

Report - Assignment 1

Foundations of Machine Learning

Prof. Vineeth N Balasubramanian

Naitik Malav | CS19BTECH11026

Deliverables:

1. I have submitted answers of question 1 and 2 in a hand-written pdf format – CS19BTECH11026_FoML.pdf
2. I have submitted two jupyter notebooks for Decision Tree Implementation.
 - a. In first notebook – Decision_Tree_Entropy.ipynb I have implemented decision tree using the information gain, binary univariate and entropy. Information gain is calculated using entropy according to the formula discussed in class. Below is the average accuracy computed using 10-fold cross validation algorithm for the provide dataset- *"wine-dataset.csv"*.
Average Accuracy – 81.19%
 - b. In second notebook – Decision_Tree_Gini.ipynb I have implemented decision tree using the information gain, binary univariate and gini index. Information gain is calculated using gini index according to the formula discussed in class. Below is the average accuracy computed using 10-fold cross validation algorithm for the provide dataset- *"wine-dataset.csv"*.
Average Accuracy – 84.92%

Conclusion:

Average Accuracy is more in case of gini index implementation in comparison to entropy. Gini Impurity is more efficient than entropy in terms of computing power.

Gini impurity it goes up to 0.5 and then it starts decreasing but in case of entropy it increases up to 1 hence, entropy requires more computational power.

The range of Entropy lies in between 0 to 1 and the range of Gini Impurity lies in between 0 to 0.5. *Hence, we can conclude that Gini Impurity is better as compared to entropy for selecting the best features.*