CS5691 : Programming Assignment 1

Linear Regression using Different Basis Functions

Team 14 :

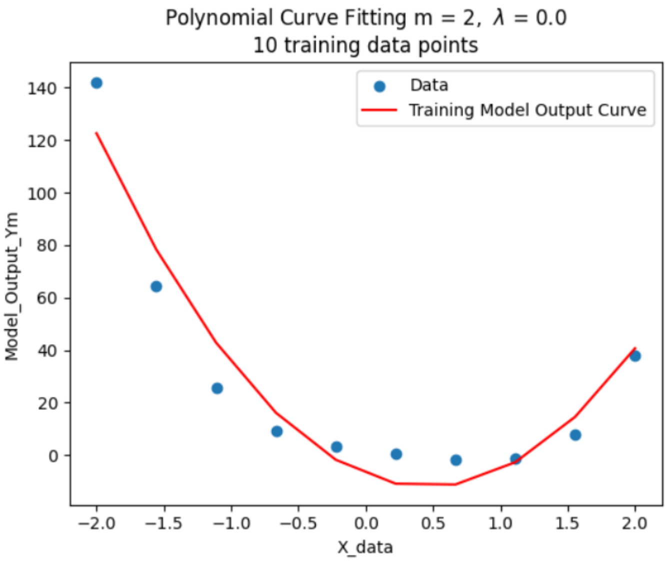
ME20B043	ME20B072	ME20B075
Balakumar R	Girish Madhavan V	Gopalakrishnan T V

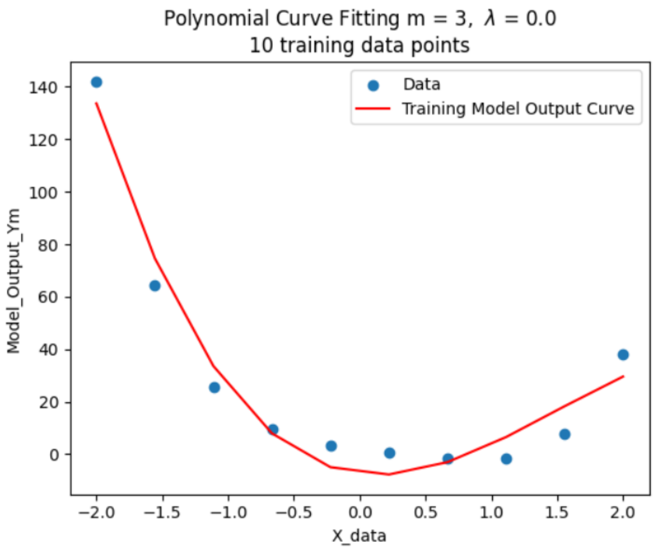
Task 1:

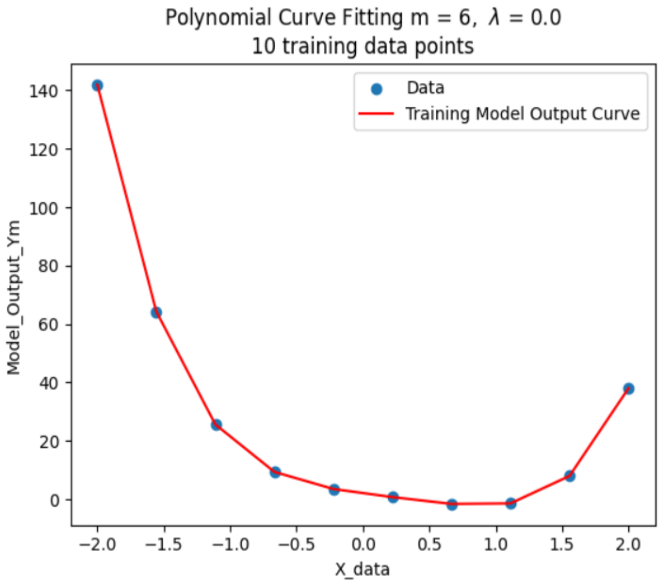
Polynomial curve fitting for Dataset 1

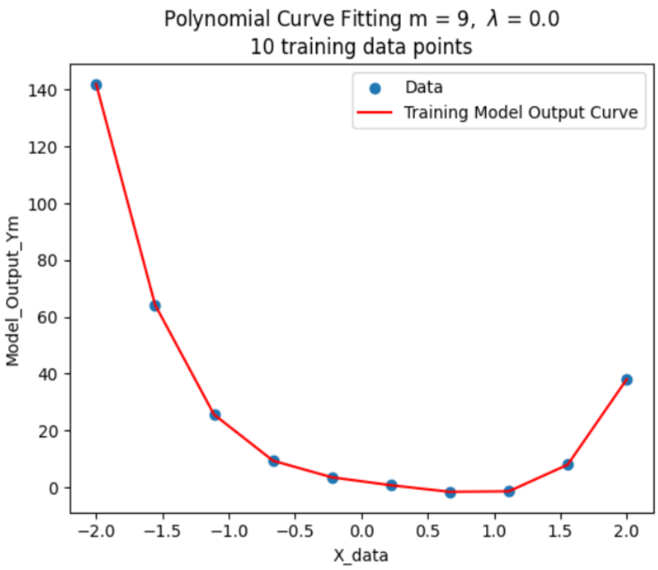
Approximated functions obtained using training datasets of different sizes (10 and 100) for different model complexities (degrees 2, 3, 6 and 9) and different values of λ.

Without regularization: 10 training points





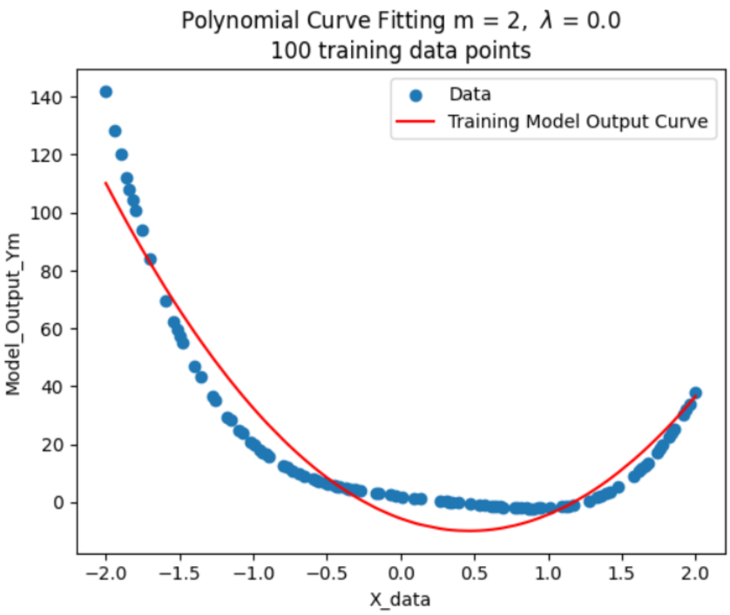


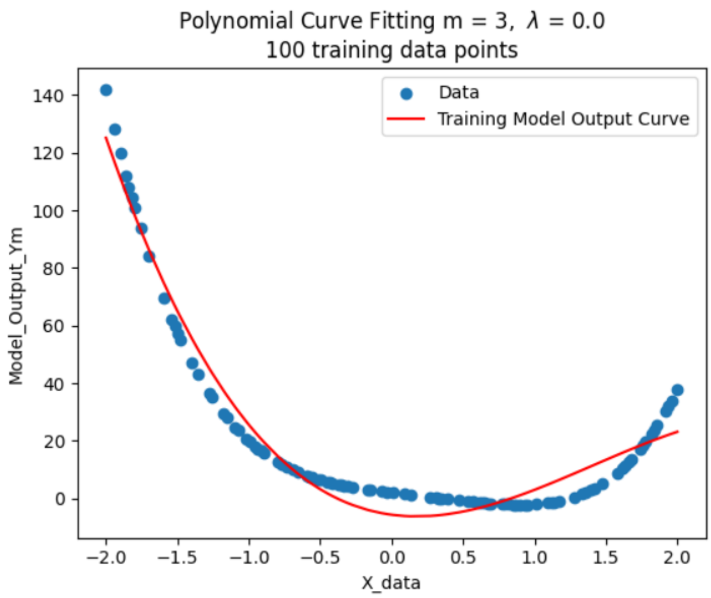


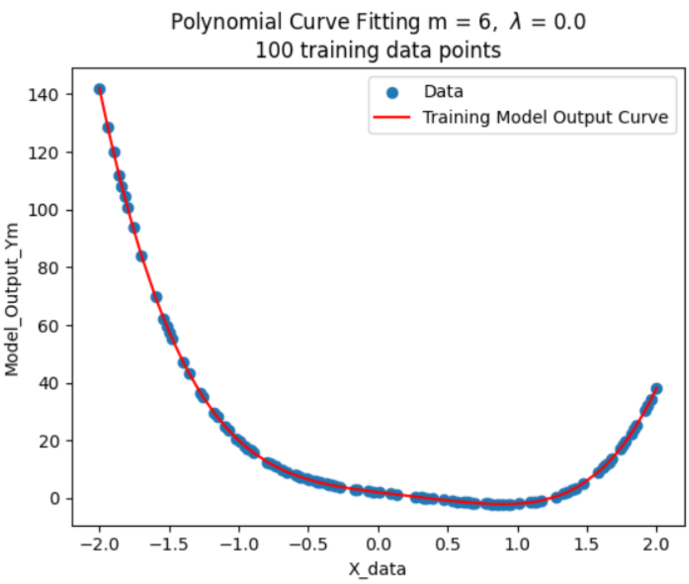
Inferences: (Size-10, no regularization):

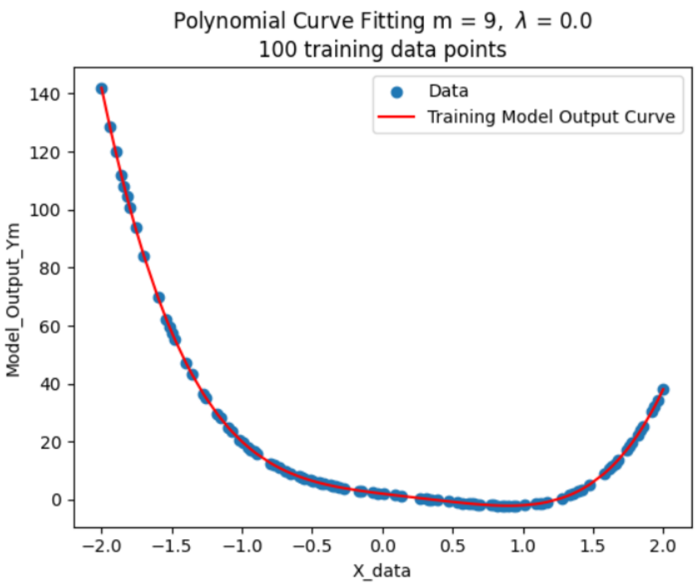
1. **Inference 1: Degree 2 Polynomial Fit**
    a. A polynomial of degree 2 exhibits low complexity and is insufficient to accurately represent the training data. It fails to capture the underlying patterns in the data, resulting in a poor fit to the training points.
2. **Inference 2: Degree 3 Polynomial Fit**
    a. Increasing the polynomial degree to 3 significantly improves the model's performance. However, it still doesn't pass through every training point, indicating that there are some deviations in the data that a third-degree polynomial cannot capture perfectly.
3. **Inference 3: Degree 6 Polynomial Fit**
    a. The use of a sixth-degree polynomial provides an excellent fit to the training data. It passes through all the training points and closely matches the underlying curve. This suggests that a polynomial of this degree is capable of modelling the data effectively.
4. **Inference 4: Degree 9 Polynomial Fit**
    a. Employing a polynomial of degree 9 results in a curve that passes through all training points, seemingly providing a perfect fit. However, it's essential to be cautious about overfitting, as using such high complexity models (degree 9 in this case) with only 10 training points may lead to an overly intricate model that doesn't generalize well to new, unseen data.

Without regularization: 100 training points





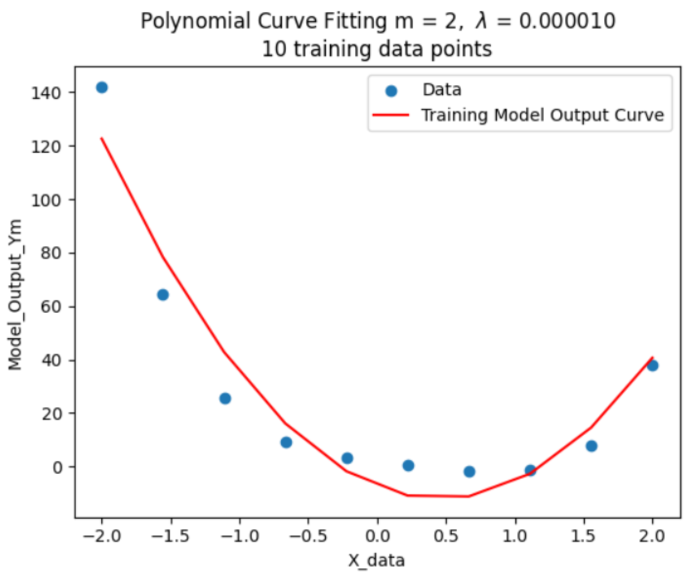


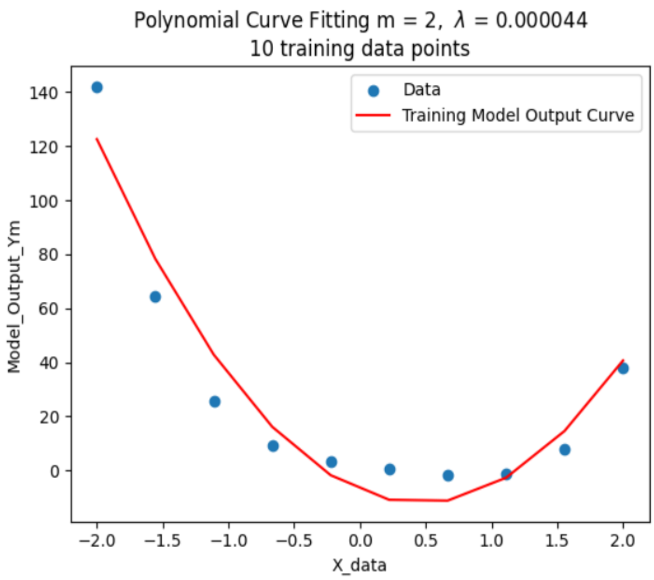


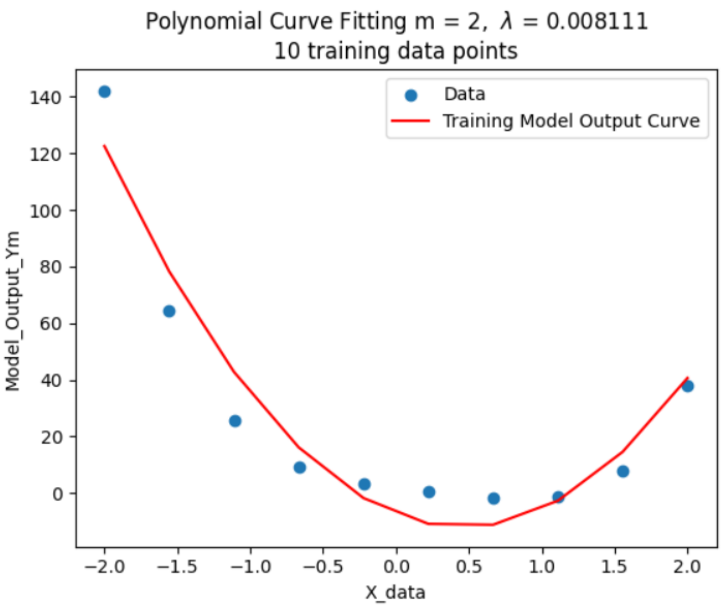
Inferences: (Size-10, no regularization):

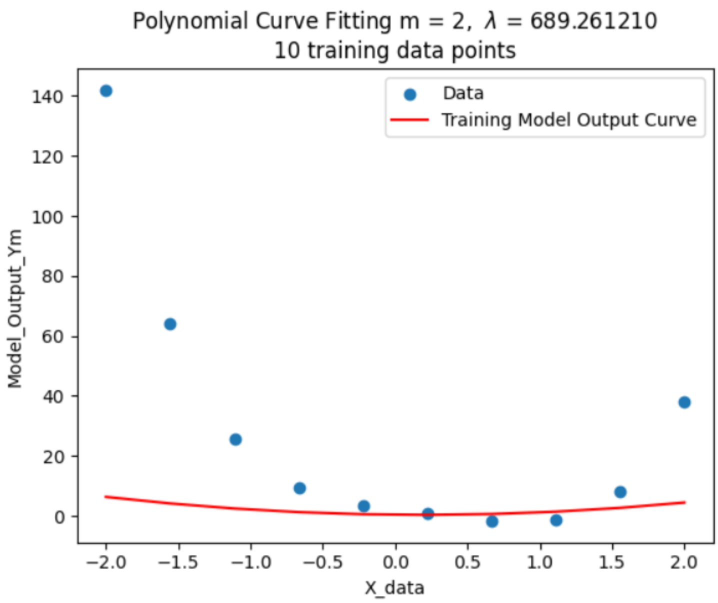
1. **Inference 1: Degree 2 Polynomial Fit**
   - Similar to the previous scenario, a degree-2 polynomial model struggles to approximate the training data effectively. It lacks the necessary complexity to capture the underlying patterns.
2. **Inference 2: Degree 3 Polynomial Fit**
   - Once again, the degree-3 polynomial model delivers a reasonably good performance, although it doesn't provide a perfect fit. It falls short of capturing all the intricacies in the data but still offers a reasonable approximation.
3. **Inference 3: Degree 6 Polynomial Fit**
   - The use of a degree-6 polynomial model results in an outstanding fit to the training data. It passes through all the training points and closely follows the underlying curve. This demonstrates the model's capacity to capture complex relationships in the data effectively.
4. **Inference 4: Degree 9 Polynomial Fit**
   - Surprisingly, even with a high degree of complexity (degree 9), the model performs exceptionally well in this scenario. It not only fits the training data accurately but also avoids overfitting. This is attributed to the ample amount of training data provided, which enables the model to resemble the performance of the degree-6 model while accommodating more intricate patterns.

