ID3 Algorithm - Complete Illustration

Consider the dataset $\mathcal{D}$ specified by the following table:

index	X	Y	Z	class
1	1	0	1	Yes
2	1	1	0	Yes
3	1	0	1	Yes
4	1	0	1	Yes
5	0	1	1	No
6	1	0	0	No
7	0	1	0	No
8	0	1	0	No
9	1	1	1	No
10	0	0	1	Yes

There are three attributes $\mathcal{A} = \{X, Y, Z\}$ and two classes $C = \{\text{Yes}, \text{No}\}$. Each attribute has possible values 0 and 1. Let us use indices 1 - 10 to denote elements of the dataset $\mathcal{D}$. That is, write $\mathcal{D} = \{1, \ldots, 10\}$.

Let me demonstrate the execution of the algorithm ID3 with impurity decrease (Gini) to select the best-classifying attributes in every call of ID3.

The algorithm proceeds as follows:

ID3($\mathcal{D}, \mathcal{A}$)

- At line 2 create the node τ_1 (see Image 1). [1]
- No "if" condition is satisfied, so we continue on line 10. On line 10, identify the best classifying attribute:
  - To compute the impurity decrease, we need to compute $Gini(\mathcal{D})$ for $\mathcal{D} = \{1, \ldots, 10\}$ as follows:
    - $p_{\text{Yes}} = p_{\text{No}} = 1/2$
    - $Gini(\mathcal{D}) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (1/2)^2 - (1/2)^2 = 0.5$
  - Consider $X \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, X)$ as follows:
    - Consider value 1 of X. Then $\mathcal{D}_1 = \{1, 2, 3, 4, 6, 9\}$.

index	X	Y	Z	class
1	1	0	1	Yes
2	1	1	0	Yes
3	1	0	1	Yes
4	1	0	1	Yes
6	1	0	0	No
9	1	1	1	No

[1] Note that the nodes of the tree are numbered sequentially so that each node gets a unique index. The indices do not correspond to the indices assigned by the pseudocode.

- Thus $p_{\text{Yes}} = 4/6$ and $p_{\text{No}} = 2/6$
- $Gini(\mathcal{D}_1) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (4/6)^2 - (2/6)^2 = 0.444$

* Consider value 0 of X. Then $\mathcal{D}_0 = \{5, 7, 8, 10\}$.

index	X	Y	Z	class
5	0	1	1	No
7	0	1	0	No
8	0	1	0	No
10	0	0	1	Yes

- Thus $p_{\text{Yes}} = 1/4$ and $p_{\text{No}} = 3/4$
- $Gini(\mathcal{D}_0) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (1/4)^2 - (3/4)^2 = 0.375$

$$\begin{aligned} ImpDec(\mathcal{D}, X) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\, Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\, Gini(\mathcal{D}_0) \\ &= 0.5 - (6/10) \cdot 0.444 - (4/10) \cdot 0.375 = 0.083 \end{aligned}$$

– Consider $Y \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, Y)$ as follows:

* Consider value 1 of Y. Then $\mathcal{D}_1 = \{2, 5, 7, 8, 9\}$.

index	X	Y	Z	class
2	1	1	0	Yes
5	0	1	1	No
7	0	1	0	No
8	0	1	0	No
9	1	1	1	No

- Thus $p_{\text{Yes}} = 1/5$ and $p_{\text{No}} = 4/5$
- $Gini(\mathcal{D}_1) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (1/5)^2 - (4/5)^2 = 0.320$

* Consider value 0 of Y. Then $\mathcal{D}_0 = \{1, 3, 4, 6, 10\}$.

index	X	Y	Z	class
1	1	0	1	Yes
3	1	0	1	Yes
4	1	0	1	Yes
6	1	0	0	No
10	0	0	1	Yes

- Thus $p_{\text{Yes}} = 4/5$ and $p_{\text{No}} = 1/5$
- $Gini(\mathcal{D}_0) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (4/5)^2 - (1/5)^2 = 0.320$

$$\begin{aligned} ImpDec(\mathcal{D}, Y) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\, Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\, Gini(\mathcal{D}_0) \\ &= 0.500 - (5/10) \cdot 0.320 - (5/10) \cdot 0.320 = 0.180 \end{aligned}$$

– Consider $Z \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, Z)$ as follows:

* Consider value 1 of Z. Then $\mathcal{D}_1 = \{1, 3, 4, 5, 9, 10\}$.

index	X	Y	Z	class
1	1	0	1	Yes
3	1	0	1	Yes
4	1	0	1	Yes
5	0	1	1	No
9	1	1	1	No
10	0	0	1	Yes

· Thus $p_{\text{Yes}} = 4/6$ and $p_{\text{No}} = 2/6$

· $Gini(\mathcal{D}_1) = 1 - p^2_{\text{Yes}} - p^2_{\text{No}} = 1 - (4/6)^2 - (2/6)^2 = 0.444$

