CT5165/CT5170 Additional Entropy Examples

Patrick Mannion

September 2021

1 Example 1: Calculate the maximum entropy for a binary classification problem

In a binary classification problem, there are just two classes, say class A and class B. The maximum entropy will occur when there is an even mixing of the classes in our dataset S.

Say we have a dataset called S_1 with 6 instances where the classes are mixed in equal proportions, $S_1 = \{A, A, A, B, B, B\}$.

Recall that entropy is calculated as:

$$Ent(S) = \sum -p_i \log_2 p_i \tag{1}$$

where S is the dataset and p_i is the proportion of class i in the dataset.

Firstly we need to calculate the proportions of all the classes in S_1:

$$p_A = \frac{3}{6} = \frac{1}{2}$$

$$p_B = \frac{3}{6} = \frac{1}{2}$$

Next we calculate the entropy of S using Eqn 1:

$$Ent(S_1) = \sum -p_i \log_2 p_i$$

$$Ent(S_1) = -p_A \log_2 p_A - p_B \log_2 p_B$$

$$Ent(S_1) = -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right)$$

$$Ent(S_1) = -\frac{1}{2} \times (-1) - \frac{1}{2} \times (-1) = 1$$

Therefore the maximum entropy for a binary classification problem is 1.

2 Example 2: Show that the entropy is less than the maximum when the classes are mixed in unequal proportions

We will show this using a binary classification problem.

From above, we already know that the maximum entropy is 1 in a binary classification problem

Say we have a dataset S_2 with 6 instances where the classes are mixed in unequal proportions, $S_2 = \{A, A, B, B, B, B\}$.

Again calculate the proportions of all the classes in S_2:

$$p_A = \frac{2}{6} = \frac{1}{3}$$

$$p_B = \frac{4}{6} = \frac{2}{3}$$

Next calculate the entropy of S_2 using Eqn 1 as before:

$$Ent(S_2) = \sum -p_i \log_2 p_i$$

$$Ent(S_2) = -p_A \log_2 p_A - p_B \log_2 p_B$$

$$Ent(S_2) = -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right)$$

$$Ent(S_2) = -\frac{1}{3} \times (-1.585) - \frac{2}{3} \times (-0.585) = 0.918$$

As 0.918 is less than the maximum entropy of 1, we can conclude that this dataset S_2 has less than the maximum entropy.

3 Example 3: Show that the entropy is zero when the dataset contains only a single class

Say we have a dataset S_3 with 6 instances where there is only one class present, $S_3 = \{A, A, A, A, A, A\}$.

Calculate the proportions of all the classes present in S:

$$p_A = \frac{6}{6} = 1$$

Next calculate the entropy of S_3 using Eqn 1 as before:

$$Ent(S_3) = \sum -p_i \log_2 p_i$$

$$Ent(S_3) = -p_A \log_2 p_A$$

$$Ent(S_3) = -1 \log_2 (1)$$

$$Ent(S_3) = -1 \times 0 = 0$$

4 Example 4: Compare the information gain of two different possible partitions of a dataset

Say we have a dataset called S_1 with 6 instances where the classes are mixed in equal proportions, $S_1 = \{A, A, A, B, B, B\}$. From Example 1 we already know that the entropy of this set is $Ent(S_1) = 1$.

We have two options for partitioning the dataset:

1. partition into $S_{1,left} = \{A, A, B\}$ and $S_{1,right} = \{B, B, A\}$. In the left partition $p_A = \frac{2}{3}$ and $p_B = \frac{1}{3}$, giving entropy of 0.918. In the right partition $p_A = \frac{1}{3}$ and $p_B = \frac{2}{3}$, giving entropy of 0.918 (as per Example 2 above).

2. partition into $S_{1,left} = \{A, A, A\}$ and $S_{1,right} = \{B, B, B\}$. In the left partition $p_A = 1$, giving entropy of 0. In the right partition $p_B = 1$, giving entropy of 0 (as per Example 3 above).

Recall that information gain (IG) may be calculated as:

$$Gain = Ent(S) - \sum \frac{|S_v|}{|S|} Ent(S_v) \tag{2}$$

where $\frac{|S_v|}{|S|}$ is the proportion of the dataset that has been assigned to a new partition.

As the sizes of the partitions in both option 1 and option 2 are the same, we can calculate the proportions of the dataset in the left and right partitions as:

$$\frac{|S_{1,left}|}{|S_1|} = \frac{3}{6} = \frac{1}{2}$$

$$\frac{|S_{1,right}|}{|S_1|} = \frac{3}{6} = \frac{1}{2}$$

We can calculate the IG for option 1 as:

$$Gain = Ent(S_1) - \sum \frac{|S_v|}{|S_1|} Ent(S_v)$$

$$Gain = Ent(S_1) - \frac{|S_{1,left}|}{S} Ent(S_{1,left}) - \frac{|S_{1,right}|}{S} Ent(S_{1,right})$$

$$Gain = 1 - \frac{1}{2} \times 0.918 - \frac{1}{2} \times 0.918$$

$$Gain = 0.082$$

We can calculate the IG for option 2 as:

$$Gain = Ent(S_1) - \sum \frac{|S_v|}{|S_1|} Ent(S_v)$$

$$Gain = Ent(S_1) - \frac{|S_{1,left}|}{S} Ent(S_{1,left}) - \frac{|S_{1,right}|}{S} Ent(S_{1,right})$$

$$Gain = 1 - \frac{1}{2} \times 0 - \frac{1}{2} \times 0$$

$$Gain = 1$$

Comparing the two options, we see that option 1 gives a reduction in entropy of 0.082, whereas option 2 gives a reduction in entropy of 1. Therefore we should prefer option 2 as it leads to a much greater reduction in entropy.