Decision Tree for basic formulas

Machine Learning Course A.A. 22/23

Filippo Fantinato 2041620

October 24, 2022

Contents

List of Tables

List of Figures

1 Exercise

Using the algorithm ID3, compute the decision tree corresponding to the task of realizing simple Boolean formulas (AND, OR, XOR, ...) in n variables.
The formulas I'm using are the following:

$$p_{[c=\cdot]} = \frac{|S_c|}{|S|}, E[S] = -p_- \log_2(p_-) - p_+ \log_2(p_+)$$

Figure 1: Entropy for binary classification

$$G(S, a = \cdot) = E[S] - \sum_{v \in V(a)} \frac{|S_{[a=v]}|}{|S|} \cdot E[S_{[a=v]}]$$

Figure 2: Information gain with entropy

1.1 AND

X	Y	Z	$X \wedge Y \wedge Z$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Table 1: And truth table

Here below you can find the D.T. building procedure following the algorithm ID3 exploiting entropy for the optimal attribute choice:

- Attributes $A = \{X, Y, Z\}$, Samples $S = [7-, 1+]$, total entropy $E[S] = -\frac{7}{8} \log_2 \frac{7}{8} - \frac{1}{8} \log_2 \frac{1}{8} = 0.54$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating the first node I need to compute the attribute that maximizes the information gain:
  - X:
  $$S_{X=0} = [4-, 0+], S_{X=1} = [3-, 1+]$$
  $$E[S_{X=0}] = 0, E[S_{X=1}] = 0.81$$
  $$G(S, a = X) = 0.54 - \frac{4}{8} \cdot 0 - \frac{4}{8} \cdot 0.81 = 0.135$$

- Y: analogous procedure to X

$$G(S, a = Y) = 0.135$$

- Z: analogous procedure to X

$$G(S, a = Z) = 0.135$$

The information gain is the same for each attribute, so I decide to choose X as optimal attribute.

- Given $X = 0$, Attribute $A = \{Y, Z\}$, Samples $S_{X=0} = [4-, 0+]$, total entropy $E[S_{X=0}] = 0$
Since all the samples belong to the same class, then the label of the this node is 0

- Given $X = 1$, Attribute $A = \{Y, Z\}$, Samples $S_{X=1} = [3-, 1+]$, total entropy $E[S_{X=1}] = 0.81$
Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain:
  - Y:

$$S_{X=1,Y=0} = [2-, 0+], S_{X=1,Y=1} = [1-, 1+]$$
$$E[S_{X=1,Y=0}] = 0, E[S_{X=1,Y=1}] = 1$$
$$G(S_{X=1}, a = Y) = 0.81 - \frac{2}{4} \cdot 0 - \frac{2}{4} \cdot 1 = 0.31$$

  - Z: analogous procedure to Y

$$G(S_{X=1}, a = Z) = 0.31$$

The information gain is the same for each attribute, so I decide to choose Y as optimal attribute.

- Given $X = 1, Y = 0$, Attribute $A = \{Z\}$, Samples $S_{X=1,Y=0} = [2-, 0+]$, total entropy $E[S_{X=1,Y=0}] = 0$
Since all the samples belong to the same class, then the label of the this node is 0

- Given $X = 1, Y = 1$, Attribute $A = \{Z\}$, Samples $S_{X=1,Y=1} = [1-, 1+]$, total entropy $E[S_{X=1,Y=1}] = 1$
Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain, but A is a singleton and so I choose Z as optimal attribute.

