

Decision Tree Implementation

Accuracy on Entropy function: 81.1861%

Cross-Validation:

Accuracy on Entropy function: 82.0328%

Improvement Strategies:

Accuracy on Gini Impurity : 81.5951%  
Accuracy on Cross Validation Gini impurity : 82.1144%

Entropy is a criteria to measure the impurity of data. Gini, a criterion to minimize the misclassification of data. Entropy slower to compute because of log function. Gini will tend to find the largest class, and “entropy” tends to find groups of classes that make up approx. 50% of the data. Entropy tries to maximize the mutual information.

Both Gini and Entropy has almost same result and we can see in the below figure.

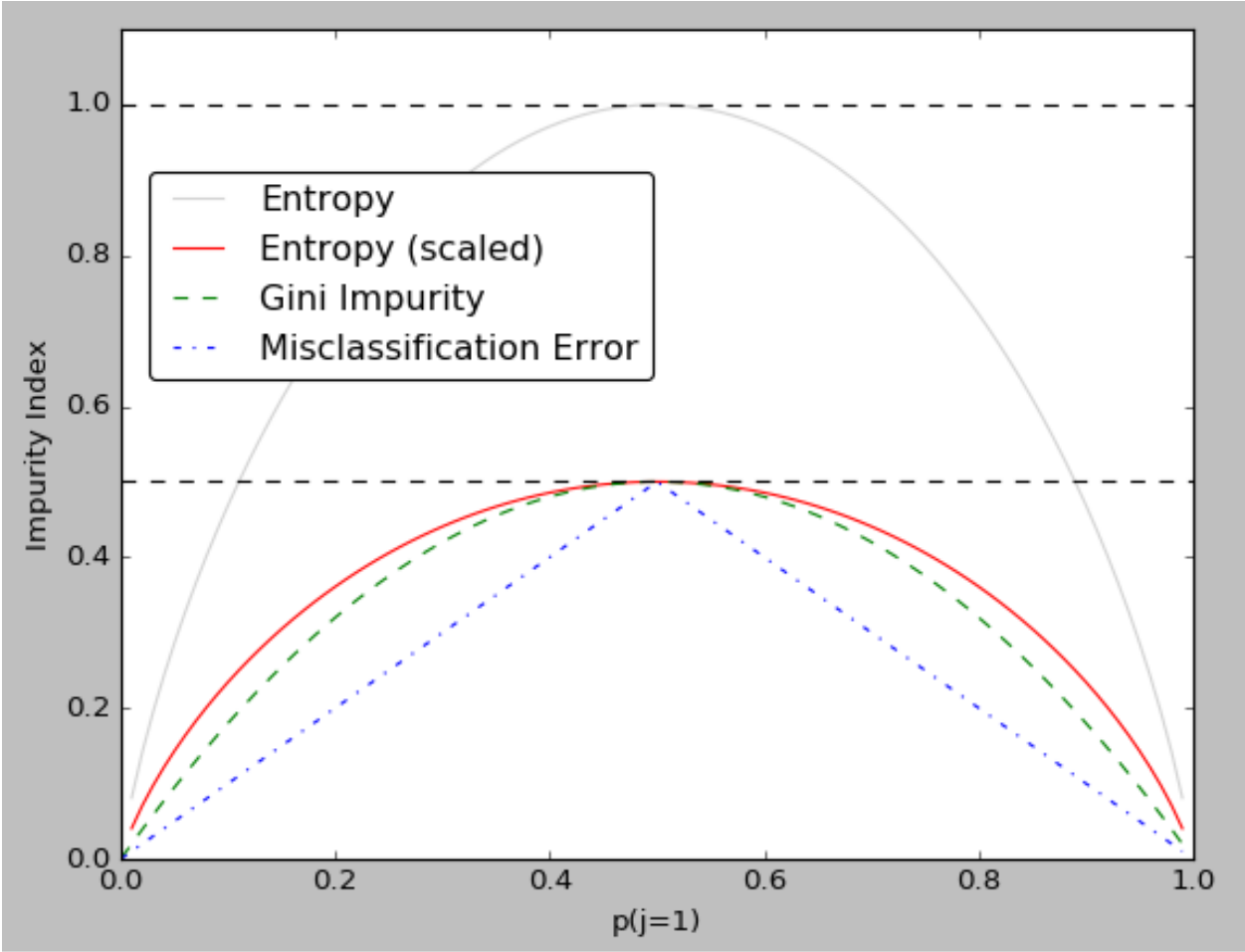


Figure 1: Gini vs Entropy

Our result is also displaying the same.