Chapter 8, Exercise 10 (p. 334)

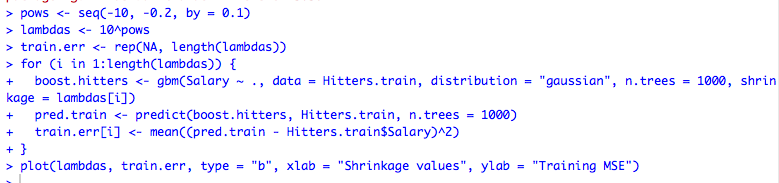
a)

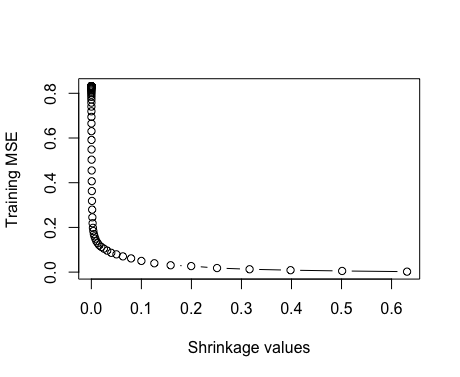
Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.37.15 PM.png

b)

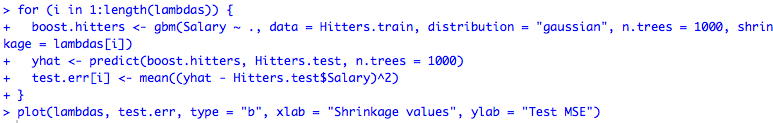
Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.37.28 PM.png

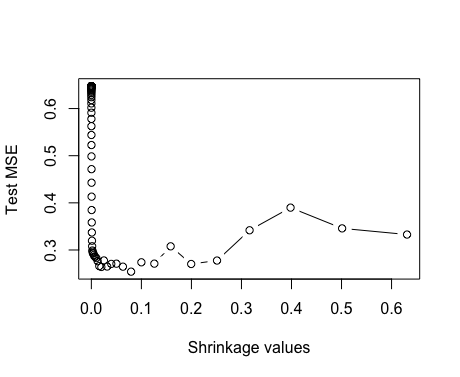
c) Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.44.11 PM.png

Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.44.11 PM.pngMacintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.44.11 PM.png



d)



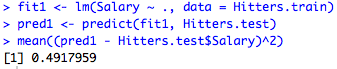


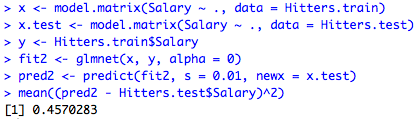
Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.48.20 PM.png

Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 10.48.32 PM.png

Test MSE for boosting is 0.25 and lambda = 0.08

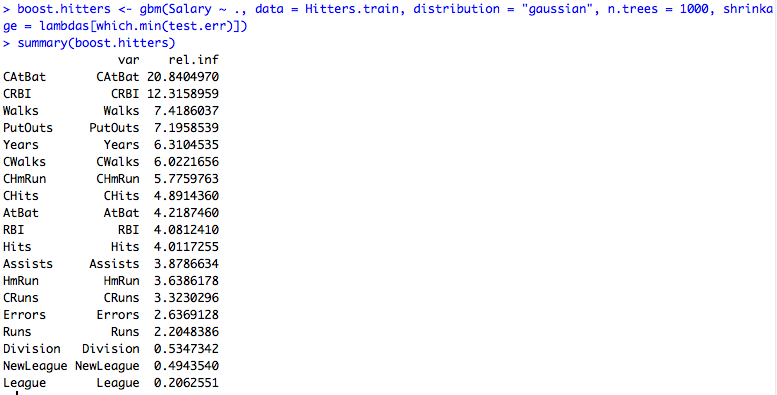
e)



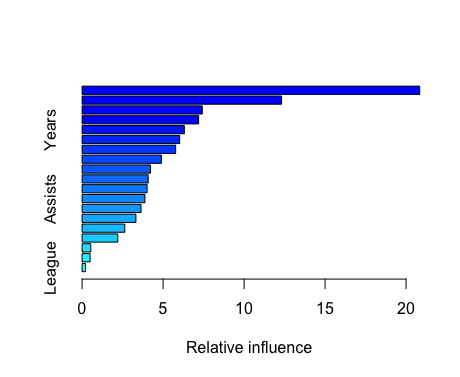
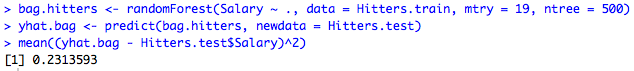