- Given $X = 1, Y = 1, Z = 0$, Attribute $A = \emptyset$, Samples $S_{X=1,Y=1,Z=0} = [1-, 0+]$
Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

- Given $X = 1, Y = 1, Z = 1$, Attribute $A = \emptyset$, Samples $S_{X=1,Y=1,Z=1} = [0-, 1+]$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 1

Graphically, the decision tree looks like:

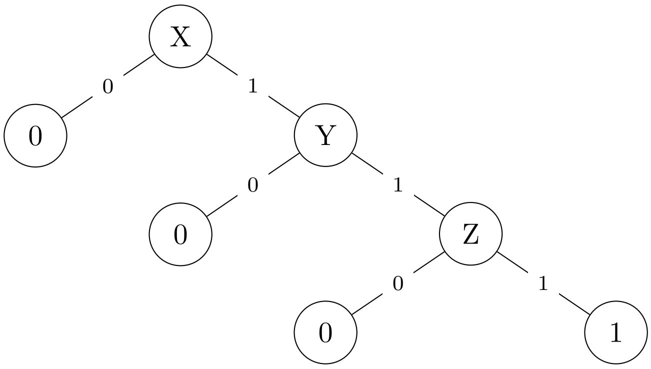

Figure 3: D.T. AND

1.2 OR

x	y	z	$x \vee y \vee z$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

Table 2: Or truth table

Here below you can find the D.T. building procedure following the algorithm ID3 exploiting entropy for the optimal attribute choice:

- Attributes $A = \{X, Y, Z\}$, Samples $S = [1-, 7+]$, total entropy $E[S] = -\frac{1}{8}\log_2\frac{1}{8} - \frac{7}{8}\log_2\frac{7}{8} = 0.54$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating the first node I need to compute the attribute that maximizes the information gain:
  - X:
$$S_{X=0} = [1-, 3+], S_{X=1} = [0-, 4+]$$
$$E[S_{X=0}] = 0.81, E[S_{X=1}] = 0$$
$$G(S, a = X) = 0.54 - \frac{4}{8}\cdot 0 - \frac{4}{8}\cdot 0.81 = 0.135$$

- Y: analogous procedure to X

$$G(S, a = Y) = 0.135$$

- Z: analogous procedure to X

$$G(S, a = Z) = 0.135$$

The information gain is the same for each attribute, so I decide to choose X as optimal attribute.

- Given $X = 1$, Attribute $A = \{Y, Z\}$, Samples $S_{X=1} = [0-, 4+]$, total entropy $E[S_{X=1}] = 0$
Since all the samples belong to the same class, then the label of the this node is 1

- Given $X = 0$, Attribute $A = \{Y, Z\}$, Samples $S_{X=0} = [1-, 3+]$, total entropy $E[S_{X=0}] = 0.81$
Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain:
  - Y:

$$S_{X=0,Y=0} = [1-, 1+], S_{X=0,Y=1} = [0-, 2+]$$
$$E[S_{X=0,Y=0}] = 1, E[S_{X=0,Y=1}] = 0$$
$$G(S_{X=0}, a = Y) = 0.81 - \frac{2}{4} \cdot 0 - \frac{2}{4} \cdot 1 = 0.31$$

  - Z: analogous procedure to Y

$$G(S_{X=0}, a = Z) = 0.31$$

The information gain is the same for each attribute, so I decide to choose Y as optimal attribute.

- Given $X = 0, Y = 1$, Attribute $A = \{Z\}$, Samples $S_{X=0,Y=1} = [0-, 2+]$, total entropy $E[S_{X=0,Y=1}] = 0$
Since all the samples belong to the same class, then the label of the this node is 1

- Given $X = 0, Y = 0$, Attribute $A = \{Z\}$, Samples $S_{X=0,Y=0} = [1-, 1+]$, total entropy $E[S_{X=0,Y=0}] = 1$
Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain, but A is a singleton and so I choose Z as optimal attribute.