With regularization: 10 training points





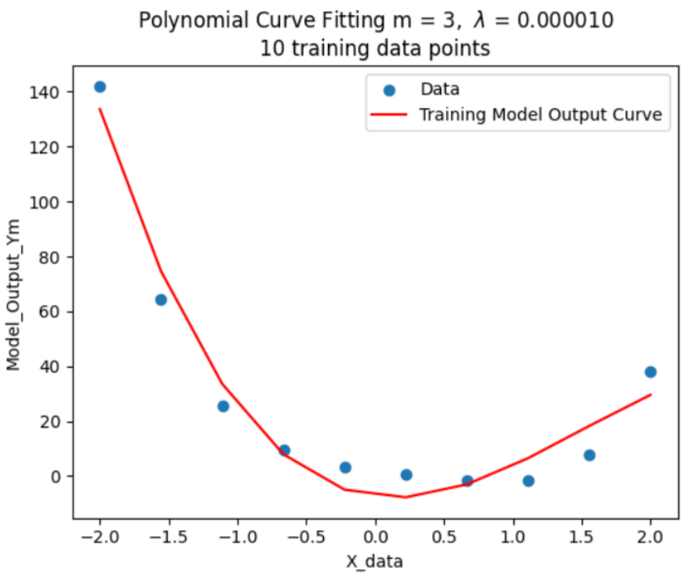


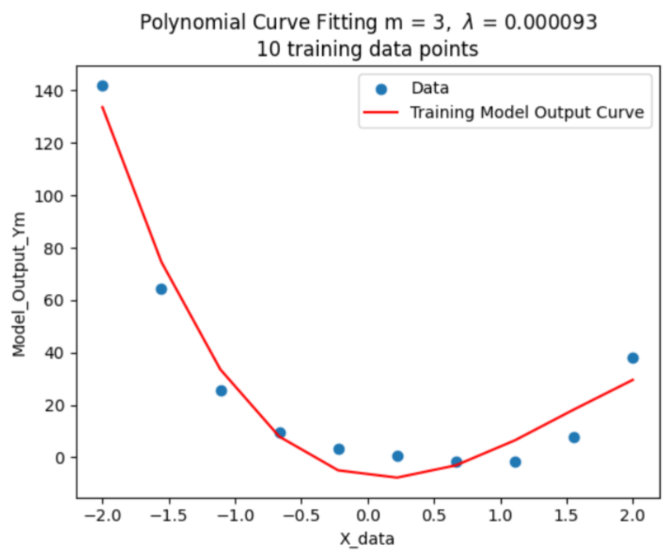


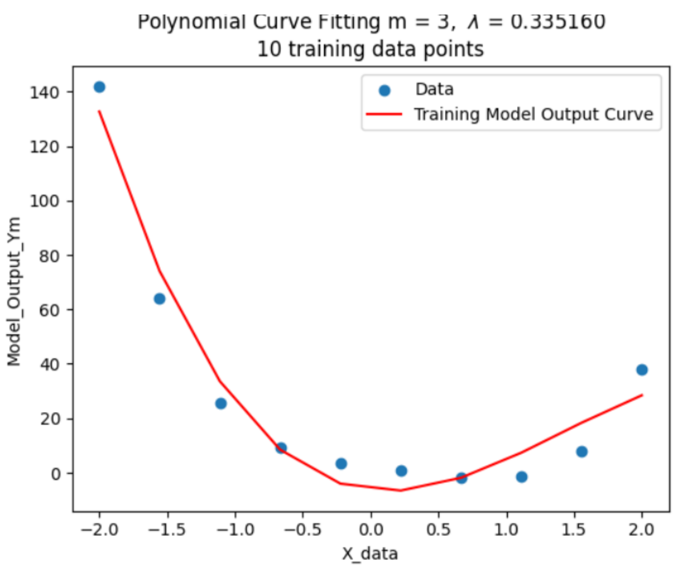
Small Lambda results in negligible regularization, making the regularized curve nearly identical to the non-regularized curve.

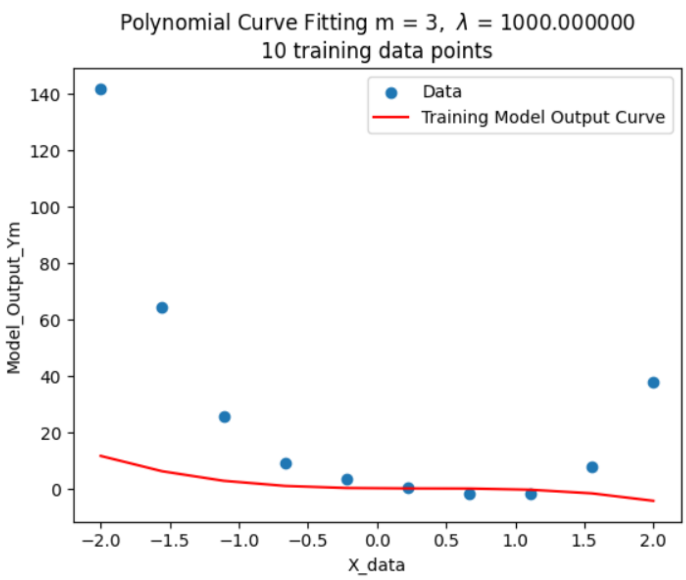
Large Lambda significantly affects regularization, causing noticeable deviations between the regularized and non-regularized curves. This helps prevent overfitting but may lead to underfitting.

Regularization alone can't compensate for overly simple models or small training datasets. To fit training points accurately, a balance between regularization, model complexity, and dataset size is crucial

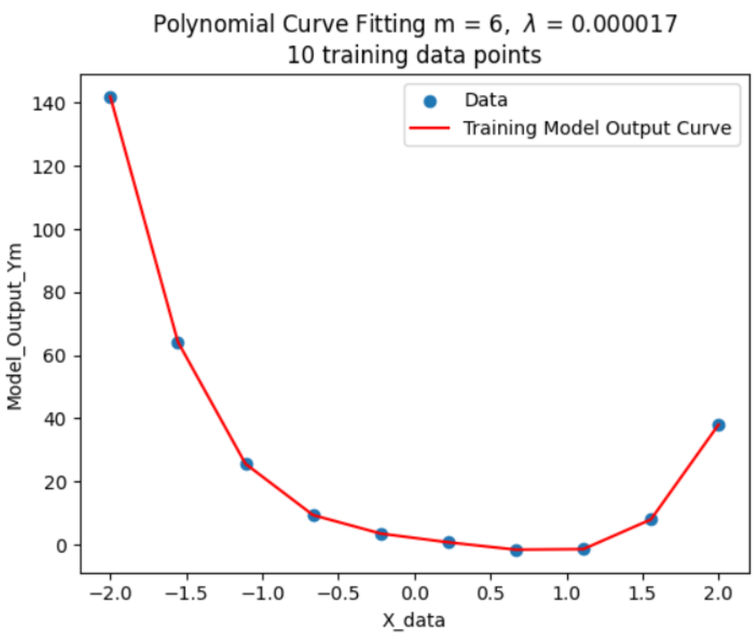


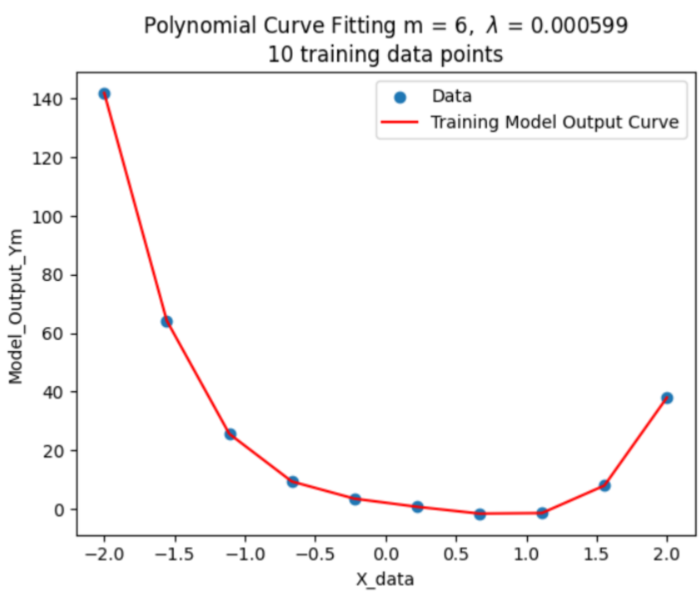


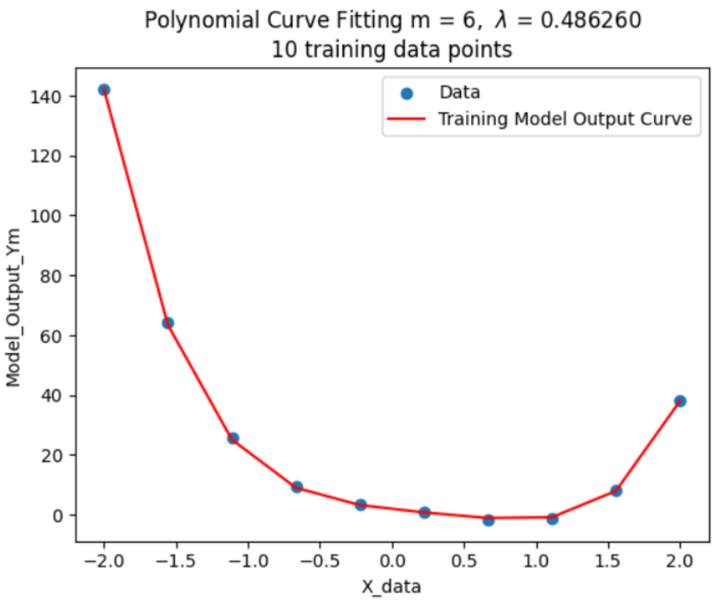


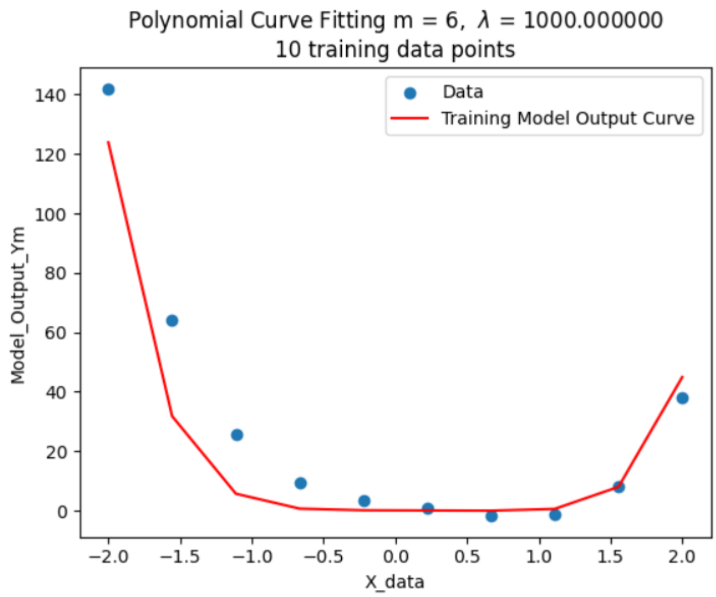


Small Lambda values have minimal regularization impact, while very large Lambda values lead to underfitting, causing a loss of valuable training data information.





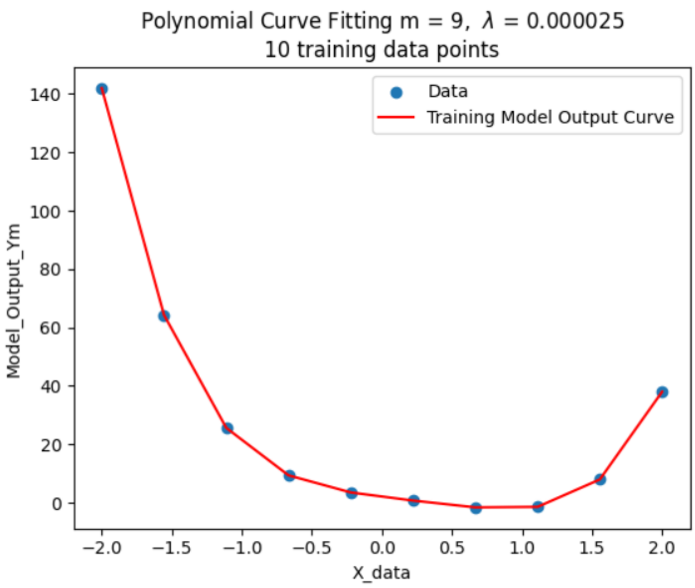


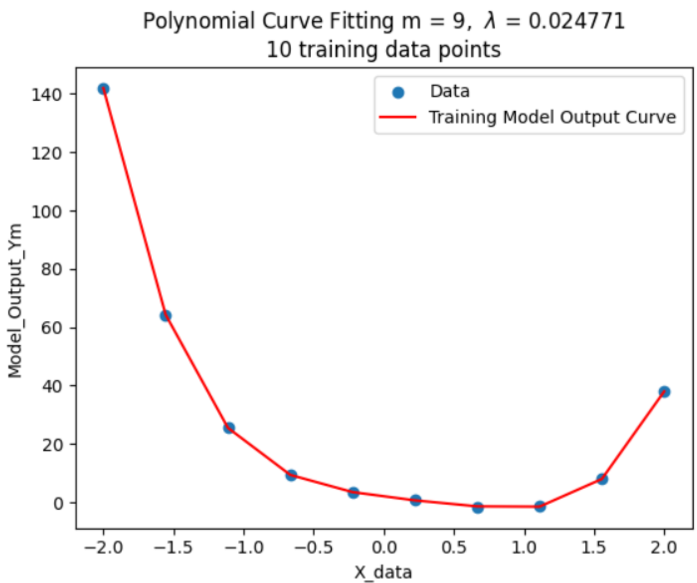


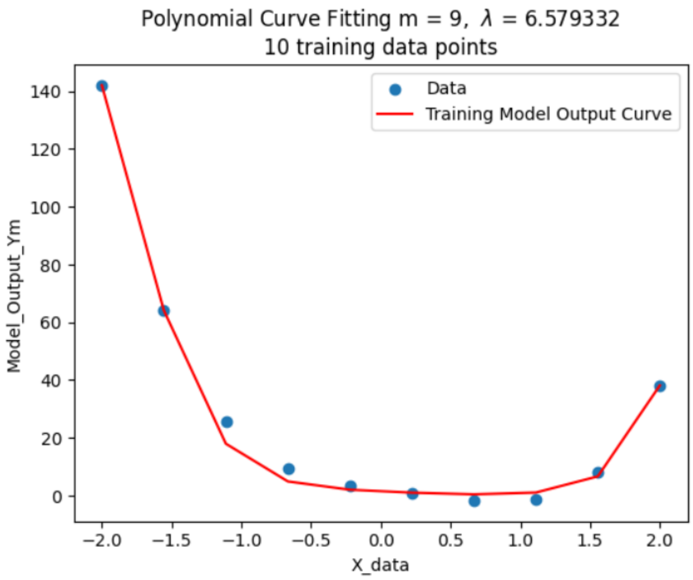
Degree 6 appears to be the optimal degree for the curve, fitting the training data perfectly without any regularization.

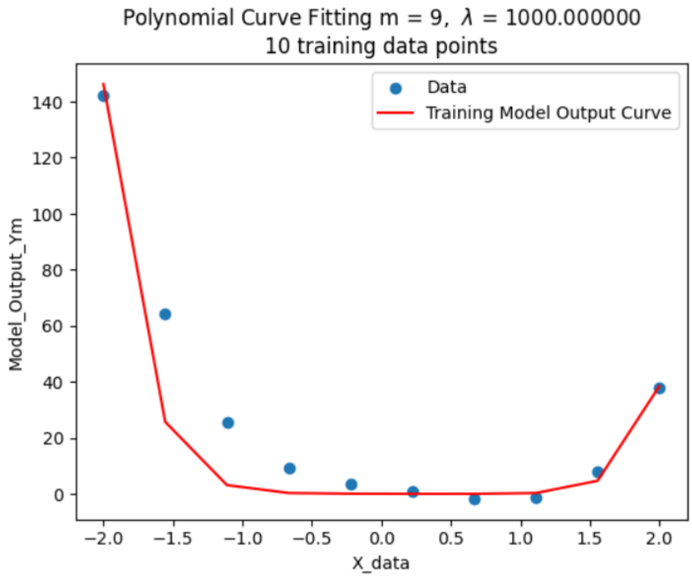
When using moderate and very low lambda values, the curve remains essentially unchanged. This suggests that the non-regularized curve is smooth and exhibits low training errors.

However, when a very large lambda is applied, information loss occurs, and the curve under fits the training points.









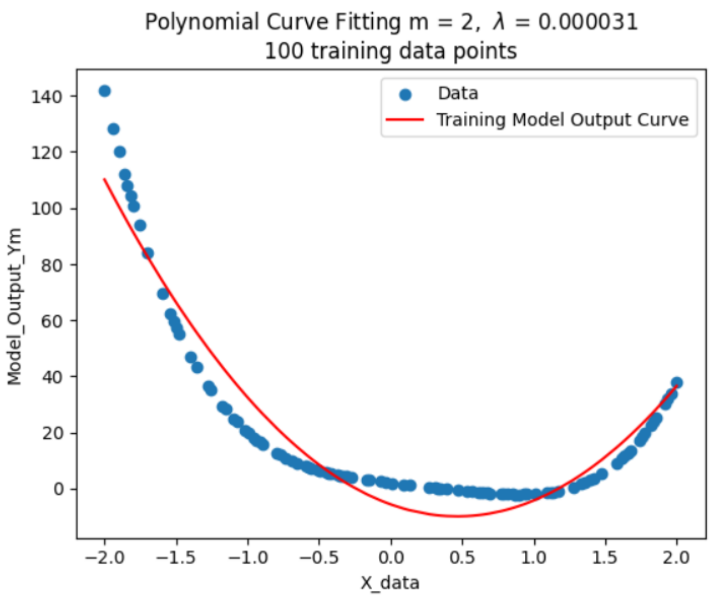
This case highlights the critical role of regularization, especially when dealing with limited training data. Complex models often tend to overfit, making regularization essential.

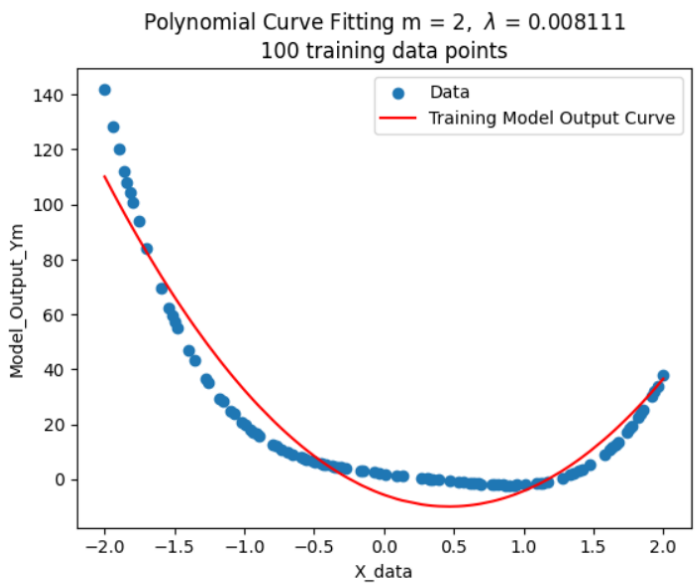
Regularization penalizes the fluctuations or roughness of the fitted curve. In this scenario, the regularized curve is notably smoother than the non-regularized counterparts.

