1 Decision Tree Classifier

1.1 Entropy and Information Gain

The entropy is a measure of impurity or diversity used in decision trees. A lower entropy indicates a purer node. The entropy of a dataset S is defined as:

$$H(S) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

where p_i is the proportion of class i in the dataset.
The information gain of an attribute A with respect to dataset S is defined as:

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v)$$

where:

- Values(A) is the set of all possible values of attribute A,
- S_v is the subset of S for which attribute A has value v,
- $|S_v|$ and $|S|$ are the sizes of sets S_v and S respectively.

1.2 Gini Index

The Gini index is a measure of impurity or diversity used in decision trees. A lower Gini index indicates a purer node. The Gini index for attribute value $a = a_j$ is defined as:

$$Gini(a = a_j) = 1 - \sum_{i=1}^{c} (p(i \mid j))^2$$

where:

- c is the number of classes,
- $p(i \mid j)$ is the probability of class i given the attribute value a_j.

The overall Gini index for an attribute a is a weighted average of the Gini indices for each of its values:

$$Gini(a) = \sum_{i=1}^{m} \frac{n_i}{n} Gini(a = a_i)$$

where:

- m is the number of distinct values of attribute a,
- n_i is the number of instances where $a = a_i$,
- n is the total number of instances,
- $Gini(a = a_i)$ is the Gini index for value a_i.

2 Decision Tree Regressor

2.1 Variance and Variance Reduction

In decision tree regressors, variance is used to measure the impurity of a node. The variance of a set of target values S is given by:

$$Var(S) = \frac{1}{|S|} \sum_{i=1}^{|S|} (y_i - \bar{y})^2$$

where:

- y_i is the target value of instance i,
- $\bar{y}$ is the mean of all target values in set S,
- $|S|$ is the number of instances in the set.

The variance reduction after splitting a dataset S into subsets $S_1, S_2, \ldots, S_k$ based on an attribute is calculated as:

$$\text{VarianceReduction}(S, A) = Var(S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} Var(S_i)$$

where:

- $Var(S)$ is the variance of the original dataset,
- $Var(S_i)$ is the variance of subset S_i,
- $|S_i|$ and $|S|$ are the sizes of subset S_i and the full set respectively.