

Measures of impurity

$$\hat{p}_{mk} = \frac{1}{N_m} \cdot \sum_{x_i \text{ in leaf } m} \underbrace{I(y_i = k)}_{\substack{\text{times a particular} \\ \text{class appears in that leaf}}}$$

\hat{p}_{mk} : into leaf, k class
 N_m : N total objects

Class k appears in leaf m

Misclassification error

$$1 - \hat{p}_{mk(m)}; k(m) = \text{most common}; k$$

most common class of that particular leaf

0 \rightarrow perfect purity

0.5 \rightarrow no purity

Gini index (not to be confused with the Gini coeff)

$$\sum_{k \neq k'} \hat{p}_{mk} \cdot \hat{p}_{mk'} = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk}) = 1 - \sum_{k=1}^K \hat{p}_{mk}^2$$

one - sum of probs that you belong to any other class

Again,

0 \rightarrow perfect purity

95 \rightarrow no purity

Deviance / information gain

$$- \sum_{k=1}^K \hat{p}_{mk} \log_2 \hat{p}_{mk}$$

0 \rightarrow perfect purity

1 \rightarrow no purity

0	0	0	1
0	0	0	0
0	0	0	0
0	0	0	0

Misclassification: $1/16 = 0,06$

Gini: $1 - [(1/16)^2 + (15/16)^2] = 0,12$

Information: $- [1/16 \cdot \log_2 (1/16) + 15/16 \cdot \log_2 (15/16)] = 0,34$

0	0	●	●
0	0	●	●
0	0	●	●
0	0	●	●

Misclassification: $8/16 = 0,5$

$$E_{ini} : 1 - [(8/16)^2 + (8/16)^2] = 0,5$$

Information gain:

$$- [1/16 \cdot \log_2(1/16) + 15/16 \cdot \log_2(15/16)] = 1$$