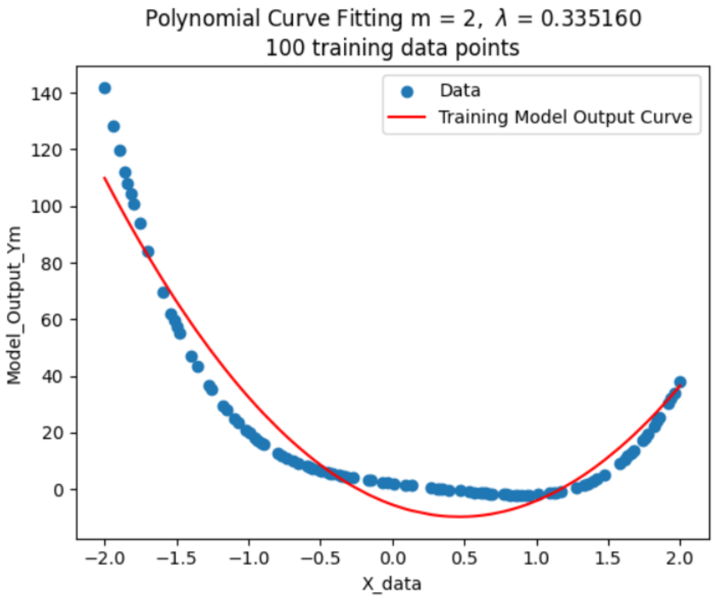
The value of the regularization parameter (Lambda) determines the quality of curve fitting. Optimal Lambda values, result in a curve that fits training points perfectly while remaining smooth and offering good generalization.

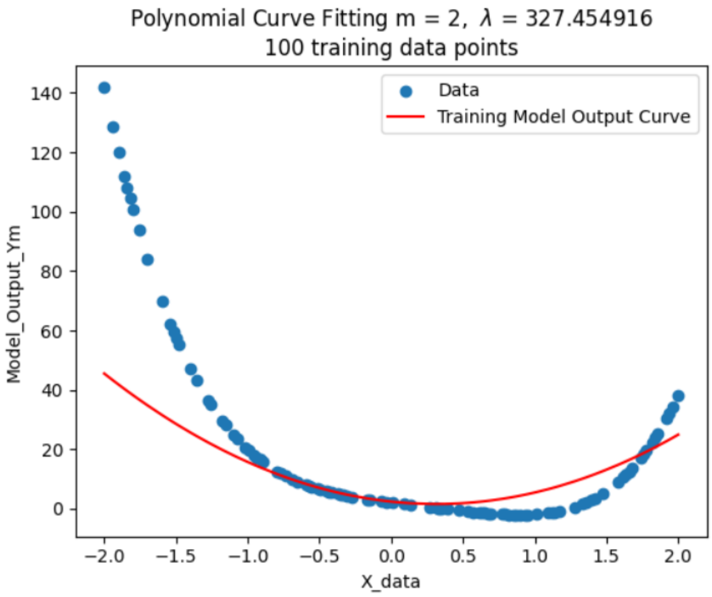
A very low Lambda value leads to overfitting, as it places little importance on regularization. Extremely high Lambda values cause underfitting, where the curve doesn't even pass through all training points.

With regularization: 100 training points







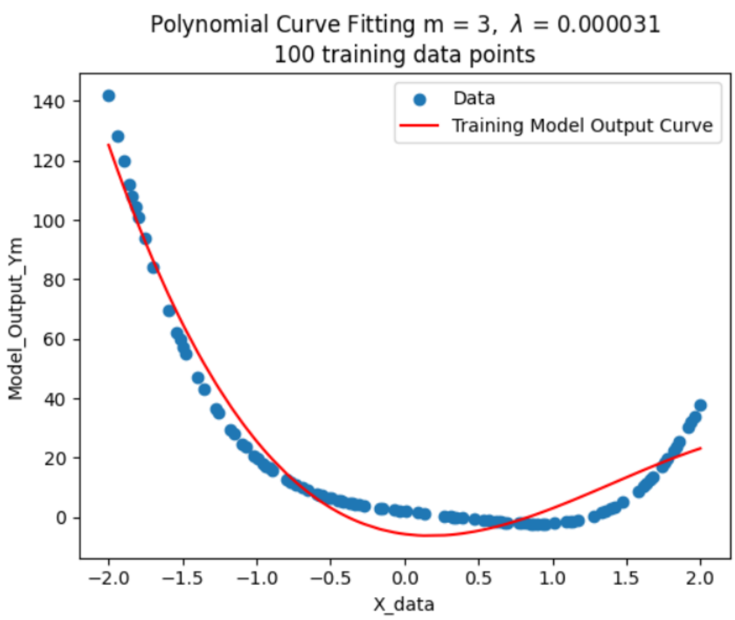


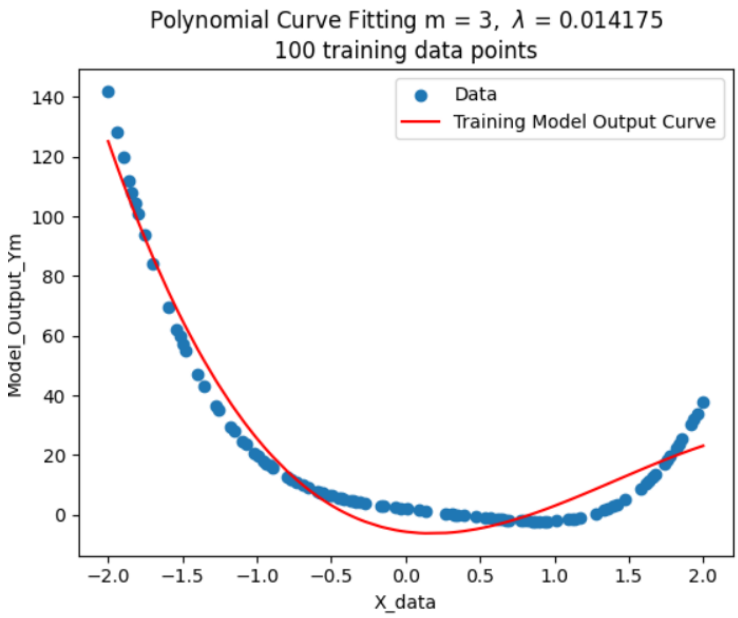
When dealing with 200 training points, there is ample data to effectively fit a model with a degree of around 20.

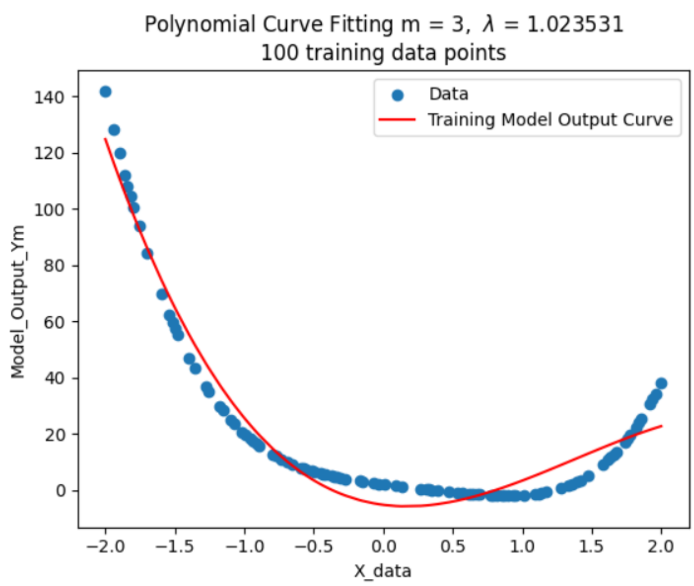
Opting for a low-complexity model, like degree=2, underutilised the training data and doesn't harness its full potential.

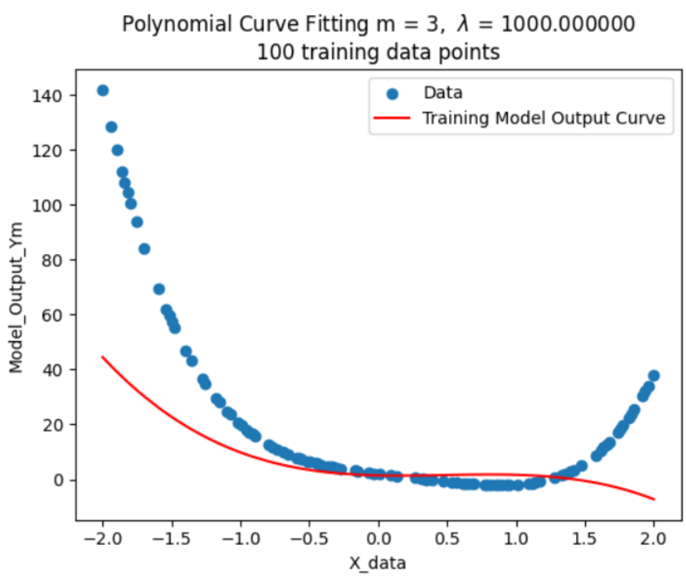
In such cases, regularization alone can't improve fitting because the model lacks the initial complexity required to represent the underlying complexity of the data distribution.

Consequently, this results in underfitting due to a mismatch between the model's simplicity and the complex nature of the training data distribution.





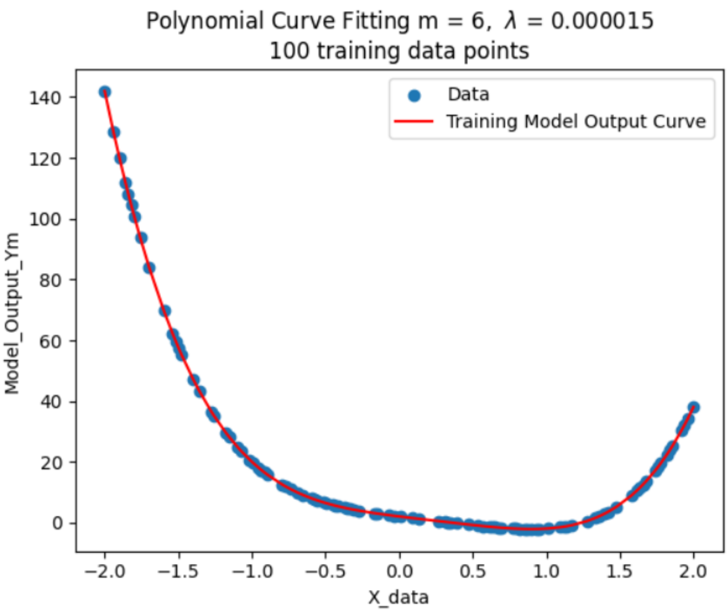


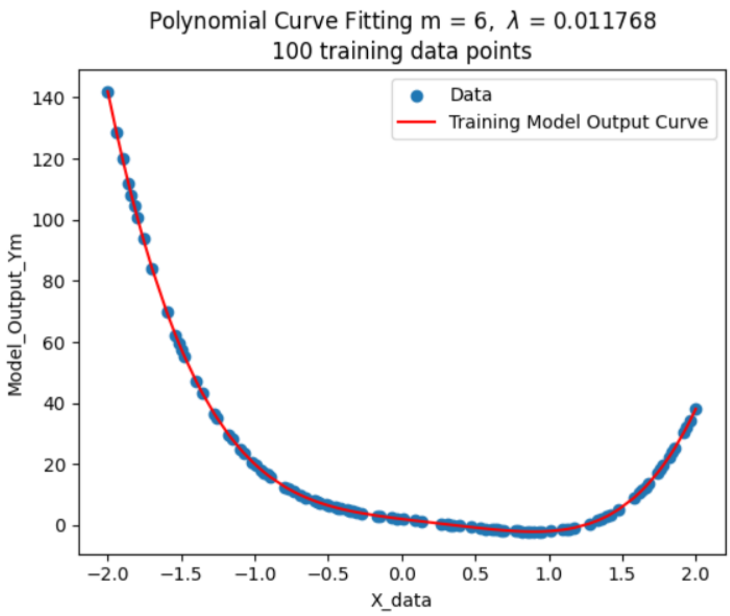


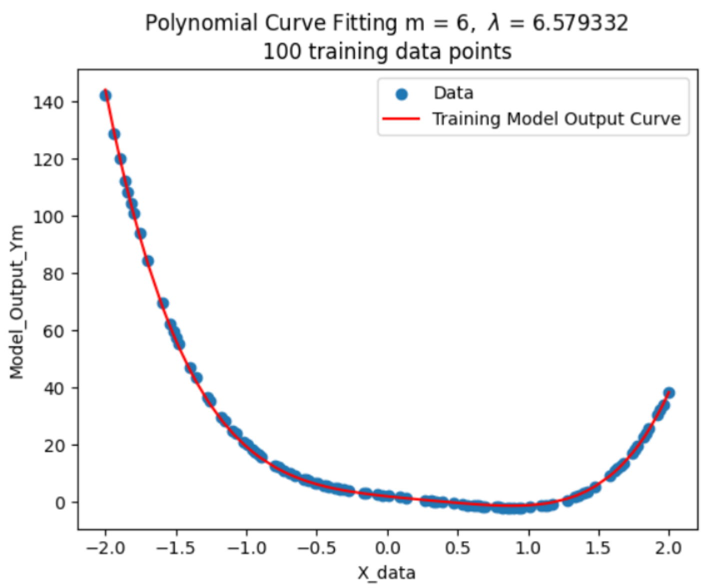
A polynomial model of degree 3 performs noticeably better than a degree 2 model but still falls short of fully capturing the training data.

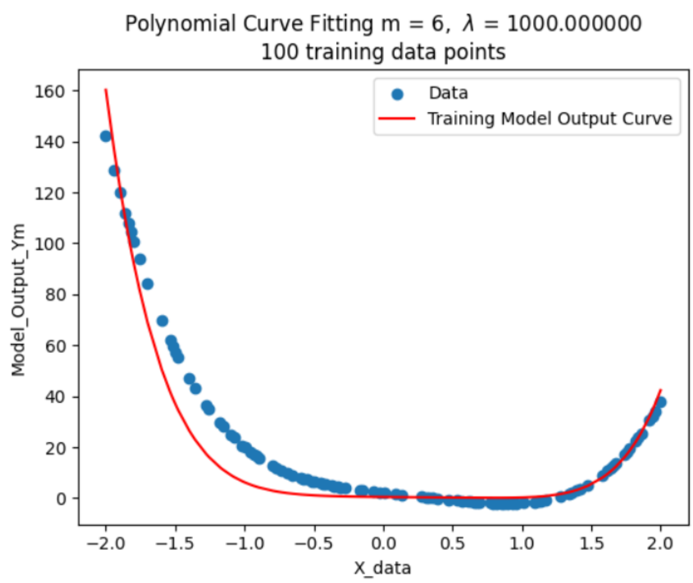
When a model under fits the data, regularization alone is insufficient to enhance its performance.

In this context, a very high lambda value exacerbates underfitting, emphasizing the need to balance model complexity and regularization for optimal results.





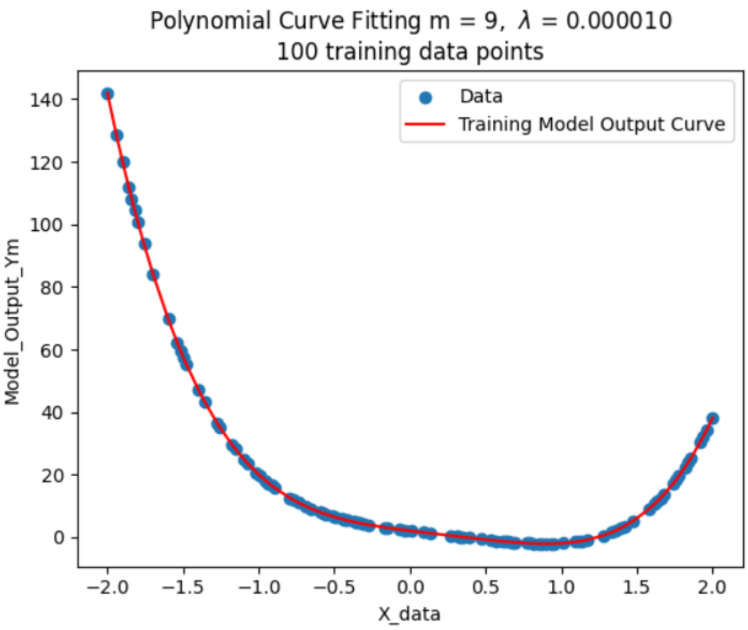


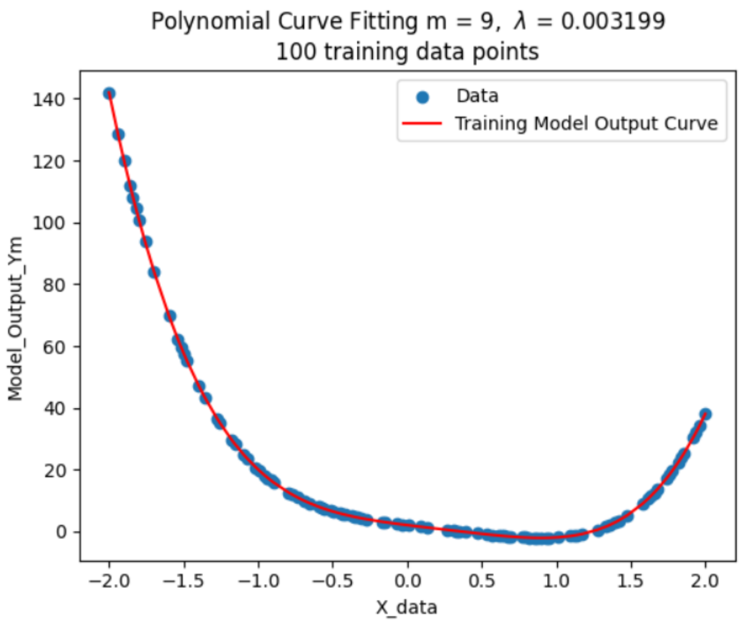


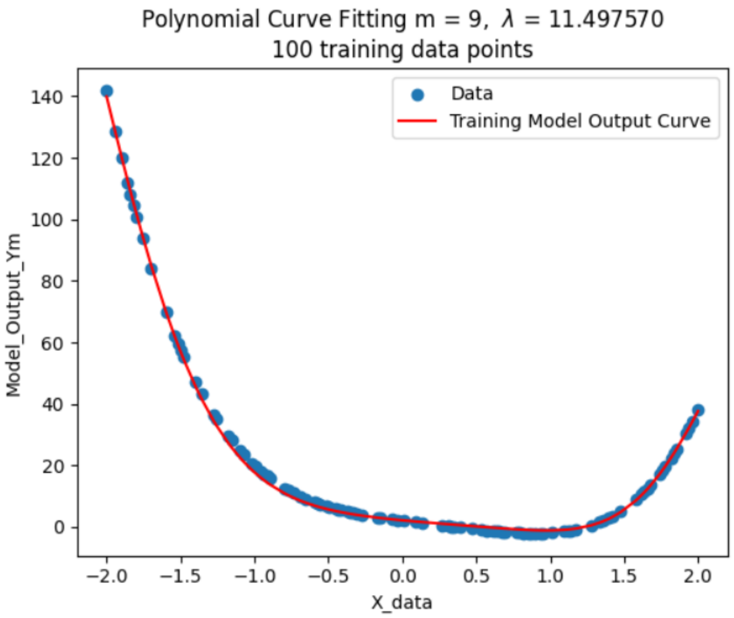
The training points follow a curve of degree 6, and the non-regularized model fits them perfectly.