Test MSE for boosting (0.25) is lower than test MSE for linear and ridge regression (0.49, 0.45 respectively)

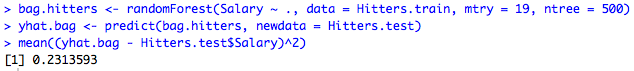
f)



Catbat is the most important by far (rel. inf is highest)

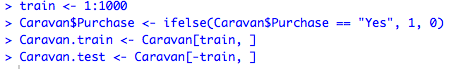


g)



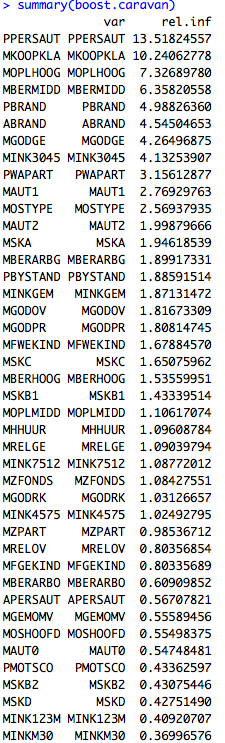
Test MSE is 0.23, which is lower than test MSE for boosting (0.25)

Chapter 8, Exercise 11 (p. 335)  
a)

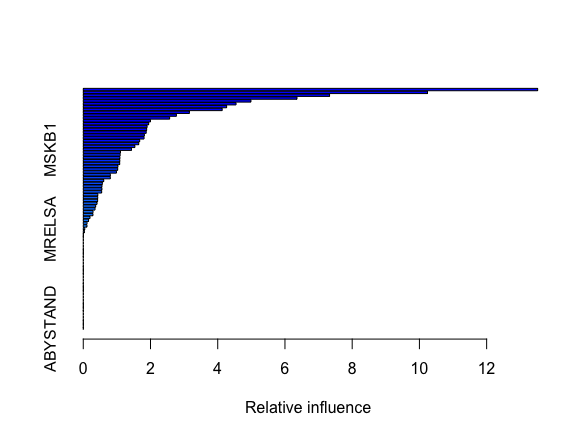


.b)

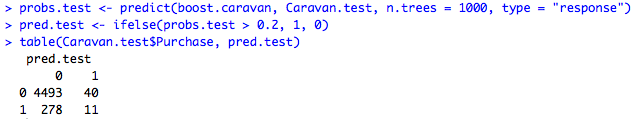
Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-11 at 11.16.43 PM.pngc)



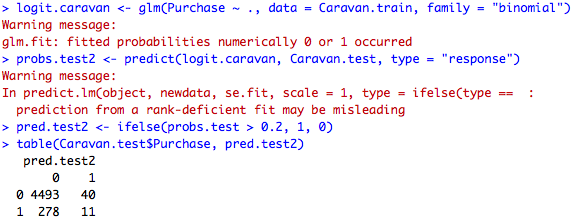
The top 3 most important variables are PPERSAUT, MKOOPKLA, and MOPLHOOG (due to rel. inf)



c)



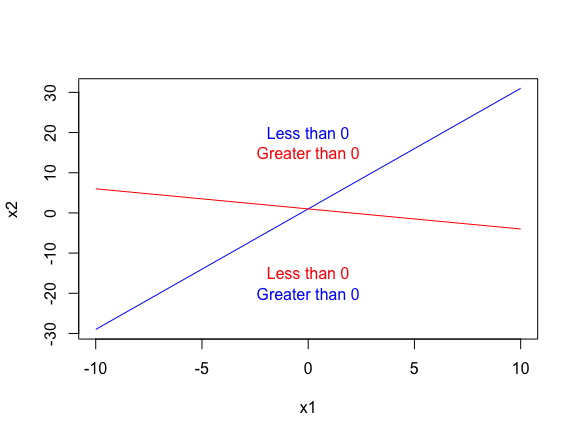
For boosting, the percentage of people buying is 11/(11+40) = 0.2156



For logistic regression, the percentage of people buying is 11/(11+40) = 0.2156

Chapter 9, Exercise 1 (p. 368)

