

GINI

INDEX

Proportion of observations
in the m th leaf of K th class.

$$G = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk})$$

Diagram illustrating the components of the Gini Index formula:

- \hat{p}_{mk} : Proportion of observations in the m th leaf of the k th class. (Indicated by a red arrow from the text above and an orange arrow from "leaf" below).
- $1 - \hat{p}_{mk}$: Complement of the proportion of observations in the m th leaf of the k th class. (Indicated by an orange arrow from "leaf" below).
- k : Index of the class. (Indicated by a black arrow from "Class" below).

Used at each node
to decide which
feature is best
to split on.

The smaller
the value of G
the more purity
there is in the
node.

Measure of purity
in tree based methods