

Name: Alvee Chowdhury

Answer 1

All the code are in the R script file.

	model train RMS	model test RMSE	model number
Model 1	1.4483400	1.4247811	4
Model 2	1.1596770	1.1197802	7
Model 3	0.5070891	0.5269903	13
Model 4	0.5044712	0.5332113	18

We see that ‘Model 1’ is the least flexible model, and ‘Model 4’ is the most flexible model. Moreover, we see that the Train RMSE decrease as flexibility increases. The Test RMSE is smallest for ‘Model 3’, so it is the model we believe to perform the best on data not used to train the model. In general, ‘Model 1’ and ‘Model 2’ have high values of Train RMSE and high values of Test RMSE. Hence underfitting. Model 4 seem to have overfitting. To conclude, Model 3 is the best model out of these 4.

Answer 2

All code is in R script file.

Sumarising the models:

	model train RMS	model test RMSE	number of parameters
fit_smaller	4.675465	4.767746	13
fit	2.641685	2.813286	91
fit_larger	2.622012	2.833941	96

We see that ‘fit_smaller’ is the least flexible model, and ‘fit_larger’ is the most flexible model. Moreover, we see that the Train RMSE decrease as flexibility increases. The Test RMSE is smallest for ‘fit’, so it is the model we believe to perform the best on data not used to train the model. In general, ‘fit_smaller’ has high values of Train RMSE and Test RMSE. Hence underfitting. fit_larger seem to have overfitting. To conclude, fit is the best model out of these 3.

Answer 3

The table below summarizes the results which displays RMSE (both Train and Test), k, and whether or not that value of the tuning parameter appears to be overfitting, under-fitting, or the “best” value of the tuning parameter:

k	Train RMSE	Test RMSE	Fit?
5	1.65	2.16	Over
10	1.70	2.08	Over
15	1.79	2.05	Best
20	1.93	2.06	Under
25	2.02	2.14	Under
30	2.28	2.36	Under
35	2.60	2.67	Under
40	2.96	2.99	Under
45	3.27	3.29	Under
50	3.58	3.57	Under

Answer 4

The table below summarizes our findings:

k	test RMSE	Scaled?
1	0.6839884	No
1	0.7214409	Yes
5	0.5369142	No
5	0.5467248	Yes
25	0.5157672	No
25	0.5081259	Yes

Answer 5

Model	model test RMSE
Linear Model	3.155381
KNN Model	2.960250

For the KNN model I used k= 5 and scaled the x-data. Clearly from the test RMSE data we can see that the KNN performs better. It has a lower Bias-variance tradeoff so clearly performs better than the linear model.