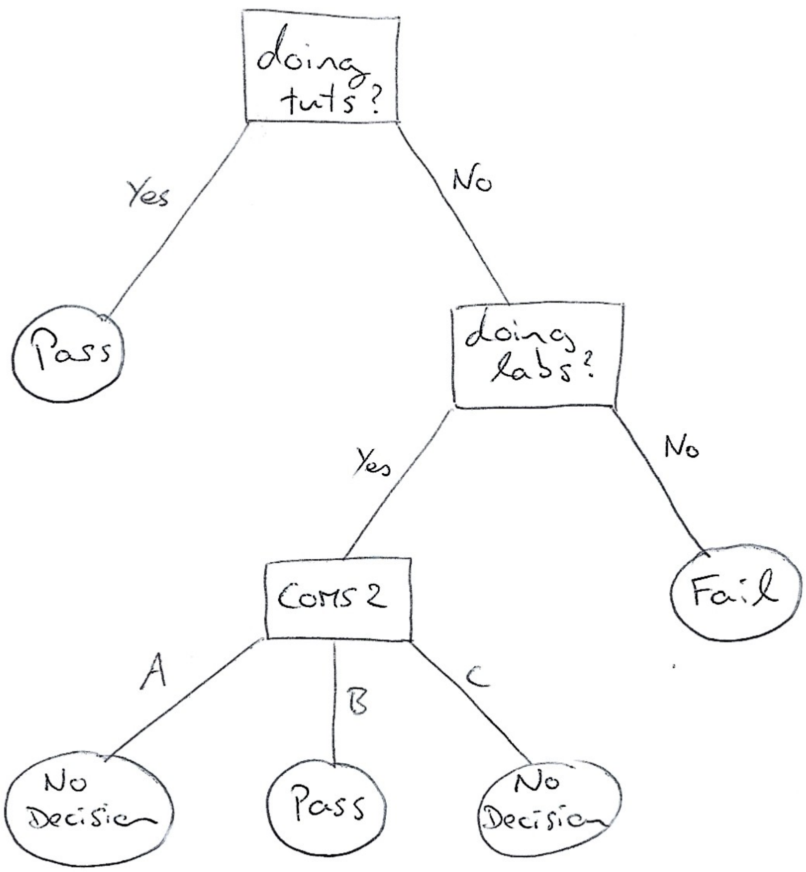
COMS 4030A

Adaptive Computation and Machine Learning

EXERCISES on page 60

(1) Complete the decision tree in the above example.

Solution:

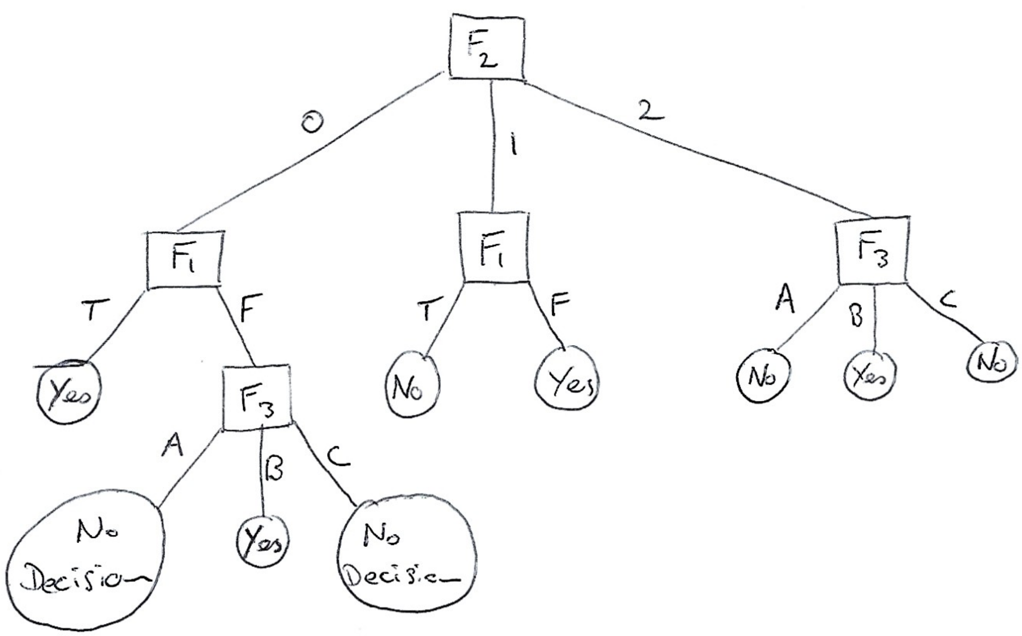


(2) Construct the decision tree for the following dataset using the entropy method described above:

$S:$

F_1	F_2	F_3	target
T	0	B	Yes
F	2	A	No
F	1	C	Yes
T	2	B	Yes
F	0	A	No
F	2	C	No
F	0	A	Yes
T	1	B	No
T	0	B	Yes
F	1	C	Yes
T	2	A	No
T	2	C	No

Solution:



