MLFA MINI-PROJECT 2

Effect of change in lambda on w in regressions

Name – Bhosale Ratnesh Sambhajirao
Roll No. – 19MF10010

Instructor – Prof. Adway Mitra

1.1 Problem Statement

Try out different kinds of regularized functions on regression problems, either of synthetic datasets or standard/public datasets. Explore the impacts of changing the “lambda” parameter on the structure of “w” and on the least-square error. In addition to ridge and LASSO, you may try out i) elastic net regularization, ii) mixed-norm or group-sparsity regularization.

1.2 Dataset Description

Dataset has been created manually. Input array is defined with angles from 60^{o} to 300^{o} converted to radians. Dataset looks similar to sine curve with some noise. The sine function using polynomial regression with powers of x from 1 to 15 has been estimated. Column for each power up to 15 has been added in the data frame.

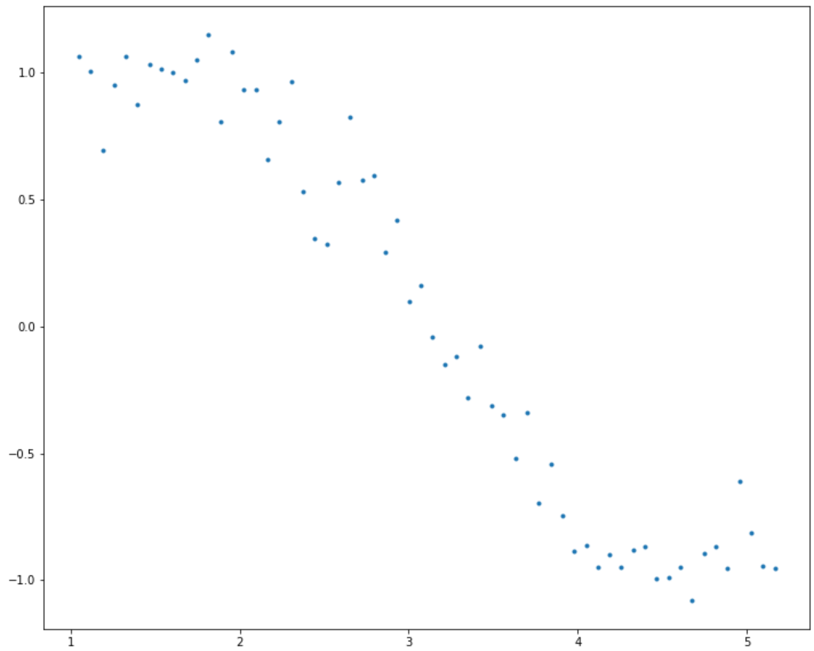

Fig. 1. Dataset

Table 1. A column for each power up to 15 in DataFrame (head)

	x	y	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_10	x_11	x_12	x_13	x_14	x_15
0	1	1.1	1.1	1.1	1.2	1.3	1.3	1.4	1.4	1.5	1.6	1.7	1.7	1.8	1.9	2
1	1.1	1	1.2	1.4	1.6	1.7	1.9	2.2	2.4	2.7	3	3.4	3.8	4.2	4.7	5.3
2	1.2	0.7	1.4	1.7	2	2.4	2.8	3.3	3.9	4.7	5.5	6.6	7.8	9.3	11	13
3	1.3	0.95	1.6	2	2.5	3.1	3.9	4.9	6.2	7.8	9.8	12	16	19	24	31
4	1.3	1.1	1.8	2.3	3.1	4.1	5.4	7.2	9.6	13	17	22	30	39	52	69

Source: Notebook

1.3 Linear Regression

1.3.1 Brief Introduction to Linear Regression

The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

$$Y = B_0 + B_1 x$$

1.3.2 Variation in Errors

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with power of polynomial regressor.

Table 2. Variation of errors with power of regressor.

Power	MAE	MSE	RMSE	R2 score
1	0.192	0.054	0.233	0.902
2	0.193	0.054	0.233	0.902
3	0.105	0.018	0.135	0.969
4	0.105	0.017	0.134	0.969
5	0.1006	0.016	0.13	0.971
6	0.098	0.016	0.128	0.972
7	0.093	0.015	0.124	0.974
8	0.093	0.015	0.123	0.974
9	0.094	0.014	0.12	0.975
10	0.094	0.014	0.12	0.975
11	0.093	0.014	0.12	0.975
12	0.093	0.014	0.12	0.975

13	0.093	0.014	0.119	0.976
14	0.088	0.013	0.114	0.978
15	0.083	0.011	0.108	0.98

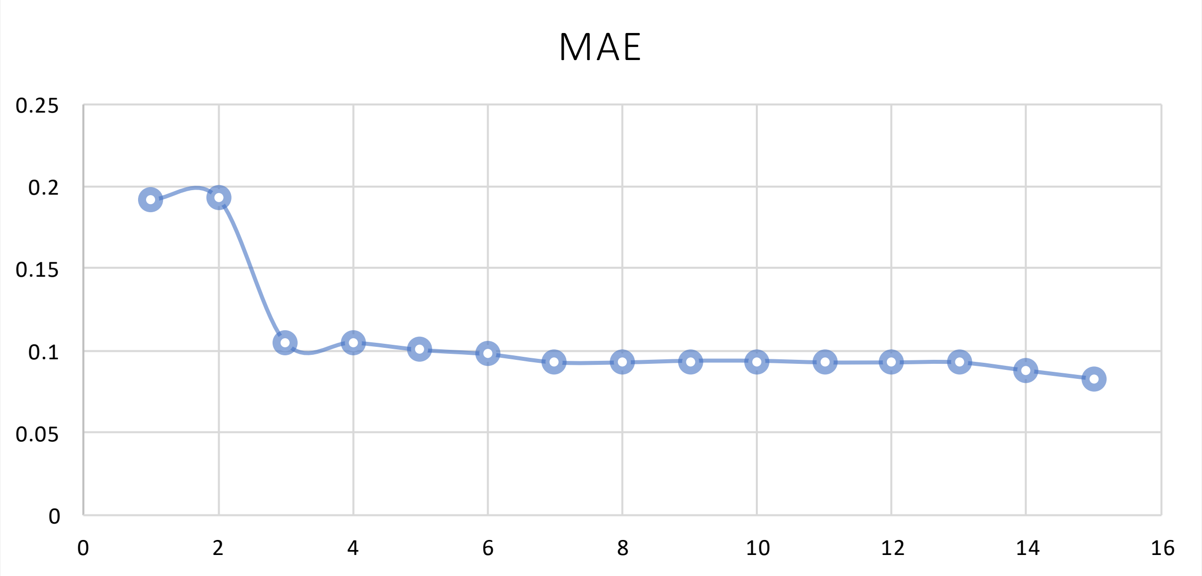

Fig. 2. Variation of Mean Absolute Error (MAE) for linear regression

As the graph shows, MAE is decreasing with the increase in model complexity.

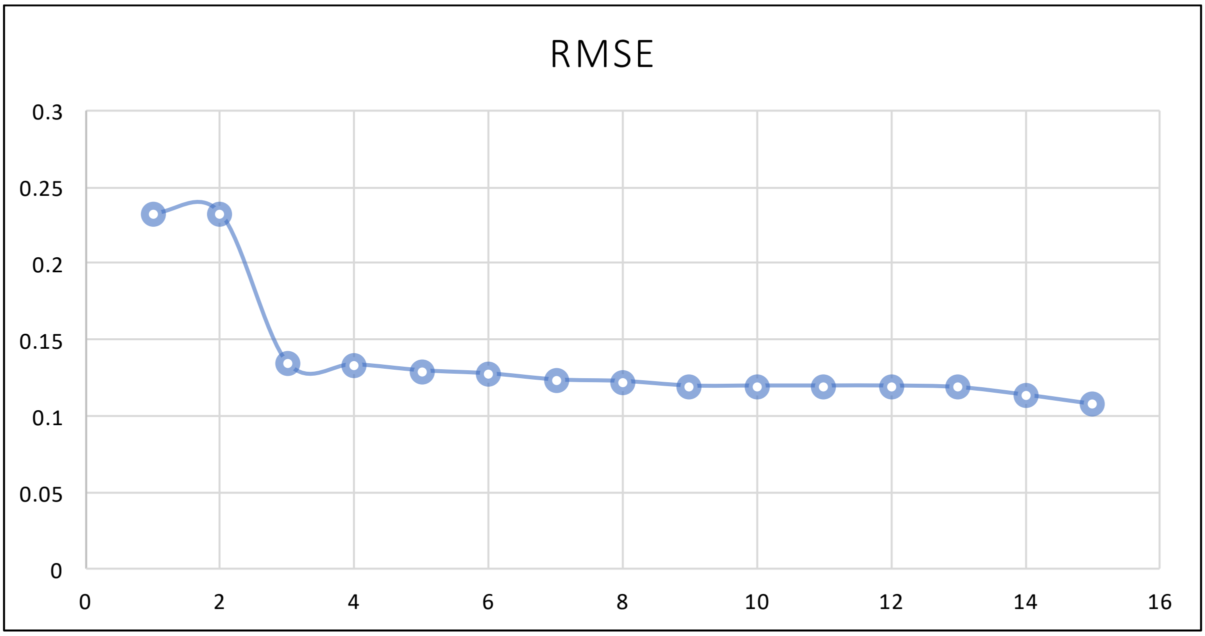

Fig. 3. Variation of Root Mean Absolute Error (RMSE) for linear regression

As the graph shows, RMSE is decreasing with the increase in model complexity.

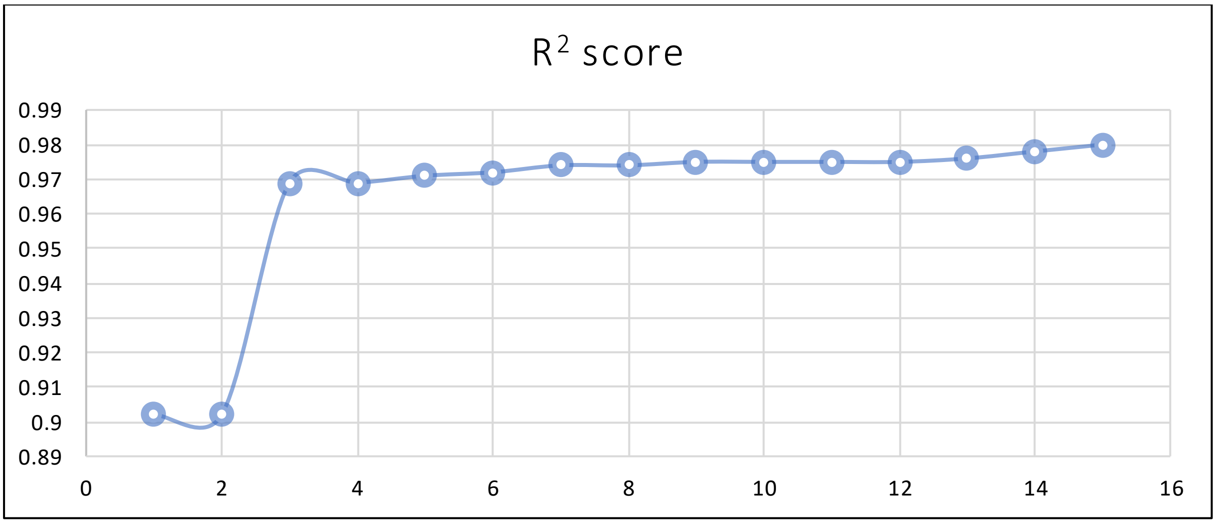

Fig. 4. Variation of R^2 Score for linear regression

As the graph shows, R^2 score is increasing with the increase in model complexity.

