Jason Melnik

09/16/2021

Homework 3

1. Does it appear from this scatterplot that these predictor variables will be helpful in determining whether patients have heart disease?
   1. Does it appear from this scatterplot that these predictor variables will be helpful in determining whether patients have heart disease?
      1. No since there’s “yes” and “no” dots all clustered together in every spot with no direct pattern in these two factors showing whether or not a person has a HD.
   2. Make another scatterplot of the patients’ ages (Age) and maximum heart rates (MaxHR), again coloring the points according to the outcome variable. Does it appear that these predictor variables will be helpful in determining whether patients have heart disease?
      1. This would help in predicting if a patient might have HD since there is a better pattern within this graph. This graph roughly shows that if you 50 and older with a maxHR of 140 and below your more at risk while if you are younger than 50 and a maxHR of 160 and below your at greater risk for HD.
   3. What do these boxplots tell you? Make similar side-by-side-boxplots for Age vs. HD and for MaxHR vs. HD. What do these boxplots tell you?
      1. This shows that if your RestBP is higher than 140 you are at a greater risk for HD. The boxplot for Age vs HD also shows that you are at a greater risk if you older than 60. MaxHR shows that you have a higher chance of HD if your MaxHR is below 150.
2. The table gave us a 67 percent accuracy in predicting whether a patient has a HD. The C5.0 decision tree scored a better accuracy than the one we did in class which scored a 51 percent.
   1. The first branch of the tree is whether “Old peak” is less than equal to 1.8. If it’s less than then the second branch is whether Slope is less than or equal to 1. If the first branch was greater than 1.8 then it determined Yes for HD.
   2. It would predict No.
   3. It incorrectly predicted 21 percent of patients not having HD even though they did.
   4. It incorrectly predicted 12 percent of patients having HD even though they did not have it.
   5. I would say a false negative would be more harmful when it comes to HD. If a doctor tells someone they don’t have a HD they will die much sooner than expected. But if a patient had a false positive the worse thing that can happen to them is that they increased their health to try and slow down HD.
   6. The new tree was successful in limiting false negatives. For the new tree the prediction accuracy is 57 percent, false positive rate is 8 percent, false negative rate is 35 percent. This tree has less prediction accuracy but a much better false positive rate.