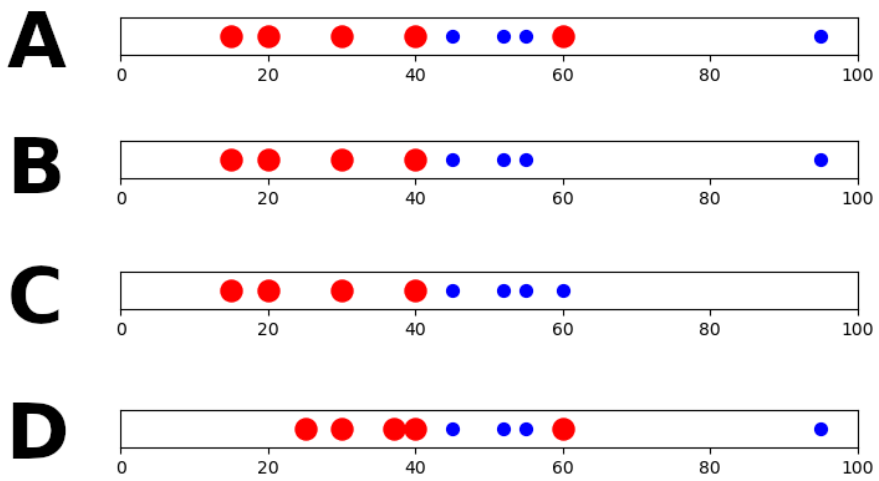
CS5228 – Tutorial 6

Classification & Regression II (Tree-Based Models)

1. **Effects of data distribution on Decision Trees.** The figure below shows 4 distributions of the values for a single feature. The color and shape of the dots reflect the class label. Since we only have two colors/sizes, the example application is a binary classification task.



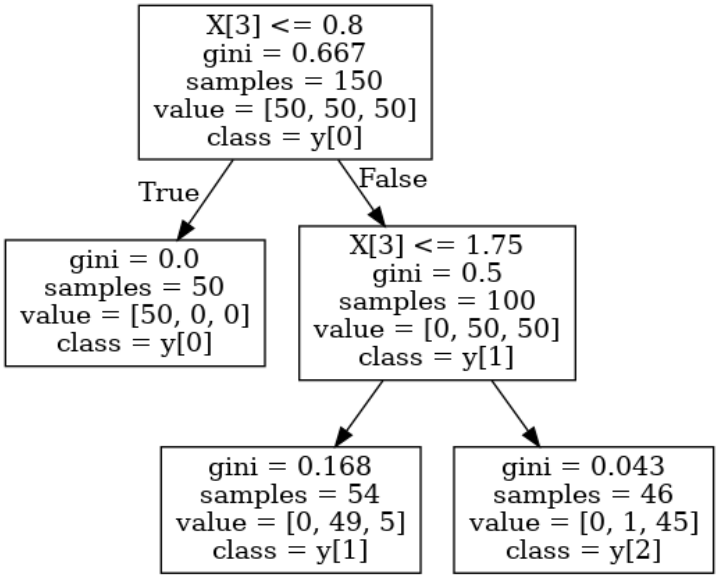
   (a) Assume a Decision Tree classifier that only performs binary splits. For each data distribution A-D, where (approximately) would the classifier split the values into 2 child nodes?

   (b) Let's assume our data points only have this one feature. Just by looking at the data distributions A-D for this single feature, what can we say about the "look" of the final decision tree (without any pre or post-pruning)?

   (c) Given the results from (a) and (b), summarize how the distribution of feature values affects the training of a Decision Tree.

   (d) You can now add a single data point of Class **Blue** into Distribution A? Where would you place this new data point to maximize the negative effect on the resulting Decision Tree in terms of the required splits?

2. **Interpreting Decision Trees.** The Decision Tree shown below has been trained over the IRIS Dataset with a maximum depth of 2. Recall that each data sample has 4 numerical features (all measures in centimeters), and is labeled with 1 out of 3 classes.



(a) What insights can you get from this Decision Tree?

(b) Assume you train the Decision Tree without any restrictions on its maximum depth. Given the Decision Tree above, which statement can you make about the full Decision Tree.

(c) How would your answer for (a) and (b) change if all the input features would have been standardized before training the Decision Tree.

(d) Assume someone gives you the optimal Decision Tree, i.e., optimal in the sense that it results in the highest accuracy (or any other suitable metric). What can we say about the root node of this optimal Decision Tree?