We would expect the models with increasing complexity to better fit the data and result in lower RSS values. This can be verified by looking at the plots generated for 6 models:

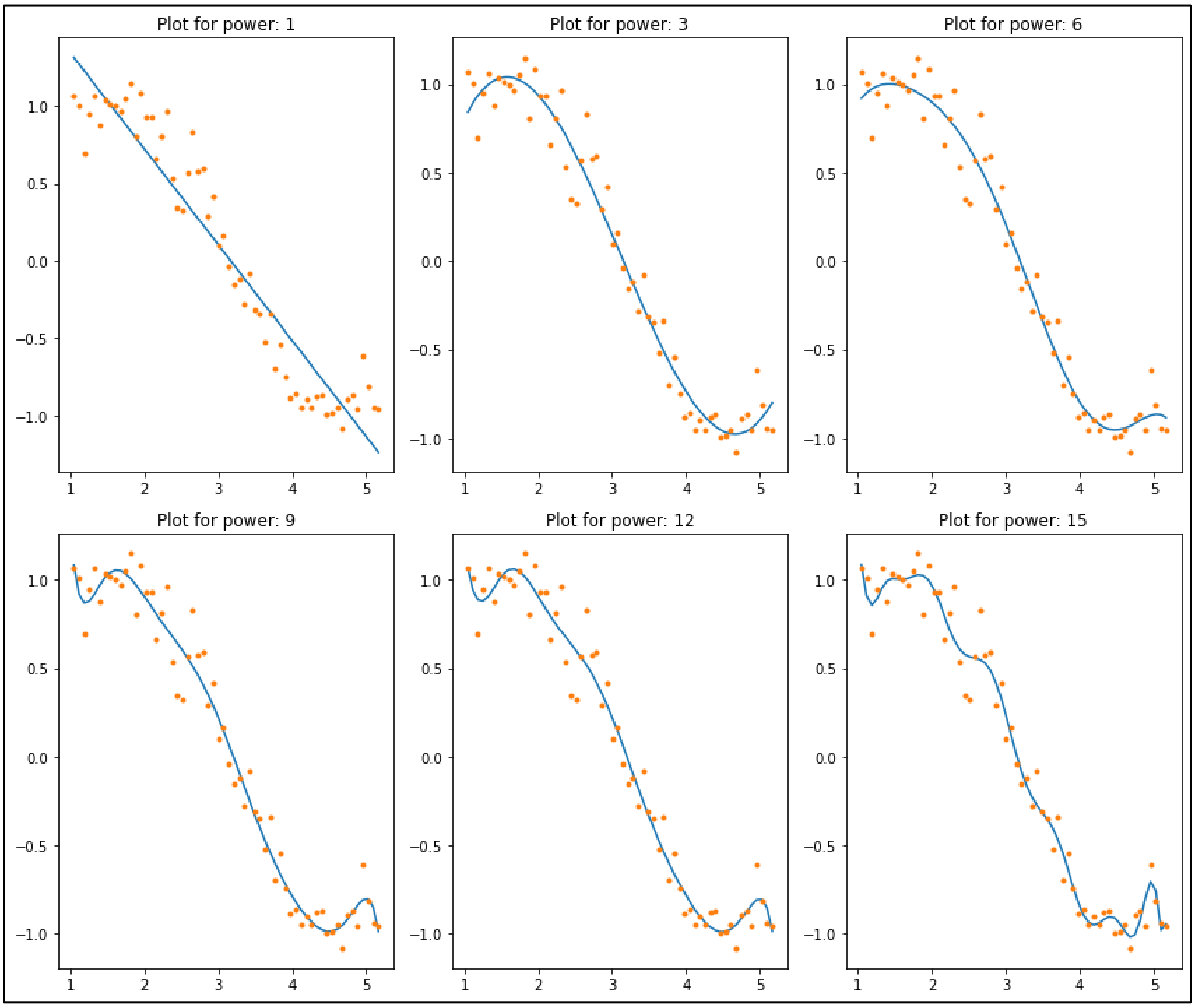

Fig. 5. Fit line with change in power.

1.3.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 3. Change in rss, intercept and w with power

	rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15
model_pow_1	3.3	2	-0.62	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_2	3.3	1.9	-0.58	-0.006	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_3	1.1	-1.1	3	-1.3	0.14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_4	1.1	-0.27	1.7	-0.53	-0.036	0.014	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_5	1	3	-5.1	4.7	-1.9	0.33	-0.021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_6	0.99	-2.8	9.5	-9.7	5.2	-1.6	0.23	-0.014	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_7	0.93	19	-56	69	-45	17	-3.5	0.4	-0.019	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_8	0.92	43	-1.40e+02	1.80e+02	-1.30e+02	58	-15	2.4	-0.21	0.0077	NaN	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_9	0.87	1.70e+02	-6.10e+02	9.60e+02	-8.50e+02	4.60e+02	-1.60e+02	37	-5.2	0.42	-0.015	NaN	NaN	NaN	NaN	NaN	NaN
model_pow_10	0.87	1.40e+02	-4.90e+02	7.30e+02	-6.00e+02	2.90e+02	-87	15	-0.81	-0.14	0.026	-0.0013	NaN	NaN	NaN	NaN	NaN
model_pow_11	0.87	-75	5.10e+02	-1.30e+03	1.90e+03	-1.60e+03	9.10e+02	-3.50e+02	91	-16	1.8	-0.12	0.0034	NaN	NaN	NaN	NaN
model_pow_12	0.87	-3.40e+02	1.90e+03	-4.40e+03	6.00e+03	-5.20e+03	3.10e+03	-1.30e+03	3.80e+02	-80	12	-1.1	0.062	-0.0016	NaN	NaN	NaN
model_pow_13	0.86	3.20e+03	-1.80e+04	4.50e+04	-6.70e+04	6.60e+04	-4.60e+04	2.30e+04	-8.50e+03	2.30e+03	-4.50e+02	62	-5.7	0.31	-0.0078	NaN	NaN
model_pow_14	0.79	2.40e+04	-1.40e+05	3.80e+05	-6.10e+05	6.60e+05	-5.00e+05	2.80e+05	-1.20e+05	3.70e+04	-8.50e+03	1.50e+03	-1.80e+02	15	-0.73	0.017	NaN
model_pow_15	0.7	-3.60e+04	2.40e+05	-7.50e+05	1.40e+06	-1.70e+06	1.50e+06	-1.00e+06	5.00e+05	-1.90e+05	5.40e+04	-1.20e+04	1.90e+03	-2.20e+02	17	-0.81	0.018

It is clearly evident that the size of coefficients increases exponentially with increase in model complexity. Large coefficient signifies that we're putting a lot of emphasis on that feature, *i.e.*, the particular feature is a good predictor for the outcome. When it becomes too large, the algorithm starts modelling intricate relations to estimate the output and ends up overfitting to the particular training data.

1.4 Ridge Regression

1.4.1 Brief Introduction to Ridge Regression

Ridge regression performs 'L2 regularization', i.e., it adds a factor of sum of squares of coefficients in the optimization objective. Thus, ridge regression optimizes the following:

$$Objective = RSS + \lambda * (sum\ of\ square\ of\ coefficients)$$

Here, λ (Lambda) is the parameter which balances the amount of emphasis given to minimizing RSS vs minimizing sum of square of coefficients. λ can take various values:

- $\lambda = 0$:
  - The objective becomes same as simple linear regression.
  - We'll get the same coefficients as simple linear regression.
- $\lambda = \infty$:
  - The coefficients will be zero. Because of infinite weightage on square of coefficients, anything less than zero will make the objective infinite.
- $0 < \lambda < \infty$:
  - The magnitude of λ will decide the weightage given to different parts of objective.
  - The coefficients will be somewhere between 0 and ones for simple linear regression.

1.4.2 Variation in Error

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with λ.

Table 4. Variation of Errors with λ.

Lambda	MAE	MSE	RMSE	R2 Score
1.00E-15	0.09432	0.01455	0.1206	0.9757
1.00E-10	0.09355	0.01537	0.124	0.9743
1.00E-08	0.09466	0.01582	0.1257	0.9736
0.0001	0.09577	0.01603	0.1266	0.9732
0.001	0.09997	0.16696	0.1292	0.9718
0.01	0.12383	0.02392	0.1546	0.9587
1	0.2702	0.09414	0.3068	0.7322
5	0.4302	0.23063	4.80E-01	-2.57E-01
10	0.48781	3.00E-01	5.48E-01	-1.68E+00
20	0.55913	3.79E-01	6.15E-01	-5.87E+00

Source: Notebook

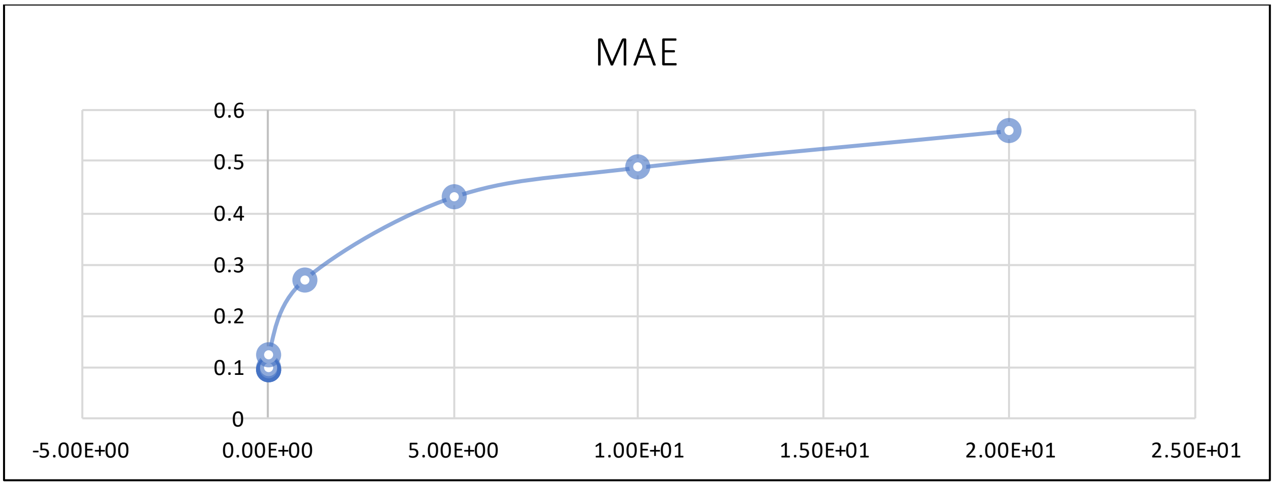

Fig. 6. Variation of Mean Absolute Error (MAE) with λ for Ridge regression

As the graph shows, MAE is increasing with the increase in λ.

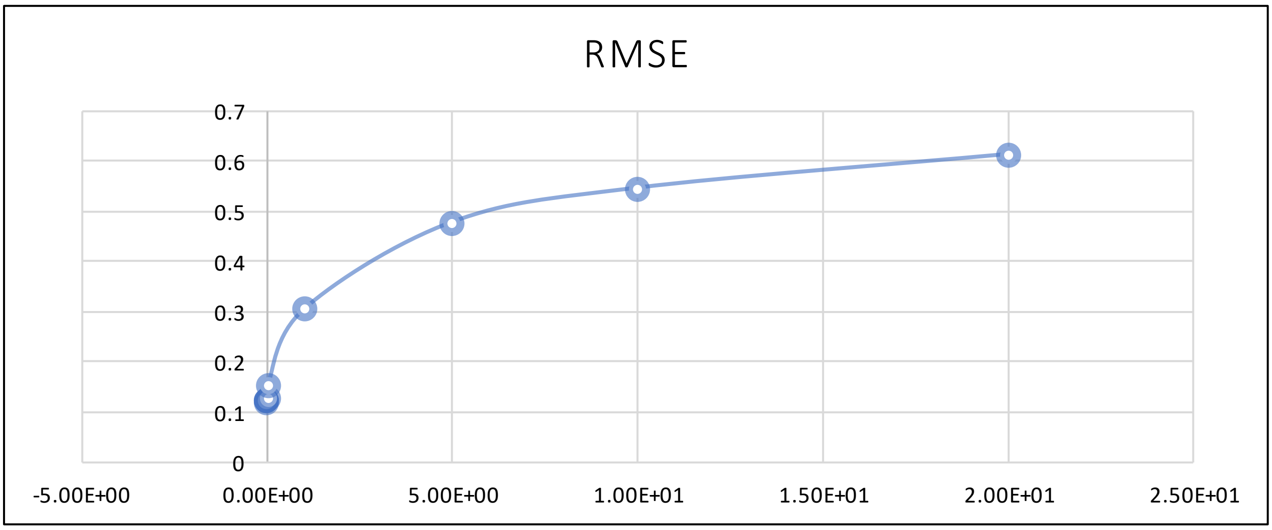

Fig. 7. Variation of Root Mean Absolute Error (RMSE) with λ

As the graph shows, RMSE is increasing with the increase in λ.

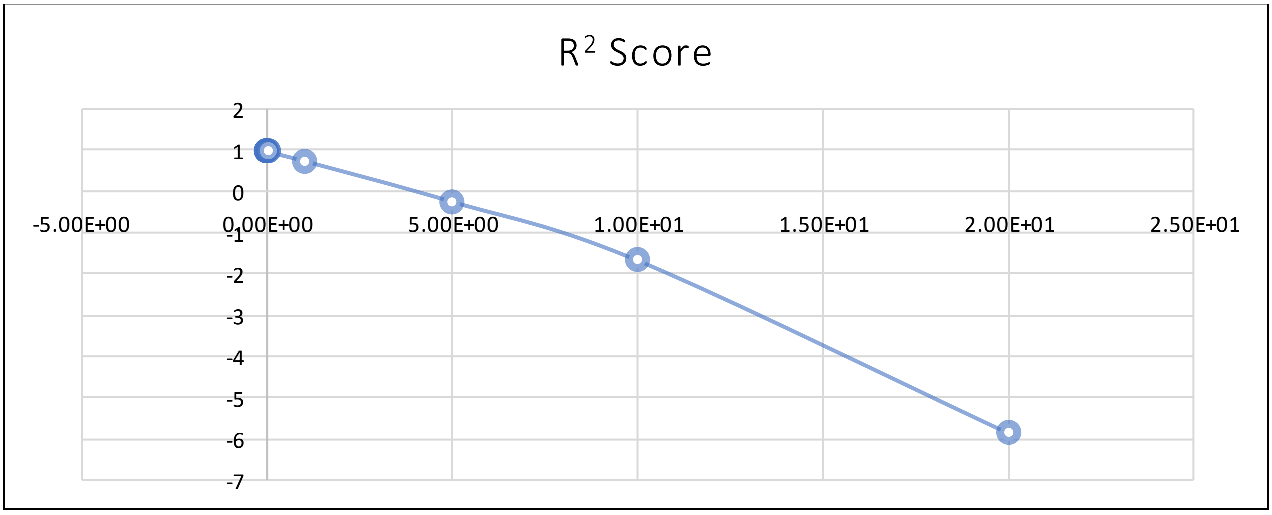

Fig. 8. Variation of R^2 Score with λ

As the graph shows, R^2 Score is decreasing with the increase in λ.

