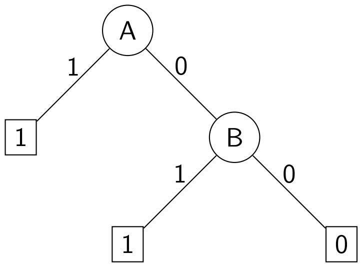
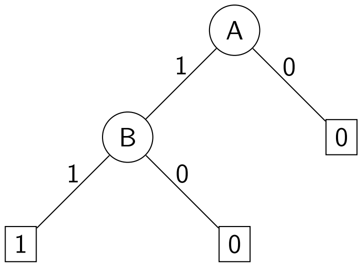
Group 6: Timm Behner, Philipp Bruckschen, Patrick Kaster, Markus Schwalb
MA-INF 4111 - Intelligent Learning and Analysis Systems: Machine Learning
Exercise Sheet 2

2. Decision Tree Representations of Boolean Functions

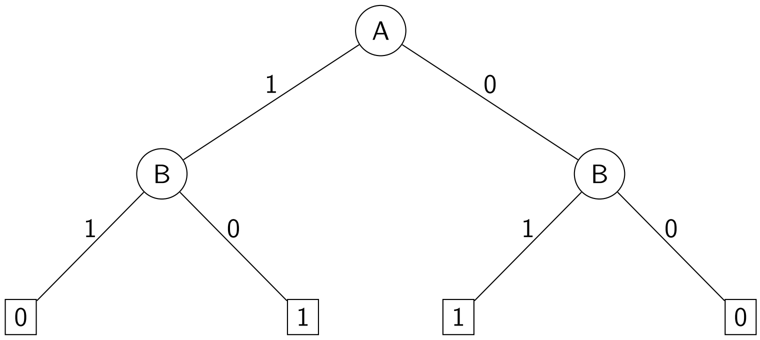
1. $f(A, B) = A \vee B$



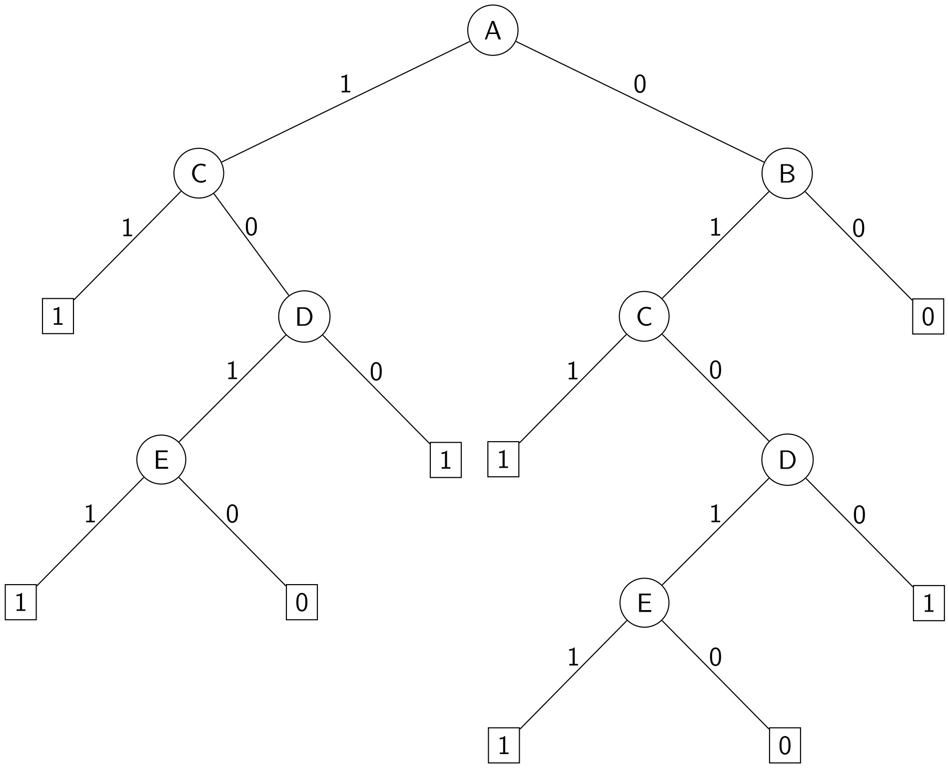
2. $f(A, B) = A \wedge B$



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



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



∎

3. Properties of the Entropy

(i)

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

We used the well known concavity of the logarithm in applying Jensen's inequality in (1).

(ii)

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

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

$$
\begin{aligned}
&= \sum_X \sum_{v \in \text{value}(Y)} P(X, Y = v) \left(\log_2 P(X) - \log_2 P(X|Y = v)\right) \\
&= \sum_X \sum_{v \in \text{value}(Y)} P(X, Y = v) \left(\log_2 \left(\frac{P(X)}{P(X|Y = v)}\right)\right) \\
&= \sum_X \sum_{v \in \text{value}(Y)} P(Y = v) P(X|Y = v) \left(\log_2 \left(\frac{P(X)}{P(X|Y = v)}\right)\right) \\
&= \sum_{v \in \text{value}(Y)} P(Y = v) \sum_X P(X|Y = v) \left(\log_2 \left(\frac{P(X)}{P(X|Y = v)}\right)\right) && (2) \\
&\leq \sum_{v \in \text{value}(Y)} P(Y = v) \left(\log_2 \left(\sum_X \frac{P(X) P(X|Y = v)}{P(X|Y = v)}\right)\right) && (3) \\
&\leq \log_2 \left(\sum_{v \in \text{value}(Y)} \sum_X \frac{P(X) P(X|Y = v) P(Y = v)}{P(X|Y = v)}\right) && (4) \\
&= \log_2 \left(\sum_{v \in \text{value}(Y)} P(Y = v) \sum_X P(X)\right) \\
&\leq \log_2 \left(\sum_{v \in \text{value}(Y)} P(Y = v)\right) \\
&= \log_2 (1) = 0
\end{aligned}
$$

So we conclude:

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

We used Jensen's inequality from task (i) from equation (2) to (3) and again from (3) to (4).

■