EXERCISES on page 66

(1) Construct the decision tree for the following dataset using

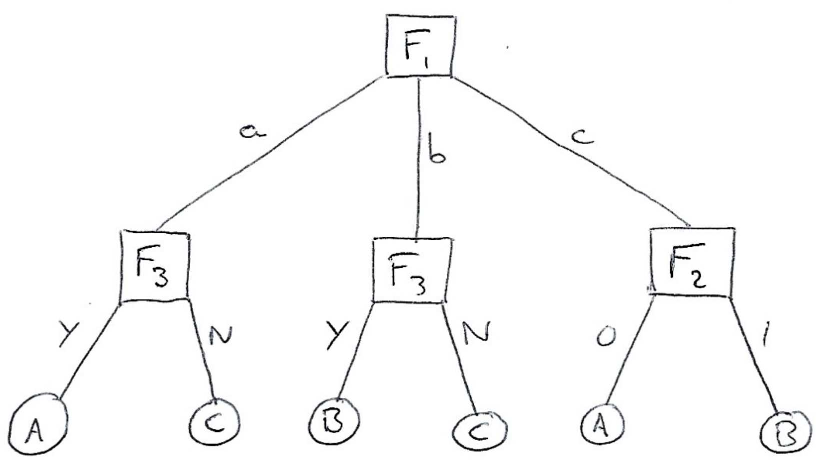
(a) the Entropy method and

(b) the Gini Impurity method.

$$S : \left[\begin{array}{c|c|c|c} F_1 & F_2 & F_3 & \text{target} \\ \hline a & 0 & Y & A \\ b & 0 & N & C \\ c & 1 & Y & B \\ b & 1 & Y & B \\ c & 0 & N & A \\ a & 0 & N & C \\ c & 1 & N & B \\ a & 1 & Y & A \\ b & 0 & Y & B \\ a & 1 & N & C \end{array}\right]$$

Solutions:

(a)



(b) is the same as (a)

(c) Find the predictions from the trees obtained in (a) and (b) for the following inputs:

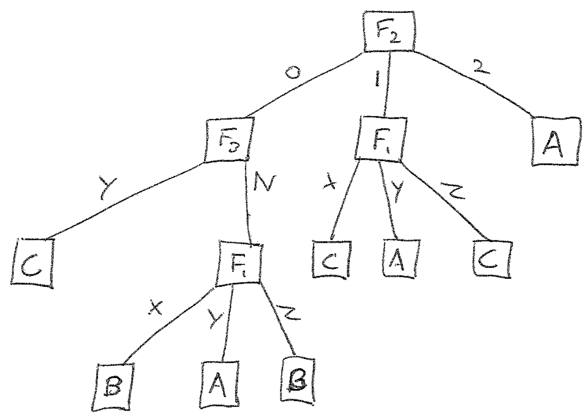
(i) $(b, 1, N)$

(ii) $(c, 1, Y)$.

Solutions:

(c) (i) C (ii) B

(2) Suppose that the decision tree below has been constructed using some training dataset.



Let the following dataset be the validation set.

Validation set :

F_1	F_2	F_3	target
X	2	Y	A
Y	0	N	B
Z	1	Y	C
Z	1	N	B
Y	2	N	A
Z	0	Y	C
X	2	N	A
Y	1	Y	A
Z	2	Y	C
X	0	N	C
Y	0	Y	B
X	0	N	B

(a) Compute the validation set error for the above decision tree.

Solution:

$\text{Error}_V = \frac{5}{12}$

(b) For each (non-leaf) node in the tree, prune the node and see if the validation set error decreases on the pruned tree.

Solutions:

If the lower branch with F_1 is pruned and replaced with a B leaf, then $\text{Error}_V = \frac{4}{12}$.

If the branch with F_3 is pruned and replaced with a B leaf, then $\text{Error}_V = \frac{3}{12}$.

If the higher branch with F_1 is pruned and replaced with a C leaf, then $\text{Error}_V = \frac{6}{12}$.

(3) Suppose a decision tree has been constructed from some dataset. Then it can be used to make predictions for new data points that are not part of the original dataset. Suppose a data point has one of its attribute values missing. Describe how a prediction could still be made in this case. For example, for the tree in Question 2, suppose the input is: $[Z, ?, Y]$. What prediction could be obtained from the tree?

Solution:

All possible branchings for the unknown value can be considered and the most common prediction amongst them can be taken as the prediction.

In the tree in Question 2, the value of F_2 is unknown, but for each of the possible values (i.e., 0, 1, 2) a prediction at a leaf can be reached. For the input $[Z, ?, Y]$, if $F_2 = 0$ the prediction is C, if $F_2 = 1$ it is C, and if $F_2 = 2$ it is A. Then one could choose C as the most common prediction.

(4) A regression tree is to be constructed for the dataset below using MSE on the targets. The attributes F_1 and F_2 can take any real value between 0 and 1.

Determine what question should be asked at the root node of the tree.

When choosing a split-point value for attributes F_1 and F_2, just consider the values 0.25, 0.5 and 0.75.

$$S = \left[\begin{array}{c|c|c|c} F_1 & F_2 & F_3 & \text{target} \\ \hline 0.8 & 0.2 & a & 2.7 \\ 0.7 & 0.1 & b & 2.3 \\ 0.2 & 0.9 & c & 1.7 \\ 0.3 & 0.9 & b & 2.1 \\ 0.3 & 0.6 & a & 2.0 \\ 0.1 & 0.4 & c & 1.3 \\ 0.4 & 0.8 & b & 2.9 \\ 0.1 & 0.3 & c & 1.4 \end{array}\right]$$

Solution:

The largest Gain comes from using F_3 at the root node.

$\text{Gain}(S, F_3) = 0.2056$

For F_1 with split-point 0.25, $\text{Gain}(S, F_1) = 0.2045$

For F_1 with split-point 0.5, $\text{Gain}(S, F_1) = 0.0653$

For F_1 with split-point 0.75, $\text{Gain}(S, F_1) = 0.0607$

For F_2 with split-point 0.25, $\text{Gain}(S, F_1) = 0.0653$

For F_2 with split-point 0.5, $\text{Gain}(S, F_1) = 0.0159$

For F_2 with split-point 0.75, $\text{Gain}(S, F_1) = 0.0205$