∗ Consider value 0 of Y. Then $\mathcal{D}_0 = \{2, 6, 7, 8\}$.

index	X	Y	Z	class
2	1	1	0	Yes
6	1	0	0	No
7	0	1	0	No
8	0	1	0	No

· Thus $p_{\text{Yes}} = 1/4$ and $p_{\text{No}} = 3/4$

· $Gini(\mathcal{D}_0) = 1 - p^2_{\text{Yes}} - p^2_{\text{No}} = 1 - (1/4)^2 - (3/4)^2 = 0.375$

$$\begin{aligned} ImpDec(\mathcal{D}, Z) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\, Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\, Gini(\mathcal{D}_0) \\ &= 0.500 - (6/10) \cdot 0.444 - (4/10) \cdot 0.375 = 0.083 \end{aligned}$$

So, the attribute Y maximizes the decrease in impurity.

- Set the decision attribute of τ_1 to Y and continue recursively by calling
  - **ID3**$(\{2, 5, 7, 8, 9\}, \{X, Z\})$ giving a node τ_2
  - **ID3**$(\{1, 3, 4, 6, 10\}, \{X, Z\})$ giving a node τ_3.
- Connect τ_1 with τ_2 by an edge assigned 1, and τ_1 with τ_3 using an edge assigned 0.

Now, let us demonstrate the recursive calls.

ID3$(\{2, 5, 7, 8, 9\}, \{X, Z\})$

Now $\mathcal{D} = \{2, 5, 7, 8, 9\}$ and $\mathcal{A} = \{X, Z\}$.

index	X	Z	class
2	1	0	Yes
5	0	1	No
7	0	0	No
8	0	0	No
9	1	1	No

- At line 2, create the node τ_2.

- No "if" condition is satisfied, so we continue on line 10. On line 10, identify the best classifying attribute:
  - We have already computed
    $$Gini(\mathcal{D}) = 0.320$$
  - Consider $X \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, X)$ as follows:
    * Consider value 1 of X. Then $\mathcal{D}_1 = \{2, 9\}$.

index	X	Z	class
2	1	0	Yes
9	1	1	No

      · Thus $p_{\text{Yes}} = 1/2$ and $p_{\text{No}} = 1/2$
      · $Gini(\mathcal{D}_1) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (1/2)^2 - (1/2)^2 = 0.500$
    * Consider value 0 of X. Then $\mathcal{D}_0 = \{5, 7, 8\}$.

index	X	Z	class
5	0	1	No
7	0	0	No
8	0	0	No

      · Thus $p_{\text{Yes}} = 0$ and $p_{\text{No}} = 1$
      · $Gini(\mathcal{D}_0) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - 0^2 - 1^2 = 0$

$$\begin{aligned} ImpDec(\mathcal{D}, X) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\,Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\,Gini(\mathcal{D}_0) \\ &= 0.320 - (2/5) \cdot 0.500 - (3/5) \cdot 0.000 = 0.120 \end{aligned}$$

  - Consider $Z \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, Z)$ as follows:
    * Consider value 1 of Z. Then $\mathcal{D}_1 = \{5, 9\}$.

index	X	Z	class
5	0	1	No
9	1	1	No

      · Thus $p_{\text{Yes}} = 0$ and $p_{\text{No}} = 1$
      · $Gini(\mathcal{D}_1) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - 0^2 - 1^2 = 0$
    * Consider value 0 of Z. Then $\mathcal{D}_0 = \{2, 7, 8\}$.

index	X	Z	class
2	1	0	Yes
7	0	0	No
8	0	0	No

      · Thus $p_{\text{Yes}} = 1/3$ and $p_{\text{No}} = 2/3$
      · $Gini(\mathcal{D}_0) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (1/3)^2 - (2/3)^2 = 0.444$

$$\begin{aligned} ImpDec(\mathcal{D}, X) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\,Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\,Gini(\mathcal{D}_0) \\ &= 0.320 - (2/5) \cdot 0.000 - (3/5) \cdot 0.444 = 0.053 \end{aligned}$$

  So, the attribute X maximizes the decrease in impurity.

- Set the decision attribute of τ_2 to X and continue recursively by calling
  - **ID3**($\{2, 9\}, \{Z\}$) giving a node τ_4
  - **ID3**($\{5, 7, 8\}, \{Z\}$) giving a node τ_5.
- Connect τ_2 with τ_4 by an edge assigned 1, and τ_1 with τ_5 using an edge assigned 0.