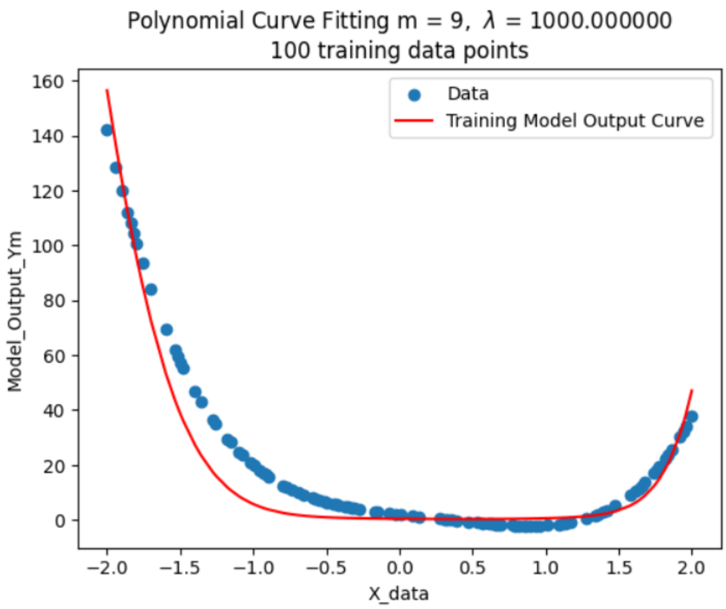
With a moderate regularisation value, the regularised curve closely aligns with the non-regularized curve. The same happens with very low regularisation values.

As Lambda values increase beyond a certain point, underfitting of the training points becomes evident.









With a degree of 9 and a substantial dataset of 1 00 training points, there is no overfitting issue as observed when there were only 10 points. The non-regularized curve fits the training points exceptionally well while maintaining smoothness.

Curves generated with appropriate lambda values closely resemble the non-regularized curve.

However, when Lambda values become exceedingly large, the result is underfitting of the training data, indicating the importance of balancing regularisation with model complexity, even with ample training data.

E_{RMS} Table:

Degree	Size	λ(Regularization)	ERMS_train	ERMS_val	ERMS_test
2	10	0	11.05514657	9.36719127	11.10589248
		1.00E-05	11.05514657	9.36717883	11.10588307
		1.00E-01	11.05900258	9.2494783	11.01436003
		11.52	18.50602344	13.3021282	10.75856484
		156	41.76083688	31.79788799	29.14223624
	100	0	8.620815109	9.05578354	8.7251331
		1.00E-05	8.620815109	9.05578344	8.72513239
		1.00E-01	8.620862616	9.05481989	8.71812385
		11.52	9.047141395	9.52397611	8.36122354
		156	18.03922708	19.44345719	15.59614254
3	10	0	7.91834128	7.16664054	8.26291714
		1.00E-05	7.91834128	7.16663178	8.2629
		1.00E-01	7.92262119	7.08555839	8.187345
		11.52	13.09958885	8.45592232	7.5354805
		156	35.96331063	25.13517152	25.1107322
	100	0	6.07102537	6.22935771	6.70746183
		1.00E-05	6.07102537	6.22935724	6.70746125
		1.00E-01	6.07107906	6.22475485	6.70163225
		11.52	6.48007079	6.27778921	6.46673031
		156	13.10108919	12.12644285	11.34387317
6	10	0	1.44E-07	3.11E-06	3.54E-06
		1.00E-05	1.55E-05	1.91E-05	1.70E-05
		1.00E-01	1.18E-01	1.36E-01	1.20E-01
		11.52	4.88E+00	4.32E+00	6.04E+00
		156	1.00E+01	8.99E+00	1.29E+01
	100	0	2.27E-06	3.52E-06	3.46E-06
		1.00E-05	2.76E-06	4.27E-06	3.36E-06
		1.00E-01	1.51E-02	1.72E-02	1.39E-02
		11.52	1.00E+00	1.01E+00	1.13E+00
		156	5.50E+00	5.16E+00	6.65E+00
9	10	0	3.18E-09	3.24E-06	3.47E-06
		1.00E-05	1.99E-04	7.04E-04	1.34E-03
		1.00E-01	1.80E-01	2.00E-01	4.18E-01

		11.52	3.69E+00	3.73E+00	4.97E+00
		156	9.40E+00	8.02E+00	1.21E+01
		0	2.26E-06	3.56E-06	3.51E-06
		1.00E-05	8.91E-06	9.53E-06	9.81E-06
		1.00E-01	5.34E-02	5.22E-02	5.76E-02
		11.52	1.10E+00	9.60E-01	1.17E+00
	100	156	4.45E+00	3.66E+00	5.13E+00

Highlighted values in the Erms_vald column represent Lambda values that yield the minimum validation error. These values are likely to be selected during extensive iterations.

In general, training errors are consistently small, and training and validation errors exhibit similarity.

The most significant improvement due to regularization over the non-regularized model is observed in the degree 9, Size 10 model. Here, the impact of appropriate Lambda values on reducing test prediction error is particularly pronounced.

Conclusion

The analysis of the provided data points reveals that they were derived from a 6th-degree curve. Notably, when a 6th-degree polynomial model is trained with only 10 points, it exhibits a remarkable characteristic: its training and test errors are closely comparable. This parity between training and test errors signifies that the model generalizes effectively to all data points, relying solely on the information encapsulated within the 10 training points.

Additionally, it is evident that this 6th-degree model precisely passes through every training point, further underscoring its ability to faithfully represent the underlying data distribution. This implies that for this particular dataset, a 6th-degree polynomial model trained with a minimal number of points demonstrates exceptional generalization and fitting performance.

Task 2:

Linear model for regression using polynomial basis functions for Datasets 2.

Dataset 2: Bivariate Dataset - 2d data

File Name	Data Points
test_14.csv	100
train50_14.csv	50
train200_14.csv	200
val_14.csv	100

Procedure:

- Gathering dataset and analysing the dimensions.
- Creating polynomial basis functions of degree **M** and dimension **d**(=2).
- Linear Regression is performed between the N x D design matrix and the N x 1 output vector to determine the regression parameters
- The estimated model is used to generate the output for validation dataset
- Based on this generated output **RMS Error** is calculated and the model with the least RMS Error is chosen as the best performing model for that dataset
- If the model was observed to be overfitting : steep decrease in Training RMSE but increase in Validation RMSE then we add regularisation term to regression

Surface Plots for N=200

Function approximation is done for the bivariate data and the following results are obtained:
When model is trained with **train200_14.csv ,** the surface plot approximated functions:

Size=200, λ=0:

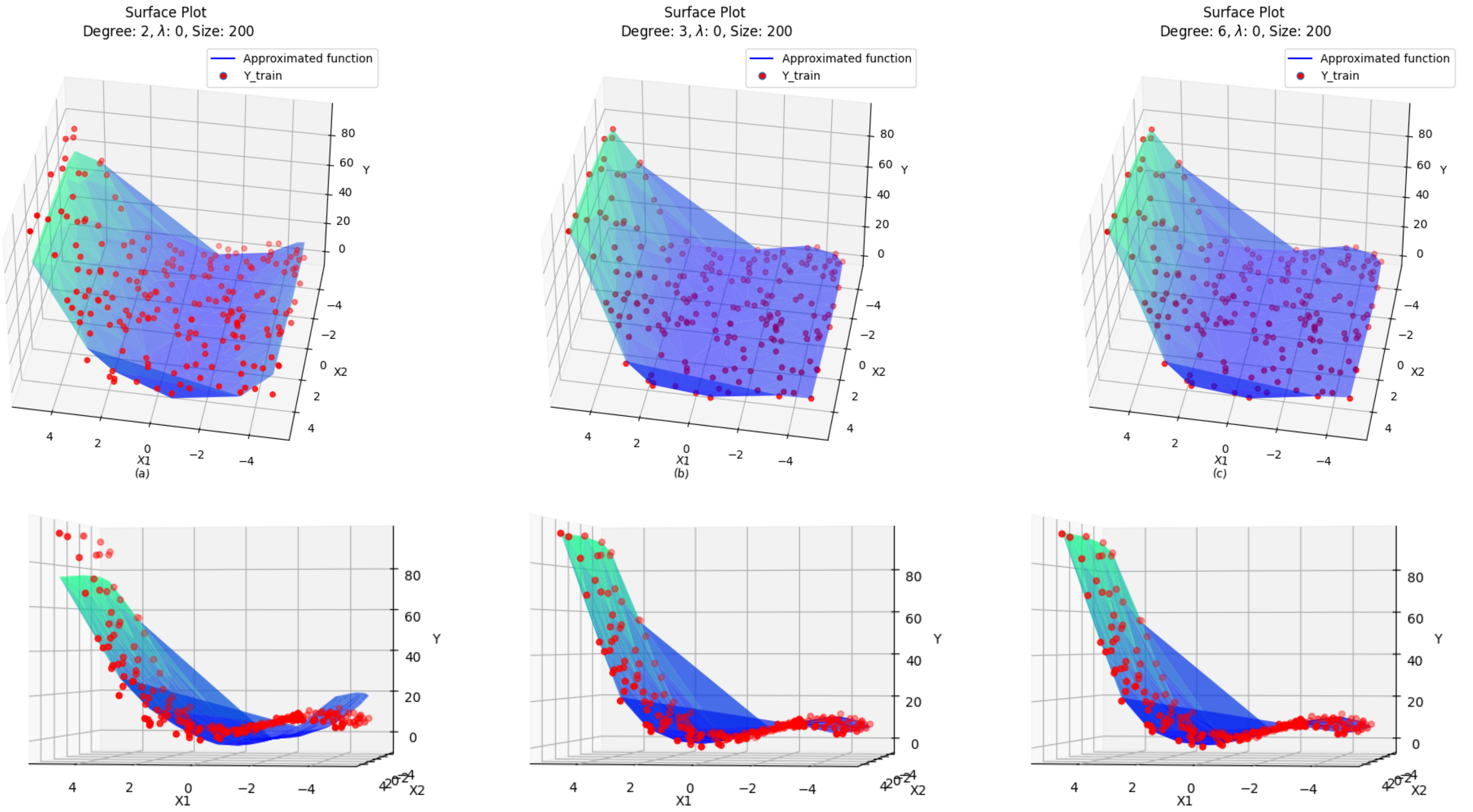
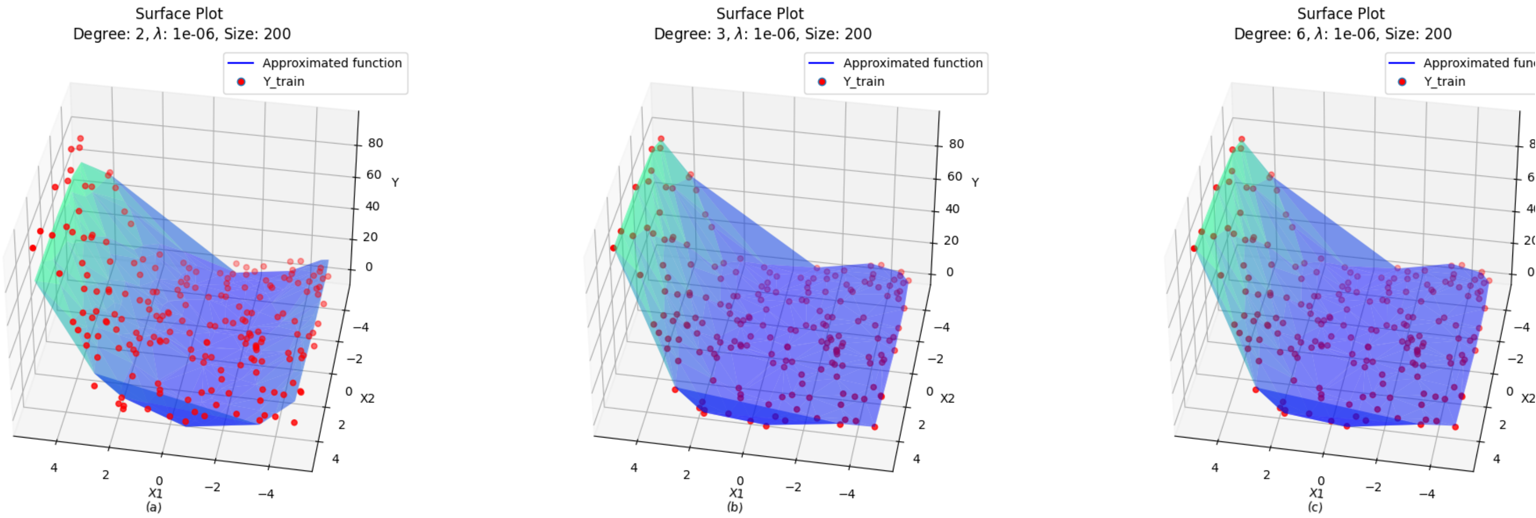


Fig 2.01: No Regularization

Size=200, λ=10^{-6}:



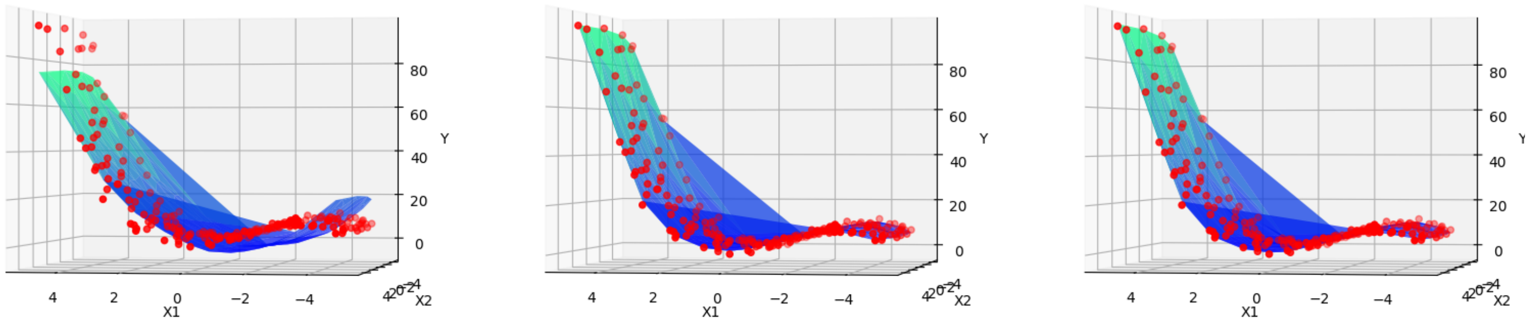


Fig 2.02: With Regularization λ=1e-6

Size=200, $\lambda=10^{-2}$:

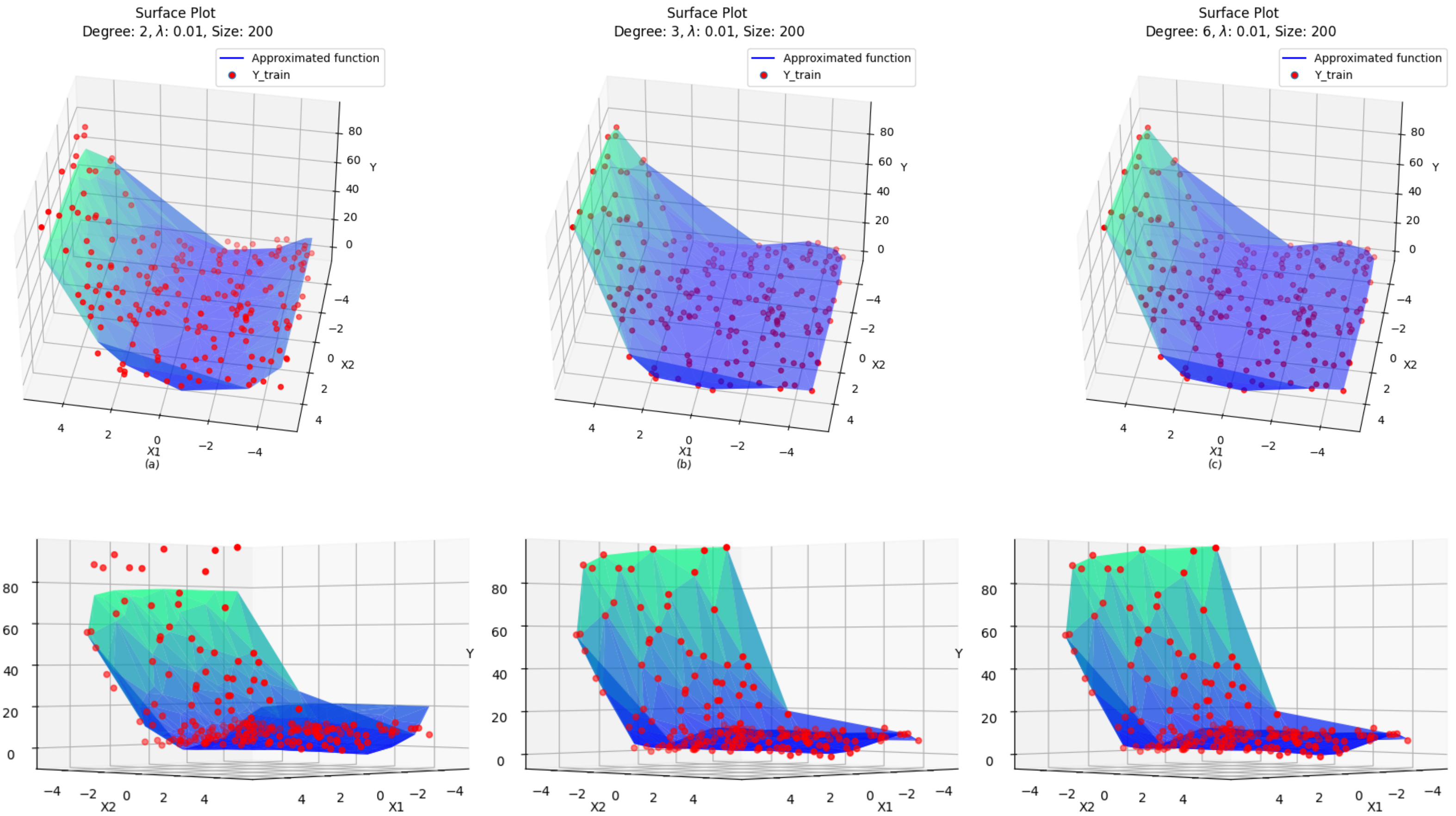
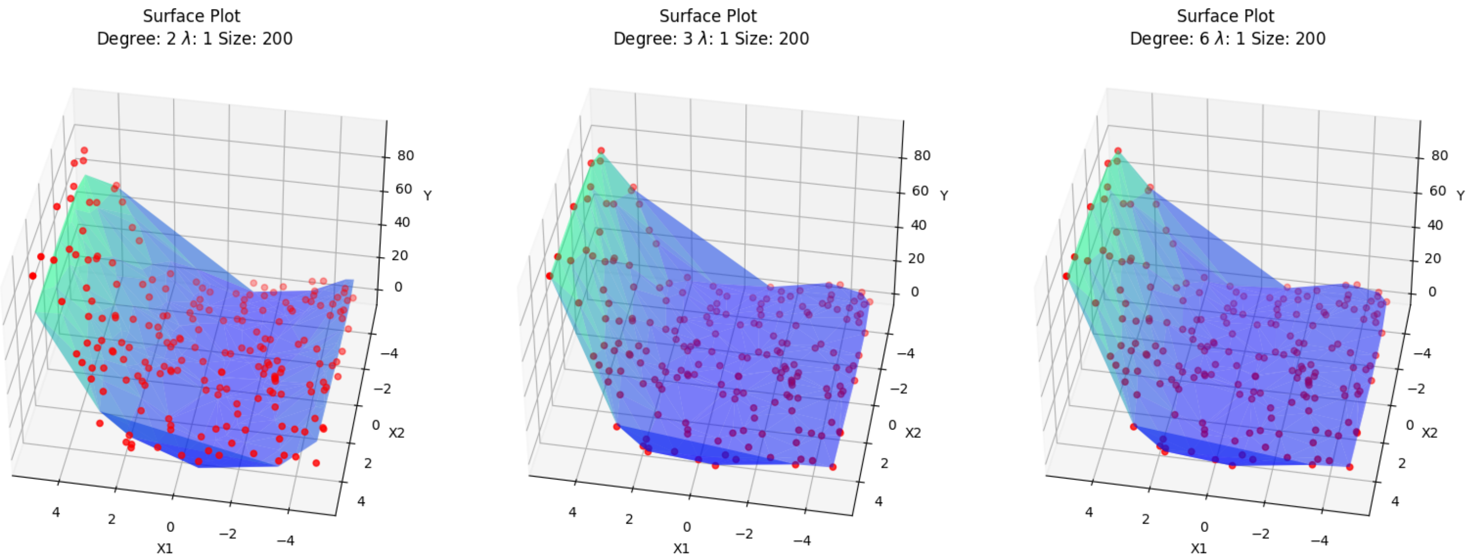


