Assignment 3

Performance Analysis of Classifiers

Prasad Pande (psp150030)

Output of the code:

```
     [,1]                      [,2]  [,3]       [,4]       [,5]       [,6]       [,7]
 [1,] "house-votes-84.data 1" "0.8" "96.55 %" "96.55 %" "96.55 %" "96.55 %" "95.4 %"
 [2,] "house-votes-84.data 2" "0.9" "95.45 %" "97.73 %" "97.73 %" "95.45 %" "88.64%"
 [3,] "car.data 1"            "0.8" "93.93 %" "93.06 %" "97.4 %"  "87.28 %" "85.55%"
 [4,] "car.data 2"            "0.9" "94.22 %" "67.63 %" "95.38 %" "89.6 %"  "84.39%"
 [5,] "SPECT.data 1"          "0.8" "87.04 %" "74.07 %" "77.78 %" "85.19 %" "55.56%"
 [6,] "SPECT.data 2"          "0.9" "74.07 %" "74.07 %" "74.07 %" "74.07 %" "66.67%"
 [7,] "wine.data 1"           "0.8" "91.67 %" "36.11 %" "97.22 %" "100 %"   "100 %"
 [8,] "wine.data 2"           "0.9" "94.44 %" "33.33 %" "94.44 %" "100 %"   "94.44%"
 [9,] "heart.data 1"          "0.8" "63.93 %" "57.38 %" "50.82 %" "62.3 %"  "45.9 %"
[10,] "heart.data 2"          "0.9" "51.61 %" "54.84 %" "51.61 %" "61.29 %" "51.61 %"
```

Description
Column 1 – Name of the dataset
Column 2 – Training set to Test set split
Column 3 – Decision Tree Accuracy
Column 4 – Perceptron Accuracy
Column 5 – Neural Network Accuracy
Column 6 – Support Vector Machine Accuracy
Column 7 – Naïve Bayesian Accuracy

Reporting Result

Dataset	Number of toltal Instances	Number of Attributes	Percepnt Split	Decision Tree	Percep-tron	Neural Networks	Support Vector Machines	Naïve Bayesian
House-Votes	436	16	80/20	96.55	96.55	96.55	96.55	95.4
House-Votes	436	16	90/10	95.45	97.73	97.73	95.45	88.64
Car Data	1729	8	80/20	93.93	93.06	97.4	87.28	85.55
Car Data	1729	8	90/10	94.22	67.63	95.38	89.6	84.39
Spect Heart Data	268	22	80/20	87.04	74.07	77.78	85.19	55.56
Spect Heart Data	268	22	90/10	74.04	74.07	74.07	74.07	66.67
Wine Data	178	13	80/20	91.67	36.11	97.22	100	100
Wine Data	178	13	90/10	94.44	33.33	94.44	100	94.44
Heart Data	303	13	80/20	63.93	57.38	50.82	62.3	45.9
Heart Data	303	13	90/10	51.61	54.84	51.61	61.29	51.61

Conclusion:

As per the experiment, it is difficult to decide which dataset it best for all kind of datasets. This depends on many factors important of which is the kind of data set we are using for the experiment.

If we are using data set which is linearly separable, a single unit of perceptron can do our job. It is very easily implementable if you have only binary classes.

For non-linear and complex dataset one can go with the multi-layer perceptrons i.e. Neural Networks or Support vector machines.

For categorical dataset decision tree can give you best accuracy but decision tree is susceptible to noise.

All these classifiers except Naïve Bayesian are discriminative classifiers because they learn boundary between them while Naïve Bayesian is a probabilistic classifier which finds the distribution of the datasets for individual classes. Naïve Bayesian can be used for any categorical dataset which gives us the distribution of data over the class variables.