Model fit with increase in λ can be visualize by following graphs:

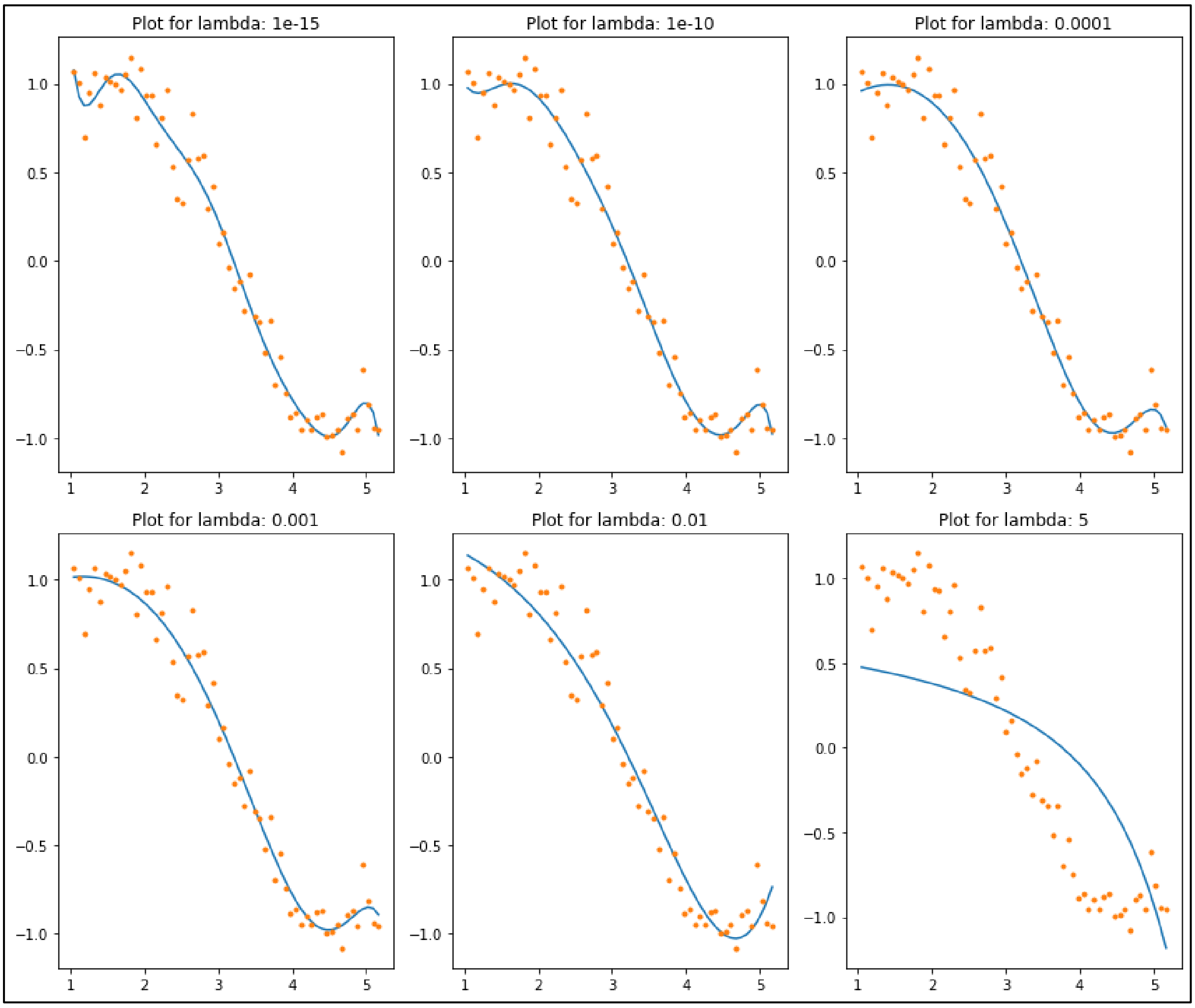

Fig. 9. Effect of increment in λ in Ridge Regression

1.4.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 5. Change in rss, intercept and w with power

	rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15
lambda_1e-15	0.87	95	-3.00E+02	3.80E+02	-2.30E+02	63	2.1	-5	0.6	0.17	-0.031	-0.0051	0.00085	0.00024	-6.10E-05	4.50E-06	-9.30E-08
lambda_1e-10	0.92	11	-29	31	-15	2.9	0.17	-0.091	-0.011	0.002	0.0064	2.40E-05	-2.00E-05	-4.20E-06	2.20E-07	2.30E-07	-2.30E-08
lambda_1e-08	0.95	1.3	-1.5	1.7	-0.68	0.039	0.016	0.0016	-0.0036	-5.40E-05	-2.90E-07	1.10E-06	1.90E-07	2.00E-08	3.90E-09	8.20E-10	-4.60E-10
lambda_0.0001	0.96	0.56	0.55	-0.13	-0.026	-0.0028	-0.0011	4.10E-05	1.50E-05	3.70E-06	7.40E-07	1.30E-07	1.90E-08	1.90E-09	-1.30E-10	-1.50E-10	-6.20E-11

lambda_0.001	1	0.82	0.31	-0.087	-0.02	-0.0028	-0.0022	1.80E-05	1.20E-05	3.40E-06	7.30E-07	1.30E-07	1.90E-08	1.70E-09	-1.50E-10	-1.40E-10	-5.20E-11
lambda_0.01	1.4	1.3	-0.088	-0.052	-0.01	-0.0014	-0.0013	7.20E-07	4.10E-06	1.30E-06	3.00E-07	5.60E-08	9.00E-09	1.10E-09	4.30E-11	-3.10E-11	-1.50E-11
lambda_1	5.6	0.97	-0.14	-0.019	-0.003	-0.0047	-7.00E-05	-9.90E-06	-1.30E-06	-1.40E-07	-9.30E-09	1.30E-09	7.80E-10	2.40E-10	6.20E-11	1.40E-11	3.20E-12
lambda_5	14	0.55	-0.059	-0.0085	-0.0014	-0.0024	-4.10E-05	-6.90E-06	-1.10E-06	-1.90E-07	-3.10E-08	-5.10E-09	-8.20E-10	-1.30E-10	-2.00E-11	-3.00E-12	-4.20E-13
lambda_10	18	0.4	-0.037	-0.0055	-0.00095	-0.0017	-3.00E-05	-5.20E-06	-9.20E-07	-1.60E-07	-2.90E-08	-5.10E-09	-9.10E-10	-1.60E-10	-2.90E-11	-5.10E-12	-9.10E-13
lambda_20	23	0.28	-0.022	-0.0034	-0.0006	-0.0011	-2.00E-05	-3.60E-06	-6.60E-07	-1.20E-07	-2.20E-08	-4.00E-09	-7.50E-10	-1.40E-10	-2.50E-11	-4.70E-12	-8.70E-13

Source: Notebook

The RSS increases with increase in lambda, this model complexity reduces. A lambda as small as 1e-15 gives us significant reduction in magnitude of coefficients. High lambda values can lead to significant underfitting.

1.5 Lasso Regression

1.5.1 Brief Introduction to Lasso Regression

Lasso regression performs L1 regularization, i.e., it adds a factor of sum of absolute value of coefficients in the optimization objective. Thus, lasso regression optimizes the following:

$$Objective = RSS + \lambda * (sum\ of\ absolute\ value\ of\ coefficients)$$

Like that of ridge, α can take various values.

- $\lambda = 0$: Same coefficients as simple linear regression
- $\lambda = \infty$: All coefficients zero (same logic as before)
- $0 < \lambda < \infty$: coefficients between 0 and that of simple linear regression

1.5.2 Variation in Error

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with λ.

Table 6. Variation of Errors with λ.

Lambda	MAE	MSE	RMSE	R^2 Score
1.00E-15	0.09509	0.01594	0.1262	0.9734
1.00E-10	0.09509	0.01594	0.1262	0.9734
1.00E-08	0.09509	0.01594	0.1262	0.9734

1.00E-05	0.09598	0.01602	0.1265	0.9732
0.0001	0.10254	0.01721	0.1311	0.9709
0.001	0.13401	0.02798	0.1672	0.9507
0.01	0.20569	0.06065	0.2462	0.8655
1	0.72200	0.61580	0.7847	0.0
5	0.72200	0.61580	0.7847	0.0
10	0.72200	0.61580	0.7847	0.0

Source: Notebook

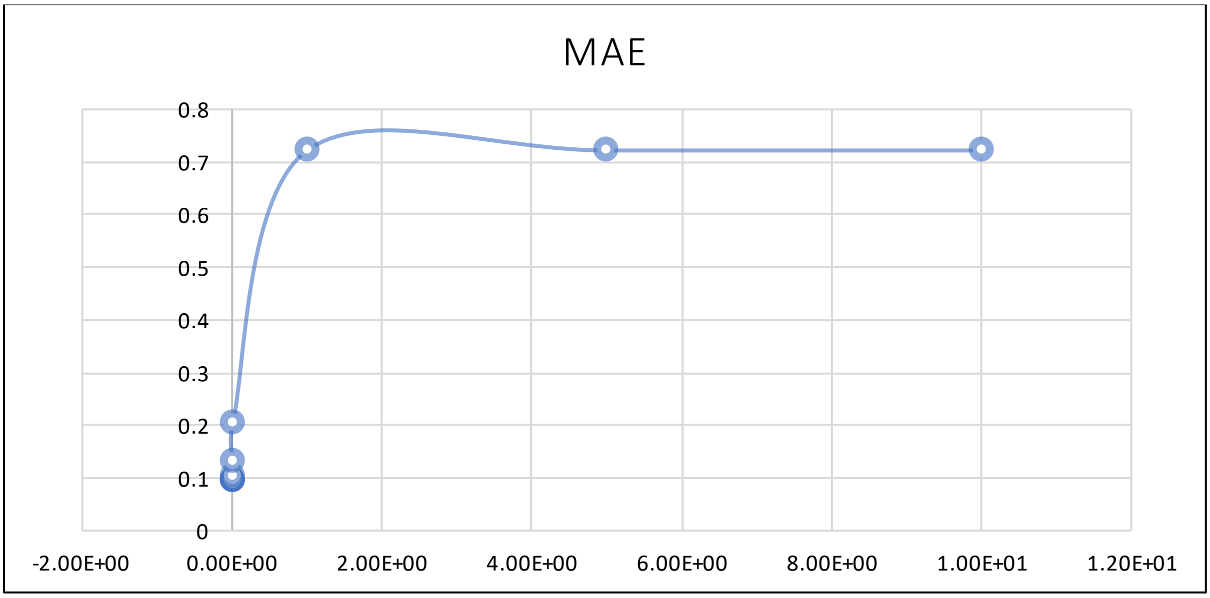

Fig. 10. Variation of Mean Absolute Error (MAE) with λ for Lasso regression

As the graph shows, MAE is increasing with the increase in λ.

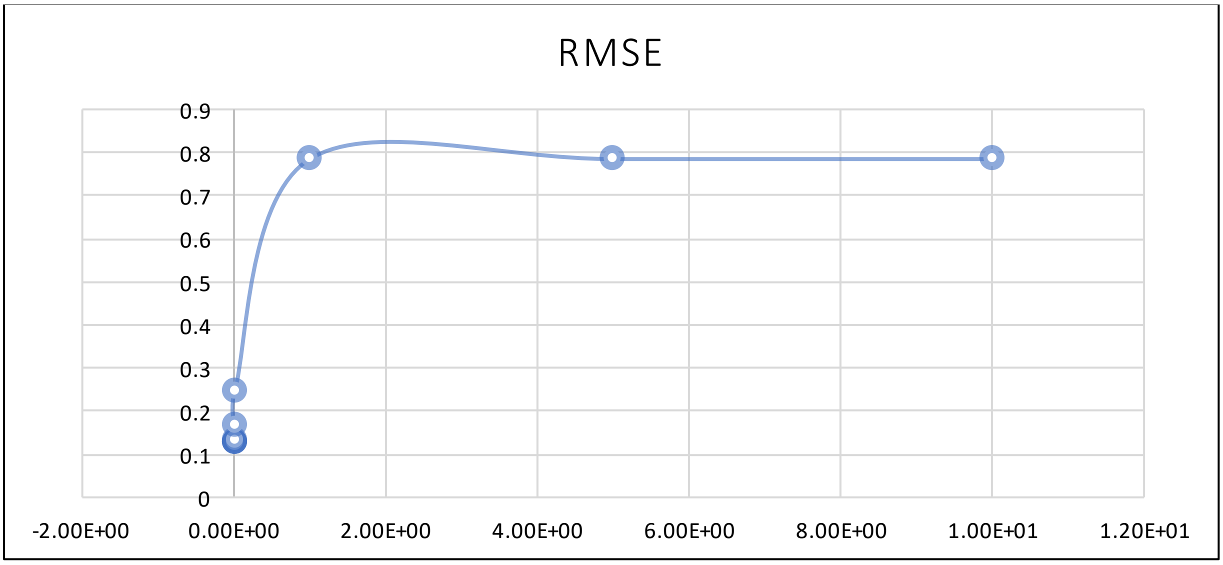

Fig. 11. Variation of Root Mean Absolute Error (RMSE) with λ

As the graph shows, RMSE is increasing with the increase in λ.

