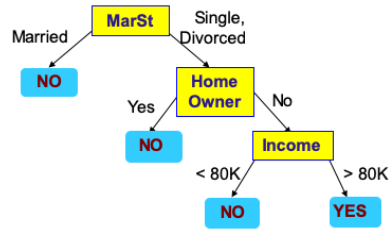


Classification and Decision Trees Activity

1. Tree Structures for Classification



A tree structure can be created to perform classification. For example, the tree on the left is used to predict whether someone who borrows money from a bank will default on the loan.

Use the tree to classify a arrived examples with the following features:

Homeowner	Marital Status	Income	Loan Will Default?
no	married	80K	??
no	divorced	93K	??

Trace the pathway through the tree for these examples. What labels (yes or no) does the tree assign?

2. Decision Tree Construction

The following is a subset of the Titanic Survivors dataset available from Kaggle <https://www.kaggle.com/c/titanic/>. Draw a decision tree that correctly classifies each item.

Pclass	Sex	Age	Survived
3	male	14	No
2	female	25	Yes
2	male	35	No
3	female	24	Yes
3	female	48	No
2	female	45	Yes
2	female	22	Yes
1	male	62	No
2	male	3	Yes
3	male	9	No
2	male	2	Yes
2	female	36	Yes

Now use your tree to classify the items in this test set:

Pclass	Sex	Age	Survived
1	female	17	
3	female	41	
3	male	17	
3	male	1	

3. Tree Reflection

In this case, it was straightforward to develop a decision tree that could perfectly classify the training set by “eyeballing” the data. Consider the case where we have thousands of elements in our training set, each with dozens of attributes.

- Given that there are potentially many possible trees that fit the training data, how might you go about deciding between them? What constitutes a “good” tree?
- How could you automate the process of tree construction?