Project 2 - Regression and CART

Group 3

3/13/2021

Group Leader: Brianna Johnson

Member Names: Marco Sousa, Ben Pfeffer, Nikita Seleznev, Brianna Johnson

Introduction to the Heart Dataset

Observing the data as its initial dataframe.

```
##   Age Sex    ChestPain RestBP Chol Fbs RestECG MaxHR ExAng Oldpeak Slope Ca
## 1  63   1      typical    145  233   1       2   150     0     2.3     3  0
## 2  67   1 asymptomatic    160  286   0       2   108     1     1.5     2  3
## 3  67   1 asymptomatic    120  229   0       2   129     1     2.6     2  2
## 4  37   1   nonanginal    130  250   0       0   187     0     3.5     3  0
## 5  41   0   nontypical    130  204   0       2   172     0     1.4     1  0
## 6  56   1   nontypical    120  236   0       0   178     0     0.8     1  0
##         Thal AHD AHDBinary
## 1      fixed  No         0
## 2     normal Yes         1
## 3 reversable Yes         1
## 4     normal  No         0
## 5     normal  No         0
## 6     normal  No         0
```

Describing the data, and (1) identify Y and X.

I imagine the feature we are interested in is AHD, which is either "Yes" or "No" in regards to having heart disease. X are the other possible features which may predict Y.

Exploratory Data Analysis

EDA

Logistic Model

Logistic Model

Perhaps construct logistic model.

Confusion Table

Generate confusion table.

Mode-Test Statistics

Calculate classification accuracy and error, sensitivity, specificity, PPV and NPV. Will need to be compared elsewhere

ROC and AUC

Generate ROC and compute AUC for each model

S sigmoid curve

Generate s-curve for Y against one attribute (you can pick any one attribute), and interpret your findings

CART Model

CART MODEL

Other subsections

Comparing Models

Citations