Yanhao Quan – Assignment 2 – Titanic

**Data**

This data set has 891 rows and 12 columns. There are 12 features which are survival dummy, ticket class, sex of the passengers, age of passengers and so on. But in this data set, not every feature has a strong relationship with survival, so I picked out some of the data from the 12 features, which is ticket class, sex of passengers, age of passengers, number of siblings/spouses aboard the Titanic, number of parents and children aboard the Titanic, and passenger fare since I assume the fare determines the purchase crowd and the higher is the fare the saver the place is. And, we have two data set, so I use one of the data sets as training data set to make our model learn and another data set as test data to exam if the result is robust.

**Method**

We use the classification tree which estimates the probability of death in the Titanic data and provides an overview of which features were the strongest predictors of death. Since there are many NA cells in the data set I fill empty numerical values with median, and transform sex into sex dummy which male = 1 and female = 0.

**Results**

After I train and test the data, we got our classification tree graph and in the tree graph we can see that the passenger with higher ticket class is more likely to survive and the passengers with lower fare are more likely to die. It does make sense because higher ticket class pair with higher fare. From the classification tree I also find that younger people are more likely to die then elder people, and the passengers with no siblings, spouses, parents, and children are more likely to survive. If we use intuition to explain this, it does make sense too, since people tend to let elder people or the ones they love to evacuate first.

However, if we look at the part of the tree below, sex is the strongest predictor of death/survival, in the graph, X [1] <= 0.5 means sex dummy equals 0 which is female, and if the passenger is female then she has higher probability to survive than male passenger, and we could see this information in this graph too.

墙上贴着海报和广告

描述已自动生成

In conclusion, lower class ticket passengers, male passengers, young passengers, passengers with families, and passengers with lower fare have larger probability to die if there is shipwreck.

Note: For ticket class, the lower the number the higher the class.