SSC 442 Lab 04

Team 7

3/31/2020

Exercise 1

Part 1.1:

fit_caps: 0.2166961 fit_selected: 0.1587043 fit_over: 0.14 fit_additive: 0.08684467

Part 1.2:

fit_caps: 0.2168794 fit_selected: 0.1589461 fit_over: 0.144 fit_additive: 0.07994084

The difference in results is trivial, thus our conclusion remains unchanged.

Part 2.1:

```
##          actual
## predicted nonspam spam
##   nonspam    2057  157
##   spam        127 1260

##
##   nonspam      spam
## 0.6064982 0.3935018
```

Part 2.2:

We determined that the fit_additive model is the best. We know that based on the rankings, overfit models will have lower training error and high testing error as opposed to underfit models which have both high testing and training error. Although having lower testing error would provide more reliable predictions when applied to unexposed data, having a low training error still helps increase testing accuracy. Sensitivity is the true positive rate, while specificity is the true negative rate. Here, sensitivity refers to nonspam ending up as spam, while specificity refers to spam ending up an nonspam. As a result, we decided that the fit_additive model would be the best option.

Exercise 2

Based on our model, age has a positive coefficient. This means that older individuals are more likely to have y = yes. Housing also has a positive coefficient. This means that individuals with housing or homeowners are more likely to have y = yes. Loan has a positive coefficient. This implies that individuals that took out loans are more likely to have y = yes. Finally, education has a positive coefficient. This suggests that individuals with some education level are more likely to have y = yes.

```
##          actual
## predicted   no  yes
##        no 2382   39
```