Notes

Gini Coefficient

$$\text{gini}(T) = 1 - \sum_{i=0}^{n} p_i^2$$

$$\text{gini}_A(T) = \frac{S_1}{S_1 + S_2}\ \text{gini}(T_1) + \frac{S_2}{S_1 + S_2}\ \text{gini}(T_2)$$

Information Entropy

$$H(X) = -\sum_i p_i \log_2(p_i)$$

$$H_A(X) = \frac{S_1}{S_1 + S_2} H(X_1) + \frac{S_2}{S_1 + S_2} H(X_2)$$

Information Gain

$$\text{Gain}(A) = H(X) - H_A(X)$$

Regression Decision Tree

It uses MSE as the split criteria, instead of Gini or Entropy.

$$\text{MSE} = \frac{1}{n} \sum (y^{(\text{i})} - \hat{y}^{(\text{i})})$$