CART Terminology

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Internal nodes | The non-terminal nodes, and the content of the nodes represent the classifier used for the next split (i.e. splitting variable and the corresponding value). |
| Parent node | The node directly above a **child node**. For example, the root node is the parent node of two first-level internal nodes. CARTs permit only binary splits; therefore, all non-terminal nodes have two child nodes. |
| Root node | The top-most node where the data is first separated into two subspaces. |
| Terminal node | Also known as a **leaf node**, which is furthermost node from the node without any **child nodes**. The content of node represents the prediction of the subspace. |

Loss functions

* The Gini index and cross-entropy are differentiable; therefore, they can be used in **numerical** optimisation.
* The Gini index and cross-entropy are more sensitive to changes in **node probability** (and thus are often used in growing a tree).

Loss functions and impurity

It is inadvisable to avoid predictors with many categories because they tend to overfit the data even though they often result in a lower impurity measure.

False positives and false negatives should not be treated equally when measuring impurity, as the worse outcome depends on the context. Therefore, a K-by-K loss matrix should be incorporated into the impurity measure as a weighting factor for different consequences if required. Note that standard loss functions treat the misclassification of false positives and false negatives equally.