Group 6: Timm Behner, Philipp Bruckschen, Patrick Kaster, Markus Schwalb
MA-INF 4111 - Intelligent Learning and Analysis Systems: Machine Learning
Exercise Sheet 2

1. TDIDT

Begin with the emtpy decision tree and search for the best split in all attributes.
Entropy of examples $\mathrm{H}(X) = 0.994$

Wind

	+	-
Weak	3	2
Strong	3	3

$$\mathrm{Gain}(X|\mathrm{Wind}) = \mathrm{H}(X) - \sum_{w \in \{\mathrm{Weak}, \neg\mathrm{Weak}\}} \mathrm{P}(\mathrm{Wind} = w)\, \mathrm{H}(X|\mathrm{Wind} = w) \tag{1}$$

$$= 0.0072 \tag{2}$$

Temperature

Compute the best splitting test for this attribute

Value	Outcome
7	+
8	+
10	+
11	+
20	+
21	-
23	+
23	-
31	-
32	-
34	-

	+	-
$20.5 >$	5	0
$20.5 \leq$	1	5

$\mathrm{H}(X|\mathrm{Temp} < 20.5) = 0$ $\mathrm{H}(X|\mathrm{Temp} \geq 20.5) = 0.65$

	+	-
$22 >$	5	1
$22 \leq$	1	4

$\mathrm{H}(X|\mathrm{Temp} < 22) = 0.65$ $\mathrm{H}(X|\mathrm{Temp} \geq 22) = 0.7219$

From this follows that the maximum gain for this attribute is $\mathrm{Gain}(X|\mathrm{Temp} \leq 20.5) = 0.6395$

Outlook

	+	-
Sunny	2	2
¬ Sunny	4	3

$\Rightarrow \mathrm{Gain}(X|\mathrm{Outlook} = \mathrm{Sunny}) = 0.0034$

	+	-
Overcast	1	2
¬ Overcast	5	3

$\Rightarrow \mathrm{Gain}(X|\mathrm{Outlook} = \mathrm{Overcast}) = 0.0494$

	+	-
Rain	3	1
¬ Rain	3	4

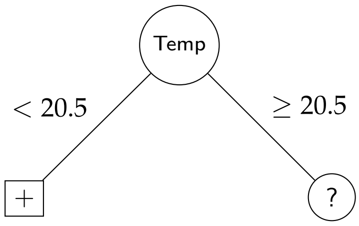
$\Rightarrow \text{Gain}(X|\text{Outlook} = \text{Rain}) = 0.072$

Now check for

$$\text{Gain}(X|\text{Outlook} \in \{\text{Rain}, \text{Overcast}\} = 0.0035 \tag{3}$$

From this follows that the maximum information gain for this attribute is $\Rightarrow \text{Gain}(X|\text{Outlook} = \text{Rain}) = 0.072$

The attribute with the maximum information gain is Temperatur (split with 20.5 $\leq$)



Search for best attribute to split in $\{\text{Overcast}, \text{Wind}\}$

Overcast

	+	-
Sunny	1	2
¬ Sunny	0	3

$\Rightarrow \text{Gain}(X|\text{Outlook} = \text{Sunny}) = 0.1901$

	+	-
Overcast	0	2
¬ Overcast	1	3

$\Rightarrow \text{Gain}(X|\text{Outlook} = \text{Overcast}) = 0.1091$

	+	-
Rain	0	1
¬ Rain	1	4

$\Rightarrow \text{Gain}(X|\text{Outlook} = \text{Rain}) = 0.0484$

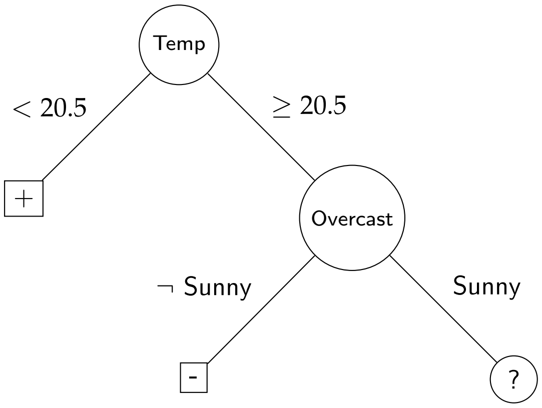
And $\text{Gain}(X|\text{Outlook} \in \{\text{Sunny}, \text{Overcast}\}) = 0.0484$