- Given $X = 0, Y = 0, Z = 1$, Attribute $A = \emptyset$, Samples $S_{X=0,Y=0,Z=1} = [0-, 1+]$
Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 1

- Given $X = 0, Y = 0, Z = 0$, Attribute $A = \emptyset$, Samples $S_{X=0,Y=0,Z=0} = [1-, 0+]$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

Graphically, the decision tree looks like:

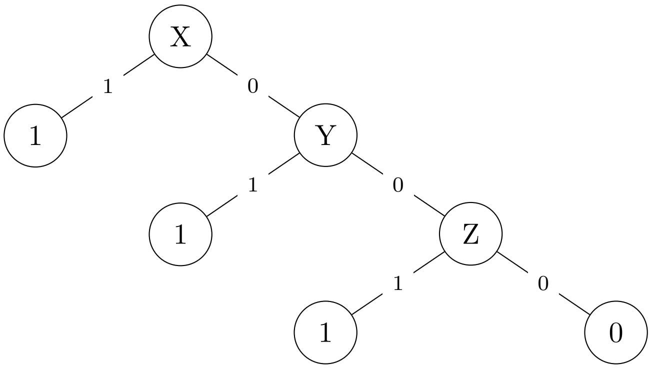

Figure 4: D.T. OR

1.3 XOR

x	y	z	$x \oplus y \oplus z$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

Table 3: XOR truth table

Here below you can find the D.T. building procedure following the algorithm ID3 exploiting entropy for the optimal attribute choice:

- Attributes $A = \{X, Y, Z\}$, Samples $S = [4-, 4+]$, total entropy $E[S] = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating the first node I need to compute the attribute that maximizes the information gain:
  - X:

$$S_{X=0} = [2-, 2+], S_{X=1} = [2-, 2+]$$
$$E[S_{X=0}] = 1, E[S_{X=1}] = 1$$
$$G(S, a = X) = 1 - \frac{4}{8} \cdot 1 - \frac{4}{8} \cdot 1 = 0$$

- Y: analogous procedure to X

$$G(S, a = Y) = 0$$

- Z: analogous procedure to X

$$G(S, a = Z) = 0$$

The information gain is the same for each attribute, so I decide to choose X as optimal attribute.

- Given $X = 0$, Attribute $A = \{Y, Z\}$, Samples $S_{X=0} = [2-, 2+]$, total entropy $E[S_{X=0}] = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating the first node I need to compute the attribute that maximizes the information gain:
  - Y:
  $$S_{X=0,Y=0} = [1-, 1+], S_{X=0,Y=1} = [1-, 1+]$$
  $$E[S_{X=0,Y=0}] = 1, E[S_{X=0,Y=1}] = 1$$
  $$G(S_{X=0}, a = Y) = 1 - \frac{2}{4} \cdot 1 - \frac{2}{4} \cdot 1 = 0$$
  - Z: analogous procedure to Y
  $$G(S_{X=0}, a = Z) = 0$$

  The information gain is the same for each attribute, so I decide to choose Y as optimal attribute.

- Given $X = 0, Y = 0$, Attribute $A = \{Z\}$, Samples $S_{X=0,Y=0} = [1-, 1+]$, total entropy $E[S_{X=0,Y=0}] = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain, but A is a singleton and so I choose Z as optimal attribute.

- Given $X = 0, Y = 0, Z = 0$, Attribute $A = \emptyset$, Samples $S_{X=0,Y=0,Z=0} = [1-, 0+]$, total entropy $E[S_{X=1,Y=0,Z=0}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

- Given $X = 0, Y = 0, Z = 1$, Attribute $A = \emptyset$, Samples $S_{X=0,Y=0,Z=1} = [0-, 1+]$, total entropy $E[S_{X=1,Y=0,Z=1}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

- Given $X = 0, Y = 1$, Attribute $A = \{Z\}$, Samples $S_{X=0,Y=1} = [1-, 1+]$, total entropy $E[S_{X=0,Y=1}] = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain, but A is a singleton and so I choose Z as optimal attribute.

