Mitchell's Book: Ex. 3.2 DTs

Machine Learning 2024-25 Course Activity

Furno Francesco - francesco.furno@studenti.unipd.it - 2139507

November 4st, 2024

Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

1. What is the entropy of this collection of training examples with respect to the target function classification?
2. What is the information gain of a_2 relative to these training examples?

Entropy

In this collection, there are 6 instances:

- 3 positive elements $\Rightarrow P_+ = \frac{3}{6} = \frac{1}{2}$
- 3 negative elements $\Rightarrow P_- = \frac{3}{6} = \frac{1}{2}$

The entropy with respect to the classification value is:

$$\begin{aligned} E(S) &= -\Big(\log_2 (P_+)_\times P_+ + \log_2(P_-) \times P_-\Big) \\ &= -\left(\log_2\left(\frac{1}{2}\right) \times \frac{1}{2} + \log_2\left(\frac{1}{2}\right) \times \frac{1}{2}\right) \\ &= -\left(-1 \times \frac{1}{2} - 1 \times \frac{1}{2}\right) \\ &= -\left(-\frac{2}{2}\right) = 1 \end{aligned}$$

We could have concluded that the entropy is 1 without calculation because the instances are perfectly balanced: there are an equal number of positive and negative elements. When the elements are perfectly distributed between the classes, the entropy function reaches its maximum value, indicating maximum uncertainty:

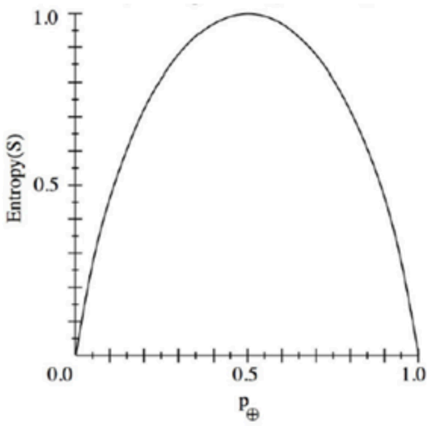

Figure 1: Entropy function

Information Gain of a_2

To find the Information Gain of a_2 we need to calculate the entropy for each possible value of this attribute. a_2 has two possible values: **T** and **F**.

Let's start with **T**. There are 4 instances where $a_2 = \mathbf{T}$.

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
5	-	F	T
6	-	F	T

In total:

- 2 positive elements $\Rightarrow P_+ = \frac{2}{4} = \frac{1}{2}$
- 2 negative elements $\Rightarrow P_- = \frac{2}{4} = \frac{1}{2}$

$$
\begin{aligned}
E(a_{2-\mathbf{T}}) &= -\left(\log_2\left(\frac{1}{2}\right) \times \frac{1}{2} + \log_2\left(\frac{1}{2}\right) \times \frac{1}{2}\right) \\
&= 1
\end{aligned}
$$

This is because the instances are perfectly balanced between the classifications.

Let's analyze **F**. There are 2 instances where $a_2 = \mathbf{F}$.

Instance	Classification	a_1	a_2
3	-	T	F
4	+	F	F

In total:

- 1 positive element $\Rightarrow P_+ = \frac{1}{2}$
- 1 negative element1 $\Rightarrow P_- = \frac{1}{2}$

$$E(a_{2-\mathbf{F}}) = -\left(\log_2\left(\frac{1}{2}\right) \times \frac{1}{2} + \log_2\left(\frac{1}{2}\right) \times \frac{1}{2}\right)$$
$$= 1$$

Again, the instances are perfectly balanced between the classifications.

Information Gain

$$G(S, a_2) = E(S) - \left(\frac{4}{6} \times E(a_{2-\mathbf{T}}) + \frac{2}{6} \times E(a_{2-\mathbf{F}})\right)$$
$$= 1 - \left(\frac{4}{6} \times 1 + \frac{2}{6} \times 1\right)$$
$$= 1 - (1)$$
$$= 0$$

The Information Gain of a_2 is 0 because, for both possible values (**T** and **F**), the instances are perfectly balanced in the classification $+$ and **-**. This means that a_2 does not provide any additional information to reduce the uncertainty of the classification.