ID3($\{2, 9\}, \{Z\}$)

Now $\mathcal{D} = \{2, 9\}$ and $\mathcal{A} = \{Z\}$.

index	Z	class
2	0	Yes
9	1	No

- At line 2, create the node τ_4
- No "if" condition is satisfied, so we continue on line 10. On line 10, identify the best classifying attribute:

  There is only Z, which is automatically selected.
- Set the decision attribute of τ_2 to Z and continue recursively by calling
  - **ID3**($\{9\}, \{\}$) giving a node τ_6
  - **ID3**($\{2\}, \{\}$) giving a node τ_7.
- Connect τ_4 with τ_6 by an edge assigned 1, and τ_4 with τ_7 using an edge assigned 0.

ID3($\{9\}, \{\}$)

Now $\mathcal{D} = \{9\}$ and $\mathcal{A} = \{\}$.

- At line 2, create the node τ_6
- The "if" at line 5 is satisfied since all elements of $\mathcal{D}$ are of class No. Assign No to τ_6 and return τ_6.

ID3($\{2\}, \{\}$)

Now $\mathcal{D} = \{2\}$ and $\mathcal{A} = \{\}$.

- At line 2, create the node τ_7
- The "if" at line 5 is satisfied since all elements of $\mathcal{D}$ are of class Yes. Assign Yes to τ_7 and return τ_7.

ID3$(\{5,7,8\},\{Z\})$

Now $\mathcal{D} = \{5,7,8\}$ and $\mathcal{A} = \{X, Z\}$.

- At line 2, create the node τ_5
- The "if" at line 5 is satisfied since all elements of $\mathcal{D}$ are of class No. Assign No to τ_5 and return τ_5.

ID3$(\{1,3,4,6,10\},\{X,Z\})$

Now $\mathcal{D} = \{1,3,4,6,10\}$ and $\mathcal{A} = \{X, Z\}$.

index	X	Z	class
1	1	1	Yes
3	1	1	Yes
4	1	1	Yes
6	1	0	No
10	0	1	Yes

- At line 2, create the node τ_3.
- No "if" condition is satisfied, so we continue on line 10. On line 10, identify the best classifying attribute:
  - We have already computed

$$Gini(\mathcal{D}) = 0.320$$

  - Consider $X \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, X)$ as follows:
    * Consider value 1 of X. Then $\mathcal{D}_1 = \{1,3,4,6\}$.

index	X	Z	class
1	1	1	Yes
3	1	1	Yes
4	1	1	Yes
6	1	0	No

      · Thus $p_{\text{Yes}} = 3/4$ and $p_{\text{No}} = 1/4$
      · $Gini(\mathcal{D}_1) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - (3/4)^2 - (1/4)^2 = 0.375$
    * Consider value 0 of X. Then $\mathcal{D}_0 = \{10\}$.

index	X	Z	class
10	0	1	Yes

      · Thus $p_{\text{Yes}} = 1$ and $p_{\text{No}} = 0$
      · $Gini(\mathcal{D}_0) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - 1^2 - 0^2 = 0$

$$\begin{aligned} ImpDec(\mathcal{D}, X) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\,Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\,Gini(\mathcal{D}_0) \\ &= 0.320 - (4/5) \cdot 0.375 - (1/5) \cdot 0.000 = 0.020 \end{aligned}$$

- Consider $Z \in \mathcal{A}$ and compute $ImpDec(\mathcal{D}, Z)$ as follows:
  - Consider value 1 of Z. Then $\mathcal{D}_1 = \{1, 3, 4, 10\}$.

index	X	Z	class
1	1	1	Yes
3	1	1	Yes
4	1	1	Yes
10	0	1	Yes

    - Thus $p_{\text{Yes}} = 1$ and $p_{\text{No}} = 0$
    - $Gini(\mathcal{D}_1) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - 1^2 - 0^2 = 0$
  - Consider value 0 of Z. Then $\mathcal{D}_0 = \{6\}$.

index	X	Z	class
6	1	0	No

    - Thus $p_{\text{Yes}} = 0$ and $p_{\text{No}} = 1$
    - $Gini(\mathcal{D}_0) = 1 - p_{\text{Yes}}^2 - p_{\text{No}}^2 = 1 - 0^2 - 1^2 = 0$

$$\begin{aligned} ImpDec(\mathcal{D}, X) &= Gini(\mathcal{D}) - (|\mathcal{D}_1|/|\mathcal{D}|)\, Gini(\mathcal{D}_1) - (|\mathcal{D}_0|/|\mathcal{D}|)\, Gini(\mathcal{D}_0) \\ &= 0.320 - (4/5) \cdot 0.000 - (1/5) \cdot 0.000 = 0.320 \end{aligned}$$

  So, the attribute Z maximizes the decrease in impurity.
- Set the decision attribute of τ_3 to Z and continue recursively by calling
  - **ID3**$(\{1, 3, 4, 10\}, \{X\})$ giving a node τ_8
  - **ID3**$(\{6\}, \{X\})$ giving a node τ_9.
- Connect τ_3 with τ_8 by an edge assigned 1, and τ_3 with τ_9 using an edge assigned 0.