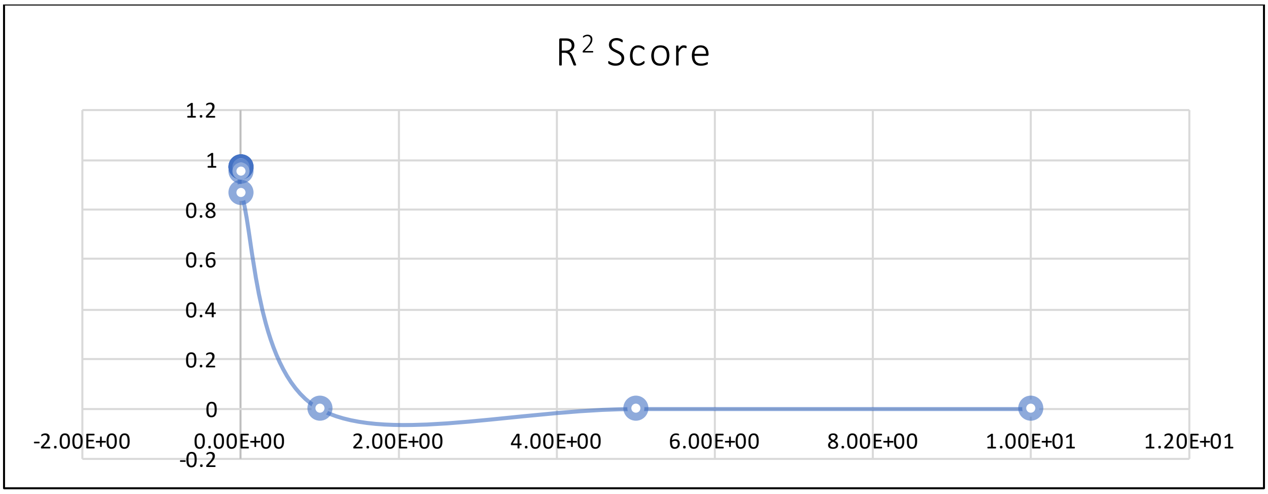

Fig. 12. Variation of R^2 Score with λ

As the graph shows, R^2 is decreasing with the increase in λ.

Model fit with increase in λ can be visualize by following graphs:

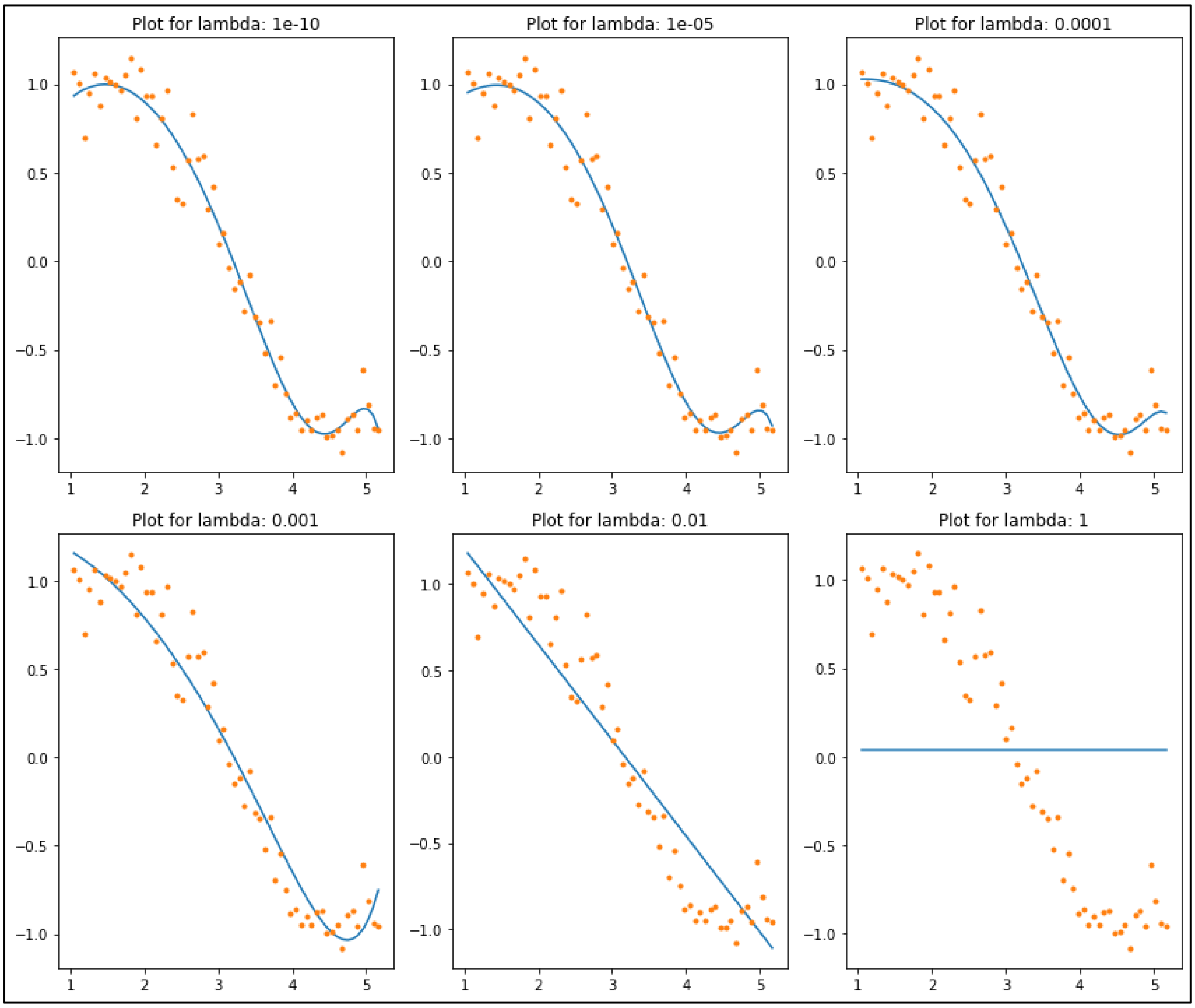

Fig. 13. Effect of increment in λ in Lasso Regression

1.5.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 7. Change in rss, intercept and w with power

	rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15
lambda_1e-15	0.96	0.22	1.1	-0.37	0.00089	0.0016	-0.00012	-6.40E-05	-6.30E-06	1.40E-06	7.80E-07	2.10E-07	4.00E-08	5.40E-09	1.80E-10	-2.00E-10	-9.20E-11
lambda_1e-10	0.96	0.22	1.1	-0.37	0.00088	0.0016	-0.00012	-6.40E-05	-6.30E-06	1.40E-06	7.80E-07	2.10E-07	4.00E-08	5.40E-09	1.80E-10	-2.00E-10	-9.20E-11
lambda_1e-08	0.96	0.22	1.1	-0.37	0.00077	0.0016	-0.00011	-6.40E-05	-6.30E-06	1.40E-06	7.80E-07	2.10E-07	4.00E-08	5.30E-09	2.00E-10	-1.90E-10	-9.30E-11
lambda_1e-05	0.96	0.5	0.6	-0.13	-0.038	0	0	0	0	7.70E-06	1.00E-06	7.70E-08	0	0	0	0	-7.00E-11
lambda_0.0001	1	0.9	0.17	0	-0.048	0	0	0	0	9.50E-06	5.10E-07	0	0	0	0	0	-4.40E-11
lambda_0.001	1.7	1.3	0	-0.13	0	0	0	0	0	0	0	0	1.50E-08	7.50E-10	0	0	0
lambda_0.01	3.6	1.8	-0.55	-0.00056	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_1	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_5	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_10	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Source: Notebook

- For the same values of lambda, the coefficients of lasso regression are much smaller as compared to that of ridge regression.
- For the same lambda, lasso has higher RSS (poorer fit) as compared to ridge regression
- Many of the coefficients are zero even for very small values of lambda

1.6 Elastic Net Regression

1.6.1 Brief Introduction

Elastic net is a popular type of regularized linear regression that combines two popular penalties, specifically the L1 and L2 penalty functions.

*Objective = RSS + λ * (sum of absolute value of coefficients) + λ *((sum of square of coefficients)*

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}|\hat{\beta}_j|),$$

where α is the mixing parameter between ridge (α = 0) and lasso (α = 1).

1.6.2 Variation in Error

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with λ.

Table 8. Variation of Errors with λ.

Lambda	MAE	MSE	RMSE	R2 Score
1.00E-15	0.09509	0.01594	0.1262	0.9734
1.00E-10	0.09509	0.01594	0.1262	0.9734

1.00E-08	0.09510	0.01594	0.1262	0.9734
1.00E-05	0.09693	0.01616	0.1271	0.9729
0.0001	0.11305	0.02027	0.1423	0.9654
0.001	0.14019	0.03017	0.1737	0.9459
0.01	0.23089	0.07095	0.2663	0.8263
1	0.72200	0.61580	0.7847	0.0
5	0.72200	0.61580	0.7847	0.0
10	0.72200	0.61580	0.7847	0.0

Source: Notebook

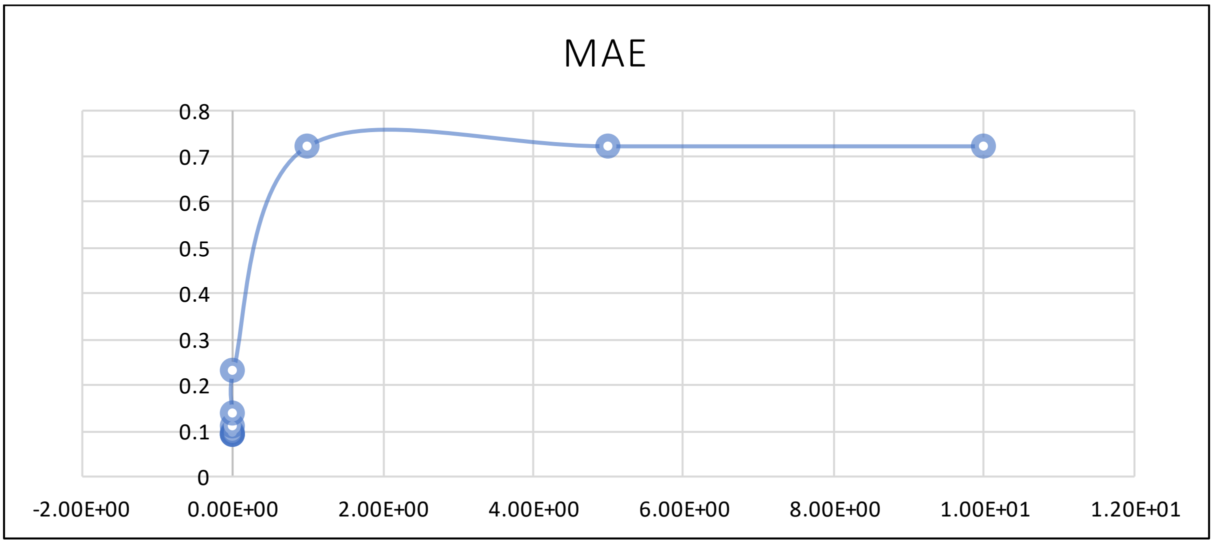

Fig. 14. Variation of Mean Absolute Error (MAE) with λ for Elastic Net regression

As the graph shows, MAE is increasing with the increase in λ.

