CS322 Fall 1999
Module 11 (Decision Tree Learning)

Assignment 11

Solution

Question 1

In electronic commerce applications we want to make predictions about what a user will do. Consider the following made-up data used to predict whether someone will ask for more information (*more_info*) based on whether they accessed from an educational domain (*edu*), whether this is a first visit (*first*), whether they have bought goods from an affiliated company (*bought*), and whether they have visited a famous online information store (*visited*).

Example	*bought*	*edu*	*first*	*visited*	*more_info*
e_1	false	true	false	false	true
e_2	true	false	true	false	false
e_3	false	false	true	true	true
e_4	false	false	true	false	false
e_5	false	false	false	true	false
e_6	true	false	false	true	true
e_7	true	false	false	false	true
e_8	false	true	true	true	false
e_9	false	true	true	false	false
e_{10}	true	true	true	false	true
e_{11}	true	true	false	true	true
e_{12}	false	false	false	false	true

We want to use this data to learn the value of *more_info* as a function of the values of the other variables.

Suppose we measure the error of a decision tree as the number of misclassified examples. The optimal decision tree from a class of decision trees is an element of the class with minimal error.

(a) Give the optimal decision tree with only one node. What is the error of this tree?

(b) Give the optimal decision tree of depth 2 (i.e., the root node is the only node with children). For each node in the tree give the examples that are filtered to that node. What is the error of this tree?

(c) Give the decision tree that is produced by the top-down induction algorithm run to completion, where we split on the attribute that reduces the error the most. For each node in the tree specify which examples are filtered to that node. As well as drawing the tree, give the tree in the format of question 2 of assignment 7 (i.e., in terms of *if* (*Att*, *Then*, *Else*)).

(d) Give two instances that don't appear in the examples above and show how they are classified. Use this to explain the bias inherent in the tree (how does the bias give you these particular predications?).

(e) How can overfitting occur in the learned network? Explain in terms of this example.

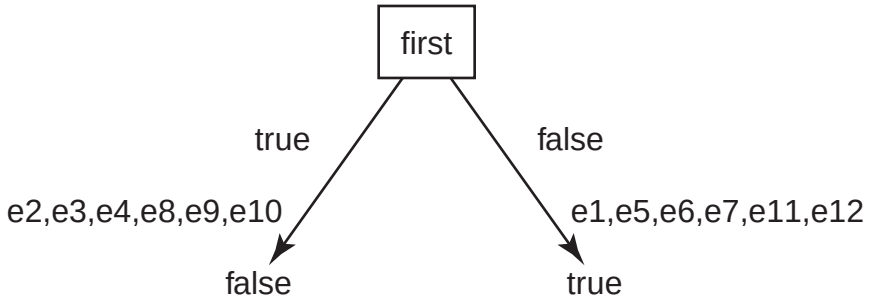

Figure 1: Classification of *more_info* for a tree of depth 2.

Solution

(a) Give the optimal decision tree with only one node. What is the error of this tree?

The optimal decision tree is the one with *more_info* = *true*. It makes 5 errors. [It gets examples e_2, e_4, e_5, e_8 and e_9 wrong.].

(b) Give the optimal decision tree of depth 2 (i.e., the root node is the only node with children). For each node in the tree give the examples that are filtered to that node. What is the error of this tree?

There is one non-leaf node, labelled with *first*. The tree is given in Figure 1. This has error of 3 (it misclassifies examples $e3$, $e10$ and $e5$).

(c) Give the decision tree that is produced by the top-down induction algorithm run to completion, where we split on the attribute that reduces the error the most. For each node in the tree specify which examples are filtered to that node. As well as drawing the tree, give the tree in the format of question 2 of assignment 7 (i.e., in terms of *if*(*Att*, *Then*, *Else*)).

There are many possible answers to this problem depending on how the arbitrary choices are resolved. All solutions have *first* at the root.

One solution is given in Figure 2.

This tree can be written as:

```
if(first,
   if(edu,
         if(bought,true,false),
         if(visited,true,false)),
   if(edu,
         true,
         if(bought,
               true,
               if(visited,false,true))))
```

(d) Give two instances that don't appear in the examples above and show how they are classified. Use this to explain the bias inherent in the tree (how does the bias give you these particular predications?).

