

1. (a) after each branch is replaced by a leaf, take the left branch as example:

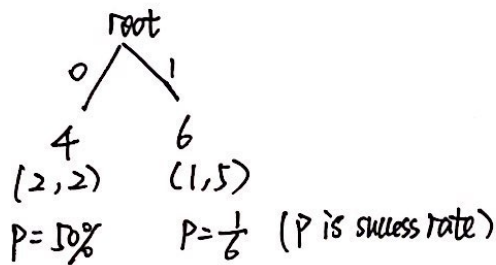
if  $p_1 > n_1$ , then there are  $n_1$  data points misclassified; similarly,

if  $p_1 < n_1$ , there will be  $p_1$  mistakes. For left branch:

$$\begin{aligned} \# \text{ training mistakes} &= \begin{cases} n_1 & (p_1 > n_1) \\ p_1 & (p_1 < n_1) \end{cases} = (p_1 + n_1) \cdot I\left(\frac{p_1}{p_1 + n_1}\right) \\ &= (p_1 + n_1) \cdot \min\left(\frac{p_1}{p_1 + n_1}, \frac{n_1}{p_1 + n_1}\right) \end{aligned}$$

$$\text{Thus, the whole training mistakes} = (p_1 + n_1) \cdot I\left(\frac{p_1}{p_1 + n_1}\right) + (p_2 + n_2) \cdot I\left(\frac{p_2}{p_2 + n_2}\right)$$

(b). Gini index on  $a_1$ .

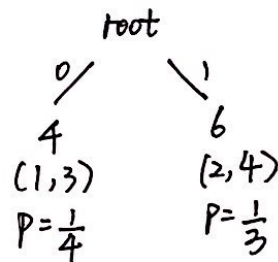


$$\text{Gini for subnode } (a_1=0) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

$$\text{for } (a_1=1) = \frac{1}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} = \frac{26}{36}$$

$$\therefore \text{Gini for split on } a_1 = \frac{4}{10} \cdot \frac{1}{2} + \frac{6}{10} \cdot \frac{26}{36} = 0.63$$

Gini index on  $a_2$

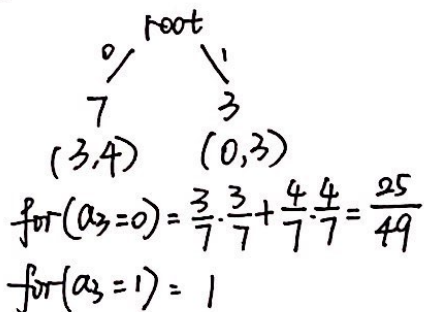


$$\text{for subnode } (a_2=0) = \frac{1}{4} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} = \frac{10}{16} = \frac{5}{8}$$

$$\text{for } (a_2=1) = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{2}{3} = \frac{5}{9}$$

$$\therefore \text{Gini for split on } a_2 = \frac{4}{10} \cdot \frac{5}{8} + \frac{6}{10} \cdot \frac{5}{9} = 0.58$$

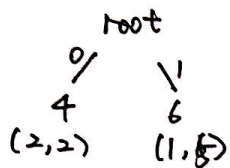
Gini index on  $a_3$ .



$$\therefore \text{on split } a_3 = \frac{7}{10} \cdot \frac{25}{49} + \frac{3}{10} \cdot 1 = 0.66,$$

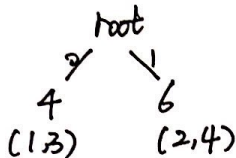
thus split on  $a_3$  yields a higher Gini score, (a more pure split). thus using Gini index will choose to split on  $a_3$ .

min-error on  $a_1$



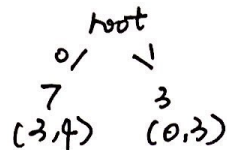
$$\text{min-error} = 4 \cdot \frac{2}{4} + 6 \cdot \frac{1}{6} = 3$$

; min-error on  $a_2$



$$\text{error} = 4 \cdot \frac{1}{4} + 6 \cdot \frac{2}{6} = 3$$

; min-error on  $a_3$



$$\text{error} = 7 \cdot \frac{3}{7} + 3 \cdot \frac{0}{3} = 3$$

using min-error impurity function, they are give same error value, thus it will choose an random attribute to split.

(c) The problem is equivalent to asking to in what condition, s.t.:

$$\min(p_1, n_1) + \min(p_2, n_2) < \min(p_1 + p_2, n_1 + n_2),$$

When  $p_1 > n_1$  and  $p_2 > n_2$  or,  $p_1 < n_1$  and  $p_2 < n_2$ , left term = right term.

While when ~~a~~ split has opposite value comparison, such as :

$$\boxed{p_1 > n_1 \text{ and } p_2 < n_2 \text{ or } p_1 < n_1 \text{ and } p_2 > n_2}, \text{ such } \# \text{ left term}$$

will be strictly smaller than right term.

(d) min-error impurity function employs a more stricter condition on making the split, that the min-error only decreases when the resulting branches have opposite majority label, it makes harder for a growing decision tree to make decision, as we can observe from (b), it's easy to appear draw conditions, when their min-error value are all the same, it will hinder the decision to be made.

## 2. Bootstrap aggregation ("bagging")

For  $i = 1 \dots N$ , Let  $X_i$  be ~~the~~ a random variable, that  $X_i = 1$  when the  $i$ th example does not appear in the replicate, and  $X_i = 0$  otherwise. Because we draw  $N$  samples with replacement, then the probability of  $X_i = 1$  is :

$$\Pr(X_i = 1) = \left(\frac{N-1}{N}\right)^N \quad \left( \text{for each sample, there is } \frac{N-1}{N} \text{ probability to choose other examples instead of } i \text{th} \right)$$

Let  $S$  be the number of distinct samples that does not appear in the replicate. Then

$$S = \sum_{i=1}^N X_i \quad (X_i \text{ are independent, for } i = 1 \dots N)$$

$$\Rightarrow E(S) = \sum_{i=1}^N \Pr(X_i = 1) = N \left(\frac{N-1}{N}\right)^N$$

$$\Rightarrow \text{the expected fraction is } \frac{E(S)}{N} = \left(\frac{N-1}{N}\right)^N = \left(1 - \frac{1}{N}\right)^N$$

$$\text{and } \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = \frac{1}{e}$$

II