CS161: FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE

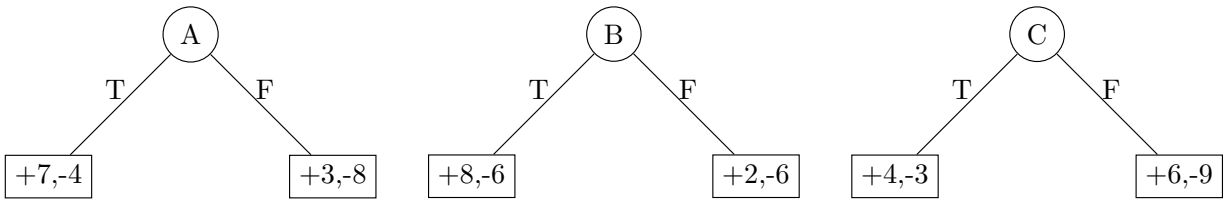
Fall 2014
Solution to Assignment 9

1. Consider the table below which represents a dataset by listing each unique example with the number of times it appears in the dataset. Construct the decision tree learned from this data by finding the most discriminating attribute at each step. Show precisely how you decided on the most discriminating attribute at each step by computing the expected entropies of the remaining attributes.

Example	Input Attributes			Class	#
	A	B	C	D	
$\mathbf{x}_1$	T	T	T	Yes	1
$\mathbf{x}_2$	T	T	F	Yes	6
$\mathbf{x}_3$	T	F	T	No	3
$\mathbf{x}_4$	T	F	F	No	1
$\mathbf{x}_5$	F	T	T	Yes	1
$\mathbf{x}_6$	F	T	F	No	6
$\mathbf{x}_7$	F	F	T	Yes	2
$\mathbf{x}_8$	F	F	F	No	2

Solution

The root of the decision tree will be A, B or C. Splitting the examples on A, B and C results in the following trees:



Using these trees, we can compute the expected entropies of A, B and C as follows:

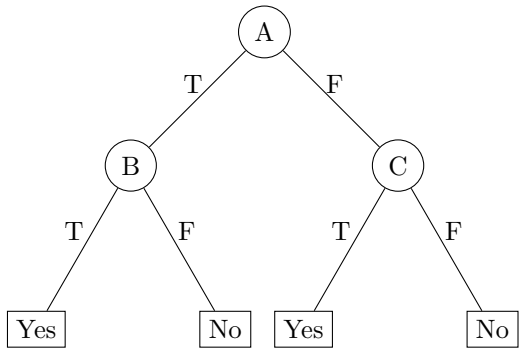
$$\begin{aligned} EE(A) = &- (7/11 \cdot \log_2(7/11) + 4/11 \cdot \log_2(4/11)) \cdot 11/22 \\ &- (3/11 \cdot \log_2(3/11) + 8/11 \cdot \log_2(8/11)) \cdot 11/22 \\ =&0.896 \end{aligned}$$

$$
\begin{aligned}
EE(B) = &- (8/14 \cdot \log_2(8/14) + 6/14 \cdot \log_2(6/14)) \cdot 14/22 \\
&- (2/8 \cdot \log_2(2/8) + 6/8 \cdot \log_2(6/8)) \cdot 8/22 \\
=&0.922
\end{aligned}
$$

