

# CART Algorithm (Breiman et al, 1984)

- Uses an alternative to *entropy* as an impurity metric: *Gini Index*

**Gini Index** – rate of misclassification

$$Gini(S) = 1 - \sum_{c \in \text{Classes}} p_c^2$$

$S$  – subset of training examples

$p_c$  – proportion of examples in  $S$  belonging to class  $c$

Higher Gini index  $\rightarrow$  more likely to misclassify

# Gini Index example

e.g. (3 yes / 3 no)

# Gini Index example

e.g. (3 yes / 3 no)

$$\text{Gini}(S) = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

# Gini Index example

e.g. (3 yes / 3 no)

$$\text{Gini}(S) = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

e.g. (4 yes / 0 no)

# Gini Index example

e.g. (3 yes / 3 no)

$$\text{Gini}(S) = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

e.g. (4 yes / 0 no)

$$\text{Gini}(S) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2 = 0$$

# Gini Index example

e.g. (3 yes / 3 no)

$$\text{Gini}(S) = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

e.g. (4 yes / 0 no)

$$\text{Gini}(S) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2 = 0$$

# Gini Index example

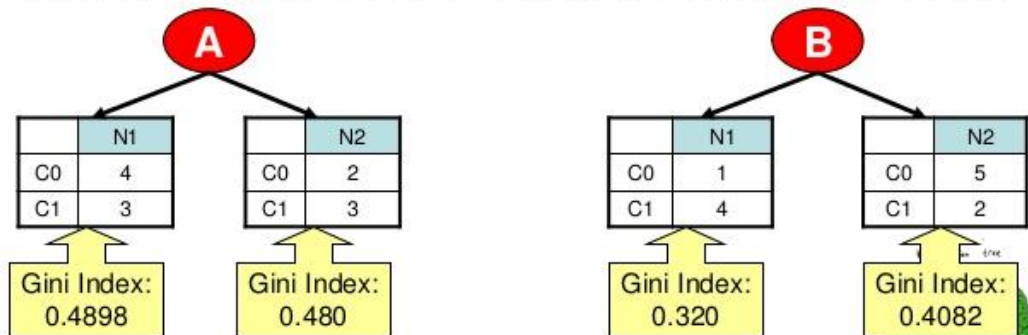
## Splitting Binary Attributes (using Gini)

Example :

	Parent
C0	6
C1	6
Gini = 0.5	

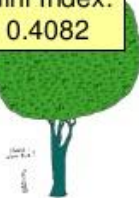
$$\begin{aligned}\text{Gini :} \\ 1 - (6/12)^2 - (6/12)^2 \\ = 0.5\end{aligned}$$

Suppose there are two ways (A and B) to split the data into smaller subset.



Which one is a better split??

Compute the **weighted average of the Gini index** of both attribute



# Information Gain

- **Information Gain** – expected reduction in the *misclassification rate* after a split on an attribute

$$Gain(S, A) = Gini(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Gini(S_v)$$

$A$  – Attribute

$Values(A)$  – possible values of  $A$

$S$  – subset of training examples

$S_v$  – subset of  $S$  for which attribute  $A$  have value  $v$



# Entropy vs Gini Index

- So which impurity measure should be used:

## **Entropy or Gini index?**

- Resulting trees are very similar in practice.
- Best advice is that it is good practice when building decision tree models to
  - Try out different impurity metrics
  - Compare results to see which suits the dataset