Decision Tree Assignment Solution

Machine Learning Course
Amirhossein Abolhassani

1 Suppose there is an attribute, "A," that consists of random values, and these values do not have any correlation with the class labels. Additionally, assume that "A" has a sufficient number of distinct values such that no two instances in the training dataset share the same value for "A." What would be the outcome if a decision tree is built using this attribute? What challenges or issues might arise in this scenario?

Answer

Part One

According to the ID3 algorithm, initially we measure the information gain of each feature and select the feature with the highest gain. (Number of target classes = k):

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

$$Entropy(S_v) = -\sum_{v} p_v \log(p_v) = -(P(S_v = 0) \log P(S_v = 0) + P(S_v = 1) \log P(S_v = 1) + \ldots + P(S_v = k) \log p(S_v = k))$$

Due to the uniqueness of this feature (being a primary key) for each sample, all terms $P(S_v = l) \log P(S_v = l)$ become equal to zero. Because either Therefore, one of the multipliers will be 0, making the entire term 0. This way we conclude:

$$Entropy(S_v) = 0$$

and this feature is chosen for the root:

$$\arg\max \{Gain(S_v) | \forall A \in \text{Header}\} = A$$

As a result, the height of the tree will be 1, and there will be as many branches as the number of values for feature A.

Part Two

If this feature has no relationship with the target feature, using this feature is completely wrong and leads to overfit. Because there is practically no room left for generalization.

2 Answer the questions according to the following dataset:

Weekend	Weather	Parents	Money	Decision (Category)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

2.1 Create a decision tree model using the given dataset to predict the value of the final column, using all other columns as input features except for the first one(weekend). Clearly explain each step of the process, including your calculations, reasoning, and decisions made while constructing the tree. What is the model's overall classification accuracy?

1 - Root Node

Decision

Cinema	Tennis	Stay in	Shopping
6	2	1	1

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(0.6 \times -0.73 + 0.2 \times -2.32 + 0.1 \times -3.32 + 0.1 \times -3.32) = 1.56$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	3	2	1	1
Poor	3	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(0.42 \times -1.25 + 0.28 \times -1.83 + 0.14 \times -2.83 + 0.14 \times -2.83) = 1.82$$

$$Entropy(S_{\text{Poor}}) = -(0.42 \times -1.25) = 0.52$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{7}{10} \times 1.82 + \frac{3}{10} \times 0.52 = 1.43$$

$$Gain(S, \text{Money}) = 1.56 - 1.43 = 0.13$$

Parents

Value	Cinema	Tennis	Stay in	Shopping
Yes	5	0	0	0
No	1	2	1	1

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Yes}}) = -(\frac{5}{5} \times 0) = 0$$

$$Entropy(S_{\text{No}}) = -(0.2 \times -2.32 + 0.4 \times -1.32 + 0.2 \times -2.32 + 0.2 \times -2.32) = 1.92$$

$$Gain(S, \text{Parents}) = Entropy(S) - \sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{5}{10} \times 0 + \frac{5}{10} \times 1.92 = 0.96$$

$$Gain(S, \text{Parents}) = 1.56 - 0.96 = 0.6$$

Weather

Value	Cinema	Tennis	Stay in	Shopping
Sunny	1	2	0	0
Windy	3	0	0	1
Rainy	2	0	1	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Sunny}}) = -(\frac{1}{3} \times -1.59 + \frac{2}{3} \times -0.59) = 0.92$$

$$Entropy(S_{\text{Windy}}) = -(0.75 \times -0.41 + 0.25 \times -2) = 0.8$$

$$Entropy(S_{\text{Rainy}}) = -(\frac{2}{3} \times -0.59 + \frac{1}{3} \times -1.59) = 0.92$$

$$Gain(S, \text{Weather}) = Entropy(S) - \sum_{v \in \text{Weather}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Weather}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{3}{10} \times 0.92 + \frac{4}{10} \times 0.8 + \frac{3}{10} \times 0.92 = 0.87$$

$$Gain(S, \text{Weather}) = 1.56 - 0.87 = 0.69$$

Picking The Best Attribute

Attribute	Information Gain
Money	0.13
Parents	0.6
Weather	0.69

The selected feature is Weather. 2 - Sunny Node

Decision

Cinema	Tennis	Stay in	Shopping
1	2	0	0

For W_1, W_2, W_{10}

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(0.33 \times -1.59 + 0.66 \times -0.59) = 0.91$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	1	2	0	0
Poor	0	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(0.33 \times -1.59 + 0.66 \times -0.59) = 0.91$$