Fig 2.03: With Regularization λ=1e-02

Size=200, λ=1:



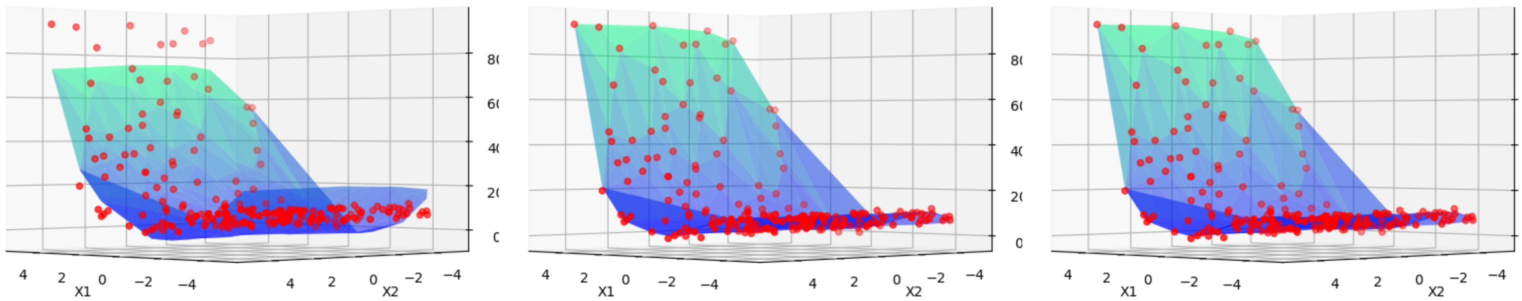


Fig 2.04: With Regularization λ=1

Size=200, λ=100:

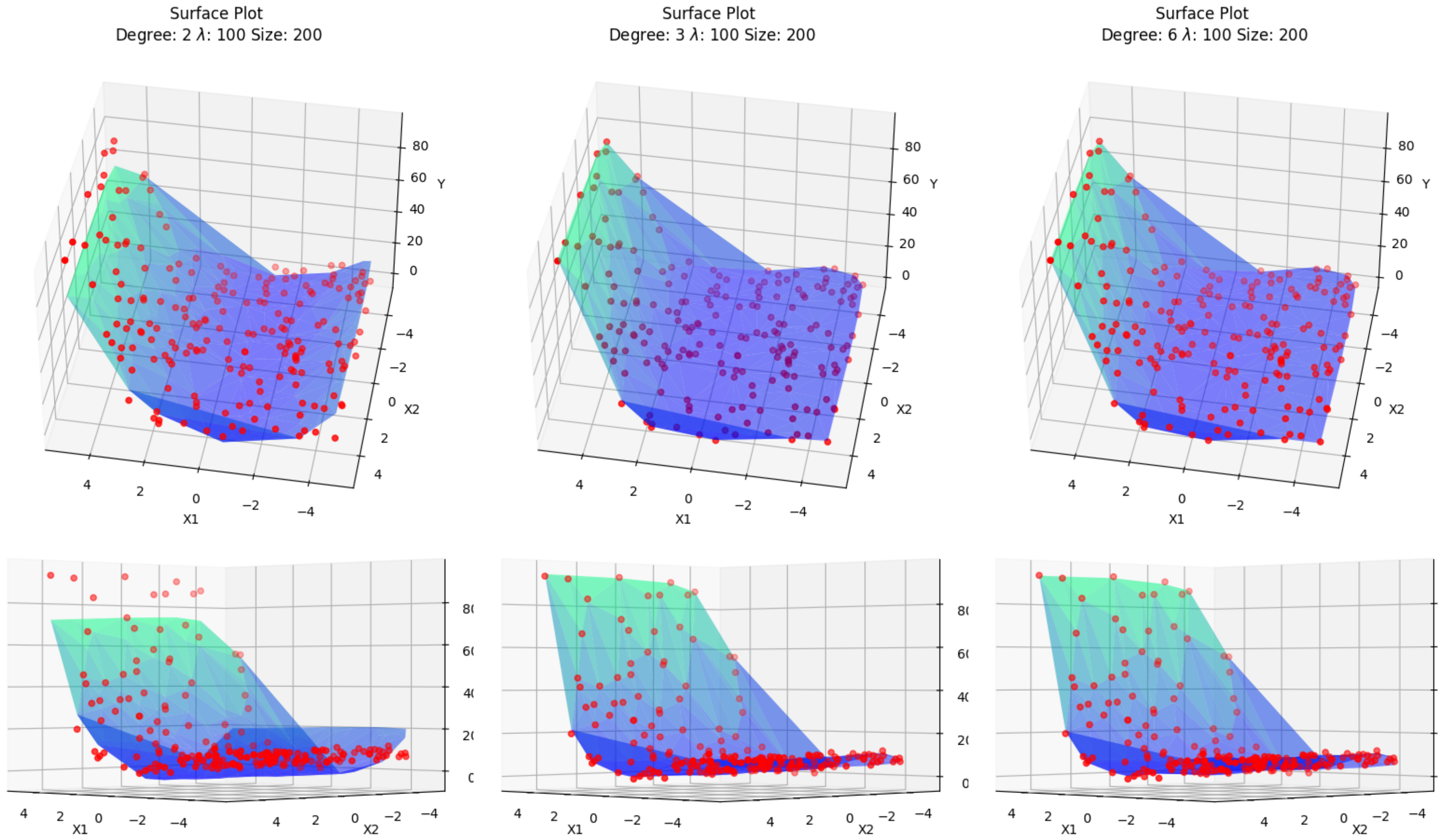


Fig 2.05: With Regularization λ=100

Overall, all graphs of degree 3 & 6 fit the data with good accuracy.

From the above graphs we can infer that with increase in regularization, the performance with cross validation improves for the model and the rate increases with increase in complexity.

This shows that plots with regularization fit data better than plots without regularization and degree **3** polynomials perform better than degree **2 & 6** in fitting the data. However, degree **6** seems similar to degree **3** approximated function.

(P.S: Regularization for degree 2 isn't required as it doesn't overfit the training sample but has been shown with other other graphs)

Surface Plots for N=50

When model is trained with **train50_14.csv** , the surface plot approximated functions:

Size=50, λ=0:

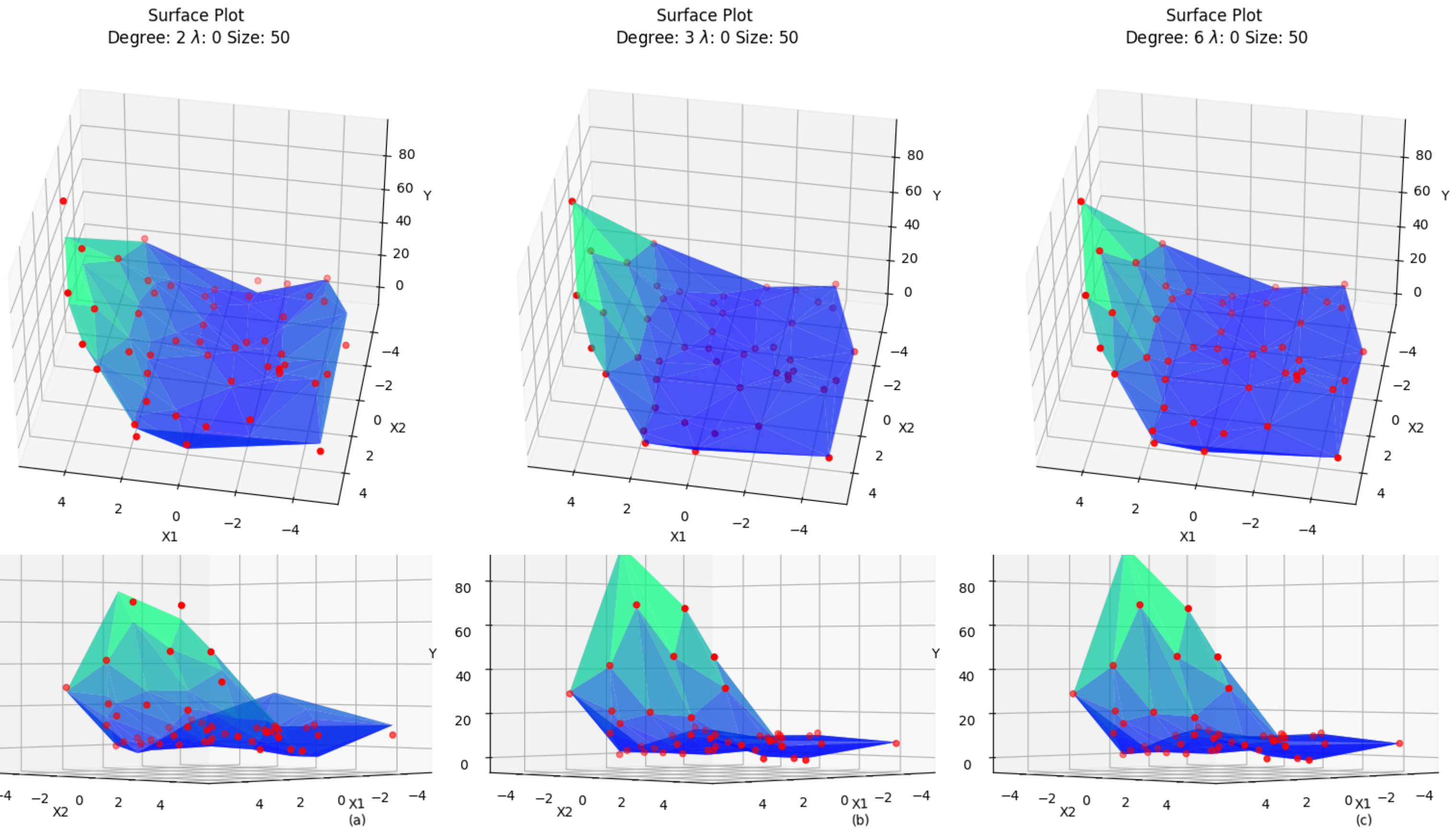
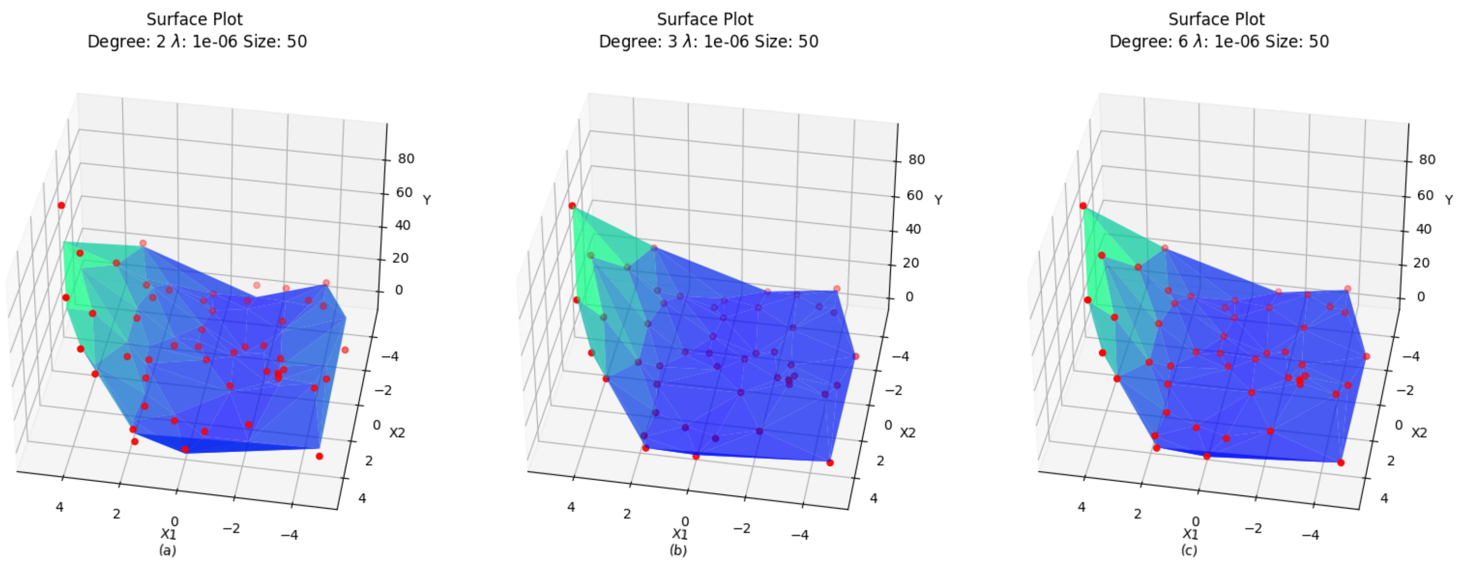


Fig 2.06: No Regularization λ=0

Size=50, λ=10^{-6}:



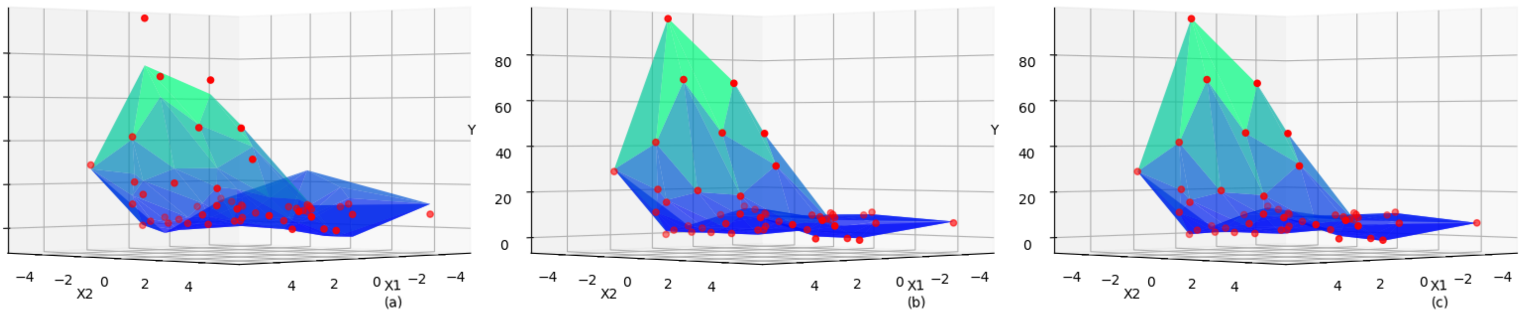


Fig 2.07: With Regularization λ=1e-6

Size=50, λ=10^{-2}:

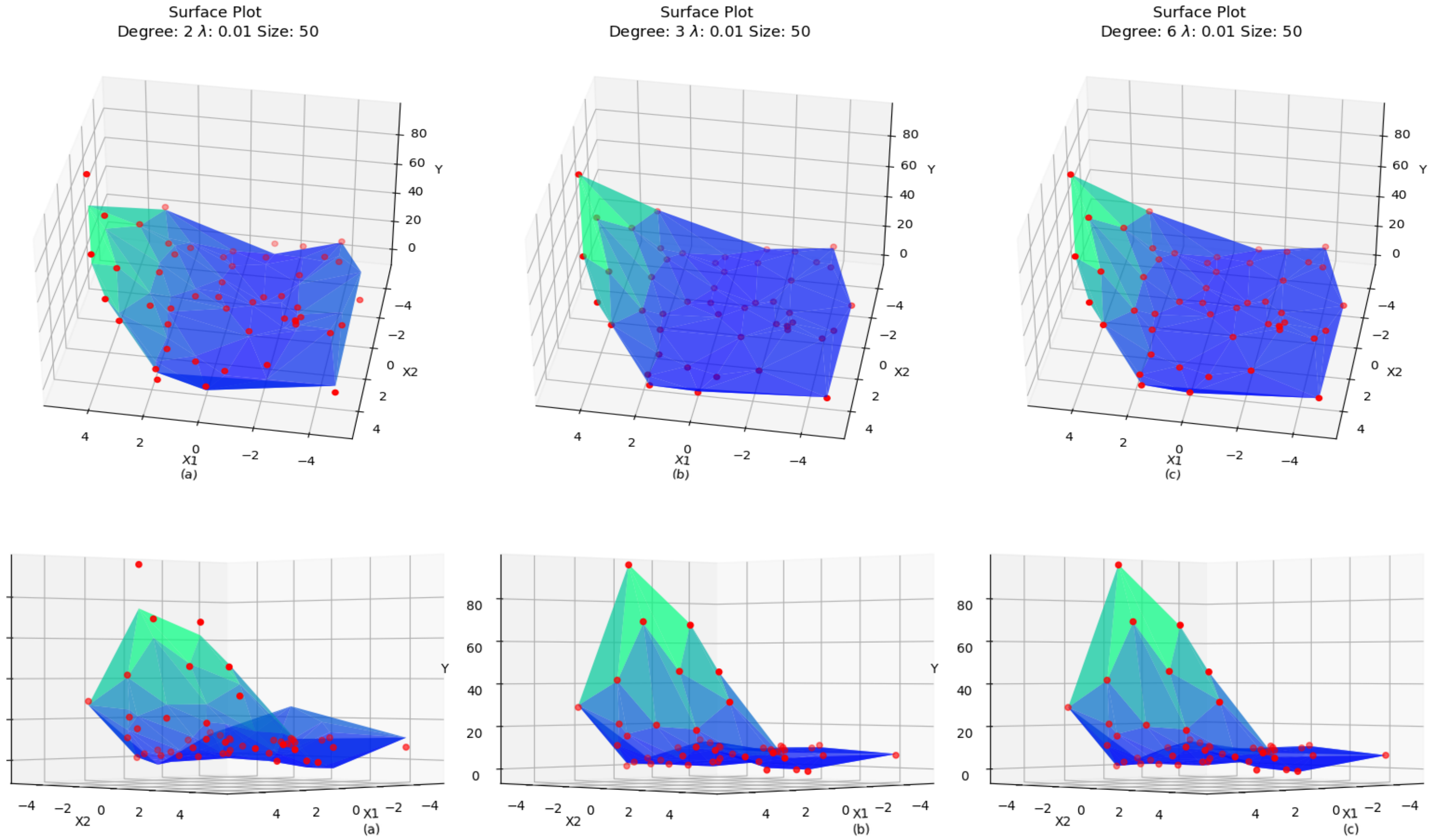


Fig 2.08: With Regularization λ=1e-2

Size=50, λ=1

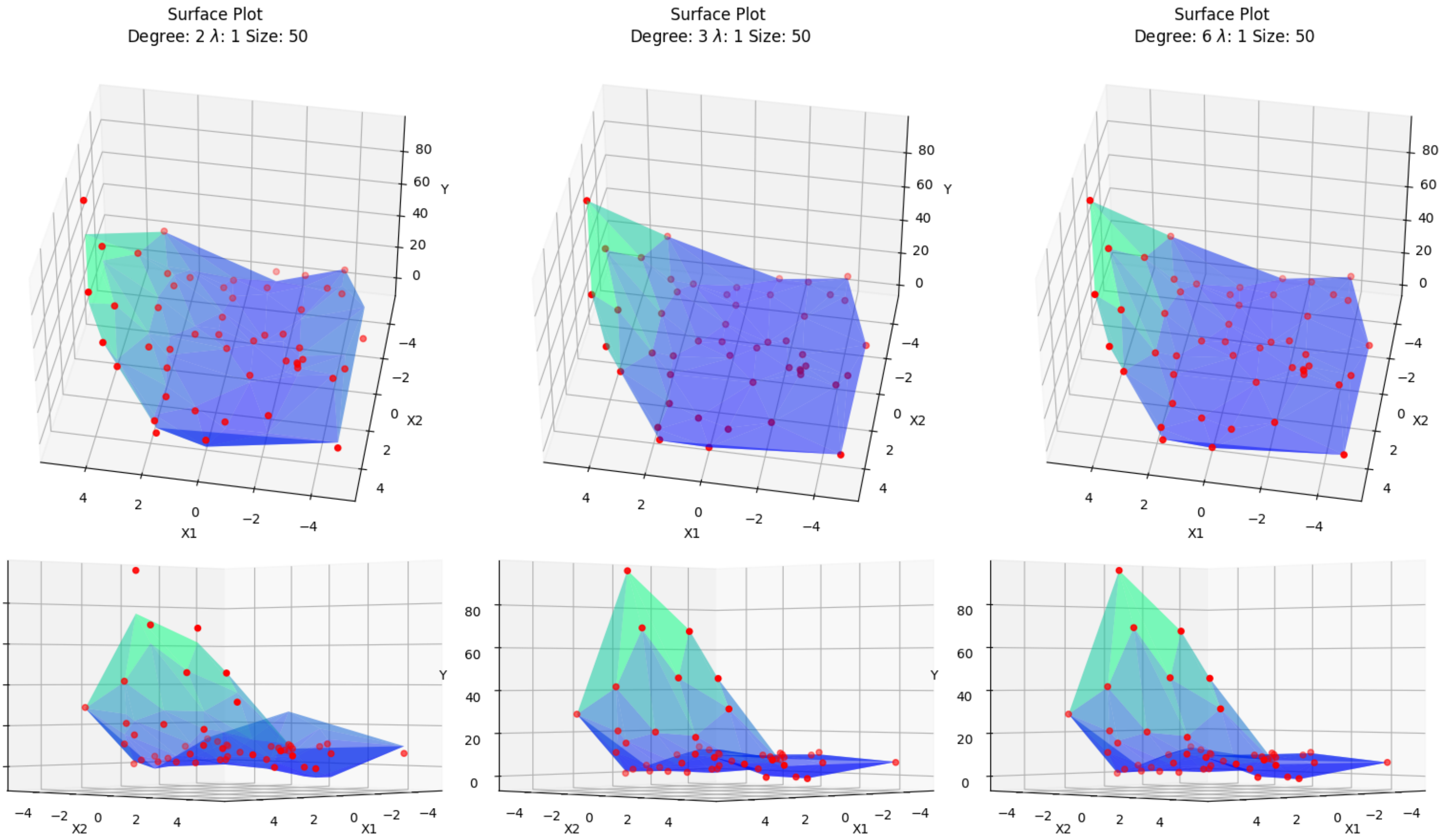


Fig 2.09: With Regularization λ=1e-2

Size=50, λ=100

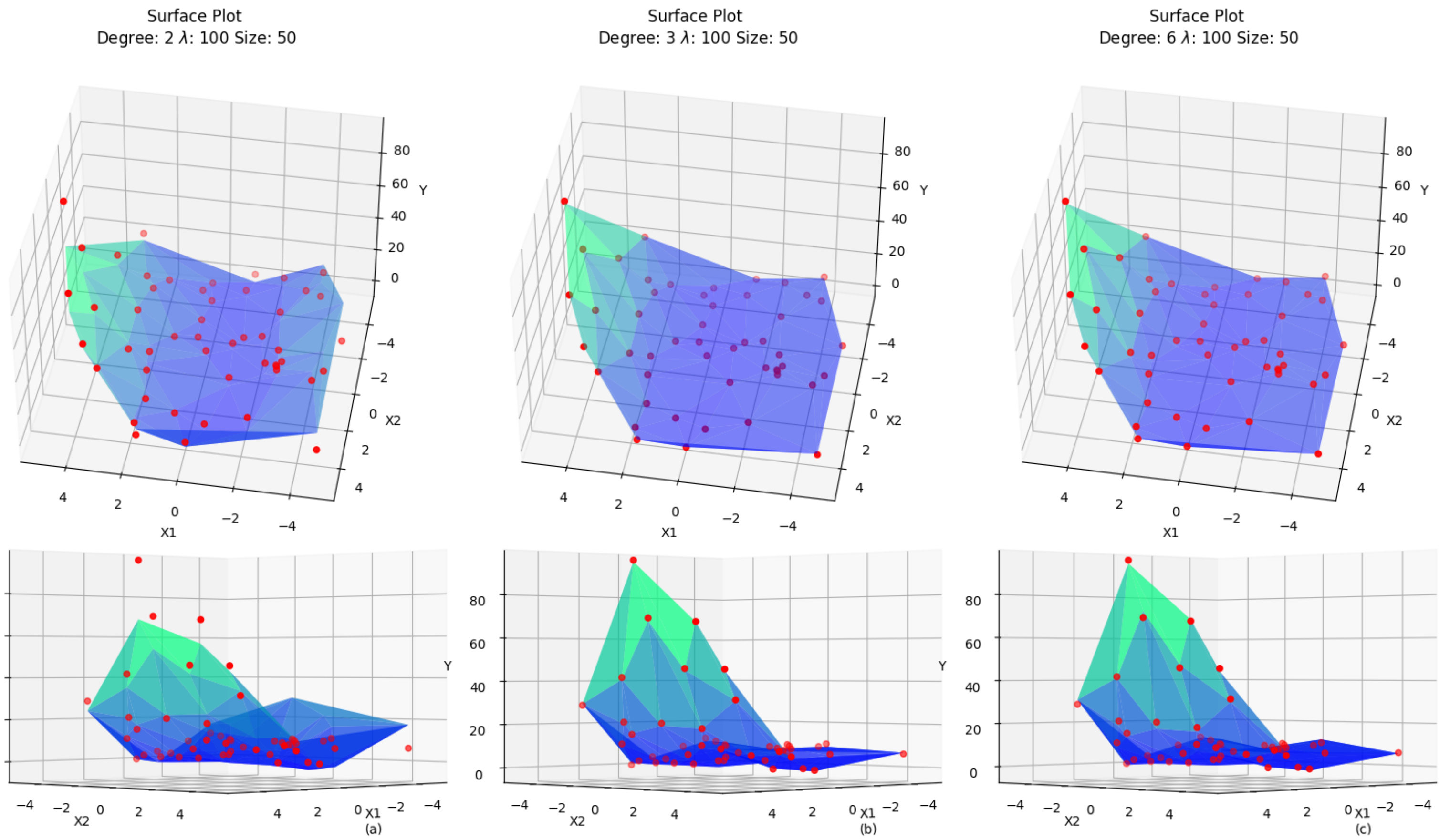


Fig 2.10: With Regularization λ=100

Overall, all graphs of degree 3 & 6 fit the data with good accuracy.

From the above graphs we can infer that with increase in regularization, the performance with cross validation improves for the model and the rate increases with increase in complexity.

This shows that plots with regularization fit data better than plots without regularization and degree **3** polynomials perform better than degree **2 & 6** in fitting the data. However, degree **6** seems similar to degree **3** approximated function.

(P.S: Regularization for degree 2 isn't required as it doesn't overfit the training sample but has been shown with other other graphs)

Best performing Model(Selected using cross validation):

Degree = 3 , λ = 0 , Size = 200

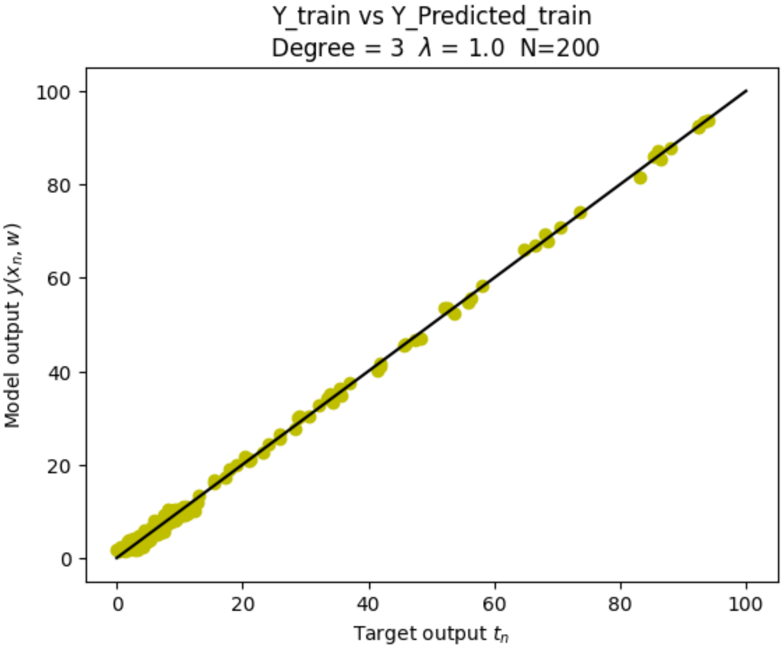
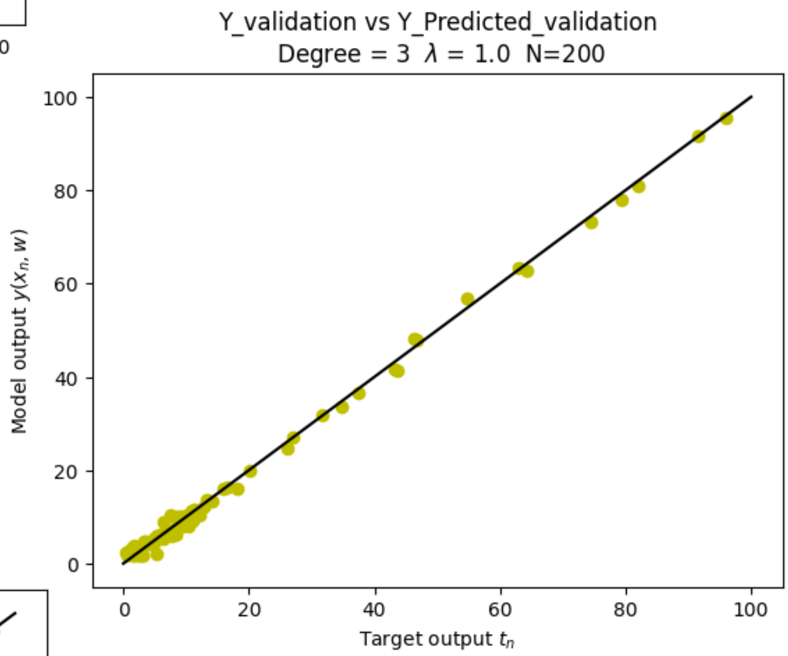


Fig 2.11: Scatter plot for training data

Fig 2.12: Scatter plot for validation data



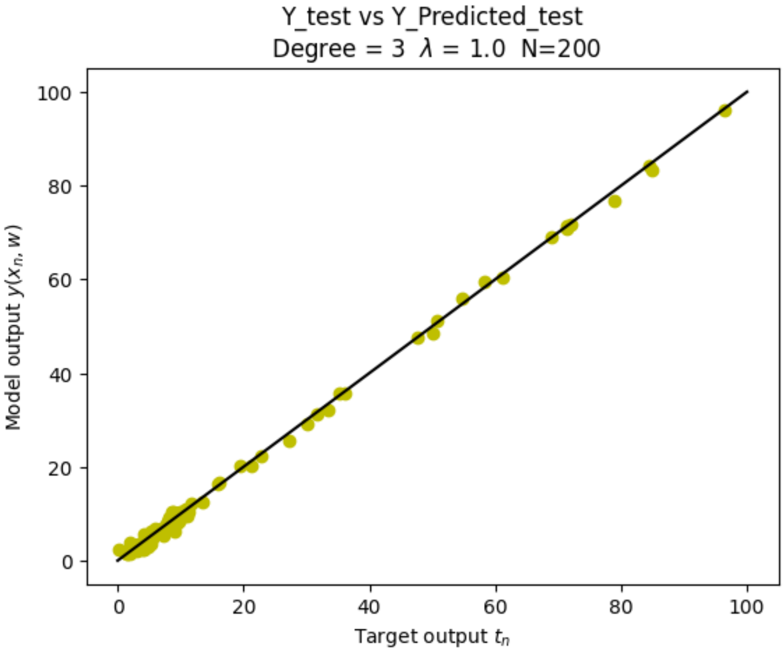


Fig 2.13: Scatter plot for test data

E_{RMS} Table:

Degree	Size	λ	ERMS_train	ERMS_val	ERMS_test
2	50	0	6.567376842	11.71779073	10.9394786
		1.00E-06	6.567376842	11.71779072	10.93947859
		1.00E-02	6.567376947	11.71774759	10.93938757
		1	6.568310968	11.71523389	10.93256487
		100	7.354425256	13.15307247	12.41347961
	200	0	7.501214773	5.284737725	4.680107642
		1.00E-06	7.501214773	5.284737724	4.680107642
		1.00E-02	7.501214776	5.284720926	4.680101433
		1	7.501244664	5.283103193	4.679516997
		100	7.590643765	5.276774231	4.7171874
3	50	0	0.726918444	1.783238224	1.570732946
		1.00E-06	0.726918444	1.783238228	1.570732951
		1.00E-02	0.7269186664	1.783273774	1.57078399
		1	0.7288375702	1.788085712	1.577595171
		100	1.115666204	2.133673807	2.010587065
	200	0	0.8891631121	0.820508344	0.6825756003
		1.00E-06	0.8891631121	0.8205083437	0.6825756007
		1.00E-02	0.8891631253	0.8205052121	0.6825795457
		1	0.8892911776	0.8202645563	0.6830458205
		100	1.058403396	0.8955491399	0.8033016838
6	50	0	0.544728718	2.770398219	2.395984812
		1.00E-06	0.544728718	2.770397815	2.395984793
		1.00E-02	0.5447318144	2.766115647	2.395660271
		1	0.5594224096	2.686863508	2.596451877
		100	1.583814447	12.18291334	8.586835929
	200	0	0.8450829126	0.8928050429	0.703755901
		1.00E-06	0.8450829126	0.8928050396	0.7037559033
		1.00E-02	0.845082958	0.8927767157	0.7037721577
		1	0.8454806924	0.8902549993	0.7055115546
		100	1.142882281	0.9676756455	0.8921194038

Table 2.1: E_{RMS} data for Task 2

Observations:

- The best performing model is selected from cross validation with minimum rms error, N=200, λ =1.0, Degree =3, E_{RMS} (CV)=0.82026.
- The models with low training error, E_{RMS} decrease with increase in regularization parameter showing that the model with no regularization is overfitting.
- We can observe that the training error for Degree = 6 and N = 200 increases as λ increases and cross validation & test error decrease as λ increases.
- The scatter plot displayed shows the correlation between the system output and predicted output which superimposes with 45° line for 99% of data showing a strong correlation
- The graphs and table(2.1), its evident that the lower value of N(50) has higher variance than higher value of N(200) keeping all other parameters the same.

Conclusion:

- **Bias-Variance Tradeoff**: The degree 2 model may have been too simple to capture the underlying patterns in the data, resulting in underfitting. On the other hand, the degree 6 model may have been too complex, leading to overfitting. The degree 3 model strikes a balance between bias and variance, making it a better choice for this dataset.
- **Regularization**: It controlled the complexity of the model and ensured it did not fit the noise in the data.
- **Model Complexity:** Degree 3 polynomials are sufficiently flexible to capture non-linear relationships in the data, but they are less prone to overfitting compared to degree 6 polynomials. This makes them a suitable choice for this task.

