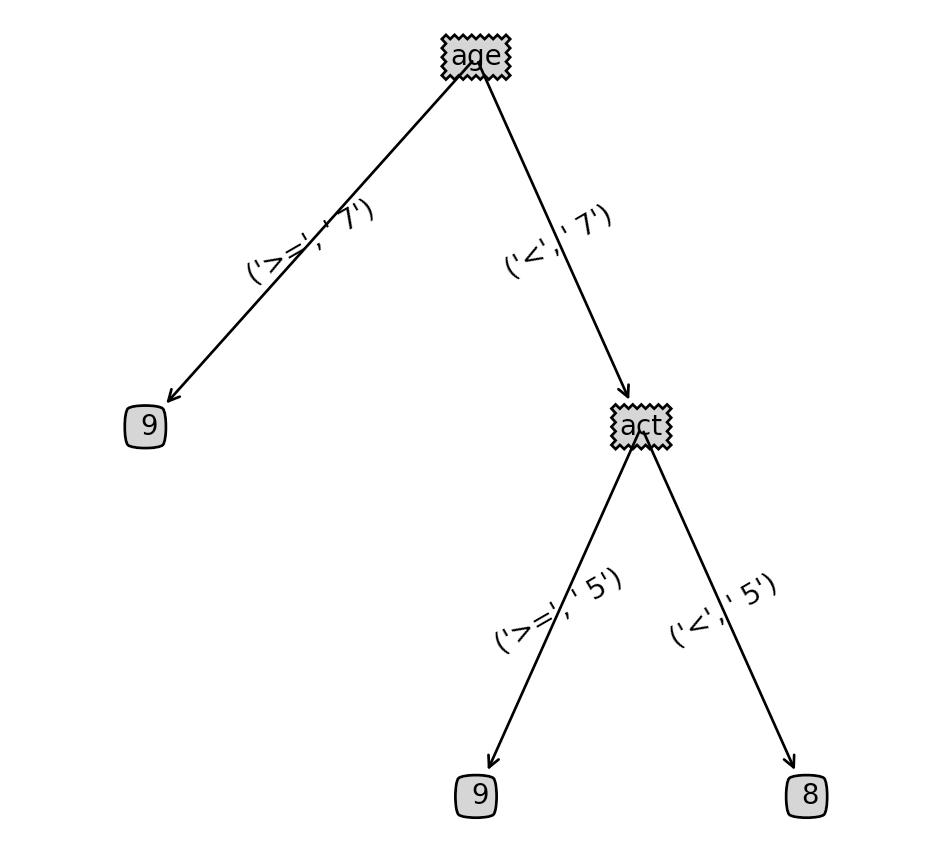
**Program3 Report**

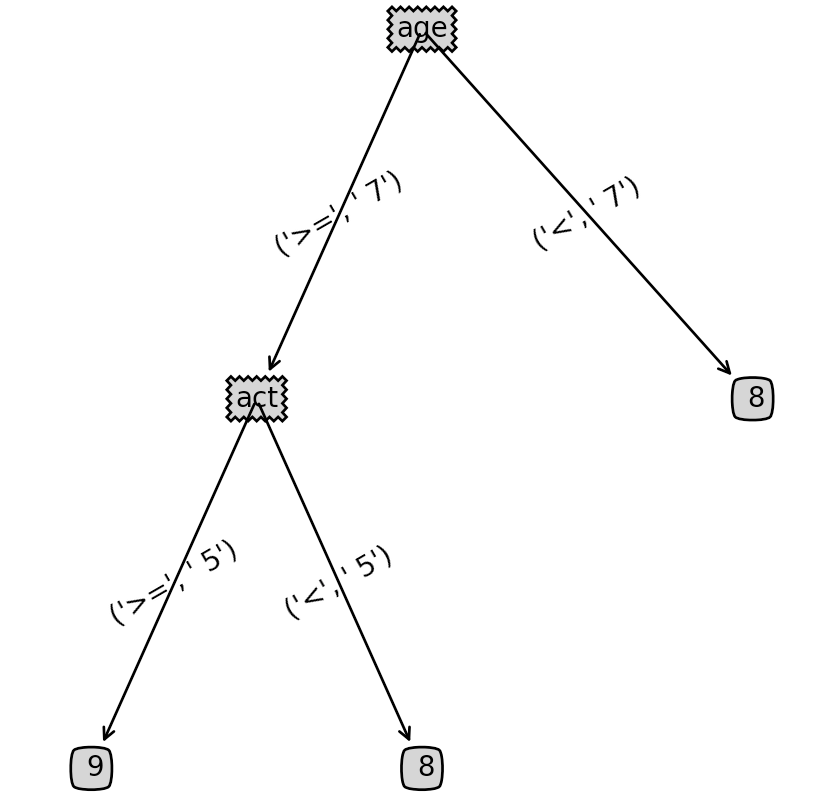
Longqian Zheng

**General Questions:**

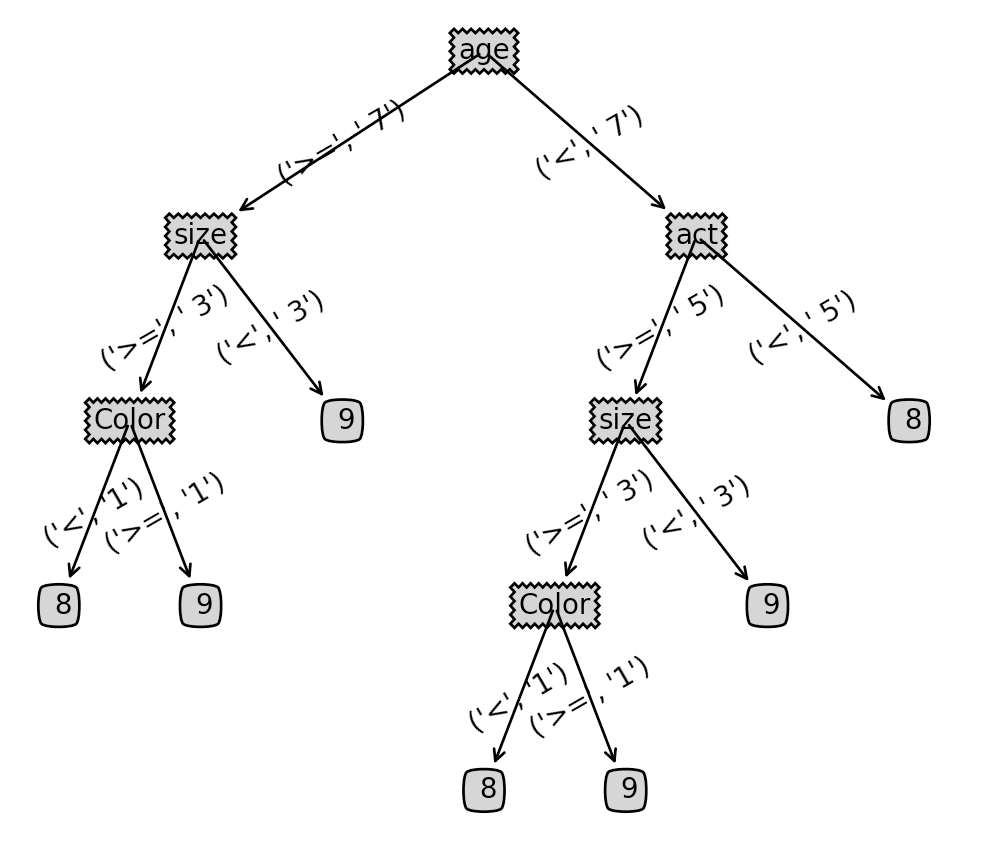
1. No. If there have missing attributes, it will make some leaf node do not have a pure class value, then it will not memorize the training data.
2. It depends on what attributes has been choose. For stocks or horses, there are too many factors need to consider, it is easy to have some hidden attributes people do not know if it can affect the result. Then the decision tree will not be accuracy on making these kinds of decisions. But if the attributes are very good, it will help to make decision.