$$Entropy(S_{\text{Poor}}) = 0$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{3}{3} \times 0.91 + \frac{0}{3} \times 0 = 0.91$$

$$Gain(S, \text{Money}) = 0.91 - 0.91 = 0$$

Parents

Value	Cinema	Tennis	Stay in	Shopping
Yes	1	0	0	0
No	0	2	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Yes}}) = -(\frac{1}{1} \times 0) = 0$$

$$Entropy(S_{\text{No}}) = -(\frac{2}{2} \times 0) = 0$$

$$Gain(S, \text{Parents}) = Entropy(S) - \sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{1}{3} \times 0 + \frac{2}{3} \times 0 = 0$$

$$Gain(S, \text{Parents}) = 0.91 - 0 = 0.91$$

Picking The Best Attribute

Attribute	Information Gain
Money	0
Parents	0.91

The selected feature is Parents. 3 - Rainy Node

Decision

Cinema	Tennis	Stay in	Shopping
2	0	1	0

For W_4, W_5, W_6

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(0.33 \times -1.59 + 0.66 \times -0.59) = 0.91$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	0	0	1	0
Poor	2	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(1 \times 0 + 1 \times 0) = 0$$

$$Entropy(S_{\text{Poor}}) = -(1 \times 0) = 0$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{1}{3} \times 0 + \frac{2}{3} \times 0 = 0$$

$$Gain(S, \text{Money}) = 0.91 - 0 = 0.91$$

Parents

Value	Cinema	Tennis	Stay in	Shopping
Yes	2	0	0	0
No	0	0	1	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Yes}}) = -(\frac{2}{2} \times 0) = 0$$

$$Entropy(S_{\text{No}}) = -(\frac{1}{1} \times 0) = 0$$

$$Gain(S, \text{Parents}) = Entropy(S) - \sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{1}{3} \times 0 + \frac{2}{3} \times 0 = 0$$

$$Gain(S, \text{Parents}) = 0.91 - 0 = 0.91$$

Picking The Best Attribute

Attribute	Information Gain
Money	0.91
Parents	0.91

The selected feature is Parents or Money. 4 - Windy Node

Decision

Cinema	Tennis	Stay in	Shopping
3	0	0	1

For W_3, W_7, W_8, W_9

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(\frac{3}{4} \times -0.41 + \frac{1}{4} \times -2) = 0.8$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	2	0	0	1
Poor	1	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(0.33 \times -1.59 + 0.66 \times -0.59) = 0.91$$

$$Entropy(S_{\text{Poor}}) = -(1 \times 0) = 0$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{3}{4} \times 0.91 + \frac{1}{4} \times 0 = 0.68$$

$$Gain(S, \text{Money}) = 0.8 - 0.68 = 0.12$$

Parents

Value	Cinema	Tennis	Stay in	Shopping
Yes	2	0	0	0
No	1	0	0	1

$$Entropy(S_\text{v}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_\text{Yes}) = -(\frac{2}{2} \times 0) = 0$$

$$Entropy(S_\text{No}) = -(\frac{1}{2} \times -1 + \frac{1}{2} \times -1) = 1$$

$$Gain(S, \text{Parents}) = Entropy(S) - \sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{2}{4} \times 1 + \frac{2}{4} \times 0 = 0.5$$

$$Gain(S, \text{Parents}) = 0.8 - 0.5 = 0.3$$

Picking The Best Attribute

Attribute	Information Gain
Money	0.12
Parents	0.3

The selected feature is Parents. 5 - (Parents - No) Node

Decision

Cinema	Tennis	Stay in	Shopping
1	0	0	1

For W_7, W_8

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(\frac{1}{2} \times -1 + \frac{1}{2} \times -1) = 1$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	0	0	0	1
Poor	1	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(1 \times 0) = 0$$

$$Entropy(S_{\text{Poor}}) = -(1 \times 0) = 0$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{1}{2} \times 0 + \frac{1}{2} \times 0 = 0$$