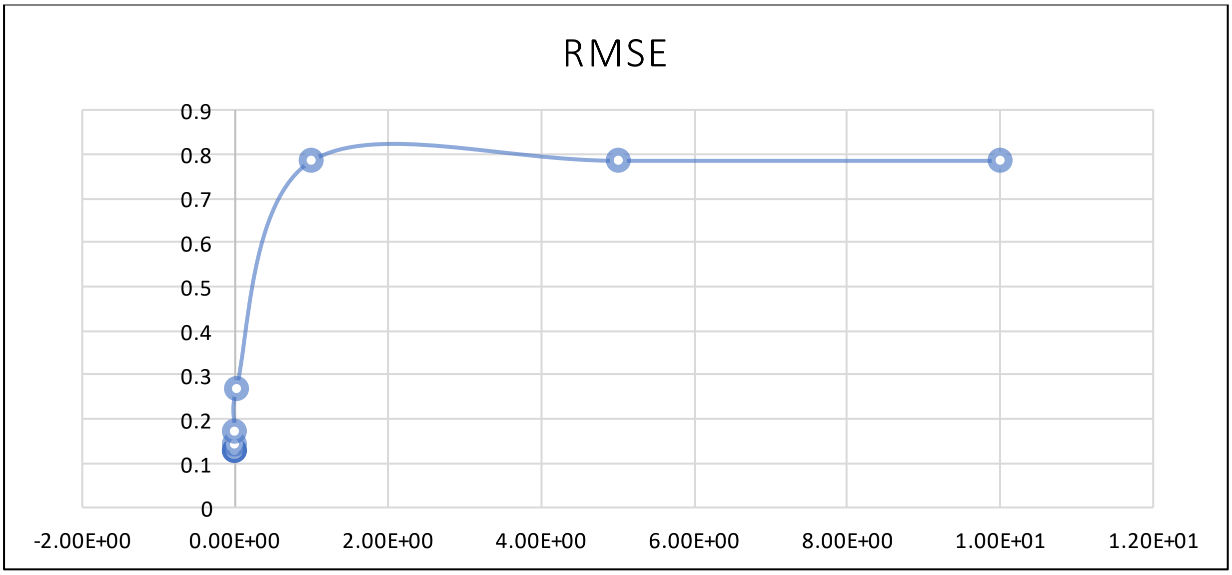

Fig. 15. Variation of Root Mean Absolute Error (RMSE) with λ

As the graph shows, RMSE is increasing with the increase in λ.

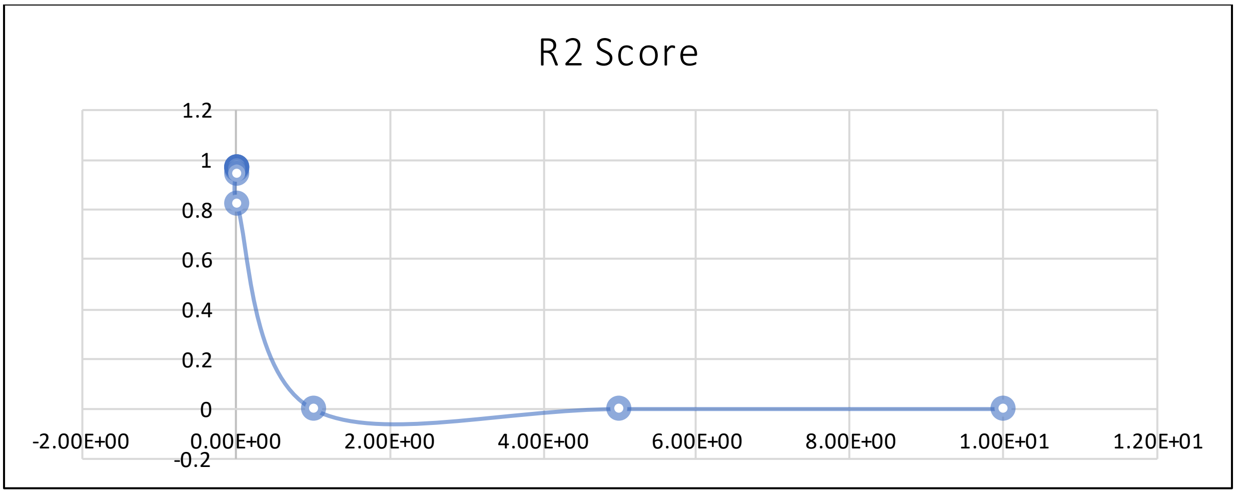

Fig. 16. Variation of R^2 Score with λ

As the graph shows, R^2 is decreasing with the increase in λ.

Model fit with increase in λ can be visualize by following graphs:

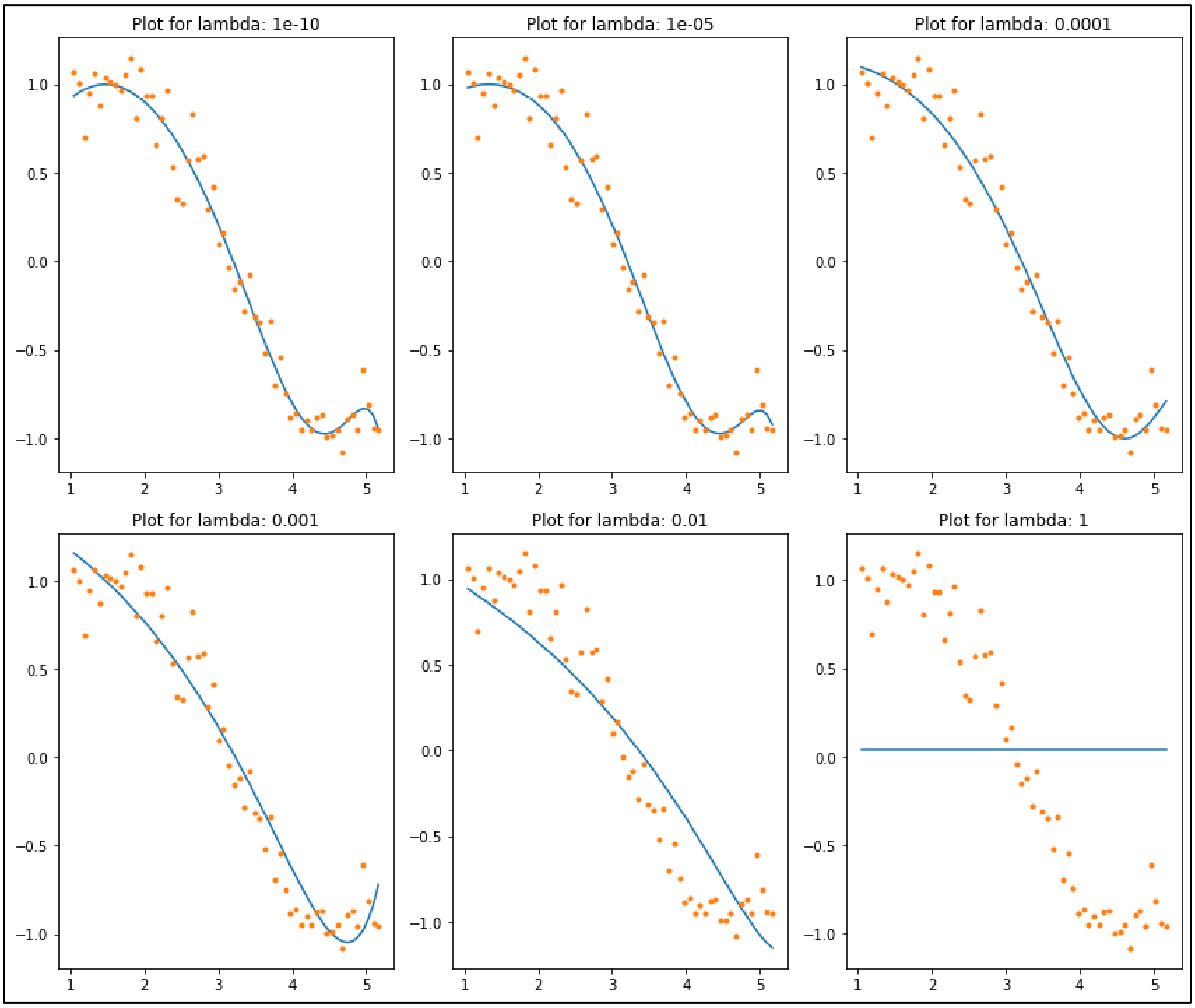

Fig. 13. Effect of increment in λ in Elastic Net Regression

1.5.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 9. Change in rss, intercept and w with λ

rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15	
lambda_1e-15	0.96	0.22	1.1	-0.37	0.00089	0.0016	-0.00012	-6.40E-05	-6.30E-06	1.40E-06	7.80E-07	2.10E-07	4.00E-08	5.40E-09	1.80E-10	-2.00E-10	-9.20E-11
lambda_1e-10	0.96	0.22	1.1	-0.37	0.00087	0.0016	-0.00012	-6.40E-05	-6.30E-06	1.40E-06	7.80E-07	2.10E-07	4.00E-08	5.40E-09	1.80E-10	-2.00E-10	-9.20E-11
lambda_1e-08	0.96	0.23	1.1	-0.36	-0.00035	0.0016	-0.0001	-6.30E-05	-6.20E-06	1.40E-06	7.80E-07	2.10E-07	4.00E-08	5.30E-09	1.90E-10	-1.90E-10	-9.20E-11
lambda_1e-05	0.97	0.67	0.44	-0.1	-0.025	-0.003	-9.40E-05	1.50E-05	1.40E-05	4.00E-06	8.10E-07	1.40E-07	1.70E-08	7.20E-10	0	-1.30E-10	-6.10E-11
lambda_0.0001	1.2	1.1	0.06	-0.063	-0.015	-0.002	-7.80E-05	0	6.40E-06	2.20E-06	5.00E-07	8.60E-08	1.10E-08	3.70E-10	0	-4.20E-11	-3.00E-11
lambda_0.001	1.8	1.4	-0.22	-0.042	-0.0064	-0.00069	-1.20E-05	0	0	0	5.90E-08	2.00E-08	4.50E-09	8.60E-10	1.40E-10	1.70E-11	1.00E-12
lambda_0.01	4.3	1.2	-0.21	-0.027	-0.0035	-0.00035	-5.40E-06	0	0	0	0	0	0	0	7.10E-12	7.80E-12	2.60E-12
lambda_1	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_5	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_10	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Source: Notebook

- Increment in value of lambda underfits the data
- For the same lambda, lasso has higher RSS (poorer fit) as compared to ridge regression, and Elastic Net regression
- Many of the coefficients are zero even for very small values of lambda

1.7 RidgeCV

1.7.1 Brief Introduction

RidgeCV implements ridge regression with built-in cross-validation of alpha parameter. It almost works in same way excepts it defaults to Leave-One-Out cross validation.

1.7.2 Variation in Error

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with λ.

Table 10. Variation of Errors with λ.

Lambda	MAE	MSE	RMSE	R2 Score
1.00E-15	5.07438	43.62436	6.60487	-0.20901
1.00E-10	0.09354	0.01537	0.12400	0.97437
1.00E-08	0.09466	0.01582	0.12578	0.97362
0.0001	0.09577	0.01603	0.12662	0.97322
0.001	0.09997	0.01669	0.12921	0.97187
0.01	0.12383	0.02392	0.15469	0.95875
1	0.27020	0.09414	0.30683	0.73228

5	0.43020	0.23063	0.48024	-0.25693
10	0.48781	0.30041	0.54810	-1.68158
20	0.55913	0.37867	0.61536	-5.87431

Source: Notebook

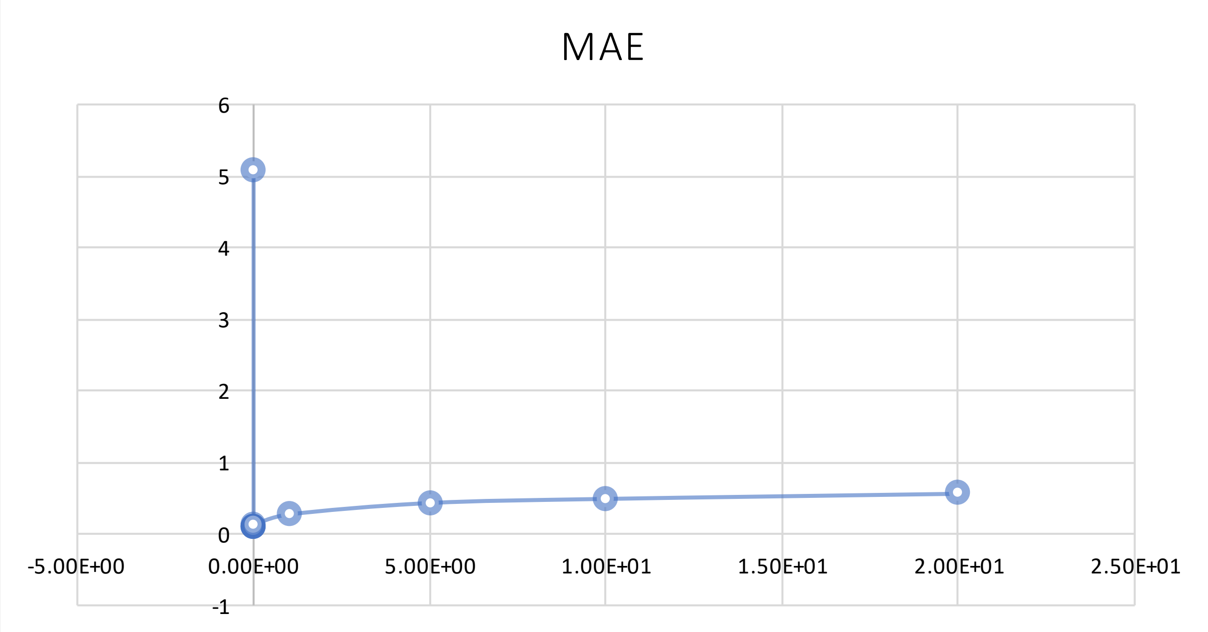

Fig. 14. Variation of Mean Absolute Error (MAE) with λ for RidgeCV

As the graph shows, MAE is first decreased and then increased with the increase in λ.

