

Extended Machine Learning Formulas

Gini Index (for Decision Trees)

1. Gini Index for a binary split:

$$Gini(S) = 1 - \sum_{i=1}^k p_i^2$$

Where: - S : Set of examples - p_i : Proportion of class i in set S - k : Number of classes

2. Weighted Gini for attribute A:

$$Gini(S, A) = \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot Gini(S_v)$$

Where: - S_v : Subset of S where attribute $A = v$



Log Loss / Binary Cross-Entropy (for Classification)

$$\text{LogLoss} = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(p^{(i)}) + (1 - y^{(i)}) \log(1 - p^{(i)}) \right]$$

Where: - m : Number of samples - $y^{(i)}$: True label (0 or 1) for sample i - $p^{(i)}$: Predicted probability for sample i



Weighted Cross-Entropy Loss (for Imbalanced Classes)

$$\text{Weighted LogLoss} = -\frac{1}{m} \sum_{i=1}^m \left[w_1 \cdot y^{(i)} \log(p^{(i)}) + w_0 \cdot (1 - y^{(i)}) \log(1 - p^{(i)}) \right]$$

Where: - w_1 : Weight for class 1 (positive) - w_0 : Weight for class 0 (negative)

This is used to penalize misclassification of rare classes more heavily.

(End of Extended Formula Reference)