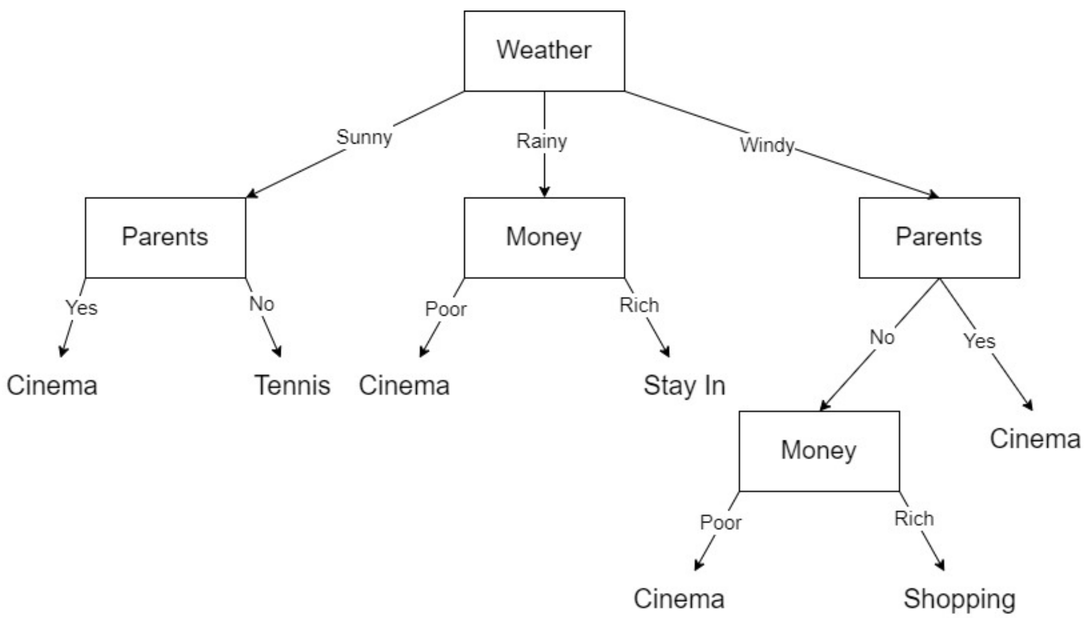
$$Gain(S, \text{Money}) = 1 - 0 = 1$$

Picking The Best Attribute

Attribute	Information Gain
Money	0.12

The selected feature is Money.

Decision Tree diagram:



Accuracy: 100%

2.2 Construct a decision tree model using only the first 6 samples from the dataset(W1 - W6). Evaluate the model's classification performance on these initial 6 samples as the training set. Then, use the model to classify the remaining samples in the dataset. What is the classification accuracy for both the training and test datasets? Discuss your findings and explain the reasons behind the observed results.

Learning with training data:

1 - Root Node

Decision

Cinema	Tennis	Stay in	Shopping
4	1	1	0

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(\frac{4}{6} \times -0.59 + \frac{1}{6} \times -2.64 + \frac{1}{6} \times -2.64) = 1.23$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	2	1	1	0
Poor	2	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(\frac{2}{4} \times -1 + \frac{1}{4} \times -2 + \frac{1}{4} \times -2) = 1.5$$

$$Entropy(S_{\text{Poor}}) = -(\frac{2}{2} \times 0) = 0$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{4}{6} \times 1.5 + \frac{2}{6} \times 0 = 1$$

$$Gain(S, \text{Money}) = 1.23 - 1 = 0.23$$

Parents

Value	Cinema	Tennis	Stay in	Shopping
Yes	4	0	0	0
No	0	1	1	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Yes}}) = -(\frac{4}{4} \times 0) = 0$$

$$Entropy(S_{\text{No}}) = -(\frac{1}{2} \times -1 + \frac{1}{2} \times -1) = 1$$

$$Gain(S, \text{Parents}) = Entropy(S) - \sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Parents}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{4}{6} \times 0 + \frac{2}{6} \times 1 = 0.33$$

$$Gain(S, \text{Parents}) = 1.23 - 0.33 = 0.9$$

Weather

Value	Cinema	Tennis	Stay in	Shopping
Sunny	1	1	0	0
Windy	1	0	0	0
Rainy	2	0	1	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Sunny}}) = -(\frac{1}{2} \times -1 + \frac{1}{2} \times -1) = 1$$

$$Entropy(S_{\text{Windy}}) = -(\frac{1}{1} \times 0) = 0$$

$$Entropy(S_{\text{Rainy}}) = -(\frac{2}{3} \times -0.59 + \frac{1}{3} \times -1.59) = 0.92$$

$$Gain(S, \text{Weather}) = Entropy(S) - \sum_{v \in \text{Weather}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Weather}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{2}{6} \times 1 + \frac{1}{6} \times 0 + \frac{3}{6} \times 0.92 = 0.79$$

$$Gain(S, \text{Weather}) = 1.23 - 0.79 = 0.44$$

Picking The Best Attribute

Attribute	Information Gain
Money	0.23
Parents	0.9
Weather	0.44

The selected feature is Parents. 2 - (Parents - No) Node

Decision

Cinema	Tennis	Stay in	Shopping
0	1	1	0

For W_2, W_5

$$Entropy(S) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S) = -(\frac{1}{2} \times -1 + \frac{1}{2} \times -1) = 1$$