ID3$(\{1, 3, 4, 10\}, \{X\})$

Now $\mathcal{D} = \{1, 3, 4, 10\}$ and $\mathcal{A} = \{X\}$.

- At line 2, create the node τ_8
- The "if" at line 5 is satisfied since all elements of $\mathcal{D}$ are of class Yes. Assign Yes to τ_8 and return τ_8.

ID3$(\{6\}, \{X\})$

Now $\mathcal{D} = \{6\}$ and $\mathcal{A} = \{X\}$.

- At line 2, create the node τ_9
- The "if" at line 5 is satisfied since all elements of $\mathcal{D}$ are of class No. Assign No to τ_9 and return τ_9.

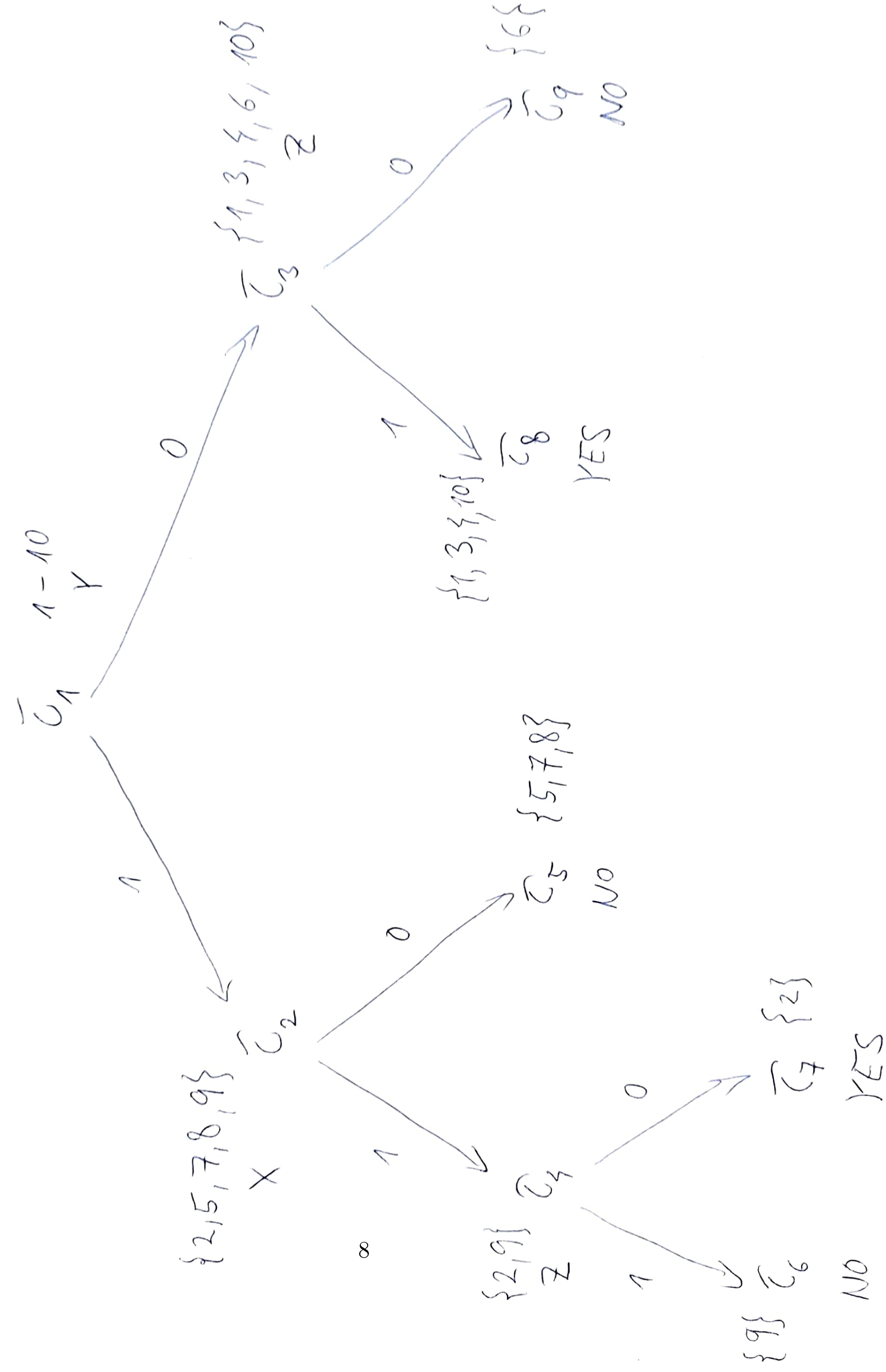

Figure 1: The tree