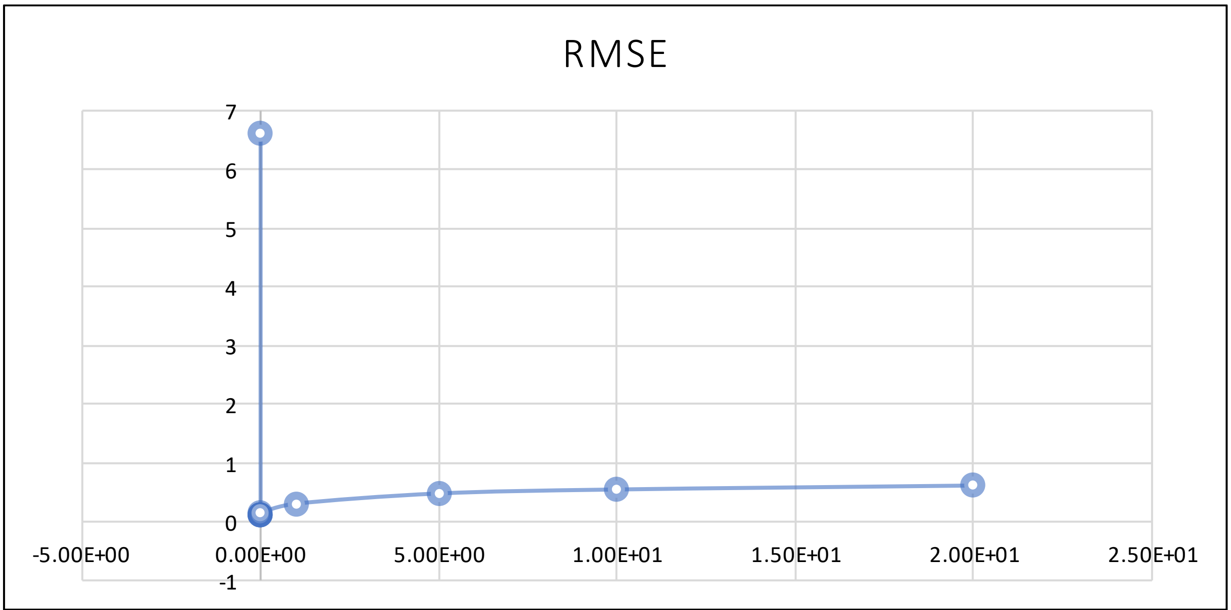

Fig. 15. Variation of Root Mean Absolute Error (RMSE) with λ

As the graph shows, RMSE is first decreased and then increased with the increase in λ.

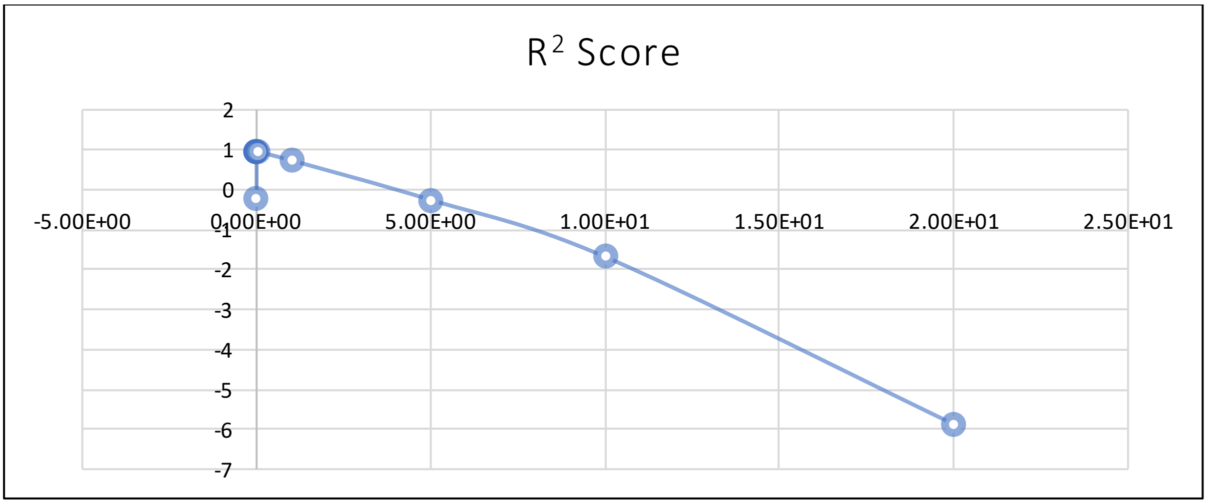

Fig. 12. Variation of R^2 Score with λ

As the graph shows, R^2 is decreasing with the increase in λ.

Model fit with increase in λ can be visualize by following graphs:

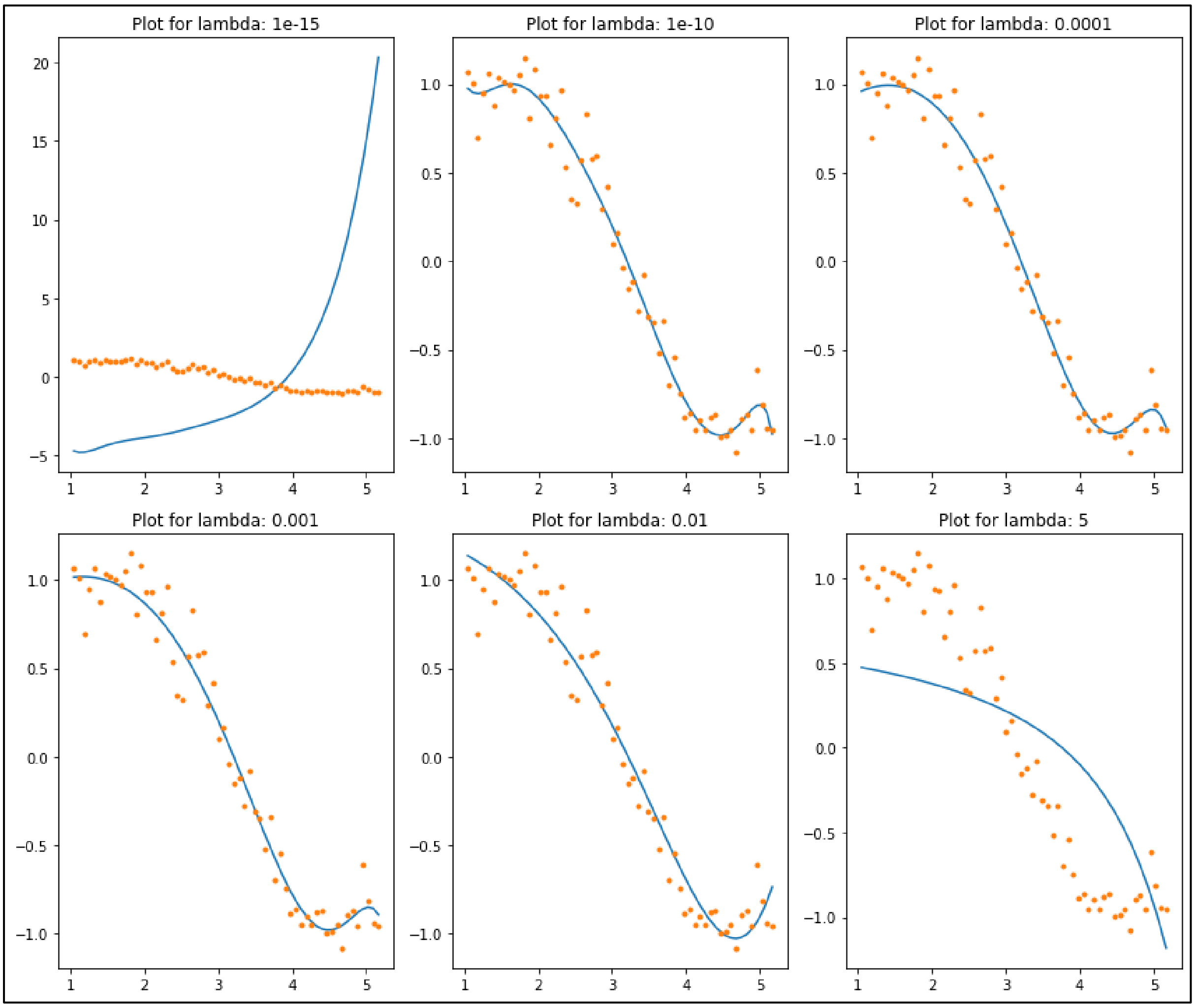

Fig. 13. Effect of increment in λ in RidgeCV

1.7.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 11. Change in rss, intercept and w with λ

	rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15
lambda_1e-15	2.60E+03	88	-3.00E+02	3.80E+02	-2.30E+02	65	0.87	-4.6	0.56	0.16	-0.026	-0.0059	0.00098	0.00019	-5.00E-05	3.30E-06	-4.50E-08
lambda_1e-10	0.92	11	-29	31	-15	2.9	0.17	-0.091	-0.011	0.002	0.00064	2.40E-05	-2.00E-05	-4.20E-06	2.20E-07	2.30E-07	-2.30E-08
lambda_1e-08	0.95	1.3	-1.5	1.7	-0.68	0.039	0.016	0.00016	-0.00036	-5.40E-05	-2.90E-07	1.10E-06	1.90E-07	2.00E-08	3.90E-09	8.20E-10	-4.60E-10
lambda_0.0001	0.96	0.56	0.55	-0.13	-0.026	-0.0028	-0.00011	4.10E-05	1.50E-05	3.70E-06	7.40E-07	1.30E-07	1.90E-08	1.90E-09	-1.30E-10	-1.50E-10	-6.20E-11
lambda_0.001	1	0.82	0.31	-0.087	-0.02	-0.0028	-0.00022	1.80E-05	1.20E-05	3.40E-06	7.30E-07	1.30E-07	1.90E-08	1.70E-09	-1.50E-10	-1.40E-10	-5.20E-11
lambda_0.01	1.4	1.3	-0.088	-0.052	-0.01	-0.0014	-0.00013	7.20E-07	4.10E-06	1.30E-06	3.00E-07	5.60E-08	9.00E-09	1.10E-09	4.30E-11	-3.10E-11	-1.50E-11
lambda_1	5.6	0.97	-0.14	-0.019	-0.003	-0.00047	-7.00E-05	-9.90E-06	-1.30E-06	-1.40E-07	-9.30E-09	1.30E-09	7.80E-10	2.40E-10	6.20E-11	1.40E-11	3.20E-12
lambda_5	14	0.55	-0.059	-0.0085	-0.0014	-0.00024	-4.10E-05	-6.90E-06	-1.10E-06	-1.90E-07	-3.10E-08	-5.10E-09	-8.20E-10	-1.30E-10	-2.00E-11	-3.00E-12	-4.20E-13
lambda_10	18	0.4	-0.037	-0.0055	-0.00095	-0.00017	-3.00E-05	-5.20E-06	-9.20E-07	-1.60E-07	-2.90E-08	-5.10E-09	-9.10E-10	-1.60E-10	-2.90E-11	-5.10E-12	-9.10E-13
lambda_20	23	0.28	-0.022	-0.0034	-0.0006	-0.00011	-2.00E-05	-3.60E-06	-6.60E-07	-1.20E-07	-2.20E-08	-4.00E-09	-7.50E-10	-1.40E-10	-2.50E-11	-4.70E-12	-8.70E-13

Source: Notebook

- For very small lambda / alpha, model is underfitting
- For lambda or alpha > 5, it is also underfitting

1.8 Lasso Least Angle Regression (LassoLARS)

1.7.1 Brief Introduction

Least-angle regression (LARS) is a regression algorithm for high-dimensional data. LARS is similar to forward stepwise regression. At each step, it finds the feature most correlated with the target. When there are multiple features having equal correlation, instead of continuing along the same feature, it proceeds in a direction equiangular between the features.

1.7.2 Variation in Error

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with λ.

Table 12. Variation of Errors with λ.