**Balloon:**

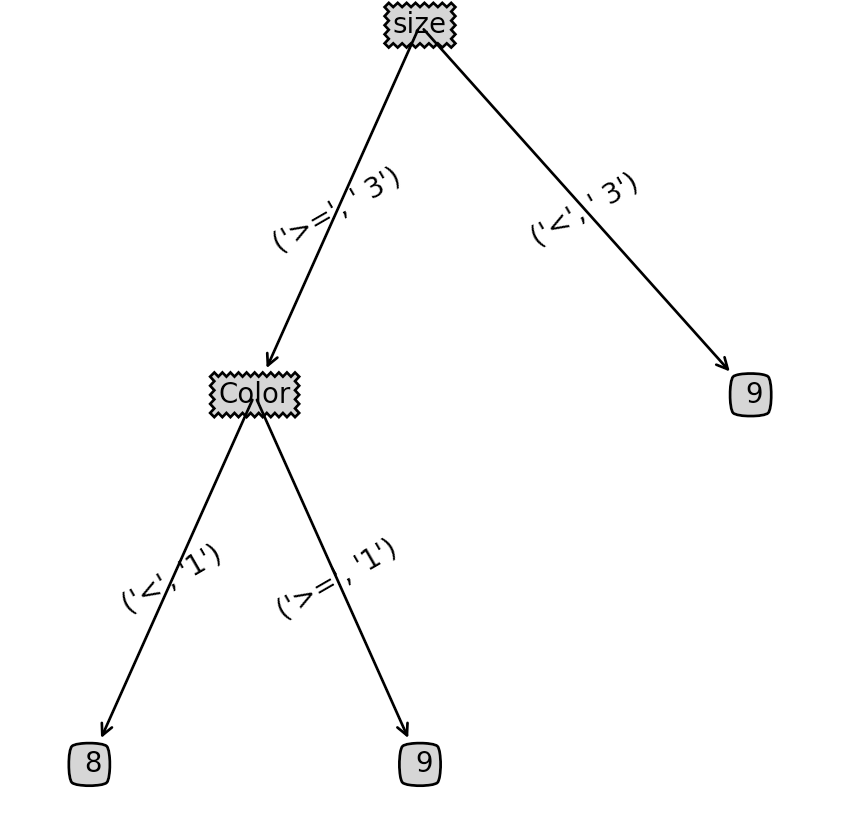
1. The 4 balloon examples:
   1. adult+stretch.data: 
   2. adult+stretch.data:



* 1. yellow-small+adult-stretch.data

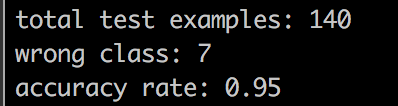


* 1. yellow-small.data

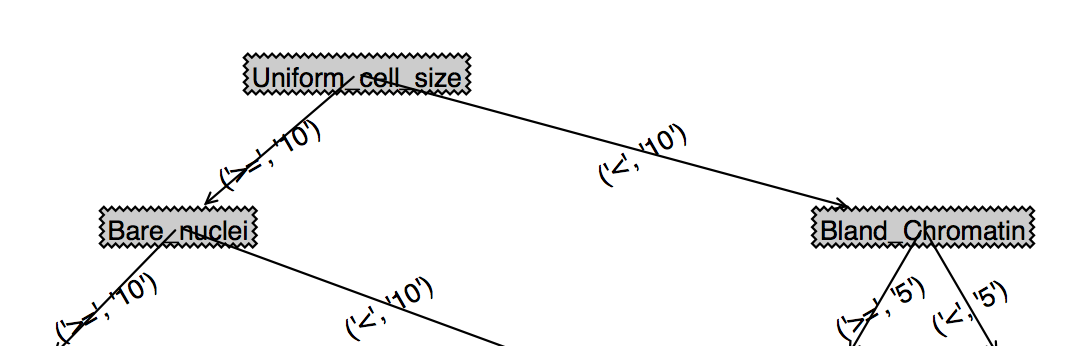


**WDBC**

1. The majority class accuracy for test set: 65.5%
2. My accuracy: 95%, much better than just choose majority class



1. The top 2 level:



**Pen**

1. The graph:
2. Because Pen2 and Pen3 miss some attributes. It depends on what feature it loses. Generally, missing attributes will cause some leaf node not belongs to one class value, so the value will be decided by the major class value of parent node, it will reduce the accuracy. But if remain features are enough to generate a full tree with all leaf node have one class value, it will not affect the accuracy.
3. Pen3. Because the training data in pen3 is better than pen2. The distribution for each value of attributes in pen3 covers more possible values it may have. In pen2, the values trend to be concentrate in some smaller range.
4. There must missing some/one important attribute in the dataset, so some of the leaf node do not have the same class value and using the major class value of parent node.