- Given $X = 0, Y = 1, Z = 0$, Attribute $A = \emptyset$, Samples $S_{X=0,Y=1,Z=0} = [0-, 1+]$, total entropy $E[S_{X=1,Y=1,Z=0}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 1

- Given $X = 0, Y = 1, Z = 1$, Attribute $A = \emptyset$, Samples $S_{X=0,Y=1,Z=1} = [1-, 0+]$, total entropy $E[S_{X=1,Y=1,Z=1}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

- Given $X = 1$, Attribute $A = \{Y, Z\}$, Samples $S_{X=1} = [2-, 2+]$, total entropy $E[S_{X=1}] = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain:
  - Y:
  $$S_{X=1,Y=0} = [1-, 1+], S_{X=1,Y=1} = [1-, 1+]$$
  $$E[S_{X=1,Y=0}] = 1, E[S_{X=1,Y=1}] = 1$$
  $$G(S_{X=1}, a = Y) = 1 - \frac{2}{4} \cdot 0 - \frac{2}{4} \cdot 1 = 0$$
  - Z: analogous procedure to Y
  $$G(S_{X=1}, a = Z) = 0$$

  The information gain is the same for each attribute, so I decide to choose Y as optimal attribute.

- Given $X = 1, Y = 0$, Attribute $A = \{Z\}$, Samples $S_{X=1,Y=0} = [1-, 1+]$, total entropy $E[S_{X=1,Y=0}] = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain, but A is a singleton and so I choose Z as optimal attribute.

- Given $X = 1, Y = 0, Z = 0$, Attribute $A = \emptyset$, Samples $S_{X=1,Y=0,Z=0} = [0-, 1+]$, total entropy $E[S_{X=1,Y=0,Z=0}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 1

- Given $X = 1, Y = 0, Z = 1$, Attribute $A = \emptyset$, Samples $S_{X=1,Y=0,Z=1} = [1-, 0+]$, total entropy $E[S_{X=1,Y=0,Z=1}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

- Given $X = 1, Y = 1$, Attribute $A = \{Z\}$, Samples $S_{X=1,Y=1} = [1-, 1+]$, total entropy $E[S_{X=1,Y=1}] = 1$
  Since the samples don't belong to the same class and the set of attributes isn't empty, by creating this node I need to compute the attribute that maximizes the information gain, but A is a singleton and so I choose Z as optimal attribute.

- Given $X = 1, Y = 1, Z = 0$, Attribute $A = \emptyset$, Samples $S_{X=1,Y=1,Z=0} = [1-, 0+]$, total entropy $E[S_{X=1,Y=1,Z=0}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 1

- Given $X = 1, Y = 1, Z = 1$, Attribute $A = \emptyset$, Samples $S_{X=1,Y=1,Z=1} = [0-, 1+]$, total entropy $E[S_{X=1,Y=1,Z=1}] = 0$
  Since the set of attributes is empty, the leaf node is returned with label the majority class, that is 0

Graphically, the decision tree looks like:

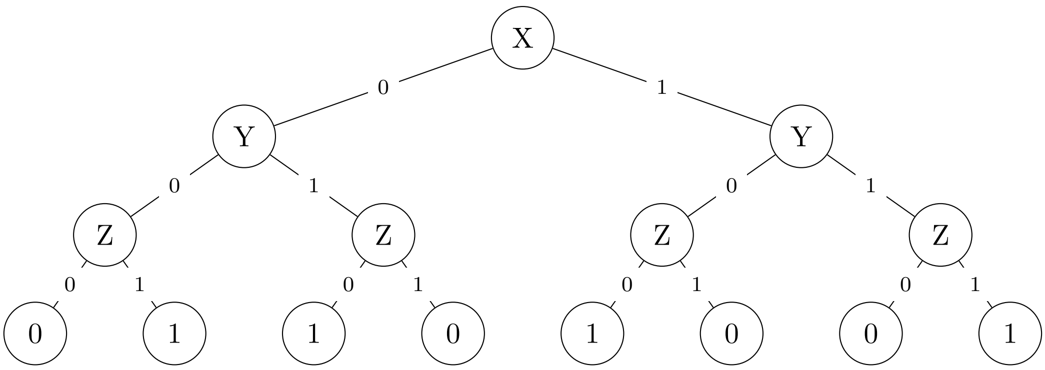

Figure 5: D.T. XOR