

Information gain:

Information gain is used to determine the amount of information a feature in a class can tell us. A decision tree classifier is used to maximize the information gain. The feature that has the highest Information gain will be chosen to split first.

General equation for information gain is,

$$\text{Information gain} = [\text{entropy}(\text{parent}) - [\text{weighted average}] * \text{entropy}(\text{children})]$$

Lets take an example of the following example to understand information gain,

Wind	Sun	Surfing
Windy	Not Sunny	No
Windy	Sunny	Yes
Not windy	Sunny	No
Not windy	Not Sunny	No

From the table data, I consider Wind and Sun as features and Surfing is the target/ label.

Now for the decision tree, we start with the target/ label – Surfing as the Parent node (NYNN)

Let's find the entropy for this,

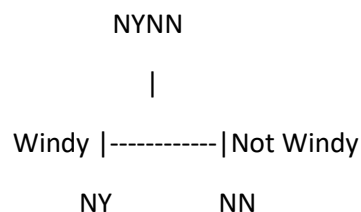
$$P(N) = \frac{3}{4} = 0.75$$

$$P(Y) = \frac{1}{4} = 0.25$$

$$\begin{aligned}\text{Entropy}(\text{Parent}) &= - (P(N)\log_2(P(N)) + P(Y)\log_2(P(Y))) = - ((0.75 * (-0.41)) + (0.25 * (-2))) \\ &= -(-0.30 - 0.5) \\ &= 0.8\end{aligned}$$

To check where the parent node is split by the decision tree algorithm, we should find the information gain from both the features wind and sun.

First let's consider wind



Entropy for not windy part of the tree split is 0 since all belong to the same class.

$$\text{Entropy}(\text{NY}) = - ((0.5) \log_2(0.5) + (0.5) \log_2(0.5)) = 1$$

Lets find the entropy of children,

Parent node – 4 (NYNN)

Left child node – 2 (NY)

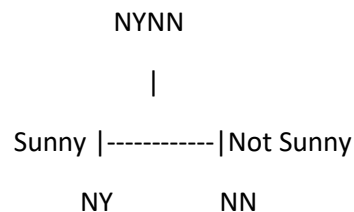
Right child node – 2 (NN)

{Weighted average}{Children} = (No of left child/ no of parent node) * left child entropy) + (No of Right child/ no of parent node) * Right child entropy)

{Weighted}{Children} = $2/4 * 1 + 2/4 (0) = \frac{1}{2} = 0.5$

Information gain from wind = $0.8 - 0.5 = 0.3$

Next let's consider Sun



Entropy for not windy part of the tree split is 0 since all belong to the same class.

Entropy (NY) = $-(0.5 \log_2(0.5) + 0.5 \log_2(0.5)) = 1$

Let's find the entropy of children,

Parent node – 4 (NYNN)

Left child node – 2 (NY)

Right child node – 2 (NN)

{Weighted}{Children} = $2/4 * 1 + 2/4 (0) = \frac{1}{2} = 0.5$

Information gain from Sun = $0.8 - 0.5 = 0.3$

Form this example, Both the features sun and wind have same information gain, hence the decision tree can pick either of the feature to make a split.

In general, the feature with the highest information gain is chosen by th decision tree to split the data.