Lambda	MAE	MSE	RMSE	R2 Score
1.00E-15	0.09493	0.01586	0.1259	0.9735
1.00E-10	0.09493	0.01586	0.1259	0.9735
1.00E-08	0.09493	0.01586	0.1259	0.9735
0.0001	0.10257	0.01725	0.1313	0.9708
0.001	0.13360	0.02788	0.1670	0.9509

0.01	0.20572	0.06065	0.2462	0.8655
1	0.72200	0.61580	0.7847	0.0
5	0.72200	0.61580	0.7847	0.0
10	0.72200	0.61580	0.7847	0.0
20	0.72200	0.61580	0.7847	0.0

Source: Notebook

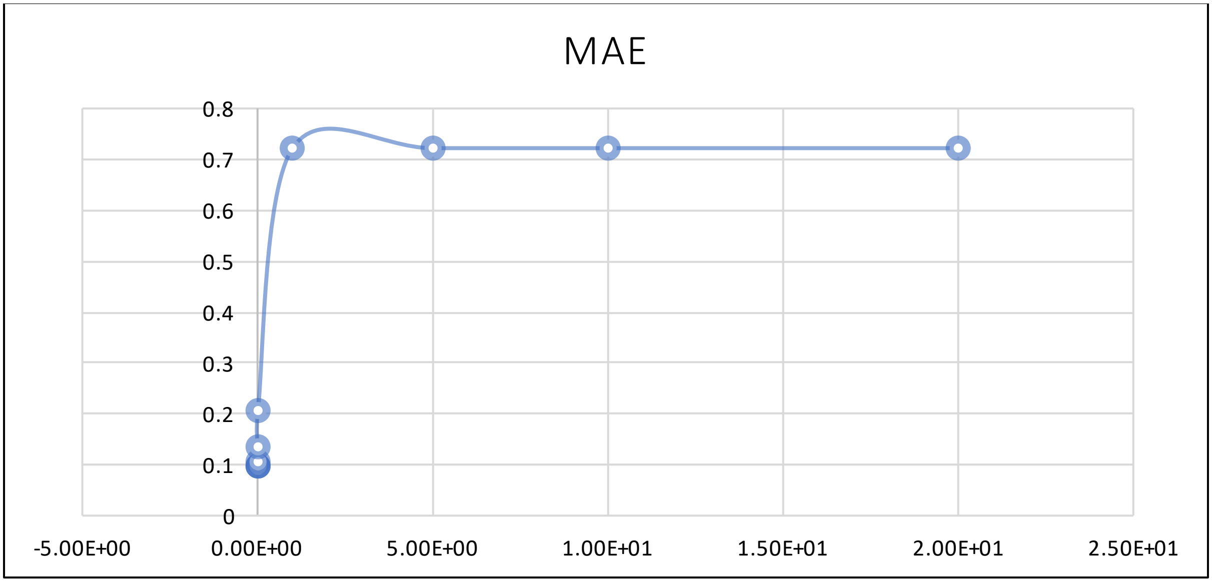

Fig. 14. Variation of Mean Absolute Error (MAE) with λ for LassoLARS

As the graph shows, MAE is increased with the increase in λ.

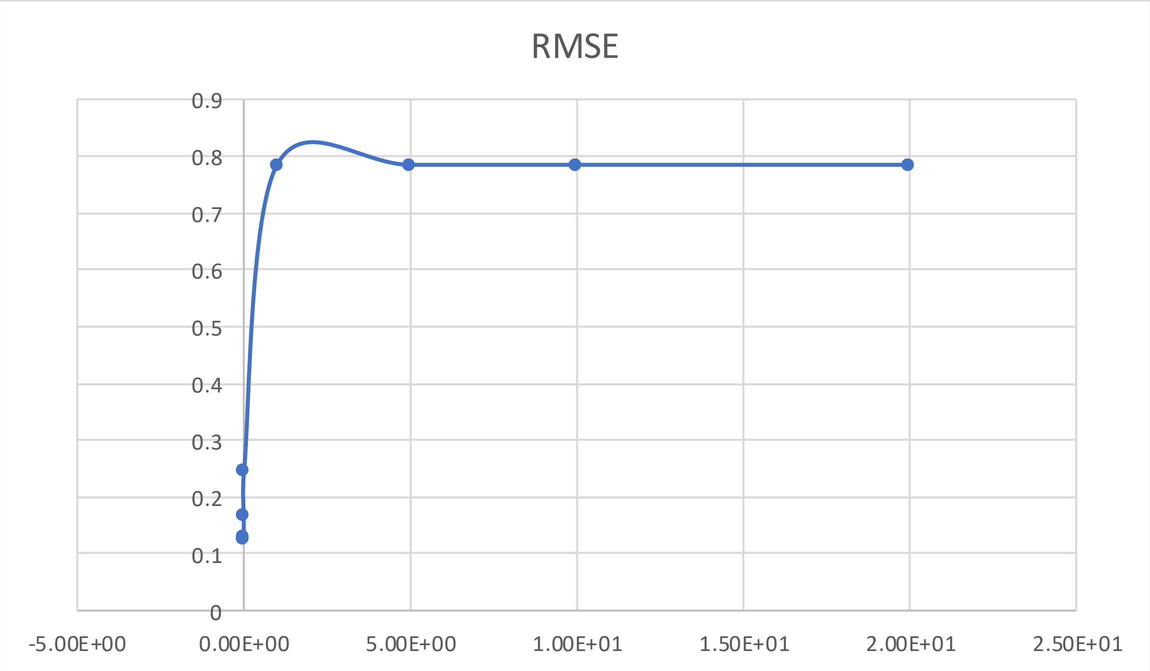

Fig. 15. Variation of Root Mean Absolute Error (RMSE) with λ

As the graph shows, RMSE increased with the increase in λ.

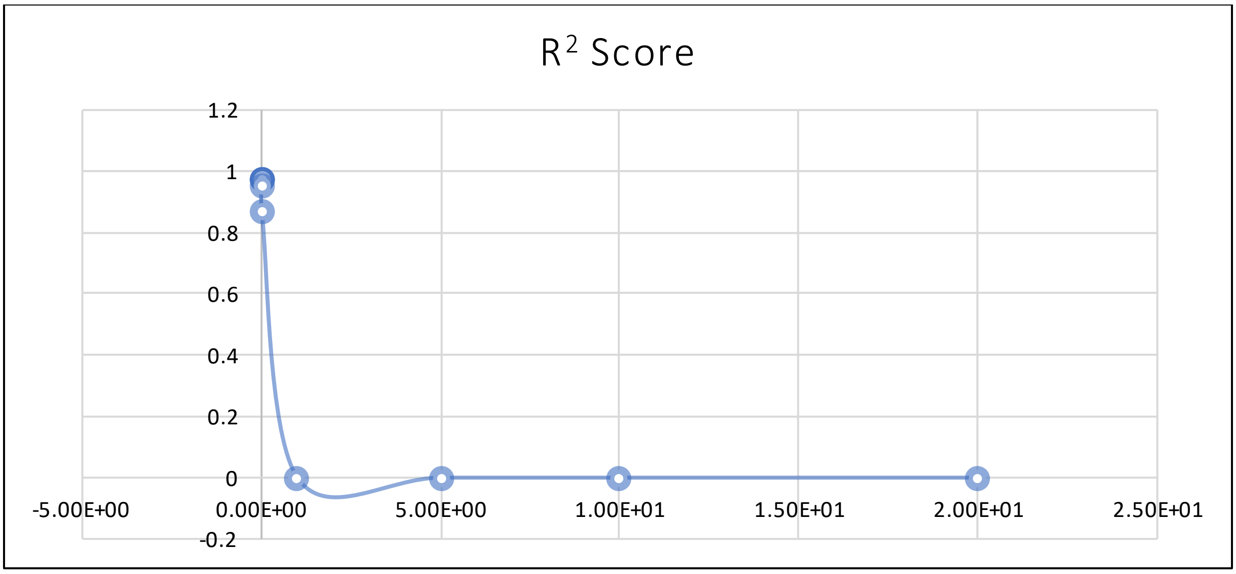

Fig. 16. Variation of R^2 Score with λ

As the graph shows, R^2 is decreasing with the increase in λ.

Model fit with increase in λ can be visualize by following graphs:

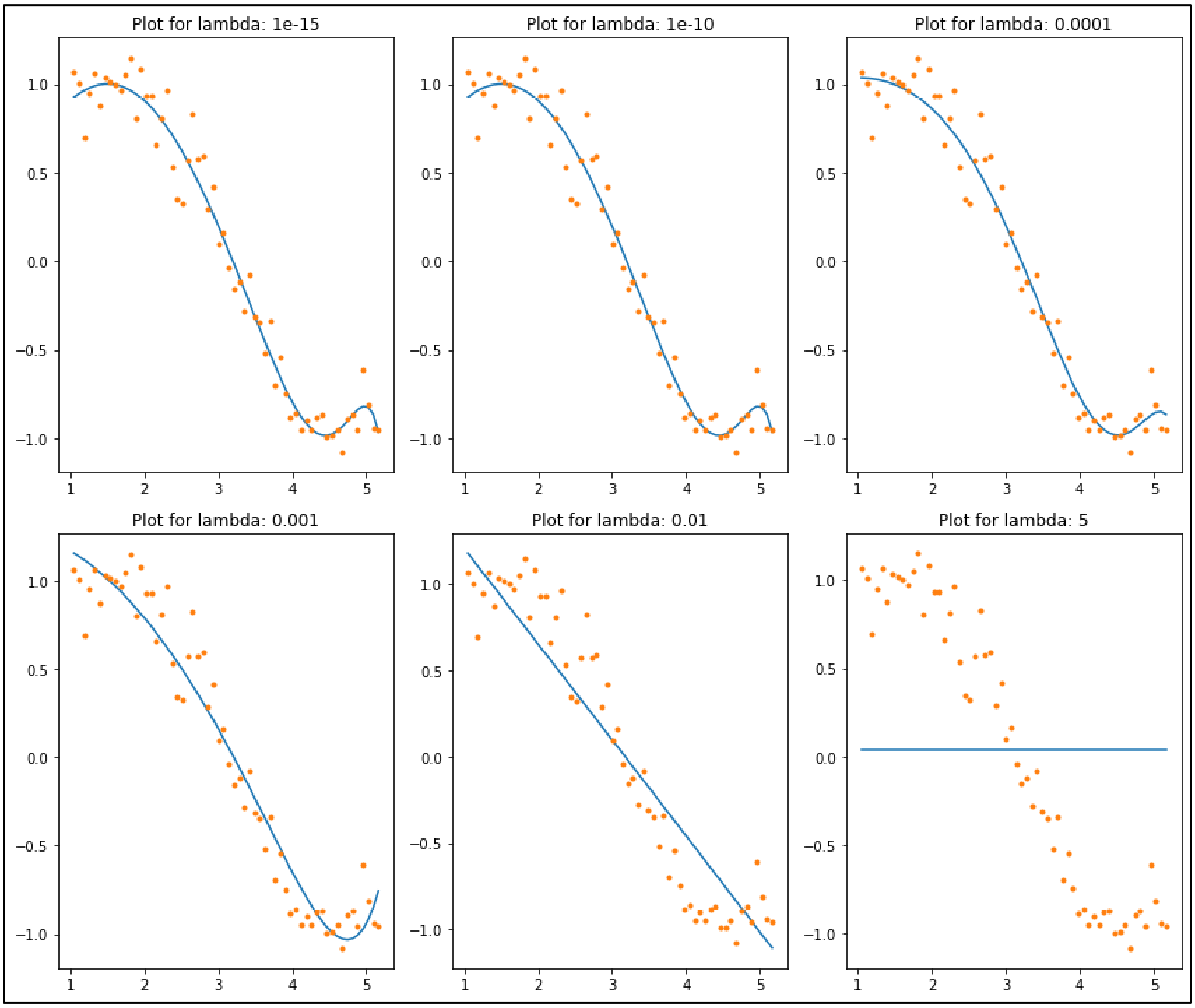

Fig. 17. Effect of increment in λ in LassoLARS

1.7.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 13. Change in rss, intercept and w with λ

	rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15
lambda_1e-15	0.95	0.36	0.67	0	-0.11	0	0.0041	0	0	0	-5.40E-06	0	0	0	1.70E-08	0	-4.50E-10
lambda_1e-10	0.95	0.36	0.67	0	-0.11	0	0.0041	0	0	0	-5.40E-06	0	0	0	1.70E-08	0	-4.50E-10
lambda_1e-08	0.95	0.36	0.67	0	-0.11	0	0.0041	0	0	0	-5.40E-06	0	0	0	1.70E-08	0	-4.50E-10
lambda_0.0001	1	0.95	0.13	0	-0.044	0	0	0	0	0	2.40E-06	0	0	0	0	0	-5.70E-11
lambda_0.001	1.7	1.3	0	-0.13	0	0	0	0	0	0	0	0	1.90E-08	0	0	0	0
lambda_0.01	3.6	1.8	-0.55	-0.00084	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_1	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_5	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_10	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lambda_20	37	0.038	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Source: Notebook

As value of lambda(alpha) increases, model is underfitting i.e., bias is increasing.

1.9 Hubber Regressor

1.9.1 Brief Introduction

The Huber Regressor is different to Ridge because it applies a linear loss to samples that are classified as outliers. A sample is classified as an inlier if the absolute error of that sample is lesser than a certain threshold.

$$\min_{w,\sigma} \sum_{i=1}^{n} \left(\sigma + H_{\epsilon} \left(\frac{X_i w - y_i}{\sigma} \right) \sigma \right) + \alpha {||w||_2}^2$$

1.9.2 Variation in Error

MAE, MSE, RMSE and R^2 score has been calculated for each index to find out the trend of errors with λ.