Wind

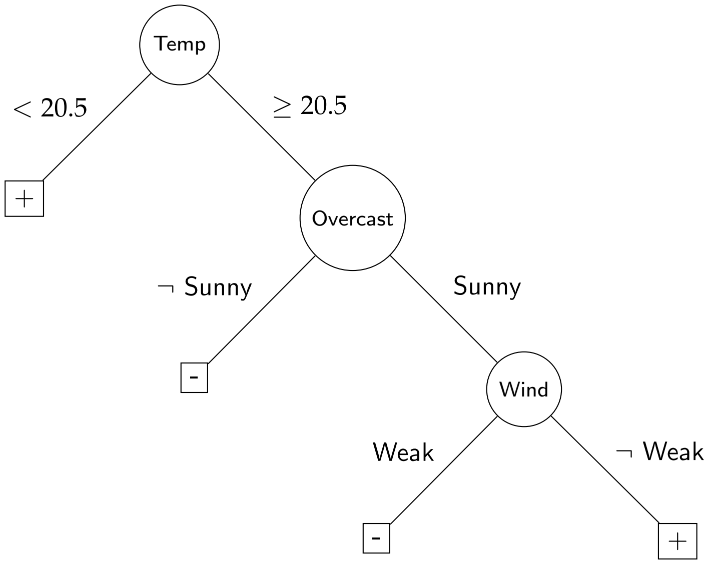
	+	-
Weak	1	4
¬ Weak	0	1

$\Rightarrow \text{Gain}(X|\text{Wind}) = 0.1091$

From this follows that the maximum information gain for this node is $\text{Gain}(X|\text{Outlook} = \text{Sunny}) = 0.1909$



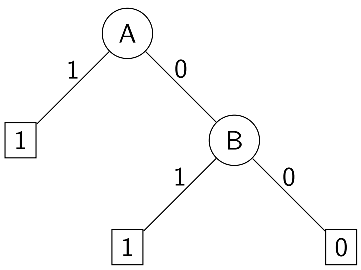
Now only the attribute Wind is left.



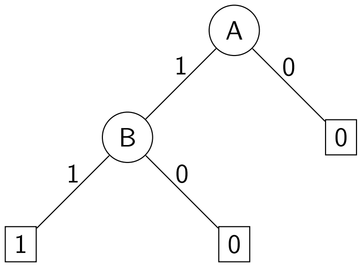
where the left right leaf of the Wind node was classified as positive, because at least half of the examples in this node were positive. ■

