Vincent Nguyen

February 16, 2014

Homework 3

TA Name: Jian Wu

Section/Lab Number: Monday 2:30

1)

Overall Error rate for a threshold of 0.5 is 0.209

False positive rate is 272/(272+2003) = 0.120

False negative rate is 356/(356+369) = 0.491

After changing marital status the overall error rate imroved. The false positive rate decreased while the false negative rate increased.

New overall error rate is 0.186

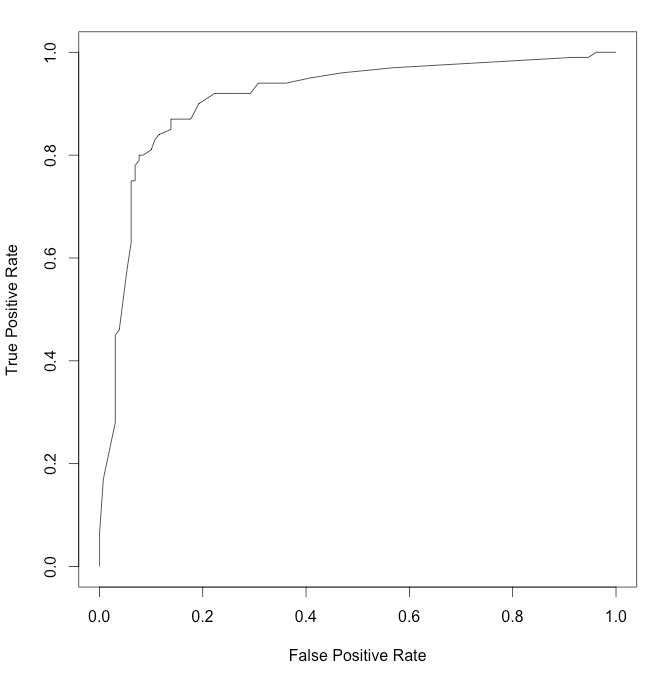
New false positive rate is 198/(198+2077) = 0.087

New false negative rate is 361/(361+364) = 0.497

3)

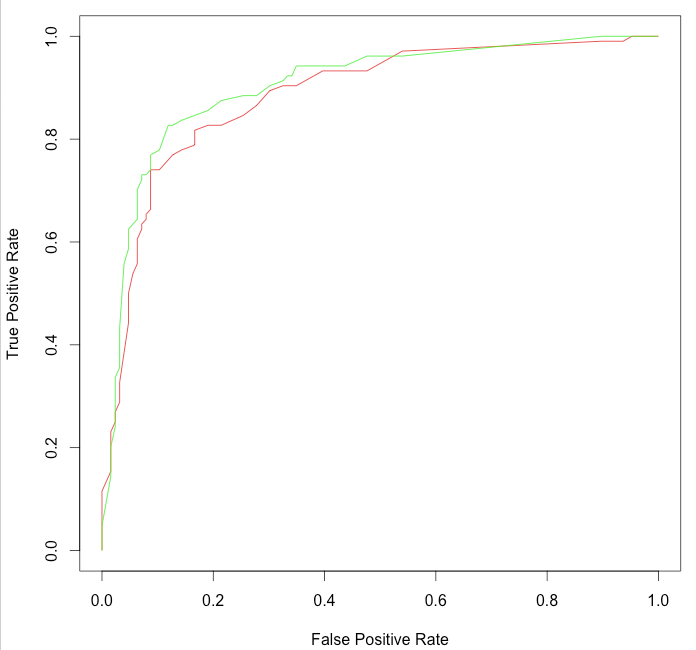
Changing the threshold generally increases the error rate for very high thresholds and very low thresholds. Which is to say that there seems to be a local minima at which errors are minimized (this occurred when the threshold value was between 0.1 and 3.4). This makes sense, because for very high thresholds, our model will tend to predict negative values (i.e will not be likely to predict positive) and produce false negatives instead of false positives while for very low threshold values (like 0.00001) our model will tend to predict positive and produce false positive errors instead of false negatives.

4)

ROC Curve 

5)

ROC Curves comparing models for homework/lab 3



ROC curve for the model that incorporates marital status and age as a single variable is shown in green while the red ROC curve is for the model that treates marital status and age as independent.

We can see that the green curve is universally better because for most of the same false positive rate values(except for aa small portion when the false positive rate is 0.58), the green curve has a higher True Positive Rate. In other words, the model that incorporates marital status and age as a single variable (as opposed to 2 independent ones) has a "higher" ROC curve