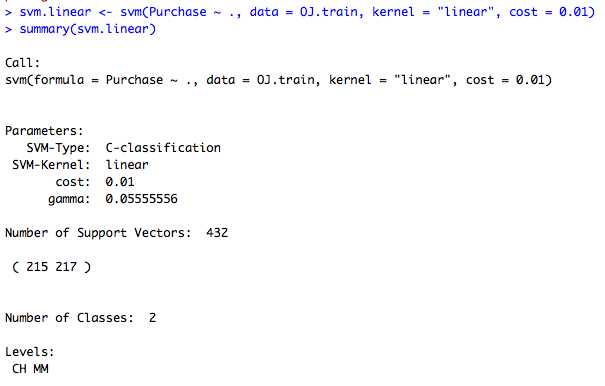
a)



Chapter 9, Exercise 8 (p. 371)

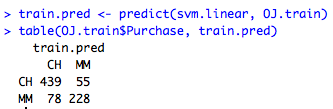
a)

Macintosh HD:Users:samlin:Desktop:Screen Shot 2017-08-12 at 4.33.45 PM.png

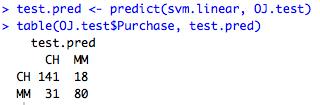
b) 

432 support vectors out of 800 observations/training points. 215 support vectors belong to CH and 217 belong to MM.

c)

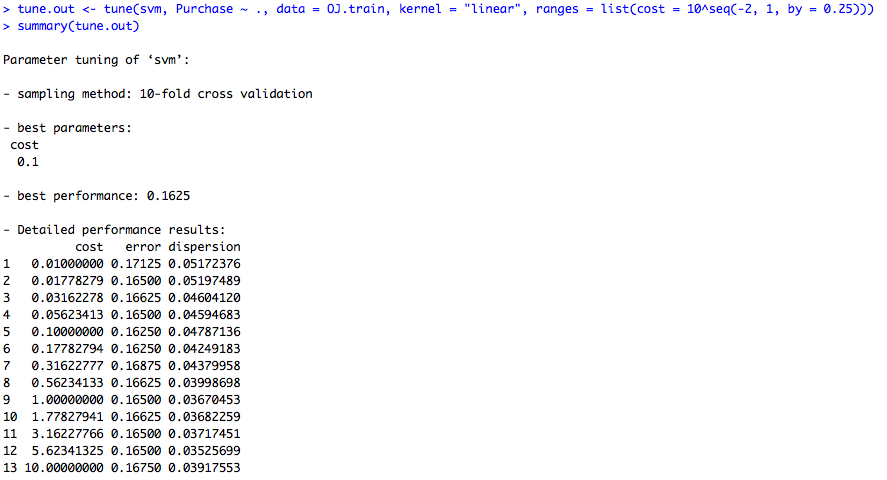


Training error rate is (78+55)/(439+78+55+228) = 133/800 = 0.16625

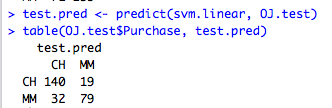


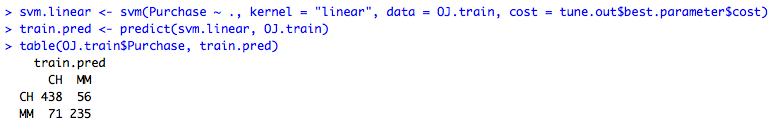
Test error rate is (31+18)/(141+18+31+80) = 49/270 = 0.18148

d) Best cost is 0.1



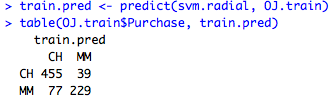
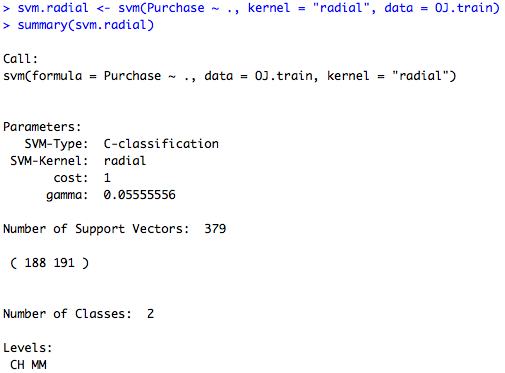
e)