There are 4 instances that don't appear in the examples above:

bought	*edu*	*first*	*visited*	*more_info*
true	true	true	true	true
true	true	false	false	true
true	false	true	true	true
false	true	false	true	true

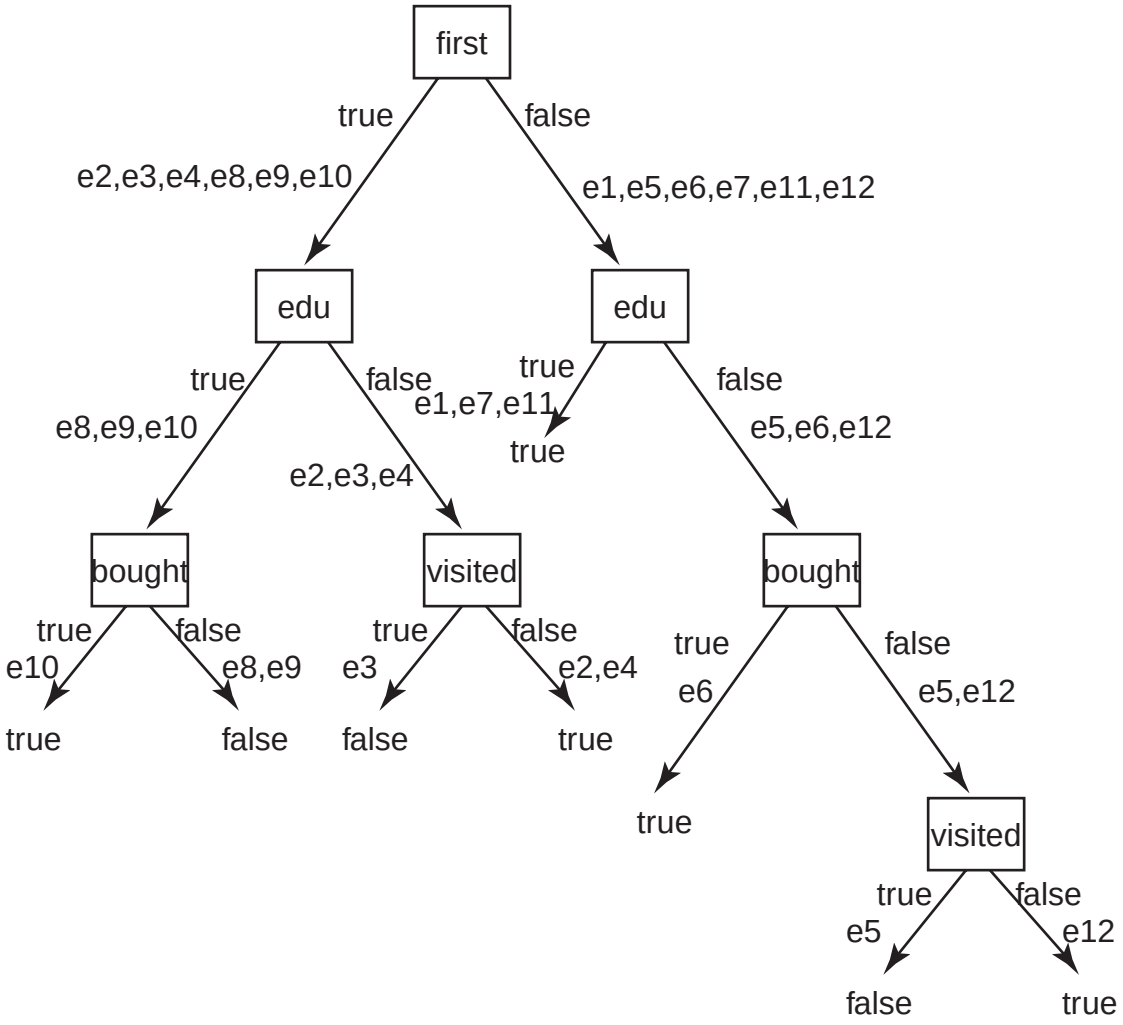

Figure 2: Classification of *more_info* for a complete tree.

The bias is that when *first* is true, *visited* is irrelevant when *edu* is true and *bought* is irrelevant when *edu* is false. When *first* is false, *more_info* is always false except for the example corresponding to $e5$.

(e) How can overfitting occur in the learned network? Explain in terms of this example.

There are two different thing that you could notice from this dataset, in particular from noticing that there is only one example that is false when *first* is false. This single example results in a complex tree for $first = false$.

The first is that on which attribute to split on is for $first = false$ is completely arbitrary. Depending on the arbitrary choice, the instance:

bought	*edu*	*first*	*visited*
false	true	false	true

will be classified differently. We can (over)fit the data to models where this instance is true and to models where this instance is false.

The second that this example $e5$ could be noisy. It may be a better model to say that the user asks for *more_info* when $first = false$, but that there is one noisy example rather than there being a complex theory about the world.