Table 14. Variation of Errors with λ.

Lambda	MAE	MSE	RMSE	R2 Score
1.00E-15	0.43125	0.34314	0.58578	-1.15179
1.00E-10	0.43125	0.34314	0.58578	-1.15179
1.00E-08	0.43125	0.34314	0.58578	-1.15179
0.0001	0.43125	0.34314	0.58578	-1.15179
0.001	0.43125	0.34314	0.58578	-1.15179

0.01	0.43125	0.34314	0.58578	-1.15179
1	0.43125	0.34314	0.58578	-1.15179
5	0.43125	0.34314	0.58578	-1.15179
10	0.43125	0.34314	0.58578	-1.15179
20	0.43125	0.34314	0.58578	-1.15179

Source: Notebook

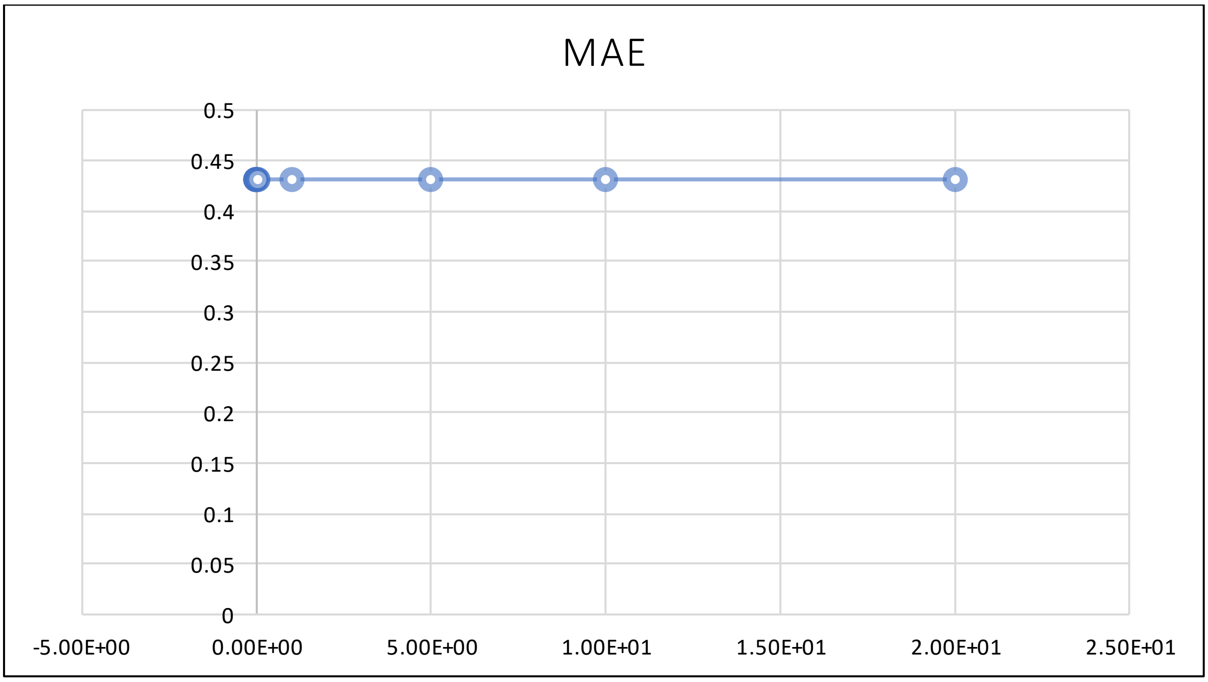

Fig. 18. Variation of Mean Absolute Error (MAE) with λ for Hubber Regressor

As the graph shows, MAE shows no change.

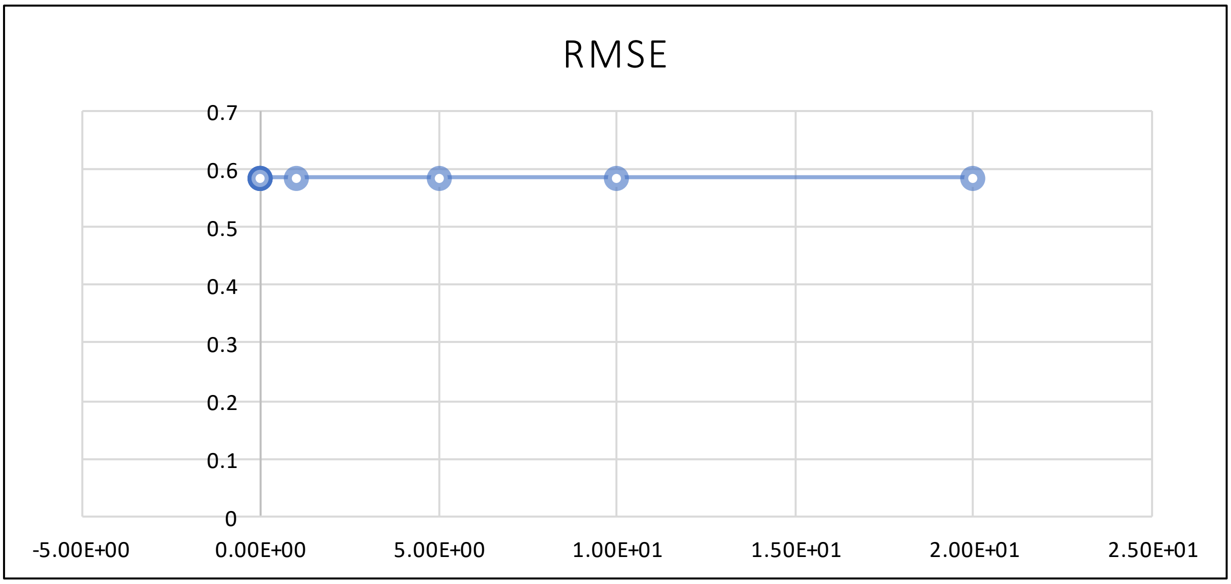

Fig. 19. Variation of Root Mean Absolute Error (RMSE) with λ

As the graph shows, RMSE shows no change.

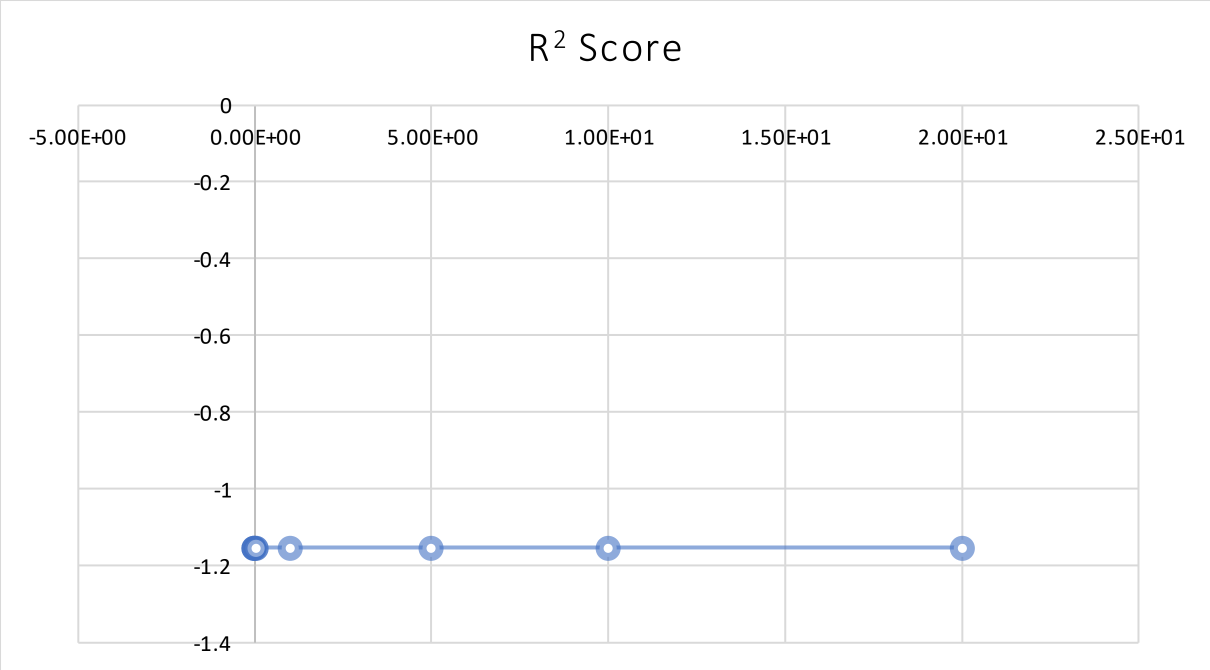

Fig. 20. Variation of R^2 Score with λ

As the graph shows, R^2 is decreasing with the increase in λ.

Model fit with increase in λ can be visualize by following graphs:

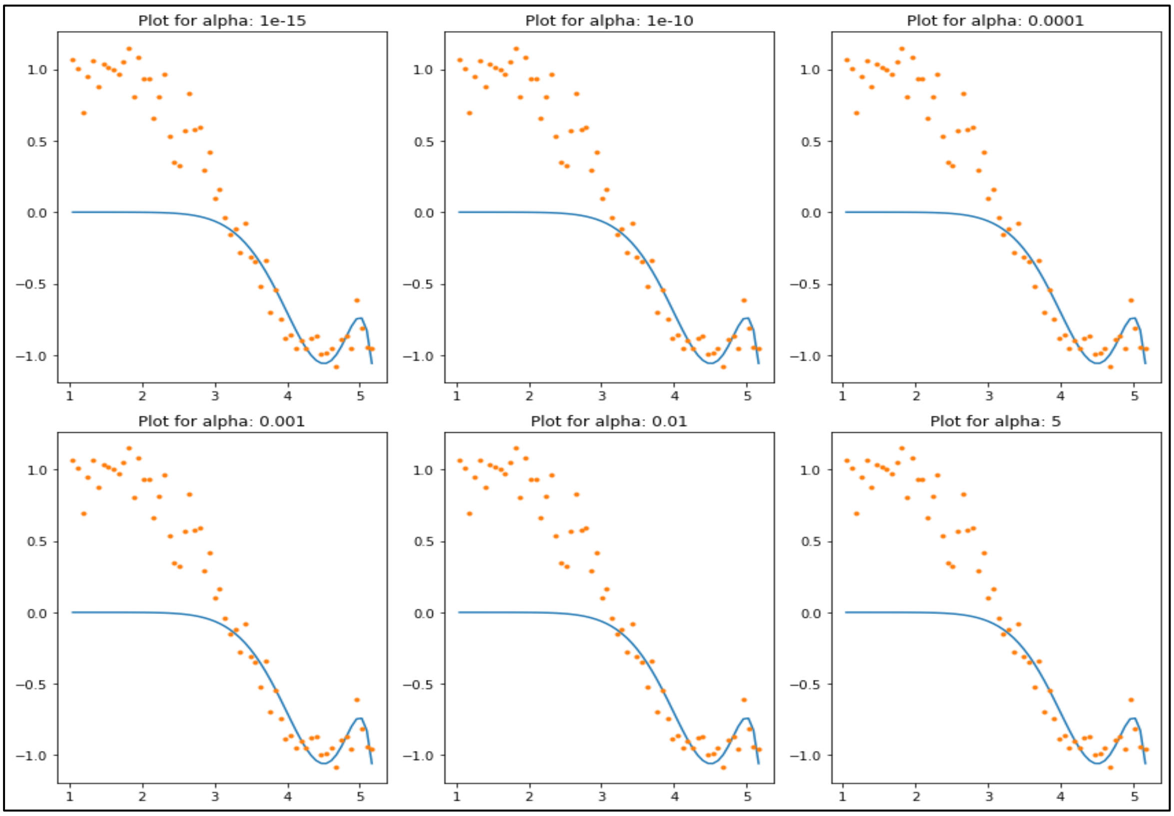

Fig. 21. Effect of increment in λ in Hubber Regressor

1.9.3 Variation in w

Changes is w or coefficients of x can be seen in the table below:

Table 15. Change in rss, intercept and w with λ

rss	intercept	coef_x_1	coef_x_2	coef_x_3	coef_x_4	coef_x_5	coef_x_6	coef_x_7	coef_x_8	coef_x_9	coef_x_10	coef_x_11	coef_x_12	coef_x_13	coef_x_14	coef_x_15	
lambda_1e-15	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_1e-10	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_1e-08	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_0.0001	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_0.001	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_0.01	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_1	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_5	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_10	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09
lambda_20	21	2.00E-13	2.70E-13	1.60E-13	-9.30E-13	-6.40E-12	-3.00E-11	-1.20E-10	-4.40E-10	-1.60E-09	-5.20E-09	-1.60E-08	-4.40E-08	-9.80E-08	-1.40E-07	6.70E-08	-7.00E-09

Source: Notebook

As value of lambda(alpha) increases, model shows no change.