Lab 4G: Growing trees

Directions: Record your responses to the lab questions in the spaces provided.

Trees vs. Lines

Our first tree

- **Why can't we just use a *linear model* to predict whether a passenger on the Titanic `survived` or not based on their `gender`?**

Viewing trees

- **Write down the labels of the two *branches*.**

- **Write down the labels of the two *leaves*.**

- **Which `gender` does the model predict will survive?**

- **Where does the plot tell you the number of people that get sorted into each leaf? How do you know?**

- **Where does the plot tell you the number of people that have been sorted *incorrectly* in each leaf?**

Leafier trees

- **Mrs. Cumings was a 38 year old female with a 1st class ticket from Cherbourg. Does the model predict that she survived?**

- **Which variable ended up not being used by `tree`?**

Tree complexity

- **How is `tree3` different from `tree2`?**

Misclassification rate

Predictions and Cross-validation

On your own

- **In your own words, explain what the *misclassification rate* is and how to calculate it.**

- **Which model (`tree1`, `tree2` or `tree3`) had the lowest misclassification rate for the `titanic_test` data?**

- **Does creating a more complex *classification tree* always lead to better predictions? Why not?**