Hunt's Algorithm

- In the Hunt's algorithm, a decision tree is grown in a recursive fashion by partitioning the training records successively into purer subsets
- Let Dt be the set of training records that are associated with node t and y = {y1, y2, · · · , yc} be the class labels. The following is a recursive definition of Hunt's algorithm.
- **Step 1**: If all the records in Dt belong to the same class yt, then t is a leaf node labeled as yt.
- **Step 2**: If Dt contains records that belong to more than one class, an attribute test condition is used to partition the records into smaller subsets. A child node is then created for each outcome of the test condition. The records in Dt are distributed to the children based upon their outcomes. This procedure is repeated for each child node.

Dt = {training records @ node t}

- If Dt = {records from different classes}

– Split Dt into smaller subsets via attribute test

– Traverse each subset with same rules

- If Dt = {records from single class yt}

– Set Node t = leaf node with class label yt

- If Dt = {} (empty)

– Set Node t = leaf node with default class label yd

- Recursively apply above criterion until ...

– No more training records left

Example

- Consider the problem of predicting whether a loan applicant will succeed in repaying (pay back) her loan on-time or become unsuccess, and delaying, default on her loan.
- The training set used for predicting borrowers who will default on their loan payments will be as follows.

binary categorical continuous class

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	**No**
2	No	Married	100K	**No**
3	No	Single	70K	**No**
4	Yes	Married	120K	**No**
5	No	Divorced	95K	**Yes**
6	No	Married	60K	**No**
7	Yes	Divorced	220K	**No**
8	No	Single	85K	**Yes**
9	No	Married	75K	**No**
10	No	Single	90K	**Yes**

Figure1

Hunt's Algorithm

- A training set for this problem can be constructed by examining the historical records of previous loan borrowers.
- In the training set shown in **Figure 1**, each record contains the personal information of a borrower along with a class label indicating whether the borrower has defaulted on her loan payments.
- The initial tree for the classification problem contains a single node with class label Defaulted = No (This means that most of the borrowers had successfully repaid their loans.) as illustrated below:

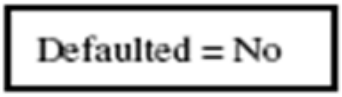


Figure 1a: Step 1

- However, the tree needs to be refined since the root node contains records from both classes.
- The records are subsequently divided into smaller subsets based on the outcomes of the Home Owner test condition, as shown in Figure below:

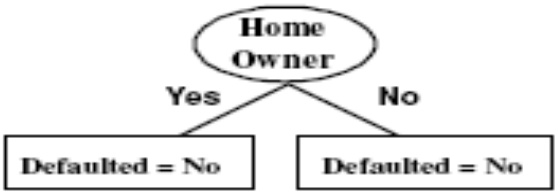


Figure 1b: Step 2

- Now we can assume that this is the best criterion for splitting the data at this point.
- The Hunt's algorithm is then applied recursively to each child of the root node.
- From the training set given in **Figure 1**, notice that all borrowers who are home owners had successfully repaid their loan. As a result, the left child of the root is a leaf node labeled as **Defaulted = No** as shown in **figure 1b**
- For the right child of the root node, we need to continue applying the recursive step of Hunt's algorithm until all the records belong to the same class. I.e. in training set Figure 1; **<Home Owner=No AND Marital Status=Married> is always No**
- This recursive step is shown in **Figures 1c** and **d** below:

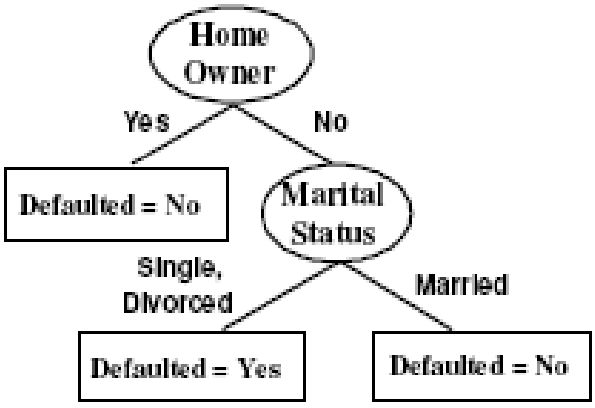


Figure1c: Step 3

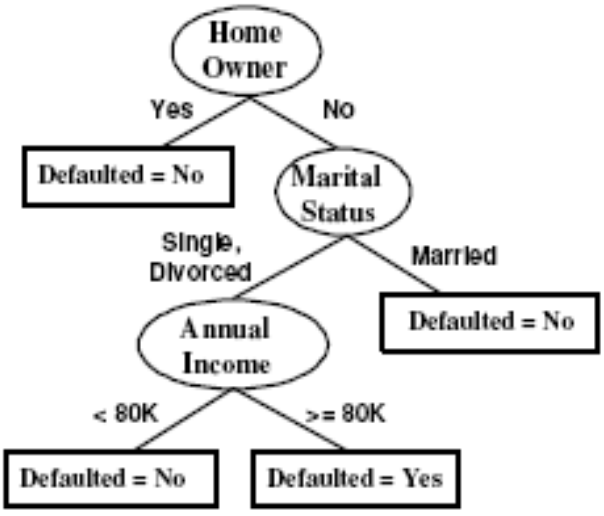
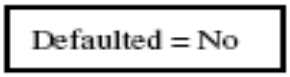


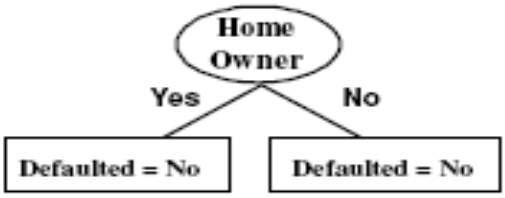
Figure 1d: step 4

- Generally, the whole diagram will be as follows

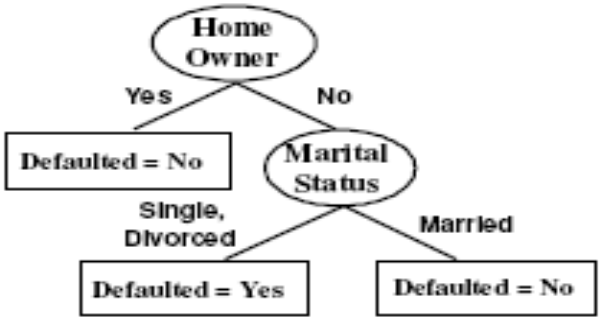
Hunt's Algorithm



(a) Step 1



(b) Step 2



(c) Step 3

(d) Step 4