# Assignment 3 – ML – Empirical study of different training models

## Dharmam Buch – DSB150030

1. We would be using the following data sets :-
   1. Wine
   2. Car
   3. Haberman
   4. Abalone
   5. Iris
2. Summary Of Observation :-

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Number Of Instance** | **Number Of Attributes** | **Split(Training/ Test)** | **Decision Tree(%)** | **Perceptron(%)** | **Neural Network(%)** | **Support Vector Machine(%)** | **Naïve Bayes Accuracy(%)** |
| Wine | 178 | 13 | 135/43 | 72 | 90.4 | 90.67 | 97.7 | 97.7 |
| Car | 1728 | 6 | 1588/140 | 93.57 | 55 | 71.42 | 80 | 60 |
| Haberman | 306 | 4 | 248/58 | 72.4 | 91.3 | 74.3 | 72.4 | 74.3 |
| Abalone | 4177 | 9 | 3978/199 | 25.6 | 42.71 | 24.2 | 25.6 | 23.6 |
| Iris | 150 | 5 | 117/33 | 93.9 | 97.9 | 100 | 97.9 | 97.9 |
| **Total** |  |  |  | 71.494 | 75.462 | 72.118 | 74.72 | 70.7 |

Note : **Abalone was a skewed data set,** for example, Rings = 1,2,3 had 15/30 training data each and rings like 28,29 had just 1 training example. Still The SVM performed better than others.

1. Explanations :-

3.1. A general study on different details can be summarized as the following.

**3.2 We were also able to prove the same in our experiment while , SVM is the most accurate, the Decision Tree is the most robust of all the models.**

3.3 **Neural network is a close second** but it is very complex and for large and complex data like abalone, it needs lots of tweaking for hidden layers and threshold.

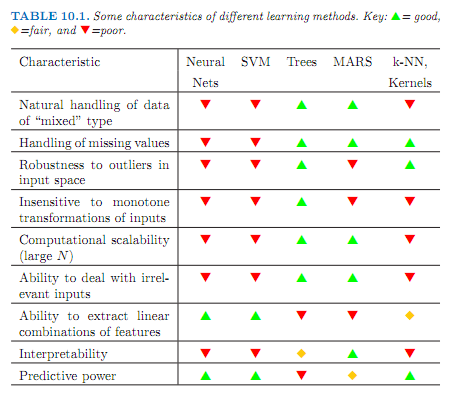
Without which it sometime fails to converge on a model with the following error.

Error in plot.nn(nn) : weights were not calculated

In addition: Warning message:

algorithm did not converge in 1 of 1 repetition(s) within the stepmax

3.4 Naïve Bayes classifier was the most fast and easy to implement, however, failed to perform when the data independence in the data set was not easily implied.



3.4 The notion that the complex and mixed data makes impacts neural network and SVM can be seen with the dip in performance in Haberman and Abalone Data sets.

3.5 Multi-class were implemented using a hidden= 0 parameter in the neural network code.

They performed better then NN in complex situation as they save a lot in back propagation and their simplicity.

**4. Hence, the experiment was fairly successful as all the concepts taught in class were numerically visualized and validated.**