2. Decision Tree Representations of Boolean Functions

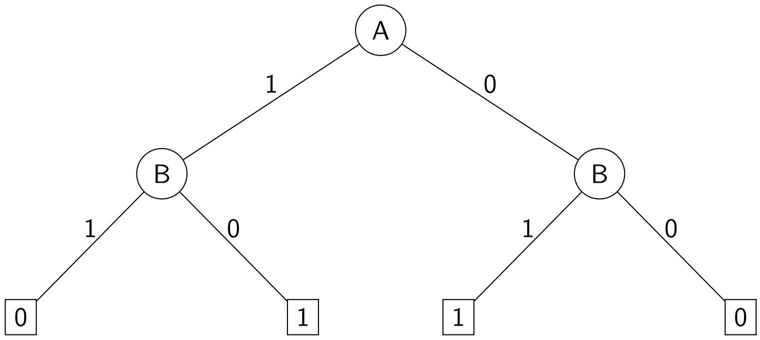
1. $f(A, B) = A \lor B$



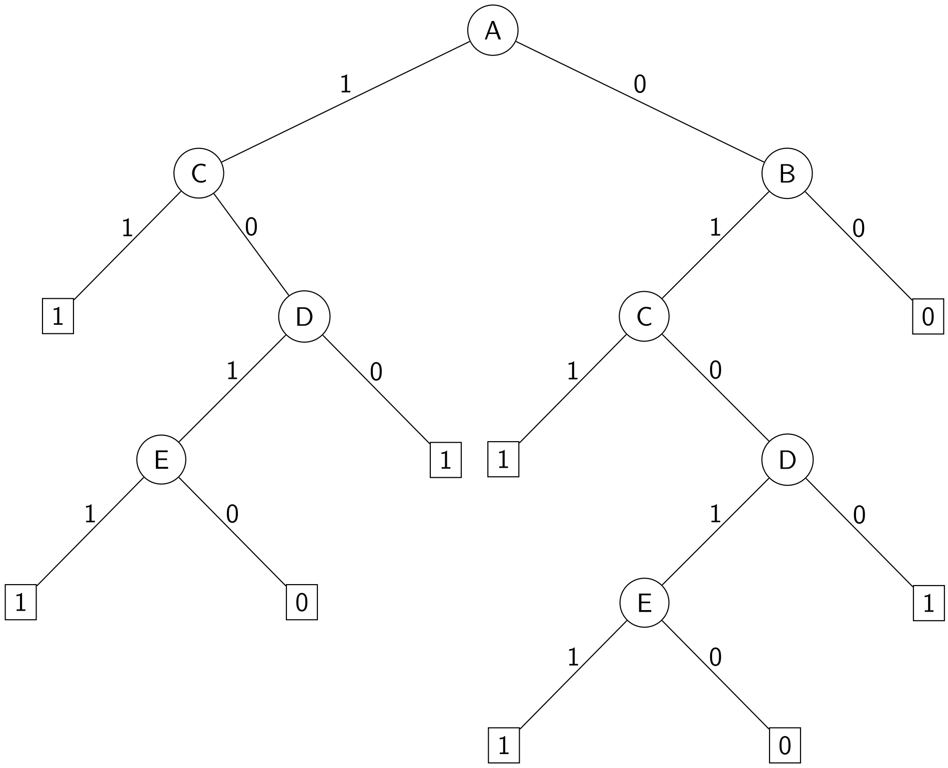
2. $f(A, B) = A \wedge B$



3. $f(A, B) = A \oplus B$



4. $f(A, B, C, D) = (A \vee B) \wedge (C \vee \neg D \vee E)$



$\blacksquare$

3. Properties of the Entropy

(i)

$$
\begin{aligned}
H(X) = H(p_1 \dots p_n) = & \sum_{i=1}^{n} -p_i \log_2 p_i \\
= & \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \\
= & \mathbb{E}\left[\log_2 (X)\right] \leq \log_2 (\mathbb{E}[X]) \qquad (4) \\
= & \log_2 \left(\sum_{i=1}^{n} p_i \frac{1}{p_i} \right) \\
= & \log_2 n
\end{aligned}
$$

We used the well known concavity of the logarithm in applying Jensen's inequality in (4).

(ii)

$$\begin{aligned}
\text{Gain}\left(X|Y\right) =& H\left(X\right) - H\left(X|Y\right) \\
=& \sum_X -P\left(X\right)\log_2 P\left(X\right) - \sum_{v\in\text{value}(Y)} P\left(Y=v\right) H\left(X|Y=v\right) \\
=& \sum_X -P\left(X\right)\log_2 P\left(X\right) \\
-& \sum_{v\in\text{value}(Y)} P\left(Y=v\right) \sum_X \left(-P\left(X|Y=v\right)\log_2 P\left(X|Y=v\right)\right)
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow -\text{Gain}\left(X|Y\right) =& \sum_X P\left(X\right)\log_2 P\left(X\right) \\
-& \sum_{v\in\text{value}(Y)} P\left(Y=v\right) \sum_X P\left(X|Y=v\right)\log_2 P\left(X|Y=v\right) \\
=& \sum_X \sum_{v\in\text{value}(Y)} P\left(X, Y=v\right)\log_2 P\left(X\right) \\
-& \sum_{v\in\text{value}(Y)} \sum_X P\left(Y=v\right) P\left(X|Y=v\right)\log_2 P\left(X|Y=v\right) \\
=& \sum_X \sum_{v\in\text{value}(Y)} P\left(X, Y=v\right)\log_2 P\left(X\right) \\
-& \sum_{v\in\text{value}(Y)} \sum_X P\left(X, Y=v\right)\log_2 P\left(X|Y=v\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_X \sum_{v \in \text{value}(Y)} P\left(X, Y = v\right)\left(\log_2 P\left(X\right) - \log_2 P\left(X|Y = v\right)\right) \\
&= \sum_X \sum_{v \in \text{value}(Y)} P\left(X, Y = v\right)\left(\log_2\left(\frac{P\left(X\right)}{P\left(X|Y = v\right)}\right)\right) \\
&= \sum_X \sum_{v \in \text{value}(Y)} P(Y = v) P\left(X|Y = v\right)\left(\log_2\left(\frac{P\left(X\right)}{P\left(X|Y = v\right)}\right)\right) \\
&= \sum_{v \in \text{value}(Y)} P(Y = v) \sum_X P\left(X|Y = v\right)\left(\log_2\left(\frac{P\left(X\right)}{P\left(X|Y = v\right)}\right)\right) && (5) \\
&\leq \sum_{v \in \text{value}(Y)} P(Y = v)\left(\log_2\left(\sum_X \frac{P\left(X\right) P\left(X|Y = v\right)}{P\left(X|Y = v\right)}\right)\right) && (6) \\
&\leq \log_2\left(\sum_{v \in \text{value}(Y)} \sum_X \frac{P\left(X\right) P\left(X|Y = v\right) P(Y = v)}{P\left(X|Y = v\right)}\right) && (7) \\
&= \log_2\left(\sum_{v \in \text{value}(Y)} P(Y = v) \sum_X P\left(X\right)\right) \\
&\leq \log_2\left(\sum_{v \in \text{value}(Y)} P(Y = v)\right) \\
&= \log_2\left(1\right) = 0
\end{aligned}
$$

So we conclude:

$$
-\text{Gain}\left(X|Y\right) \leq \log_2\left(1\right) = 0 \Leftrightarrow \text{Gain}\left(X|Y\right) \geq 0
$$

We used Jensen's inequality from task (i) from equation (5) to (6) and again from (6) to (7).

■