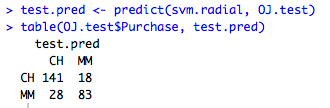


With best cost, the test error rate is (32+19)/(32+19+79+140) = 51/270 = 0.188889

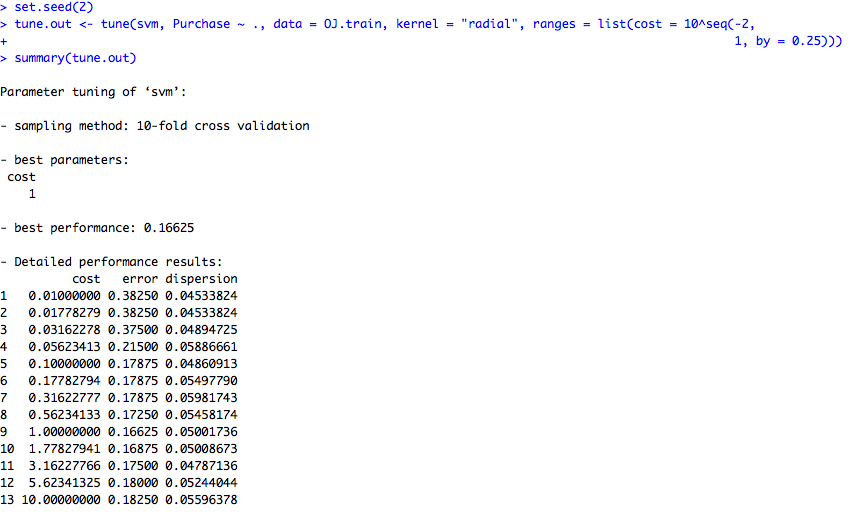
With best cost, training error rate is (71+56)/(71+56+235+438) = 127/800 = 0.15875

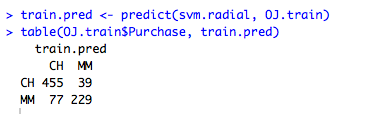
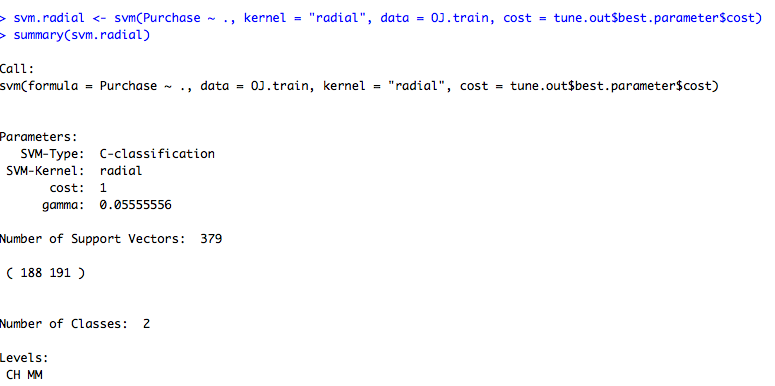
f)

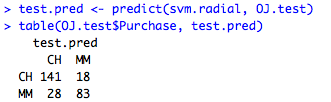




Created 379 support vectors out of the total observations with 188 belonging to CH and 191 belonging to MM. The classifier has training error of (77+39)/(455+39+77+229) = 116/800 = 0.171 and test error = (18+28)/(141+18+28+83) = (46/270) = 0.1704 which is better than linear with 15% and 18% training and testing error respectively.





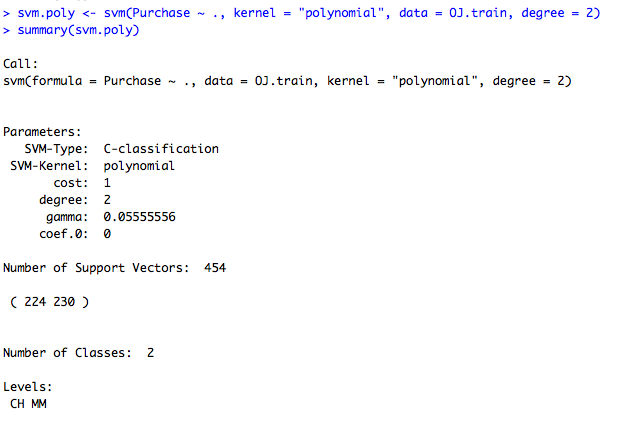


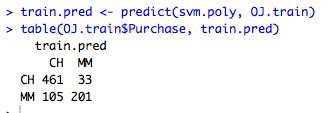
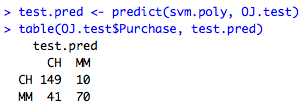
Train error: (77+39)/(455+39+77+229) = 116/800