In conclusion, based on the given dataset and the evaluation criteria (which could include metrics like **Mean Squared Error**.), the degree 3 polynomial regression model with regularization λ = 1.0 and N = **200** demonstrated superior performance compared to both the degree 2 and degree 6 models. It achieved a better balance between fitting the data's underlying patterns and avoiding overfitting, making it the preferred choice for this regression task.

above results point in the direction that the data performs well with degree **3** complexity compared to other higher complexity. It even works well with regularization which shows that the model takes up degree **3** as it surmounts the other models in test and validation errors.

Task 3 :

Linear model for regression using Gaussian basis functions for Datasets 2 and 3

Dataset 2 : Bivariate Dataset - 2d data

File Name	Data Points
test_14.csv	100
train50_14.csv	50
train200_14.csv	200
val_14.csv	100

Dataset 3 : Multivariate Dataset - 8d data

File Name	Data Points
train_data.csv	538
train_label.csv	538
val_data.csv	153
val_label.csv	153
train_data.csv	77
train_label.csv	77

Procedure :

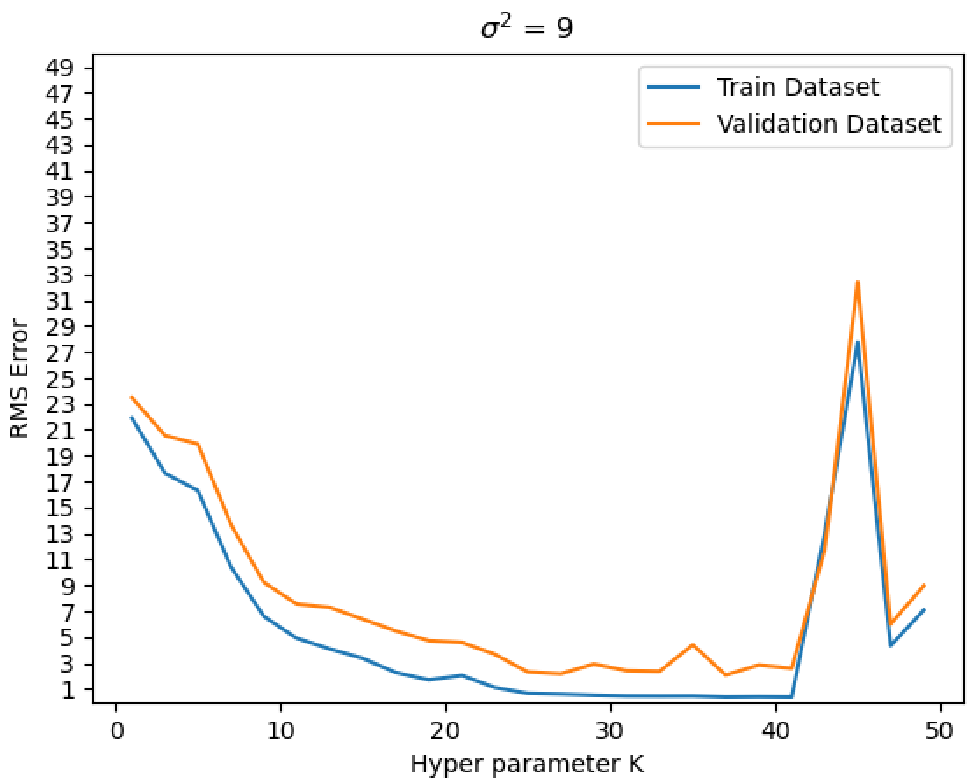
- K-means clustering is performed on the given dataset to obtain the k - means
- A design matrix of dimensions **N x K** is computed using gaussian basis functions
- Linear Regression is performed between the N x K design matrix and the N x 1 output vector to determine the regression parameters
- The estimated model is used to generate the output for validation dataset
- Based on this generated output RMS Error is calculated and the model with the least RMS Error is chosen as the best performing model for that dataset
- If the model was observed to be overfitting : steep decrease in Training RMSE but increase in Validation RMSE then we add regularisation term to regression

This report determines the best model for Bivariate Dataset [50 first then 200] followed by Multivariate Dataset

Regression without regularisation for Bivariate Dataset [50 points]

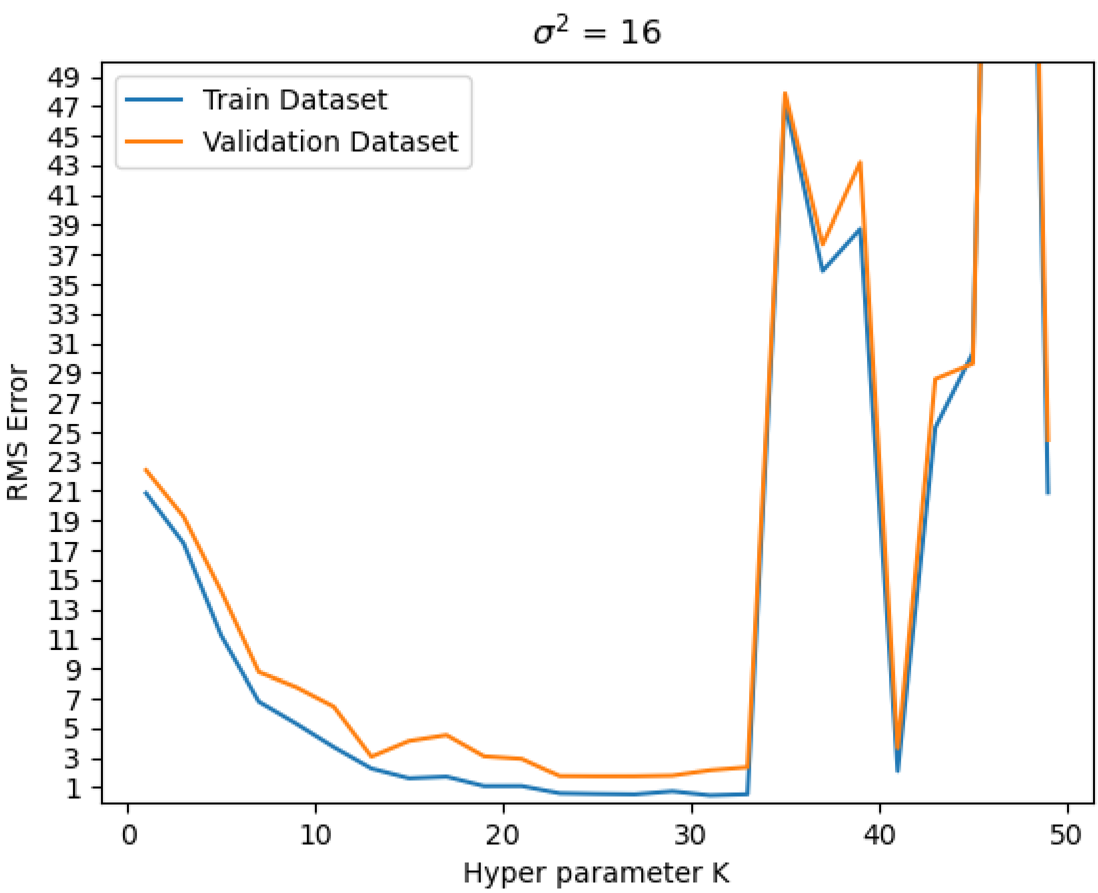
Bivariate Dataset without Regularisation : Variance : 9

```
1     | RMS train :   21.91   | RMS validation :   23.49   | RMS test :   26.10
3     | RMS train :   17.64   | RMS validation :   20.55   | RMS test :   21.90
5     | RMS train :   16.30   | RMS validation :   19.90   | RMS test :   20.67
7     | RMS train :   10.40   | RMS validation :   13.67   | RMS test :   14.96
9     | RMS train :    6.59   | RMS validation :    9.22   | RMS test :   11.21
11    | RMS train :    4.89   | RMS validation :    7.54   | RMS test :    8.99
13    | RMS train :    4.07   | RMS validation :    7.28   | RMS test :    7.97
15    | RMS train :    3.35   | RMS validation :    6.37   | RMS test :    8.27
17    | RMS train :    2.27   | RMS validation :    5.46   | RMS test :    6.14
19    | RMS train :    1.70   | RMS validation :    4.71   | RMS test :    4.64
21    | RMS train :    2.04   | RMS validation :    4.57   | RMS test :    5.22
23    | RMS train :    1.11   | RMS validation :    3.66   | RMS test :    3.45
25    | RMS train :    0.66   | RMS validation :    2.31   | RMS test :    2.90
27    | RMS train :    0.60   | RMS validation :    2.18   | RMS test :    2.48
29    | RMS train :    0.52   | RMS validation :    2.92   | RMS test :    2.79
31    | RMS train :    0.46   | RMS validation :    2.41   | RMS test :    2.11
33    | RMS train :    0.45   | RMS validation :    2.36   | RMS test :    2.50
35    | RMS train :    0.46   | RMS validation :    4.39   | RMS test :    4.74
37    | RMS train :    0.37   | RMS validation :    2.08   | RMS test :    1.72
39    | RMS train :    0.40   | RMS validation :    2.84   | RMS test :    2.70
41    | RMS train :    0.37   | RMS validation :    2.60   | RMS test :    2.41
43    | RMS train :   12.97   | RMS validation :   11.68   | RMS test :   12.86
45    | RMS train :   27.73   | RMS validation :   32.42   | RMS test :   35.55
47    | RMS train :    4.32   | RMS validation :    5.99   | RMS test :    6.41
49    | RMS train :    7.08   | RMS validation :    8.96   | RMS test :    8.63
```



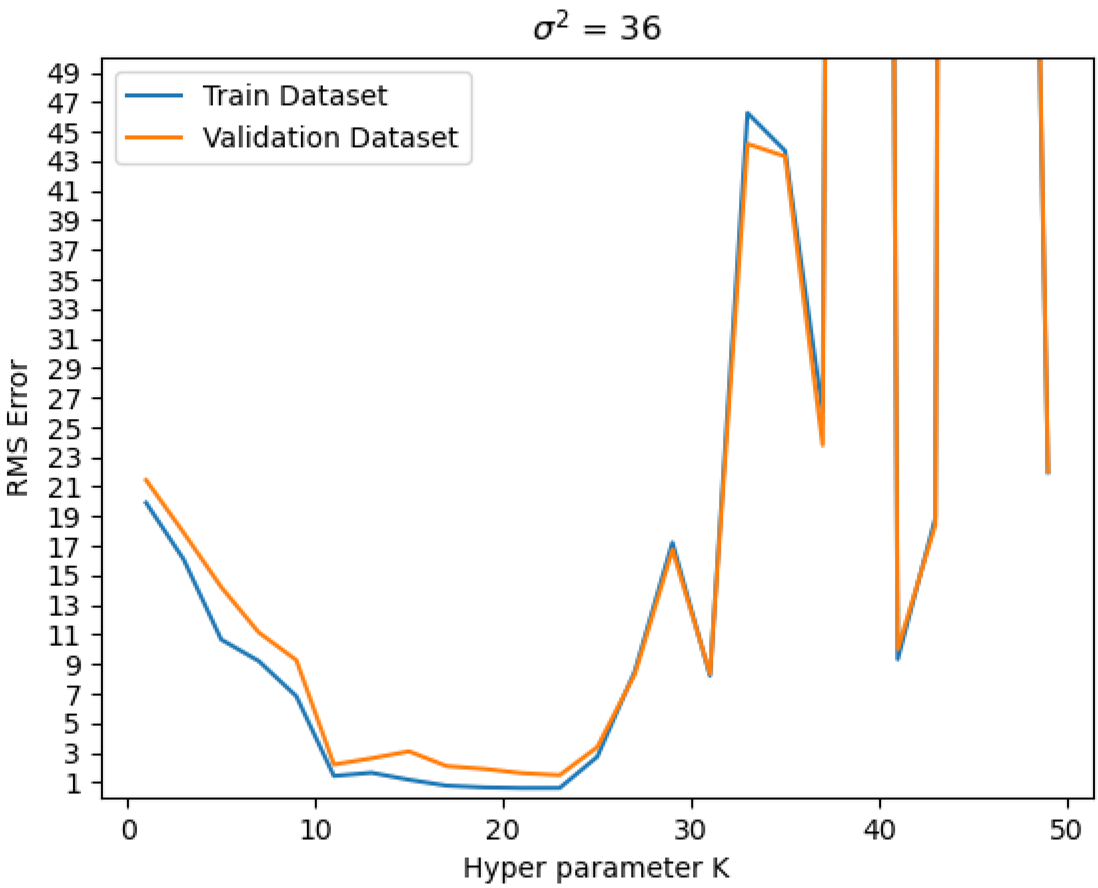
Bivariate Dataset without Regularisation : Variance : 16

1	RMS train : 20.85	RMS validation : 22.41	RMS test : 24.88
3	RMS train : 17.46	RMS validation : 19.26	RMS test : 20.67
5	RMS train : 11.23	RMS validation : 14.23	RMS test : 15.13
7	RMS train : 6.78	RMS validation : 8.77	RMS test : 10.18
9	RMS train : 5.29	RMS validation : 7.75	RMS test : 8.97
11	RMS train : 3.71	RMS validation : 6.44	RMS test : 7.11
13	RMS train : 2.26	RMS validation : 3.06	RMS test : 4.52
15	RMS train : 1.60	RMS validation : 4.12	RMS test : 3.90
17	RMS train : 1.71	RMS validation : 4.52	RMS test : 3.85
19	RMS train : 1.07	RMS validation : 3.09	RMS test : 3.66
21	RMS train : 1.08	RMS validation : 2.92	RMS test : 4.07
23	RMS train : 0.59	RMS validation : 1.75	RMS test : 1.84
25	RMS train : 0.54	RMS validation : 1.74	RMS test : 2.00
27	RMS train : 0.51	RMS validation : 1.74	RMS test : 1.97
29	RMS train : 0.72	RMS validation : 1.78	RMS test : 1.91
31	RMS train : 0.45	RMS validation : 2.15	RMS test : 1.70
33	RMS train : 0.52	RMS validation : 2.35	RMS test : 2.09
35	RMS train : 47.20	RMS validation : 47.87	RMS test : 47.51
37	RMS train : 35.88	RMS validation : 37.67	RMS test : 39.45
39	RMS train : 38.72	RMS validation : 43.20	RMS test : 39.18
41	RMS train : 2.09	RMS validation : 3.65	RMS test : 4.28
43	RMS train : 25.29	RMS validation : 28.57	RMS test : 27.03
45	RMS train : 30.34	RMS validation : 29.62	RMS test : 30.27
47	RMS train : 128.33	RMS validation : 131.28	RMS test : 130.81
49	RMS train : 20.89	RMS validation : 24.45	RMS test : 26.06



Bivariate Dataset without Regularisation : Variance : 36

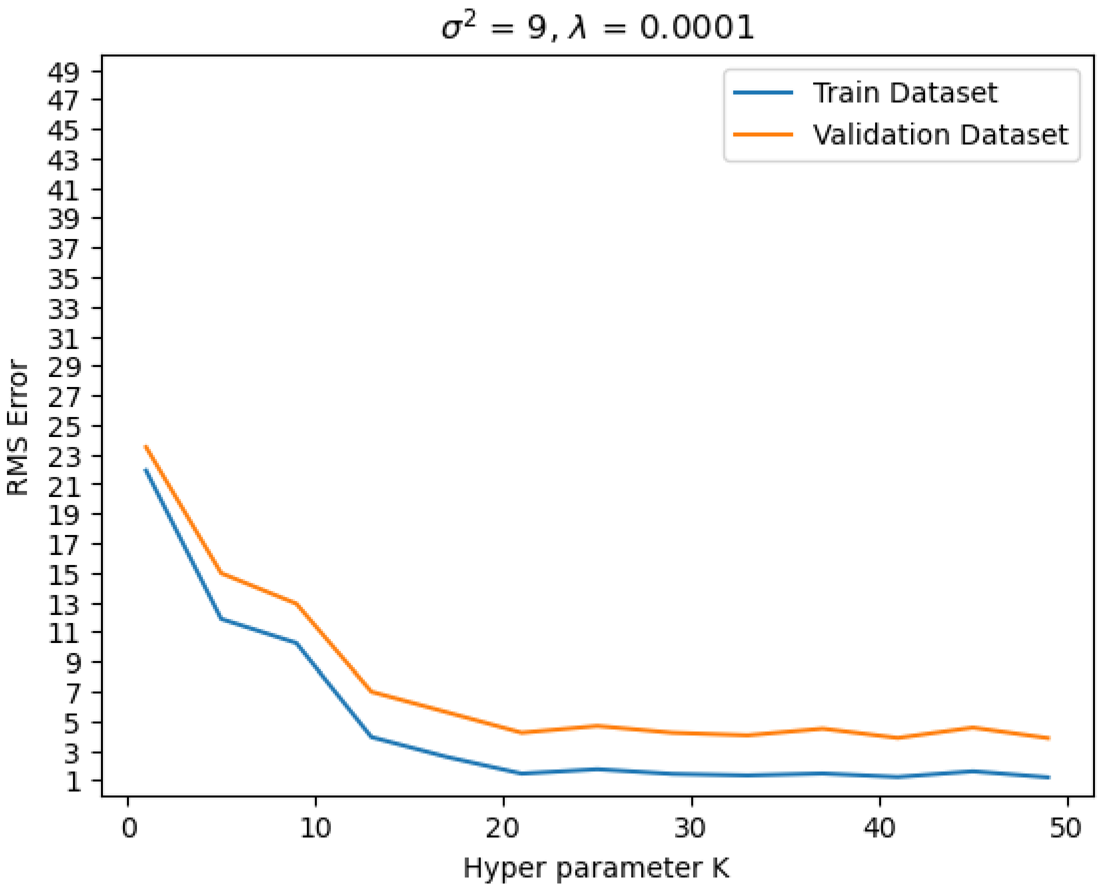
```
1     | RMS train :   19.92   | RMS validation :   21.45   | RMS test :   23.68
3     | RMS train :   16.08   | RMS validation :   17.89   | RMS test :   18.79
5     | RMS train :   10.66   | RMS validation :   14.21   | RMS test :   14.92
7     | RMS train :    9.20   | RMS validation :   11.13   | RMS test :   11.82
9     | RMS train :    6.82   | RMS validation :    9.26   | RMS test :    8.63
11    | RMS train :    1.43   | RMS validation :    2.20   | RMS test :    2.20
13    | RMS train :    1.64   | RMS validation :    2.62   | RMS test :    3.30
15    | RMS train :    1.17   | RMS validation :    3.09   | RMS test :    3.27
17    | RMS train :    0.77   | RMS validation :    2.09   | RMS test :    1.77
19    | RMS train :    0.65   | RMS validation :    1.90   | RMS test :    1.98
21    | RMS train :    0.61   | RMS validation :    1.62   | RMS test :    1.42
23    | RMS train :    0.61   | RMS validation :    1.48   | RMS test :    1.47
25    | RMS train :    2.72   | RMS validation :    3.38   | RMS test :    3.53
27    | RMS train :    8.52   | RMS validation :    8.32   | RMS test :    8.62
29    | RMS train :   17.22   | RMS validation :   16.74   | RMS test :   16.32
31    | RMS train :    8.22   | RMS validation :    8.34   | RMS test :    8.02
33    | RMS train :   46.26   | RMS validation :   44.17   | RMS test :   46.18
35    | RMS train :   43.70   | RMS validation :   43.32   | RMS test :   46.64
37    | RMS train :   26.04   | RMS validation :   23.80   | RMS test :   22.80
39    | RMS train :  419.07   | RMS validation :  416.46   | RMS test :  443.07
41    | RMS train :    9.33   | RMS validation :   10.00   | RMS test :   10.14
43    | RMS train :   18.87   | RMS validation :   18.42   | RMS test :   18.18
45    | RMS train :  589.19   | RMS validation :  551.23   | RMS test :  524.16
47    | RMS train :  149.88   | RMS validation :  153.43   | RMS test :  156.37
49    | RMS train :   21.97   | RMS validation :   22.02   | RMS test :   22.59
```