Money

Value	Cinema	Tennis	Stay in	Shopping
Rich	0	1	1	0
Poor	0	0	0	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Rich}}) = -(\frac{1}{2} \times -1 + \frac{1}{2} \times -1) = 1$$

$$Entropy(S_{\text{Poor}}) = -(0 \times 0) = 0$$

$$Gain(S, \text{Money}) = Entropy(S) - \sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Money}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{2}{2} \times 1 = 1$$

$$Gain(S, \text{Money}) = 1 - 1 = 0$$

Weather

Value	Cinema	Tennis	Stay in	Shopping
Sunny	0	1	0	0
Windy	0	0	0	0
Rainy	0	0	1	0

$$Entropy(S_{\text{v}}) = -\sum_{v \in S} p_v \log(p_v)$$

$$Entropy(S_{\text{Sunny}}) = -(\frac{1}{1} \times 0) = 0$$

$$Entropy(S_{\text{Windy}}) = 0$$

$$Entropy(S_{\text{Rainy}}) = -(\frac{1}{1} \times 0) = 0$$

$$Gain(S, \text{Weather}) = Entropy(S) - \sum_{v \in \text{Weather}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$\sum_{v \in \text{Weather}} \frac{|S_v|}{|S|} Entropy(S_v) = \frac{1}{2} \times 0 + \frac{1}{2} \times 0 = 0$$

$$Gain(S, \text{Weather}) = 1 - 0 = 1$$

Picking The Best Attribute

Attribute	Information Gain
Money	0
Weather	1

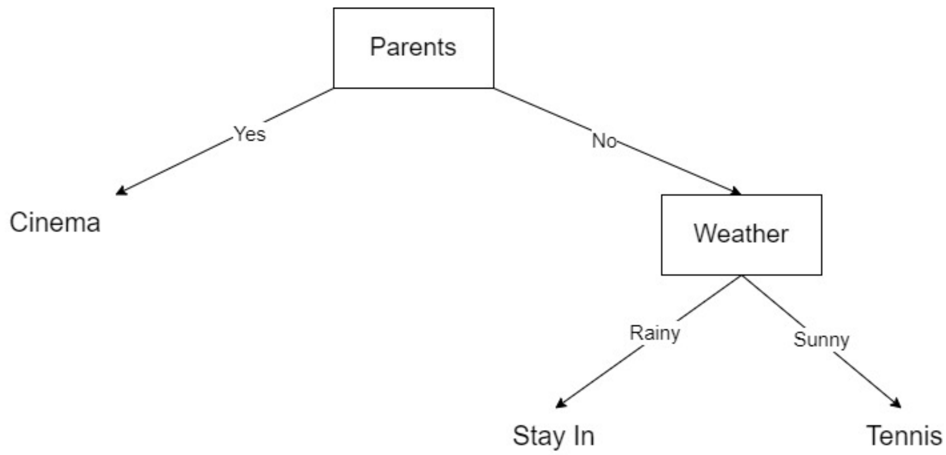
The selected feature is Weather. **Model accuracy on training data**

$$Accuracy_{\text{Train}} = \frac{\text{Correctly Labeled}}{\text{All train samples}} = \frac{6}{6} = 100\%$$

Model accuracy on test data

$$Accuracy_{\text{Test}} = \frac{\text{Correctly Labeled}}{\text{All test samples}} = \frac{2}{4} = 50\%$$

Decision Tree diagram:



It seems that one of the main reasons for the drop in accuracy on the test data was the absence of samples with different values for features in the training data. For example, no prediction can be made for the value Windy because the model has not seen it outside the dominance of the Parents feature.

2.3 In scenarios where only a limited number of labeled examples are available for training (and no extra data is available for testing or validation), propose a specific pruning technique that could be integrated into the decision tree algorithm to prevent overfitting. Justify why you believe this technique would be effective.

Since we do not have test data, we can use a different method of Post Pruning after allowing the tree to overfit. A method that could be added to the algorithm is to start from each leaf node after the tree is built. If its entropy is below a certain threshold, we move to the parent node and repeat this process until this condition is no longer met. Then, from that node where the entropy is not below the threshold, we prune its subtree to prevent overfitting and increase the model's generalization capability.