417983 | 87.10888611 |
| Fbeta Score(0.5) | 85.92628052 | 87.30556949 |
| Fbeta Score(2) | 88.5982231 | 86.91308691 |

## Data Set Nursery:

| | Naïve Bayes | AdaBoost |
|---|---|---|
| Accuracy | 100 | 100 |
| Error rate | 0 | 0 |
| Sensitivity | 100 | 100 |
| Specificity | 100 | 100 |
| Precision | 100 | 100 |
| F-1 Score | 100 | 100 |
| Fbeta Score(0.5) | 100 | 100 |
| Fbeta Score(2) | 100 | 100 |

## Data Set Mushroom:

| | Naïve Bayes | AdaBoost |
|---|---|---|
| Accuracy | 96.12365064 | 99.36211973 |
| Error rate | 3.876349362 | 0.637880275 |
| Sensitivity | 99.34944238 | 100 |
| Specificity | 99.34944238 | 100 |
| Precision | 93.68974584 | 98.80624426 |
| F-1 Score | 96.43662607 | 99.39953811 |
| Fbeta Score(0.5) | 94.76950355 | 99.04270987 |
| Fbeta Score(2) | 98.16345271 | 99.75894678 |

### Data Set Led24:

| | Naïve Bayes | AdaBoost |
|---|---|---|
| Accuracy | 88.12877264 | 88.12877264 |
| Error rate | 11.87122736 | 11.87122736 |
| Sensitivity | 88.38174274 | 88.38174274 |
| Specificity | 88.38174274 | 88.38174274 |
| Precision | 87.29508197 | 87.29508197 |
| F-1 Score | 87.83505155 | 87.83505155 |
| Fbeta Score(0.5) | 87.51027116 | 87.51027116 |
| Fbeta Score(2) | 88.16225166 | 88.16225166 |

# Parameters used

**No of iterations** - 5

*Reason:* The accuracy for mushroom got close to 99% for 5$^{th}$ iteration. The accuracy of chess and led24 did not boost much. So I have set the iterations as 5

**Laplacian Adjustment** – Added 1 to the numerator and number of distinct values in the dimension to the denominator.

*Reason:* To avoid the predicted probabilities from becoming zero.

# Does AdaBoost boost NaiveBayes

While the accuracy for nursery data set is 100 for both. AdaBoost increased accuracy for Mushroom data set but did not increase accuracy for chess and led24 significantly.

# Is NaiveBayes ensemble compatible

Looking at the results my guess is that NaiveBayes is not ensemble compatible since the accuracy for 2 data sets did not increase significantly. While I tried verifying if my guess right in Google, I read few research papers which employs a modified version of NaiveBayes[Introducing tree structure or decision stumps] with AdaBoost as their previous results shows that AdaBoost does not increase accuracy of NaiveBayes. Ref : http://onlinelibrary.wiley.com/doi/10.1111/1467-8640.00219/abstract