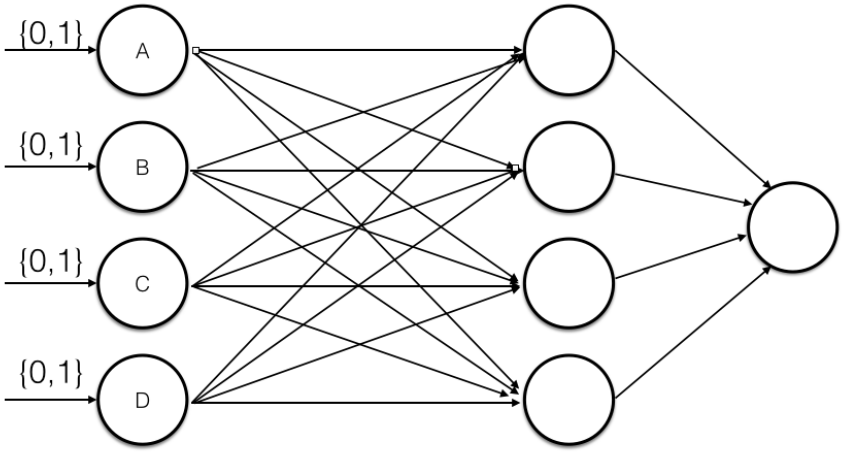
$$
\begin{aligned}
EE(C) = &- (4/7 \cdot \log_2(4/7) + 3/7 \cdot \log_2(3/7)) \cdot 7/22 \\
&- (6/15 \cdot \log_2(6/15) + 9/15 \cdot \log_2(9/15)) \cdot 15/22 \\
=&0.975
\end{aligned}
$$

A is the most discriminating attribute as its expected entropy is the smallest. Therefore, A will be the root of the decision tree.

Next, we need to find the attributes to test when A=T and A=F. In the case of A=T, assume that we split the examples on B. If B=T, all examples are positive. If B=F, all examples are negative. As its expected entopy is 0, B is the most discriminating attribute. Similarly, in the case of A=F, we can divide the positive and negative examples into two groups by splitting on C. Therefore, we can pick C as the next attribute. Final decision tree is the following:



2. Create a two layer neural network that uses the step function to implement $(A \vee \neg B) \oplus (\neg C \vee D)$, where $\oplus$ is the XOR function. You can either use the network structure provided below or another structure you construct. After drawing your network, clearly show the weights and activation function for each node. Assume inputs of $\{0, 1\}$ for each input variable. Note that solutions with more than two layers will still receive partial credit.

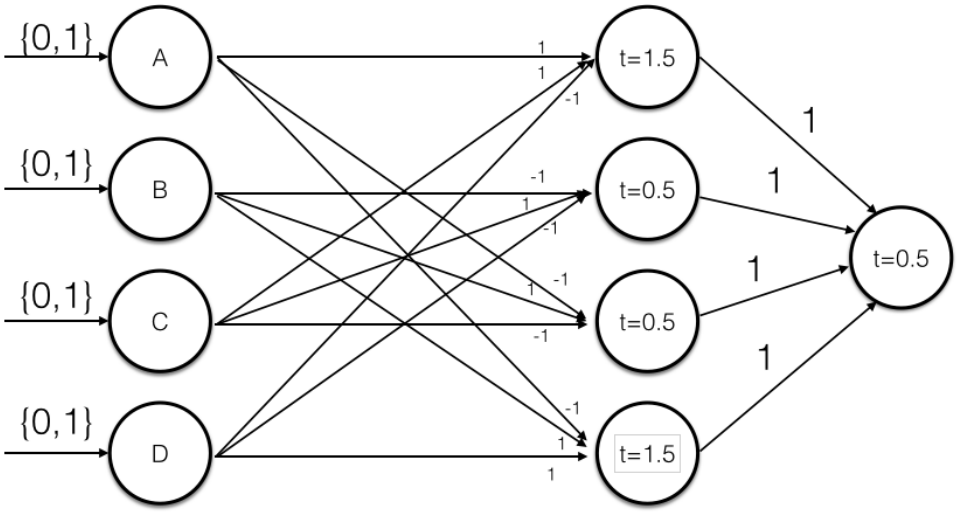


Solution

Note that as $X \oplus Y = (X \wedge \neg Y) \vee (\neg X \wedge Y)$, we can rewrite $(A \vee \neg B) \oplus (\neg C \vee D)$ as:

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

You can see below a neural network for this function with proper weights for the edges and thresholds for the step functions for the nodes.



Note that other variations of this answer are possible, e.g. an example with an edge from A to the second of the first layer nodes, with weight 0.