

Decision Trees

Kristian Wichmann

December 26, 2017

1 Simple binary decision tree

Imagine we wish to predict whether a person is more likely to die young (below 70 years old) or old based on whether they smoked and/or drank alcohol. Table 1 shows such data for 100 subjects.

The prediction can be expressed in a binary decision tree. The question is how to make such a tree.

1.1 Maximizing information gain

Recall that the *entropy* of a probability distribution is defined as follows:

$$H = - \sum_i p_i \log_2 p_i \quad (1.1)$$

Information and entropy are complementary quantities: When entropy is lowered, information is gained. So to maximize information gain is to maximize entropy loss.

Drinker	Smoker	High age	Number
Y	Y	Y	2
Y	Y	N	16
Y	N	Y	20
Y	N	N	17
N	Y	Y	5
N	Y	N	9
N	N	Y	28
N	N	N	3

Table 1: Table of the 100 subjects

1.1.1 Pre-split entropy

Before splitting the subjects according to being drinkers or smokers, there's two groups:

- Subjects with high age: $2 + 20 + 5 + 28 = 55$.
- Subjects with low age: $16 + 17 + 9 + 3 = 45$.

So the entropy is:

$$H = -\frac{55}{100} \log_2 \frac{55}{100} - \frac{45}{100} \log_2 \frac{45}{100} \quad (1.2)$$