















menu

› Course Shortcuts

› Student Lounge

› Q&A

▼ Create Decision Trees

-  Module Introduction: Create Decision Trees
-  Notations for Machine Learning
-  Build a Decision Tree
-  Explore Decision Trees
-  Measure a Partition's Purity
-  Formalize Impurity Functions
-  Determine the Minimum Entropy Split
-  CART Algorithm
-  CART Algorithm
-  CART Cheat Sheet
-  Minimize Impurity of Regression Trees
-  **Squared-loss Impurity**
-  Linear Time Regression Trees
-  Implement Regression Trees
-  Module Wrap-up: Create Decision Trees

› Choose the Right Model

› Data Science in the Wild

› Course Resources



Squared-loss Impurity

As discussed in the video, we can tweak classification trees for continuous valued labels (regression). Because labels are no longer categorical, we redefine impurity such that it captures the "spread" of values in each node. Conveniently, we can capture this spread within a set by the variance of the labels within the set. The prediction made by a regression tree for a leaf with corresponding set \mathcal{S} is simply the mean label \bar{y} . With this mean label as the predictor, the variance impurity is identical to the squared loss:

$$I(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}, y) \in \mathcal{S}} (y - \bar{y})^2 \leftarrow \text{Average squared difference from average label}$$
$$\text{where } \bar{y} = \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}, y) \in \mathcal{S}} y \leftarrow \text{Average label}$$

Finding the best split

Remember, you evaluate the quality of a split of a parent set \mathcal{S} into two sets \mathcal{S}_L and \mathcal{S}_R by the weighted impurity of the two branches:

$$\frac{|\mathcal{S}_L|}{|\mathcal{S}|} I(\mathcal{S}_L) + \frac{|\mathcal{S}_R|}{|\mathcal{S}|} I(\mathcal{S}_R)$$

In the case of the squared loss, this becomes:

$$\frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}, y) \in \mathcal{S}_L} (y - \bar{y}_{\mathcal{S}_L})^2 + \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}, y) \in \mathcal{S}_R} (y - \bar{y}_{\mathcal{S}_R})^2$$

In the CART-ID3 algorithm, we evaluate all possible splits at each parent node \mathcal{S} . As $|\mathcal{S}|$ is constant when calculating the impurity for any \mathcal{S}_L and \mathcal{S}_R , you can simply pick the split of a set \mathcal{S} into \mathcal{S}_L , \mathcal{S}_R that minimizes:

$$\sum_{(\mathbf{x}, y) \in \mathcal{S}_L} (y - \bar{y}_{\mathcal{S}_L})^2 + \sum_{(\mathbf{x}, y) \in \mathcal{S}_R} (y - \bar{y}_{\mathcal{S}_R})^2$$

☆ **Key Points**

Regression trees use a modified impurity metric appropriate for continuous label values.

This impurity quantifies the "spread" or variance of the labels in a set.

Each leaf in the final tree predicts the mean label of the training points inside the leaf.