Katie Hughes

Tim Dufala

Jasper LaFortune

Assignment 1

For this project, we implemented and evaluated a simple linear regression on a housing dataset. The goal was to learn a linear function to predict the mean housing value of an area in Boston given 13 relevant features for that area. We computed the weight vector both with and without a dummy variable in X. The weights and the SSE for each are shown in Table 1.

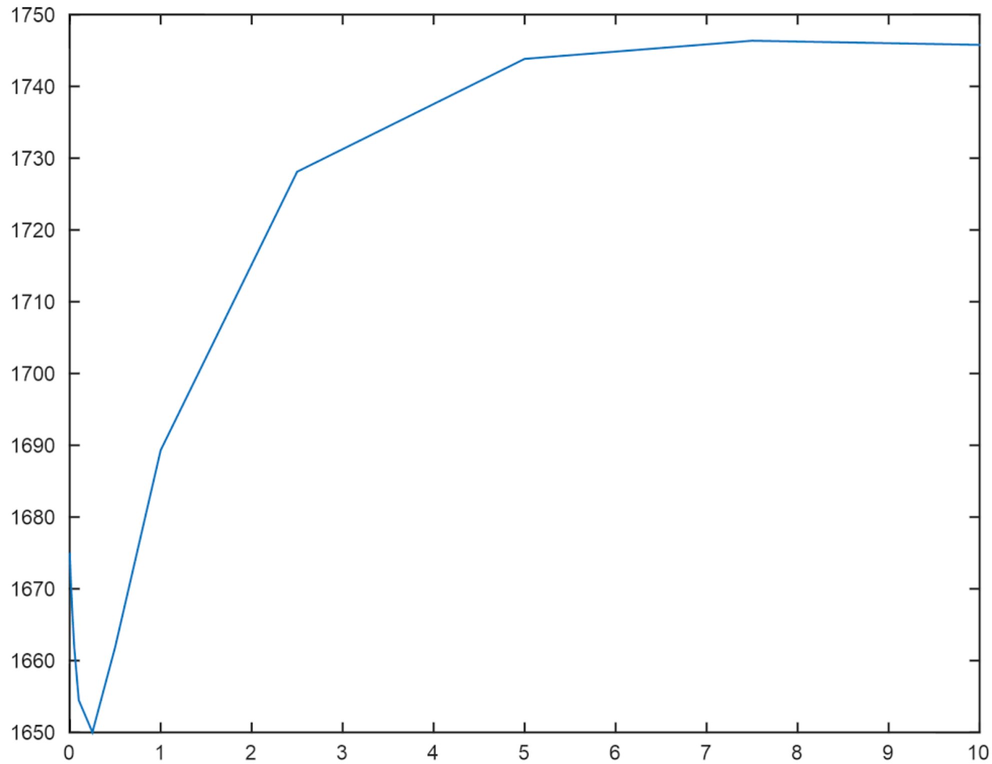
Table 1

	With Dummy	**Without Dummy**
Bias	39.5843	0
CRIM	-0.1011	-0.0979
ZN	0.0459	0.049
INDUS	-0.0027	-0.0254
CHAS	3.072	3.4509
NOX	-17.2254	-0.3555
RM	3.7113	5.8165
AGE	0.0072	-0.0033
DIS	-1.599	-1.0205
RAD	0.3736	0.2266
TAX	-0.0158	-0.0122
PTRATIO	-1.0242	-0.388
B	0.0097	0.017
LSTAT	-0.586	-0.485
SSE	**1.68E+03**	**1.80E+03**

Excluding the dummy variable confines the learned hyperplane to passing through the origin. This has the effect of skewing our prediction to fit the origin, which may not be an accurate model of the true function. The effect in this case is small but noticeable, increasing the SSE from 1.68E+03 to 1.80E+03.

This method allows overfitting by letting the weights grow unchecked. We next included a regularization term that minimized both the SSE and the norm of the weights. The relative importance of minimizing the norm of the weights was specified by a parameter λ. We performed a simple model selection on λ, plotting SSE for different chosen λ values. The results are shown in Figure 1. As λ increases, the norm of the weights decreases, and the SSE decreases until $\lambda = 0.25$, after which it increases. The varying λ values regularize the weights by penalizing high weights in the objective function. Higher λ values impose higher penalties on the weight. This penalty is achieved by causing the objective to be dependent on the weights. If λ is too large, the objective is over-dependent on the weights, so minimization of the objective relies on minimization of the weight, regardless of the model accuracy. If λ is too small, the weight is not minimized, leading to overfitting. That is, over-regularization causes minimization of the weight to be a priority over model accuracy, and under-regularization (clearly) reduces the effect of regularization on the weights, leading to overfitting. We conclude that $\lambda = 0.25$ is approximately optimal for this application.

Figure 1



1750
1740
1730
1720
1710
1700
1690
1680
1670
1660
1650
0
1
2
3
4
5
6
7
8
9
10