CSCE 633 - Homework 2

Prakhar Suryavansh

Problem 1: Information Gain

Part (1):

We are given the following training points for a classification problem:

X_1	X_2	Y
1	1	1
1	1	1
1	1	2
1	0	3
0	0	2
0	0	3

We have to calculate the information gain for both attributes X_1 and X_2.

The information gain is measures the expected reduction in entropy. It can be calculated using the formula:

$$Gain(S, X_i) = Entropy(S) - \sum_{v \in Values(X_i)} \frac{|S_v|}{|S|} Entropy(S_v), where$$

Gain(S, X_i) denotes the information gain for attribute X_i relative to a collection of examples S,

S denotes the collection of examples,

$Values(X_i)$ is the set of all possible values for attribute X_i $|S_v|$ is the subset of S, for which attribute X_i has value v

Step 1: Calculate the Entropy of Y. The overall entropy $H(Y)$ is calculated as:

$$H(Y) = -\sum_{i=1}^{3} p_i \log_2(p_i)$$

Where p_i is the probability of class i in the dataset.

From the dataset, we have the following distribution of Y:

Class 1: 2 instances

Class 2: 2 instances

Class 3: 2 instances

So

$$p_1 = \frac{2}{6}, p_2 = \frac{2}{6}, p_3 = \frac{2}{6}$$

Thus, the entropy is:

$$H(Y) = -\left(\frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6}\right)$$

Simplifying:

$$H(Y) = -3 \times \frac{2}{6}\log_2\frac{1}{3} = -3 \times \frac{2}{6} \times (-1.585) = 1.585 \text{ bits}$$

Step 2: Calculate the Entropy of Y given X_1. We split the data based on the value of X_1. When $X_1 = 1$, we have:

X_1	X_2	Y
1	1	1
1	1	1
1	1	2
1	0	3

For $X_1 = 1$, the class distribution is: - Class 1: 2 instances - Class 2: 1 instance - Class 3: 1 instance

Thus, the entropy is:

$$H(Y|X_1 = 1) = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$H(Y|X_1 = 1) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$H(Y|X_1 = 1) = 0.5 + 0.5 = 1.0 \text{ bits}$$

When $X_1 = 0$, we have:

X_1	X_2	Y
0	0	2
0	0	3

For $X_1 = 0$, the class distribution is: - Class 2: 1 instance - Class 3: 1 instance

Thus, the entropy is:

$$H(Y|X_1 = 0) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$$

$$H(Y|X_1 = 0) = 1.0 \text{ bits}$$

Step 3: Calculate the Information Gain for X_1. The information gain is given by:

$$IG(X_1) = H(Y) - \left(\frac{4}{6}H(Y|X_1 = 1) + \frac{2}{6}H(Y|X_1 = 0)\right)$$

$$IG(X_1) = 1.585 - \left(\frac{4}{6} \times 1 + \frac{2}{6} \times 1\right) = 1.585 - 1 = 0.585 \text{ bits}$$

Step 4: Calculate the Entropy of Y given X_2. When $X_2 = 1$, we have:

X_1	X_2	Y
1	1	1
1	1	1
1	1	2

For $X_2 = 1$, the class distribution is: - Class 1: 2 instances - Class 2: 1 instance Thus, the entropy is:

$$H(Y|X_2 = 1) = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)$$

$$H(Y|X_2 = 1) = 0.918 \text{ bits}$$

When $X_2 = 0$, we have:

X_1	X_2	Y
1	0	3
0	0	2
0	0	3

For $X_2 = 0$, the class distribution is: - Class 2: 1 instance - Class 3: 2 instances Thus, the entropy is:

$$H(Y|X_2 = 0) = -\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right)$$

$$H(Y|X_2 = 0) = 0.918 \text{ bits}$$

Step 5: Calculate the Information Gain for X_2. The information gain is given by:

$$IG(X_2) = H(Y) - \left(\frac{3}{6}H(Y|X_2 = 1) + \frac{3}{6}H(Y|X_2 = 0)\right)$$

$$IG(X_2) = 1.585 - \left(\frac{1}{2} \times 0.918 + \frac{1}{2} \times 0.918\right)$$

$$IG(X_2) = 1.585 - 0.918 = 0.667 \text{ bits}$$

Conclusion: Since $IG(X_2) > IG(X_1)$, we use X_2 for the first split in the decision tree.

Part 2:

Part 3:

Problem 2: Entropy

Part 1:

Part 2:

Part 3: