Name: Taiman Arham Siddiqui

CS 5900 HW Assignment 1

Implementation

1. Simple Regression:

Results:

Params: [-15682.27021631 115.41845202]

Training RMSE: 64083.51.

Training cost: 2053348364.32.

Test RMSE: 65773.19.

Test cost: 2163056355.39.

Plot of training data, test data, and linear approximation:

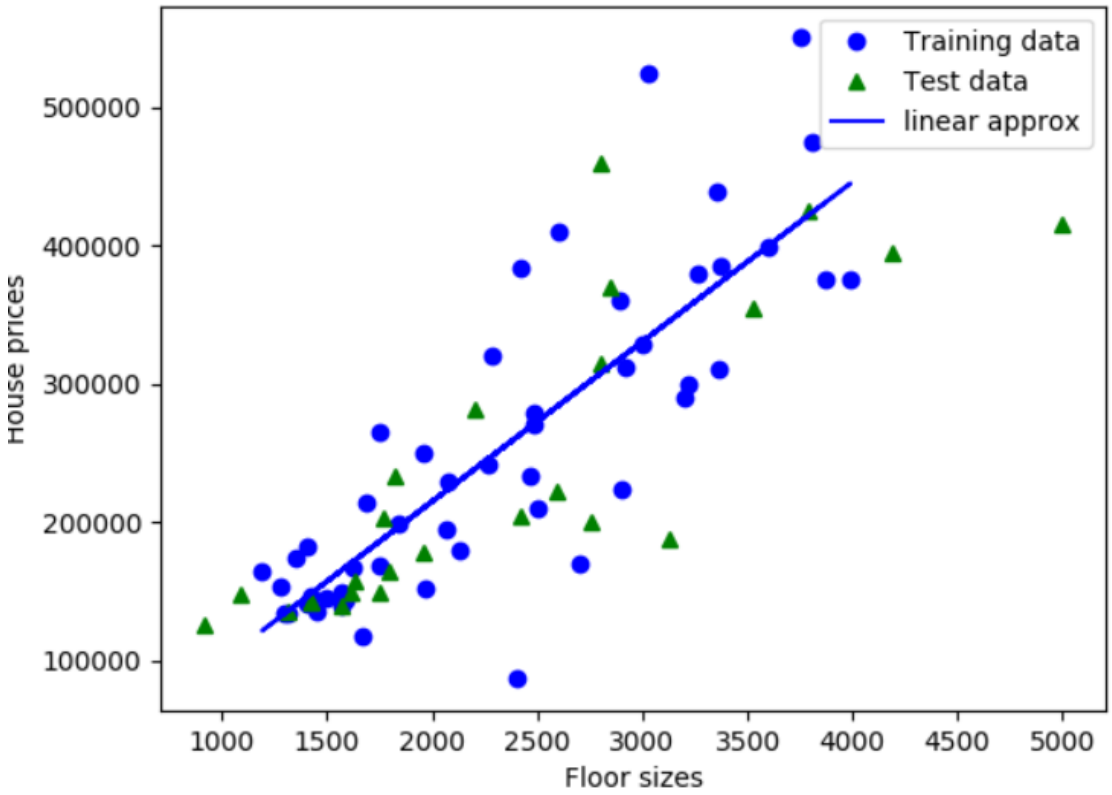

Fig: train_test_line.png

2. Multiple Regression:

Results:

Params: [-66713.84150388 96.6022094 25332.57797469 384.47514712]

Training RMSE: 61070.62.

Training cost: 1864810186.98.

Test RMSE: 58481.32.

Test cost: 1710032401.49.

The Test RMSE in this case is smaller than that of the simple case above. This is probably because of using more features to compute house price.

3. Polynomial Curve Fitting:

a) Plot of data in dataset.txt:

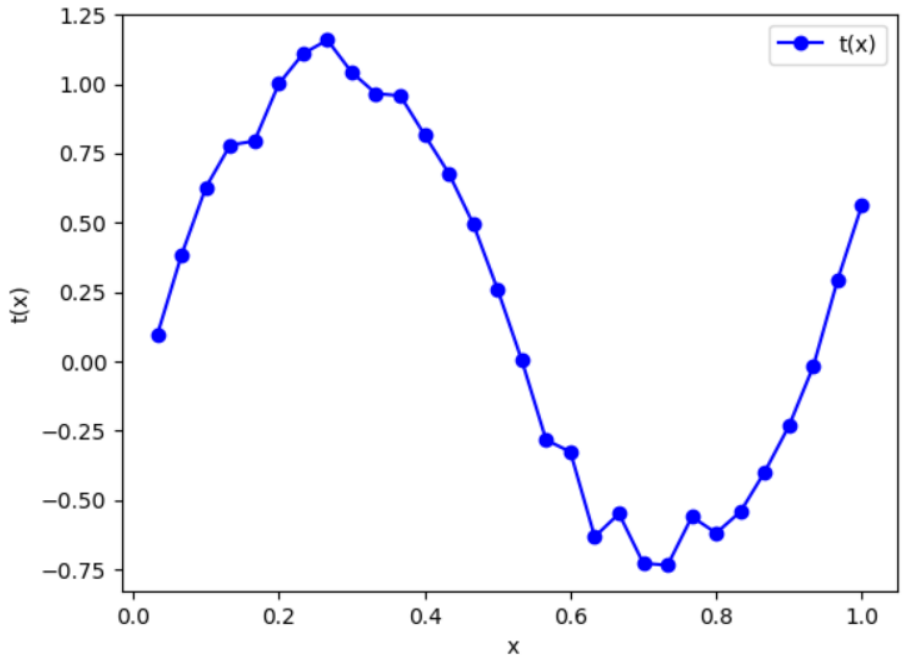

Fig: dataset.png

b) Plot of data in train.txt:

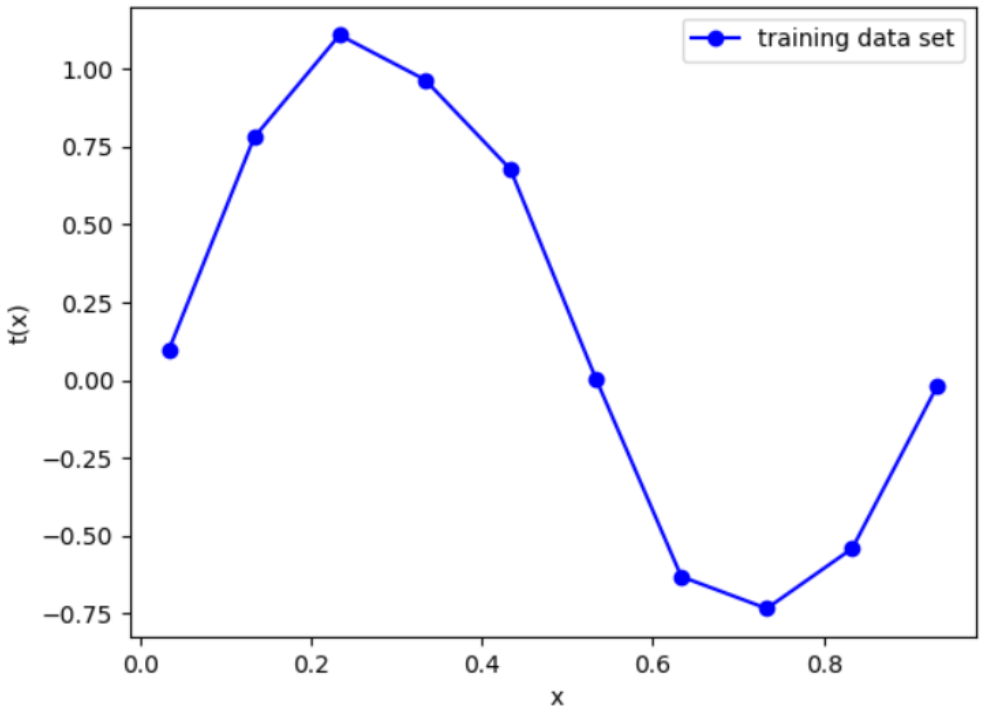

Fig: traindataset.png

Plot of data in test.txt:

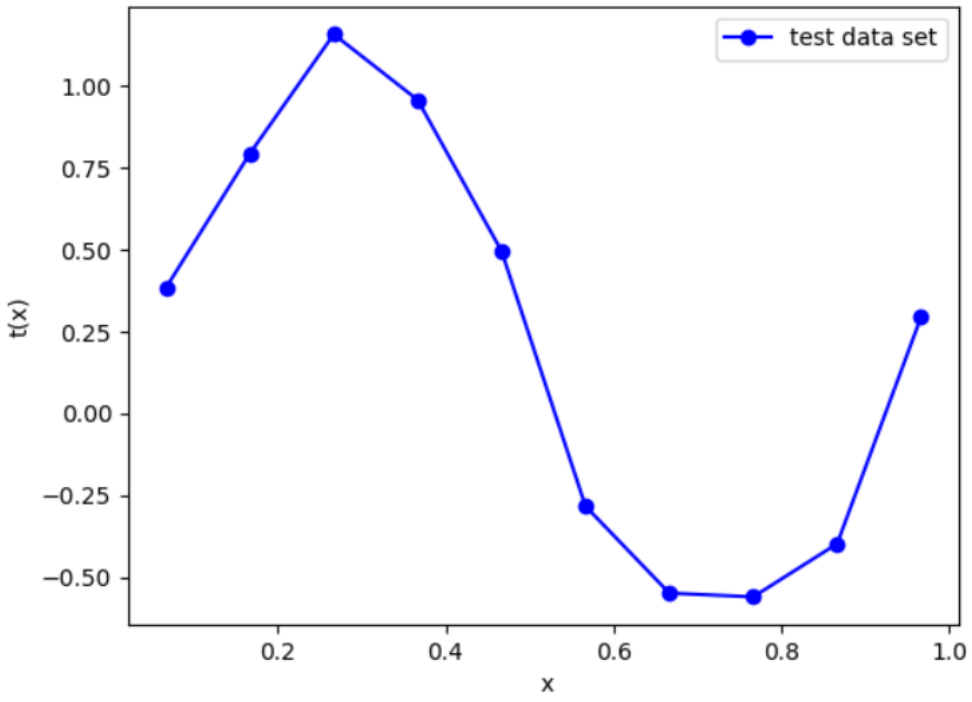

Fig: testdataset.png

Plot of data in devel.txt:

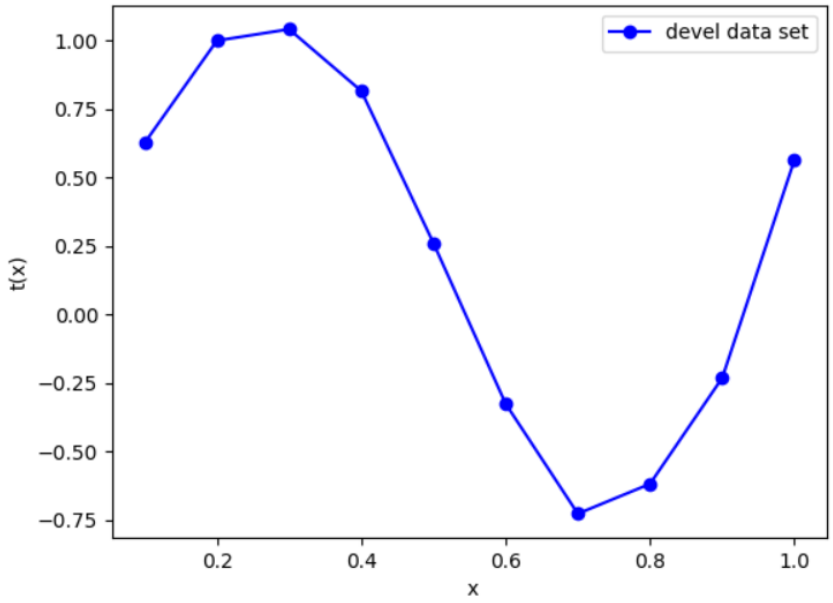

Fig: develdataset.png

c) The solution is:

$$\mathbf{w} = \left(\lambda N \mathbf{I} + \mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathbf{t}$$

[Slide 45 of Lecture 2 notes]

d)

1. Plot of RMSE values of training and test data for all values of M ∈ [0, 9]:

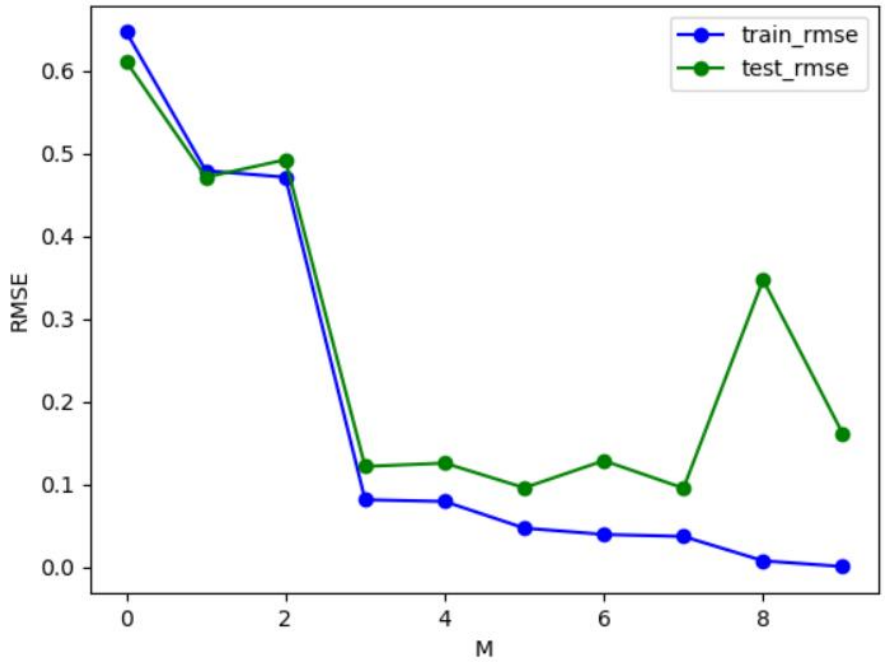

Fig: train-test-rmse-without-reg.png

We can see that the training and test RMSE plots follow a similar patter, except for M = 8,9 where the test RMSE shoots up compared to the training plot. This is because of the greater degree of overfitting for those values of M.

(Next page →)

2. Plot of training and validation RMSE values for ln λ ∈ [−50, 0] and M=9:

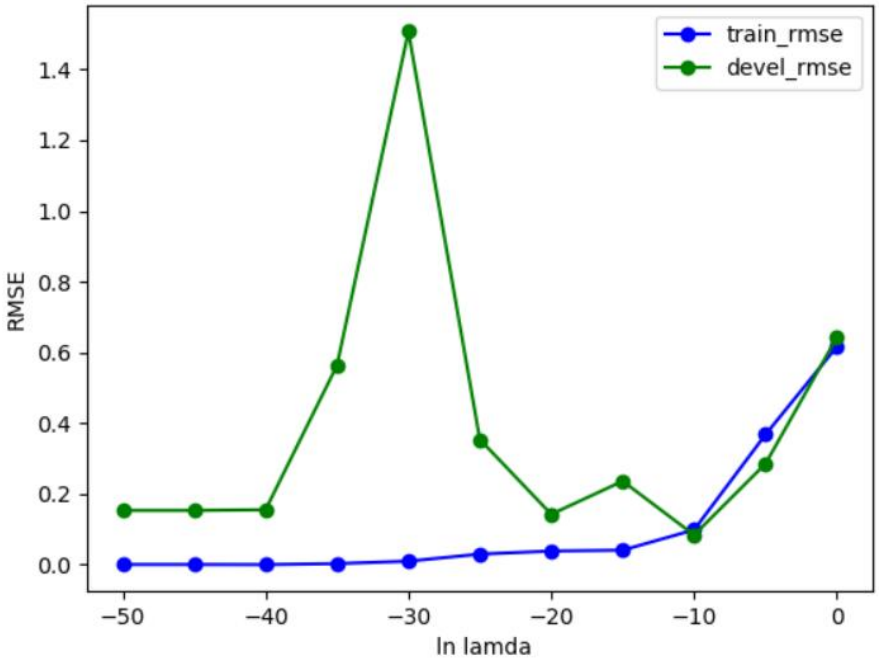

Fig: train-devel-rmse-with-reg.png

The training and validation RMSE plots follow a similar pattern except for the validation plot at ln λ = -30 where it shoots up and it seems to be an outlier case.

Test RMSE values without regularization and with regularization using parameter vector that obtained lowest RMSE on validation data:

```
Test without regularization for M = 9, RMSE: 0.16.
Test with regularization for M = 9, RMSE: 0.12.
```

As seen above, the Test RMSE obtained with regularization is lower than that obtained without regularization. This is because regularization avoids overfitting of the data, unlike the case without regularization.