

EECS545 Lecture 20 Quiz Solutions

1. Which of the following are true about Decision Trees? (Select all that apply)
- (a) Decision Trees are often simpler to understand and interpret than other models
 - (b) Decision Tree learning is stable under variations in data (e.g., sampling training data)
 - (c) Decision Trees can be used for multi-class classification problems
 - (d) Decision Trees are good for learning outputs that depend on multiple inputs, such as XOR, parity, or multiplexer problems

Solution: (a) and (c).

2. Which of the following are true about ensemble methods? (Select all that apply)
- (a) Using an Adaboost over a single weak learner reduces bias
 - (b) Using an averaging-based ensemble over a single model reduces variance
 - (c) Averaging-based ensemble methods are guaranteed to improve performance (e.g., generalization error) if individual models (weak classifiers) in the ensemble are independent.
 - (d) AdaBoost doesn't work well with depth-1 decision trees

Solution: (a) and (b).

3. Suppose we have two independent variables X and Y . What is the information gain $IG(X, Y)$?

Solution: 0. Intuitively, we cannot gain any information about X from Y (or vice-versa) because they are independent.

Notice that $H(X) = H(X|Y = y) = H(X|Y)$ because $p(x|y) = p(x)$. Hence, $IG(X, Y) = H(X) - H(X|Y) = 0$.

4. Suppose we roll two fair 6-sided dice, dice A and dice B. Suppose the sum of the two dice is 4. What is the information gain on dice A if we know the value of dice B? I.e. What is $IG(A, B)$?

Solution:

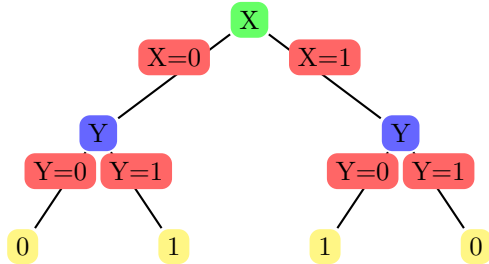
$$H(A) = -(1/3) * \log(1/3) - (1/3) * \log(1/3) - (1/3) * \log(1/3)$$

$H(A|B)$ must be 0, because the value of B completely determines the value of A.

$$IG(A|B) = H(A) = 1.585$$

5. Suppose we would like to solve the XOR problem with a decision tree. I.e. Given $X, Y \in \{0, 1\}$, we would like to predict $\text{XOR}(X, Y)$. What is the minimum number of nodes in the decision tree needed to predict XOR with 100% accuracy? Include leaf nodes in your count.

Solution: 7. We will need the maximum possible of nodes for a 2-feature decision tree. 1 root node, 2 middle-nodes, and 4 leaf nodes.



This tikz picture was draw using ChatGPT (with some light editing)!

Queries: “Can you draw me a tikz picture of a decision tree classifying the XOR problem?” Output:

```
\begin{tikzpicture}[grow=right,level distance=1.5cm]
\tikzstyle{every node}=[fill=red!60,rounded corners]
\tikzstyle{edge from parent}=[red,-o,thick,draw]
\tikzstyle{edge from parent path}=[red,-o,thick,draw]
\tikzstyle{level 1}=[sibling distance=4cm,thick]
\tikzstyle{level 2}=[sibling distance=2cm,thick]
\node [fill=green!60]{XOR}
child {node [fill=blue!60] {AND}
child {node [fill=yellow!60] {0}}
child {node [fill=yellow!60] {1}}
}
child {node [fill=blue!60] {AND}
child {node [fill=yellow!60] {1}}
child {node [fill=yellow!60] {0}}
};
\end{tikzpicture}
```

I further queried ChatGPT to add edges, and relabel some nodes. Amazing!