1. On the first round, we must consider all the attribute choices. We thus have:

$$\begin{aligned}
ENT(A, d) &= ENT(a, d) + ENT(\bar{a}, d) \\
&= Pr(a)\big[- Pr(d|a) \cdot \log_2(Pr(d|a)) - Pr(d|a) \cdot \log_2(Pr(d|a))\big] \\
&\quad + Pr(\bar{a})\big[- Pr(d|\bar{a}) \cdot \log_2(Pr(d|\bar{a})) - Pr(d|\bar{a}) \cdot \log_2(Pr(d|\bar{a}))\big] \\
&= \frac{12}{23}\left[-\frac{3}{12}\log_2\left(\frac{3}{12}\right) - \frac{9}{12}\log_2\left(\frac{9}{12}\right)\right] + \frac{11}{23}\left[-\frac{7}{11}\log_2\left(\frac{7}{11}\right) - \frac{4}{11}\log_2\left(\frac{4}{11}\right)\right] \\
&\approx 0.876
\end{aligned}$$

$$\begin{aligned}
ENT(B, d) &= ENT(b, d) + ENT(\bar{b}, d) \\
&= Pr(b)\big[- Pr(d|b) \cdot \log_2(Pr(d|b)) - Pr(d|b) \cdot \log_2(Pr(d|b))\big] \\
&\quad + Pr(\bar{b})\big[- Pr(d|\bar{b}) \cdot \log_2(Pr(d|\bar{b})) - Pr(d|\bar{b}) \cdot \log_2(Pr(d|\bar{b}))\big] \\
&= \frac{9}{23}\left[-\frac{2}{9}\log_2\left(\frac{2}{9}\right) - \frac{7}{9}\log_2\left(\frac{7}{9}\right)\right] + \frac{14}{23}\left[-\frac{8}{14}\log_2\left(\frac{8}{14}\right) - \frac{6}{14}\log_2\left(\frac{6}{14}\right)\right] \\
&\approx 0.899
\end{aligned}$$

$$\begin{aligned}
ENT(C, d) &= ENT(c, d) + ENT(\bar{c}, d) \\
&= Pr(c)\big[- Pr(d|c) \cdot \log_2(Pr(d|c)) - Pr(d|c) \cdot \log_2(Pr(d|c))\big] \\
&\quad + Pr(\bar{c})\big[- Pr(d|\bar{c}) \cdot \log_2(Pr(d|\bar{c})) - Pr(d|\bar{c}) \cdot \log_2(Pr(d|\bar{c}))\big] \\
&= \frac{16}{23}\left[-\frac{6}{16}\log_2\left(\frac{6}{16}\right) - \frac{10}{16}\log_2\left(\frac{10}{16}\right)\right] + \frac{7}{23}\left[-\frac{4}{7}\log_2\left(\frac{4}{7}\right) - \frac{3}{7}\log_2\left(\frac{3}{7}\right)\right] \\
&\approx 0.964
\end{aligned}$$

Since splitting on A results in the lowest entropy and thus greatest information gain, we split on that first. This results in

$$\left\{ \begin{array}{ll} \{+x_5, +x_7, -x_6, -x_8\}, & \text{for } A = F \\ \{+x_1, +x_2, -x_3, -x_4\}, & \text{for } A = T \end{array} \right\}$$

At this point we have completed our search, since $ENT(C|\bar{a}) = ENT(B|a) = 0$. Therefore, our resulting tree is:
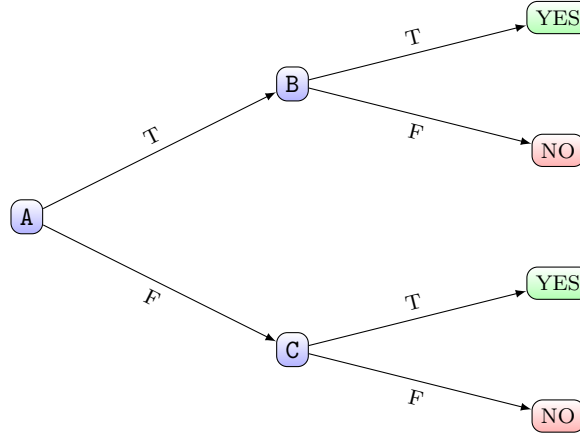
2. We assume that all nodes utilize the step function with threshold t as labeled on the node. These are therefore of the form

$$g(I) = \left\{ \begin{array}{ll} 1, & \text{for } I \geq t \\ 0, & \text{for } I < t \end{array} \right\}$$

Green arrows represent a weight of one, and red represent a weight of -1.
We can break our XOR into the sentence

$$(A \wedge C \wedge -D) \vee (-B \wedge C \wedge -D) \vee (-A \wedge -C \wedge -D) \vee (B \wedge -C \wedge D)$$

We use each of the four subsentences as one of the internodes; thus our network is: