ASSIGNMENT REPORT

STATISTICAL LEARNING, IIT MADRAS

Ridge Regression, Backward Stepwise Regression and Kernel Regression

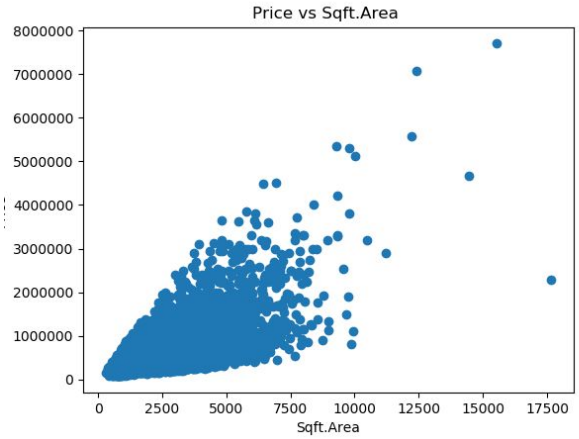
Faculty

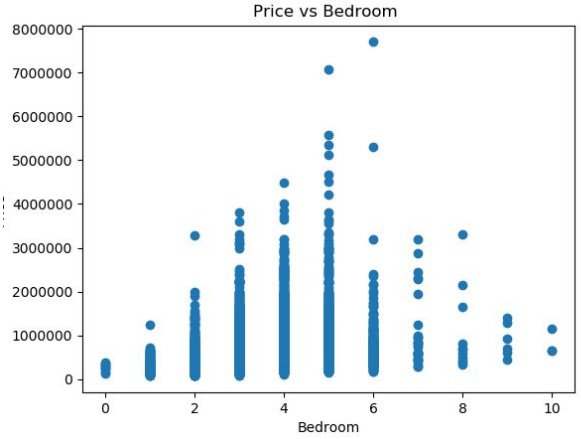
Prof. Sukhendu Das

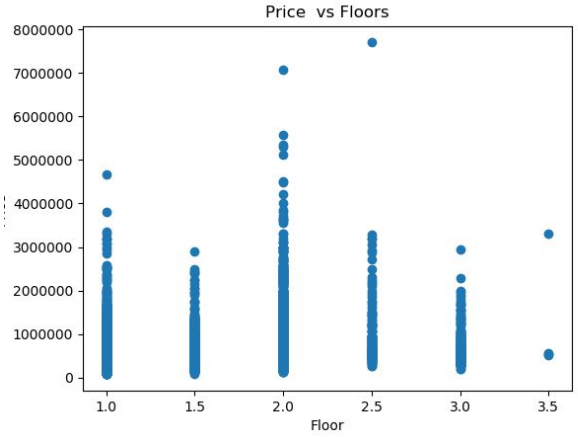
Group 5 :

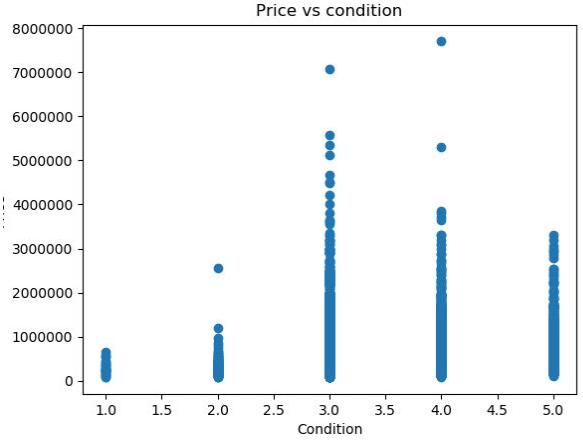
Vivek Kumar Agrawal (CS17M049)
Manas Ranjan Sahoo (CS17M024)

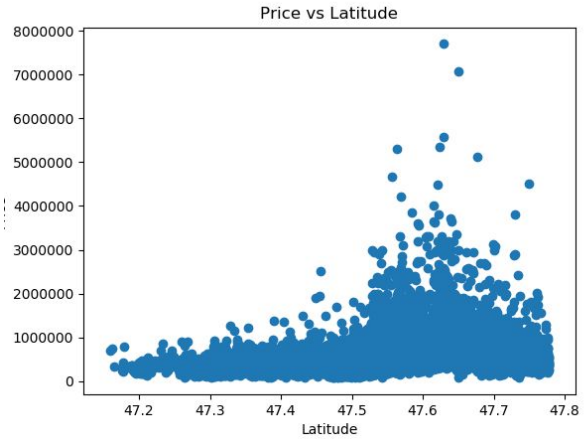
Data Interpretation:

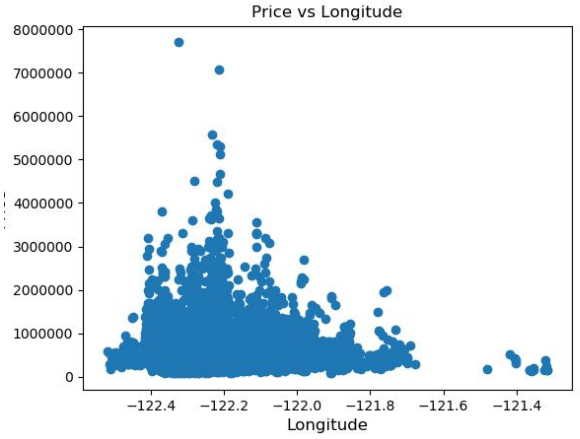












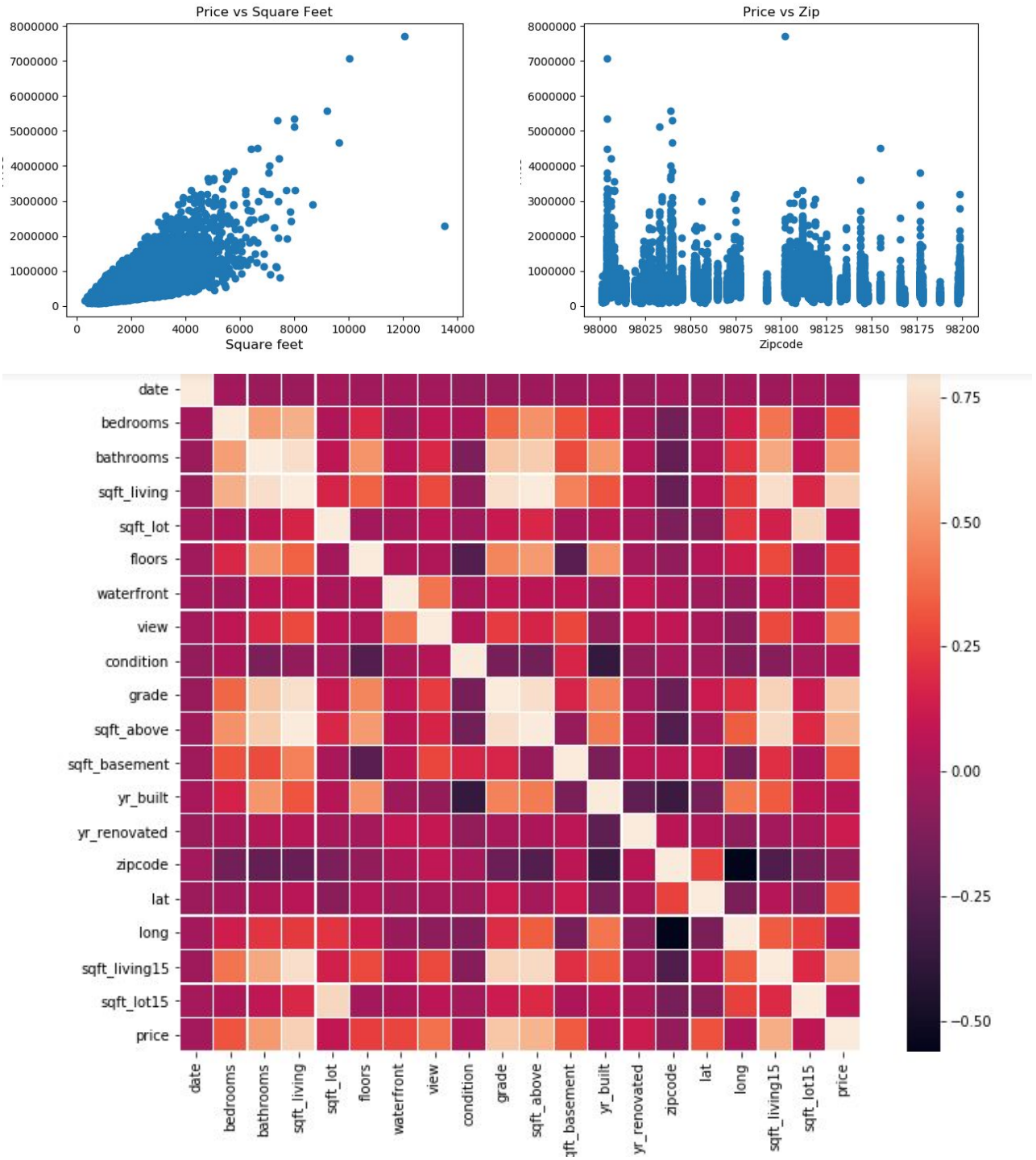


Fig: Correlation Matrix representing correlation among features

Ridge Regression:

Computed 10-fold Cross Validation Error for different Regularizer Coefficient (λ) and plotted the one with least error. Then, out of all the plotted ones, selected the least one as λ for our Ridge Regression.

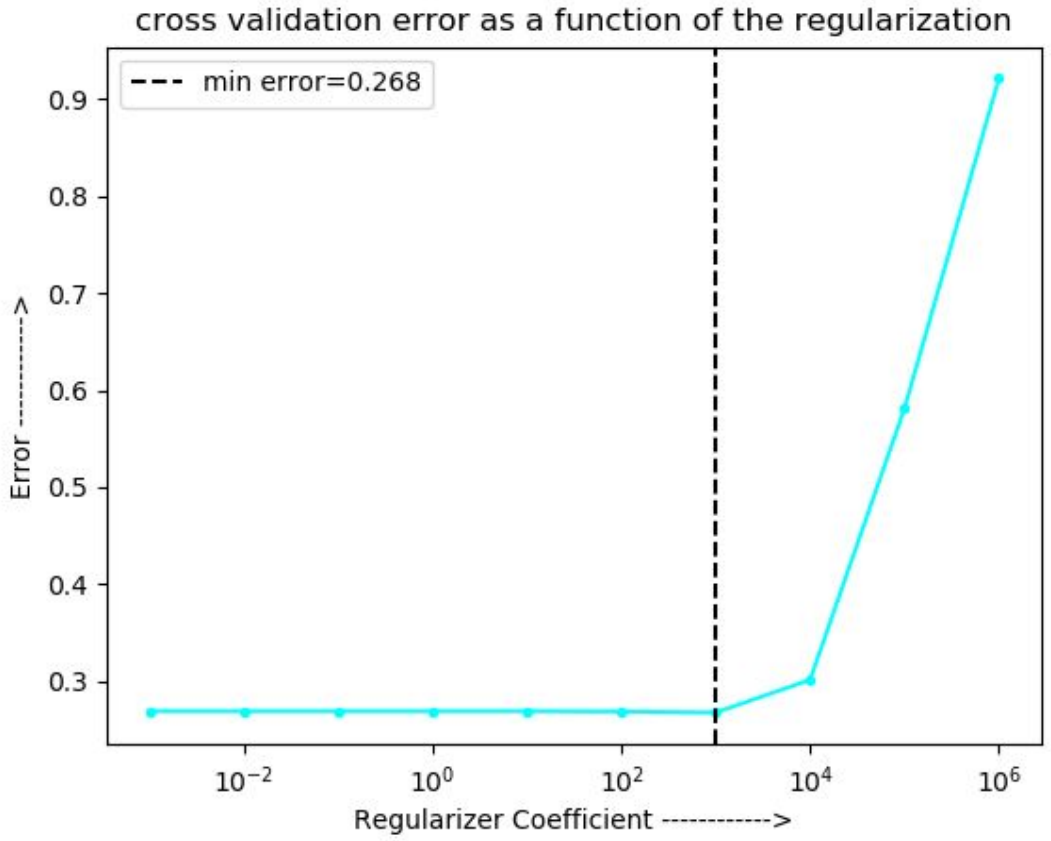


Fig 1.1 Observations:

1. Minimum val error obtained is **0.268** with **λ = 1000** (Marked with dotted line).
2. There is an increase in error with increase in **λ** value.

Plot of Error vs Steps of 10-fold cross validation for λ = 1000 is given in Fig 1.2 below.

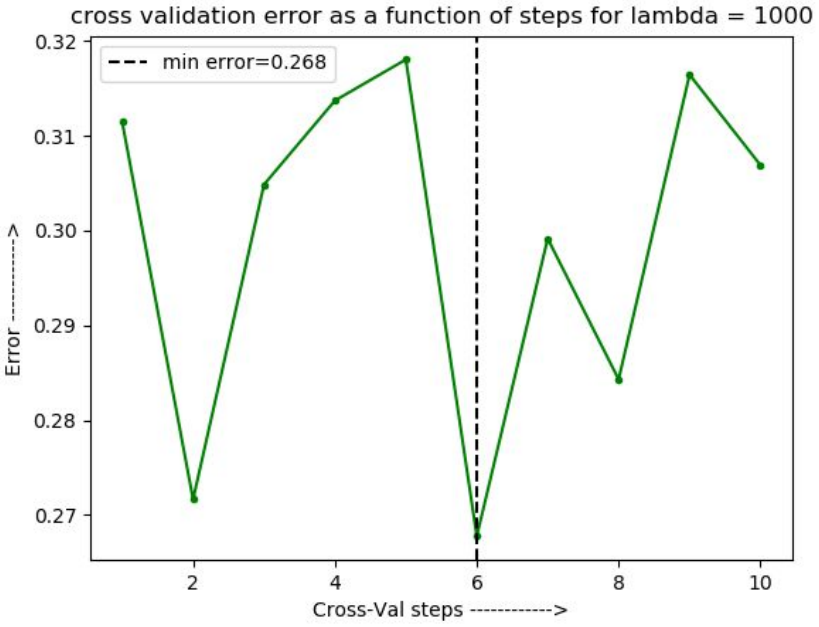


Fig:1.2 Observations:

1. Since the Validation data and Train data is different in each step of 10-fold cross validation, we can see that error is different at each step for same λ (= 1000 used at each step).
2. Minimum error was obtained at step 6.

Plot of Weights (top 5) Vs Regularizer coefficient (λ) is given in Fig:1.3

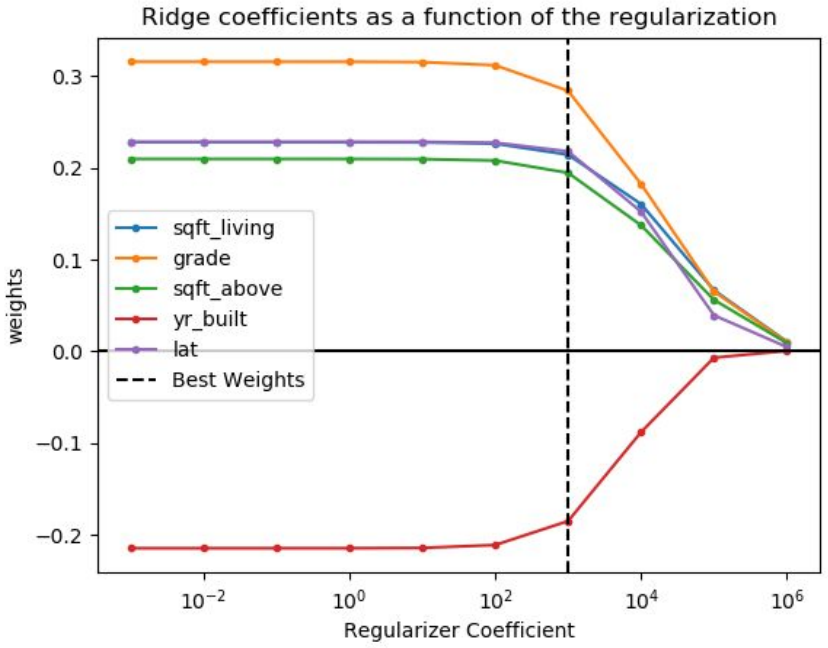


Fig:1.3 Observations:

1. The value of weights reduces with increase in **λ** and at very high **λ** weights shrinks to 0. Hence at this point the predictors are contributing nothing for making a prediction.
2. The best model was obtained at **λ = 1000** with least error (dotted line on plot).

Scattering Actual Price vs Predicted Price of Test data in Fig:1.4 below

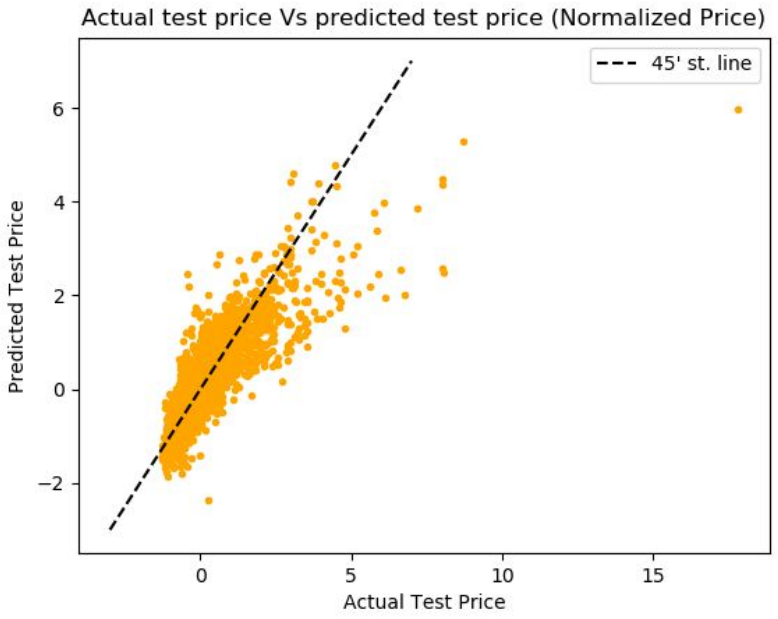


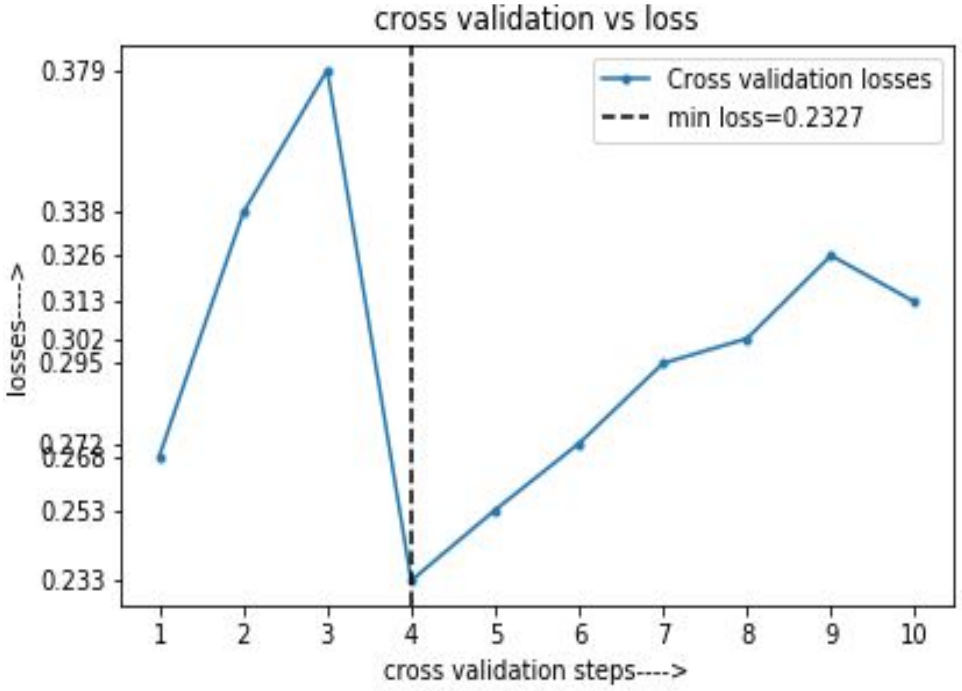
FIG:1.4

Fig:1.4 Observations:

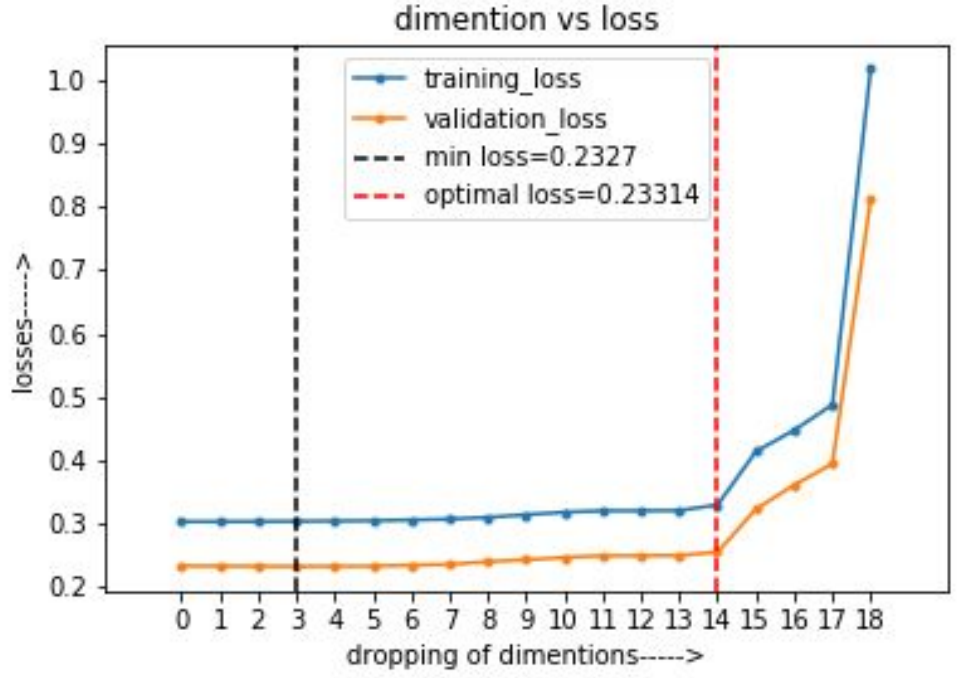
1. The most accurate scatter will be along the 45° line.
2. Hence, this deviation of scatter from dotted line shows that there is some error in predictions.
3. Test error obtained was **0.308**.

Backward Stepwise Regression:

We have performed 10-fold Cross Validation for selecting model(training and validation data) and choose the model which gave least validation loss.



After choosing model we performed backward stepwise regression using p-value as the selection criteria. So we dropped the features which has maximum p-value one at a time and iterated till all features are dropped and choose the coefficients which gave minimum validation loss.



Observations :

1) We got minimum validation loss as .2327 after dropping 3 dimensions and as we go on dropping features both training and validation loss started increasing. So we chose this point as minimum point.(Black dotted line in Fig 2.2)
2) After dropping 14 features, we incurred extra penalty 0.0004, which is very less considering performance gain. So, we chose this point in Fig 2.2 as the optimal point.(Red dotted line in Fig 2.2)
3) We incurred **0.31 loss with minimum point weights and 0.34 loss with optimal point weights** using test data. So, there is a trade off between performance and accuracy.

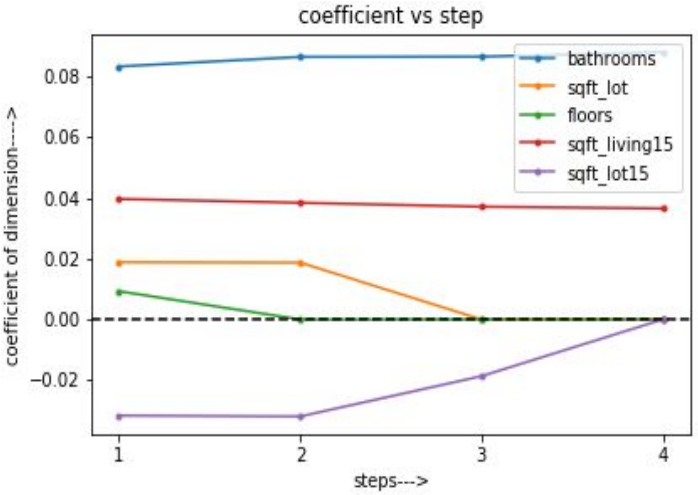


Fig. 2.3 For min point

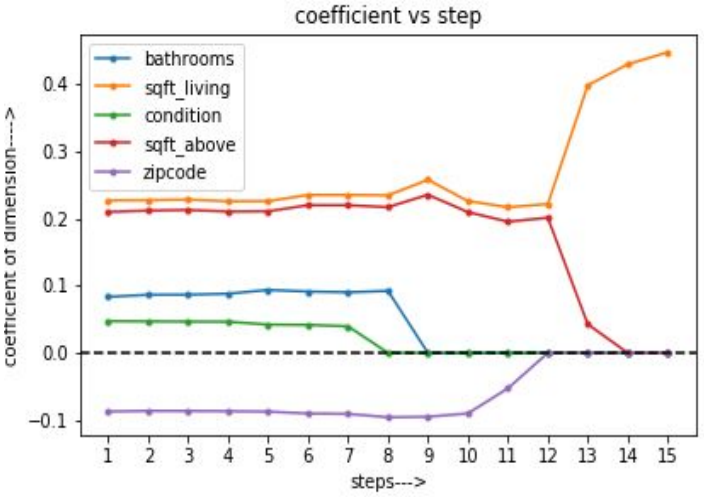


Fig. 2.4 For optimal point

4) We see much variations in coefficients of optimal point than min point.(Fig 2.3, 2.4))

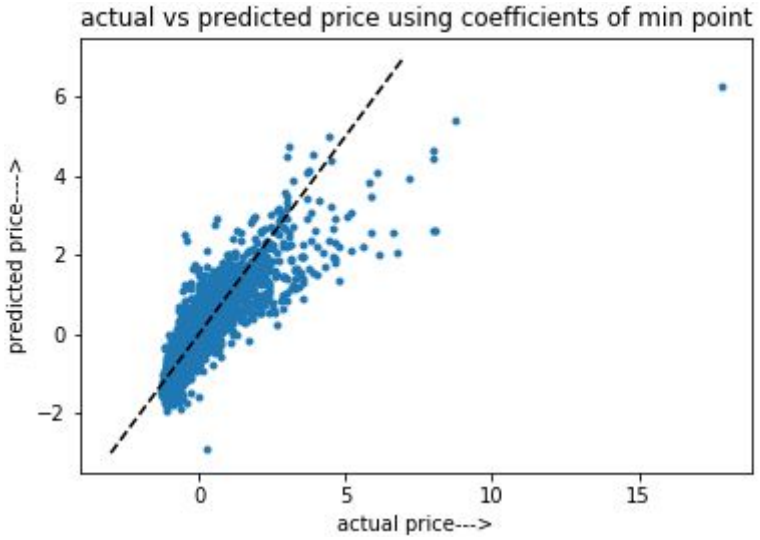


Fig. 2.5

Kernel Regression:

This method is Non-Linear method of Regression. We have used Gaussian Kernel for making predictions. In this method, the prediction is based on neighbouring points similar to KNN. But here the neighbours are weighted according to the Mahalanobis distance.

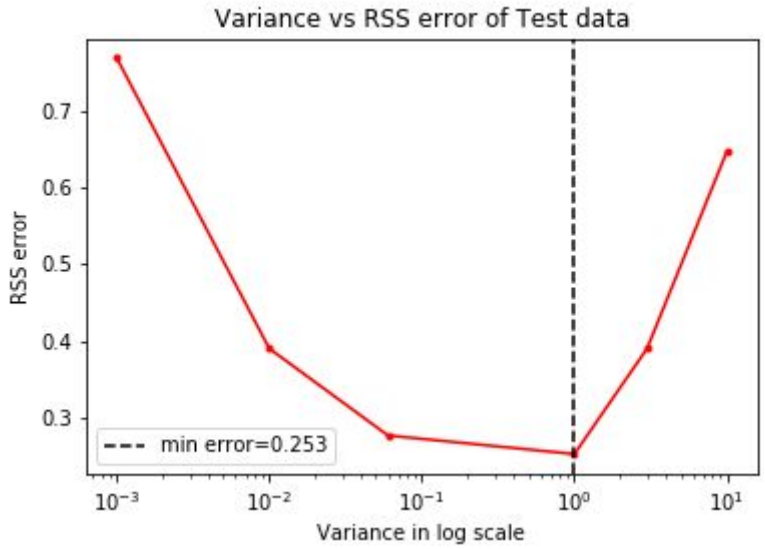


Fig:3.1

Fig:3.1 Observations

1. The covariance matrix used here is a diagonal matrix with same variance along the diagonal i.e., **Σ = variance*I**, here, **I** is identity matrix. This done because **Σ** is a hyper-parameter in kernel regression.
2. The minimum test error of **0.253** was obtained for **variance** = 1, hence, **Σ = I**.

Scattering Actual Price vs Predicted Price of Test data in Fig:3.2 below

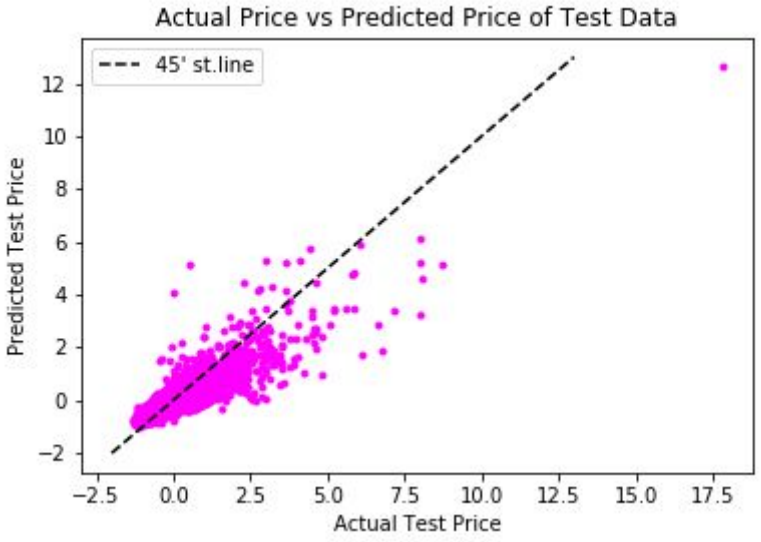


Fig:3.2

Conclusions:

1. On comparing all the three Regression Methods, the least test error was obtained for Kernel Regression followed by Ridge Regression and the most error for Backward Stepwise Regression.

2. It is because Kernel Regression is a non-linear method while the other two are linear. Hence, Kernel Regression was able to fit more accurately than others.

3. Test Data Scores:

Regression Method	**Test Loss**
Ridge Regression	0.31
Backward StepWise Regression	0.34
Kernel Regression	0.25