Test error :(28+18)/(141+18+28+83) = 46/270

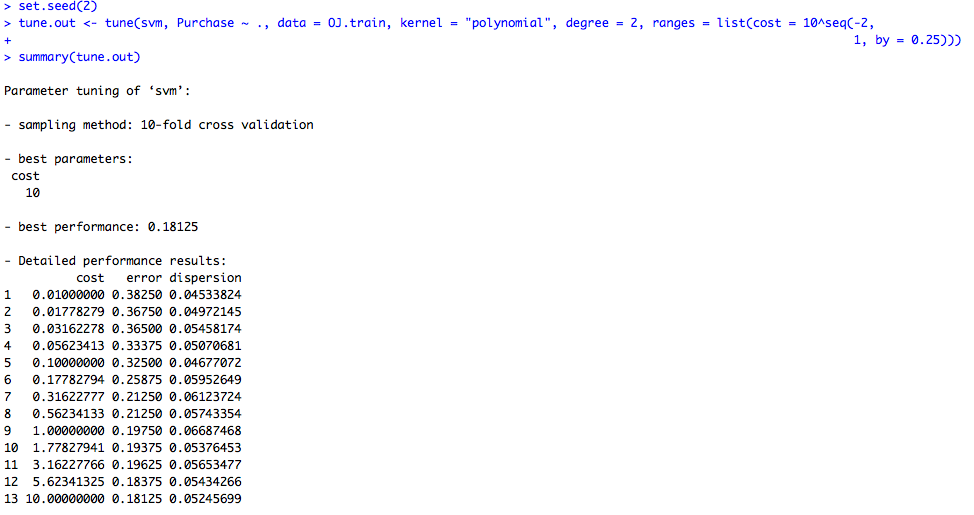
Both are same as above, so we can conclude tuning does not reduce the error rates.

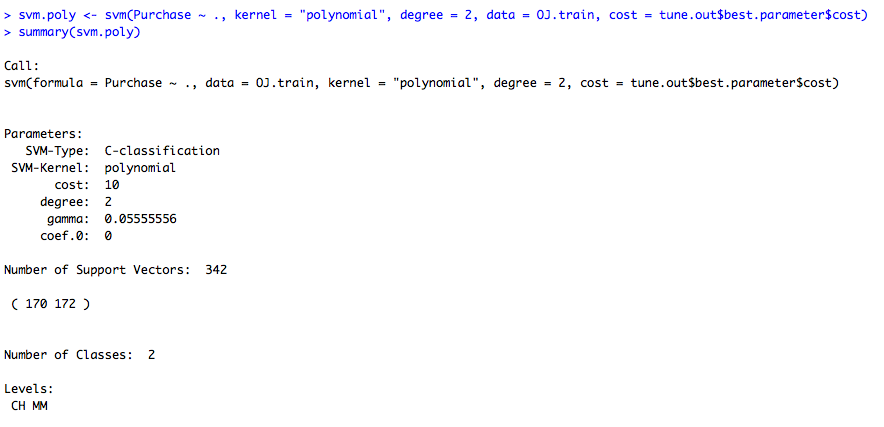
g)

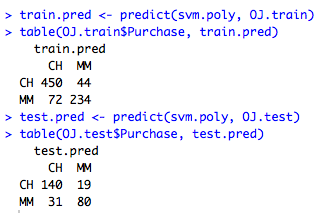




Polynomial kernel with default gamma creates 454 support vectors with 224 belonging to CH and 230 belonging to MM. The classifier has a test error of (41+10)/(149+10+70+41) = 51/270 = 0.1725 and training error of (105+33)/(461+33+105+201) = 138/800 = 0.1889. This is the same as the linear kernel.







Training error rate: (72+44)/(450+44+72+234) = 96/800 = 0.145

Test error rate: (31+19)/(140+80+31+19) = 50/270 = 0.1851

Conclusion: tuning reduces training and training error compared to 0.1725 and 0.18889 training and test error.

h)

The radial basis kernel produces the minimum classification error on test and training errors compared to others.