1. THINKING ABOUT DECISION TREES

PROBLEM 1. Recall the shapes decision tree example in the slides.

- **a:** Write down the decision tree for the shapes data in Fig. 1 if the root node was shape instead of color.
- **b:** Will the two trees make the same predictions for unseen shapes/color combinations?
- **c:** Show that multiple structurally different decision trees can represent the same Boolean function of two or more variables (think about what it means for two trees to be structurally different).

PROBLEM 2. Suppose an example is missing the value of an attribute. What can we do at training time? Or, what are different methods to "complete the example"?

PROBLEM 3. Consider non-boolean features, that is, what do we do if features can take mulitple values?

2. Implementing a Decision Tree

PROBLEM 4. Implement a decision tree for data from a previous homework using information gain to evaluate splits. Think about how costly it is to build the decision tree, and to use it to classify examples. Think about how you might want to divide your data for training vs testing. We will re-visit your decision tree program when we talk about cross-validation.

3. Feedback on slides

Let me know what you found easy and hard to understand on the slides. Any comments or criticisims are appreciated.