Regression with regularisation for Bivariate Dataset [50 points]

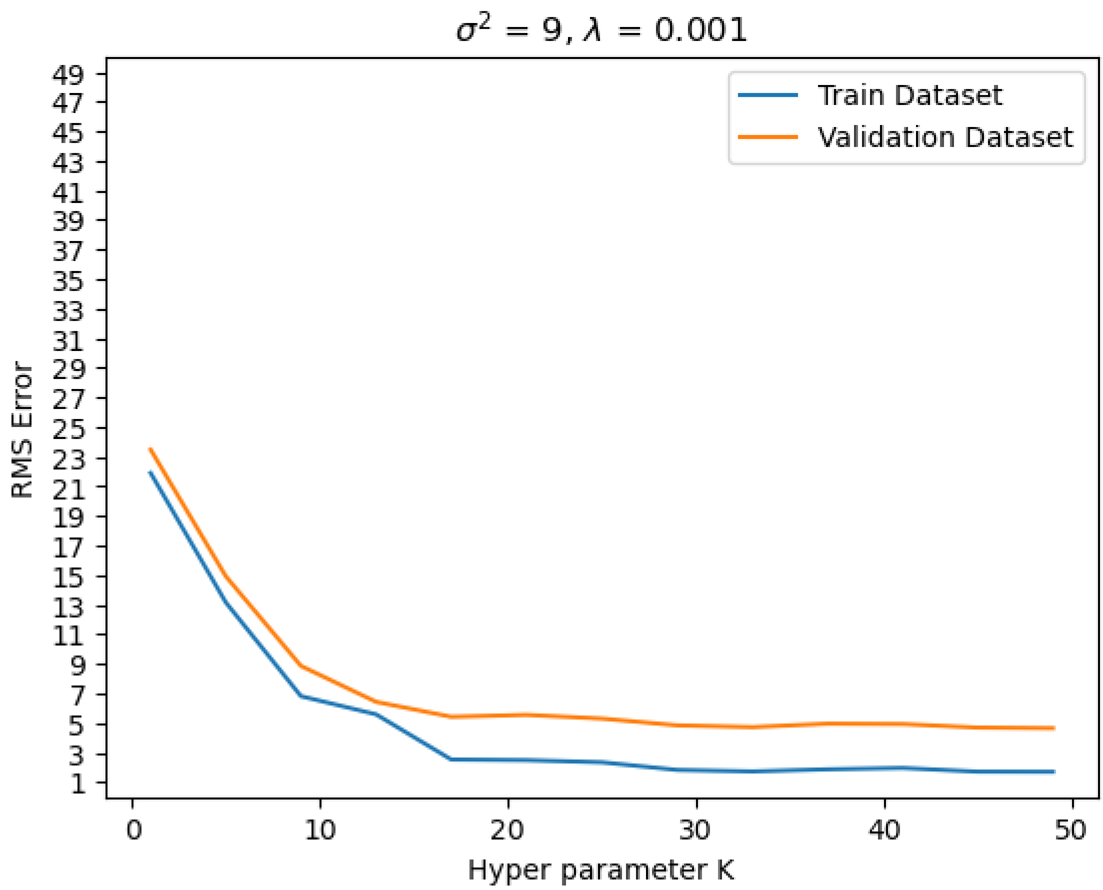
Bivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.0001

```
Regulariation Parameter : 0.0001 | Variance : 9
Model RMS Errors for 50 datapoints :

1     | RMS train :    21.91   | RMS validation :    23.49    | RMS test :    26.10
5     | RMS train :    11.87   | RMS validation :    14.97    | RMS test :    15.97
9     | RMS train :    10.25   | RMS validation :    12.91    | RMS test :    14.15
13    | RMS train :     3.91   | RMS validation :     6.98    | RMS test :     8.16
17    | RMS train :     2.57   | RMS validation :     5.60    | RMS test :     6.35
21    | RMS train :     1.44   | RMS validation :     4.21    | RMS test :     5.51
25    | RMS train :     1.74   | RMS validation :     4.67    | RMS test :     5.69
29    | RMS train :     1.42   | RMS validation :     4.21    | RMS test :     5.35
33    | RMS train :     1.32   | RMS validation :     4.03    | RMS test :     4.94
37    | RMS train :     1.45   | RMS validation :     4.48    | RMS test :     5.21
41    | RMS train :     1.22   | RMS validation :     3.86    | RMS test :     4.76
45    | RMS train :     1.60   | RMS validation :     4.56    | RMS test :     5.20
49    | RMS train :     1.19   | RMS validation :     3.85    | RMS test :     4.79
```



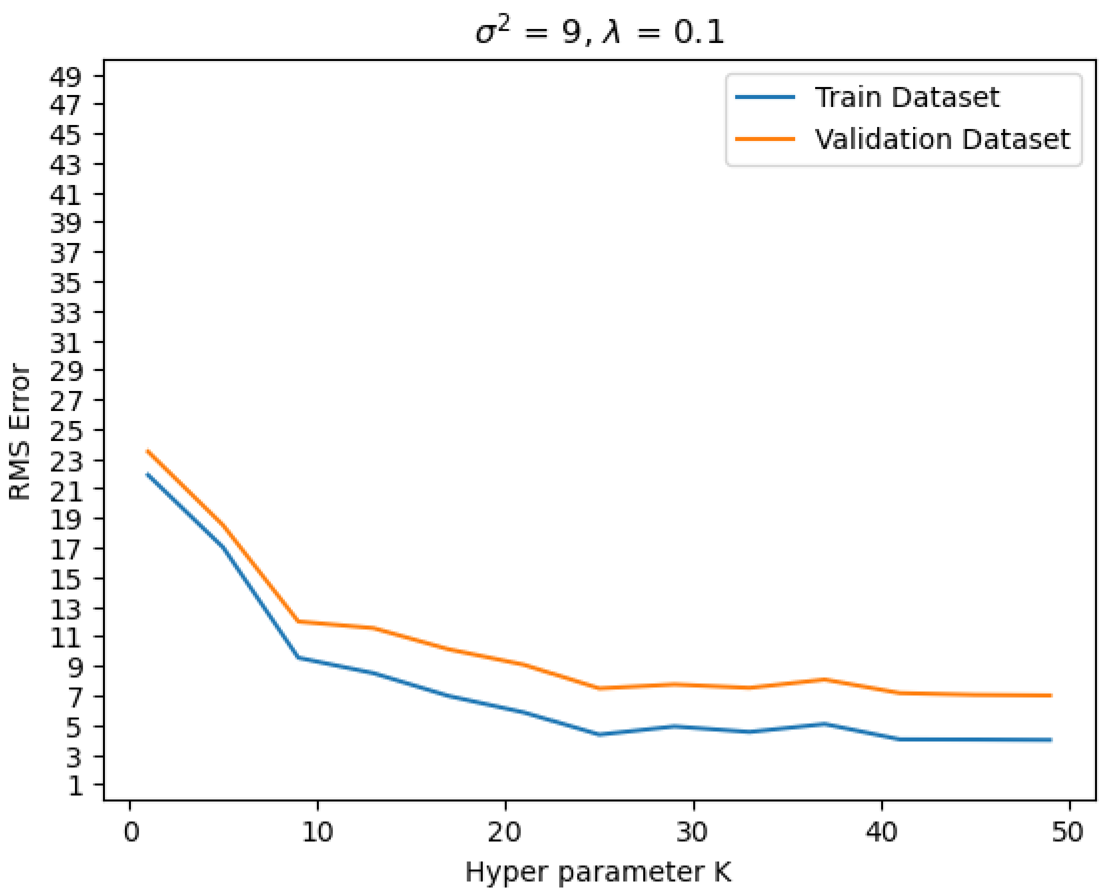
Bivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.001

```
Regulariation Parameter : 0.001 | Variance : 9
Model RMS Errors for 50 datapoints :

1       | RMS train :    21.91    | RMS validation :    23.49     | RMS test :    26.10
5       | RMS train :    13.16    | RMS validation :    14.93     | RMS test :    16.79
9       | RMS train :     6.82    | RMS validation :     8.86     | RMS test :     9.31
13      | RMS train :     5.59    | RMS validation :     6.43     | RMS test :     7.22
17      | RMS train :     2.53    | RMS validation :     5.42     | RMS test :     7.54
21      | RMS train :     2.49    | RMS validation :     5.55     | RMS test :     7.57
25      | RMS train :     2.34    | RMS validation :     5.30     | RMS test :     6.86
29      | RMS train :     1.83    | RMS validation :     4.84     | RMS test :     5.84
33      | RMS train :     1.73    | RMS validation :     4.72     | RMS test :     5.78
37      | RMS train :     1.87    | RMS validation :     4.95     | RMS test :     5.85
41      | RMS train :     1.96    | RMS validation :     4.93     | RMS test :     6.21
45      | RMS train :     1.71    | RMS validation :     4.69     | RMS test :     5.61
49      | RMS train :     1.70    | RMS validation :     4.65     | RMS test :     5.55
```



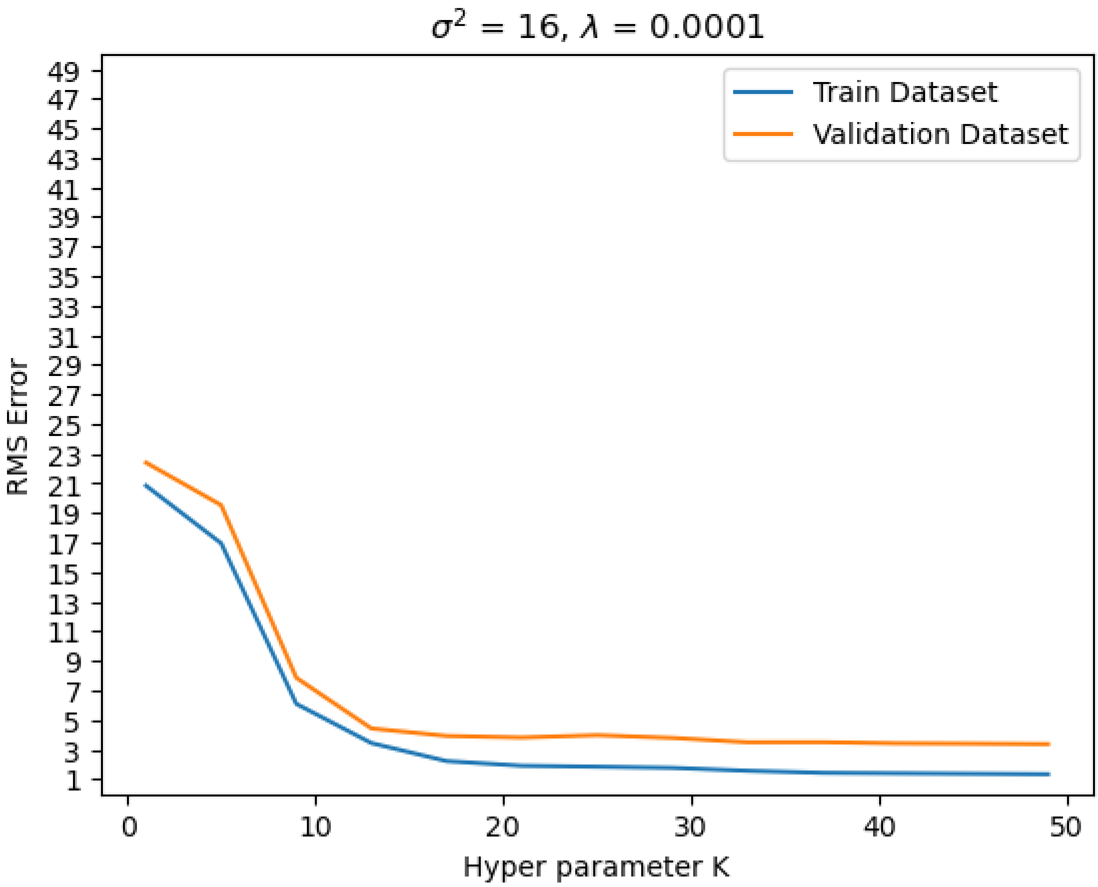
Bivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.1

```
Regulariation Parameter : 0.1 | Variance : 9
Model RMS Errors for 50 datapoints :

1      | RMS train :    21.91    | RMS validation :    23.50    | RMS test :    26.10
5      | RMS train :    17.01    | RMS validation :    18.51    | RMS test :    19.77
9      | RMS train :     9.55    | RMS validation :    12.00    | RMS test :    14.31
13     | RMS train :     8.51    | RMS validation :    11.56    | RMS test :    12.39
17     | RMS train :     6.97    | RMS validation :    10.12    | RMS test :    11.54
21     | RMS train :     5.85    | RMS validation :     9.08    | RMS test :    10.93
25     | RMS train :     4.35    | RMS validation :     7.48    | RMS test :     9.26
29     | RMS train :     4.91    | RMS validation :     7.74    | RMS test :     9.68
33     | RMS train :     4.53    | RMS validation :     7.51    | RMS test :     9.35
37     | RMS train :     5.07    | RMS validation :     8.08    | RMS test :     9.66
41     | RMS train :     4.03    | RMS validation :     7.16    | RMS test :     8.93
45     | RMS train :     4.02    | RMS validation :     7.04    | RMS test :     8.82
49     | RMS train :     3.99    | RMS validation :     7.00    | RMS test :     8.78
```



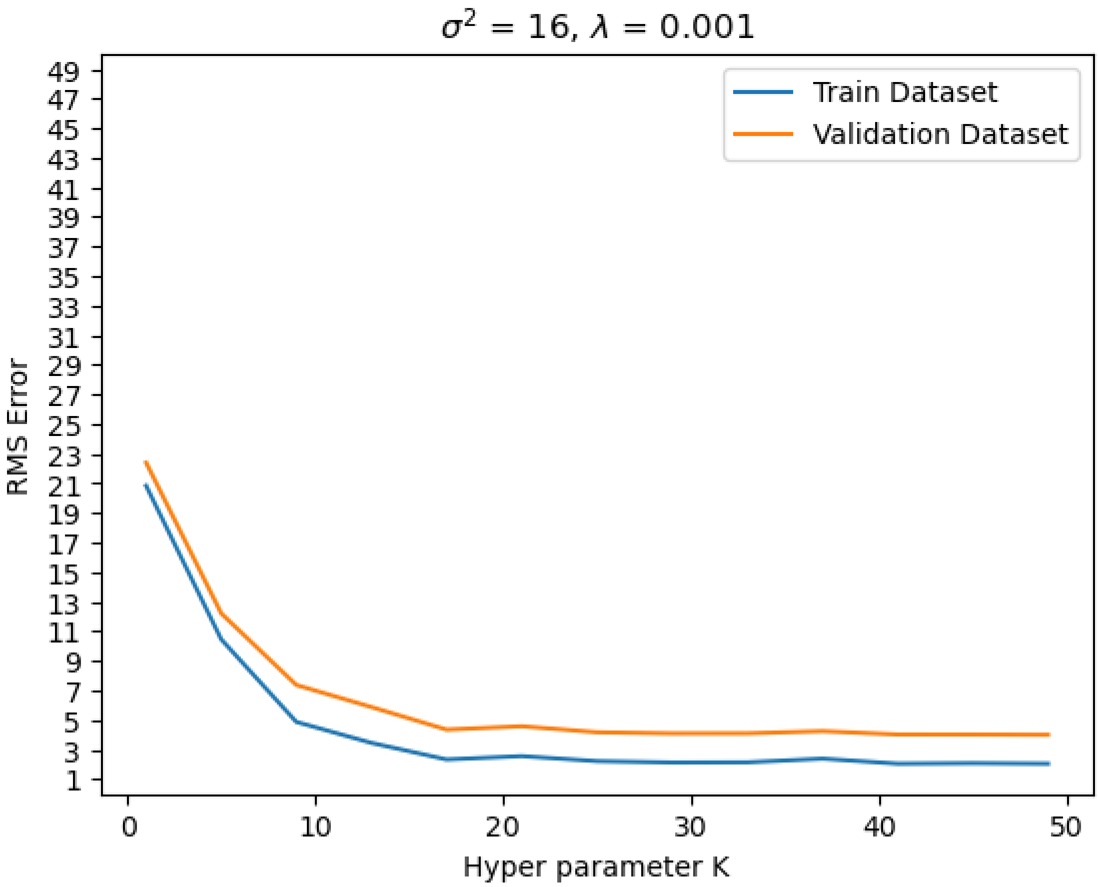
Bivariate Dataset with Regularisation : Variance : 16 and Regularisation : 0.0001

```
Regulariation Parameter : 0.0001 | Variance : 16
Model RMS Errors for 50 datapoints :

1     | RMS train :    20.85   | RMS validation :    22.41    | RMS test :    24.88
5     | RMS train :    16.95   | RMS validation :    19.53    | RMS test :    20.73
9     | RMS train :     6.09   | RMS validation :     7.86    | RMS test :     9.23
13    | RMS train :     3.44   | RMS validation :     4.43    | RMS test :     5.75
17    | RMS train :     2.22   | RMS validation :     3.92    | RMS test :     5.64
21    | RMS train :     1.91   | RMS validation :     3.82    | RMS test :     4.99
25    | RMS train :     1.84   | RMS validation :     3.97    | RMS test :     5.02
29    | RMS train :     1.77   | RMS validation :     3.79    | RMS test :     4.73
33    | RMS train :     1.57   | RMS validation :     3.50    | RMS test :     4.57
37    | RMS train :     1.43   | RMS validation :     3.50    | RMS test :     4.16
41    | RMS train :     1.40   | RMS validation :     3.43    | RMS test :     4.12
45    | RMS train :     1.37   | RMS validation :     3.40    | RMS test :     4.10
49    | RMS train :     1.34   | RMS validation :     3.37    | RMS test :     4.00
```



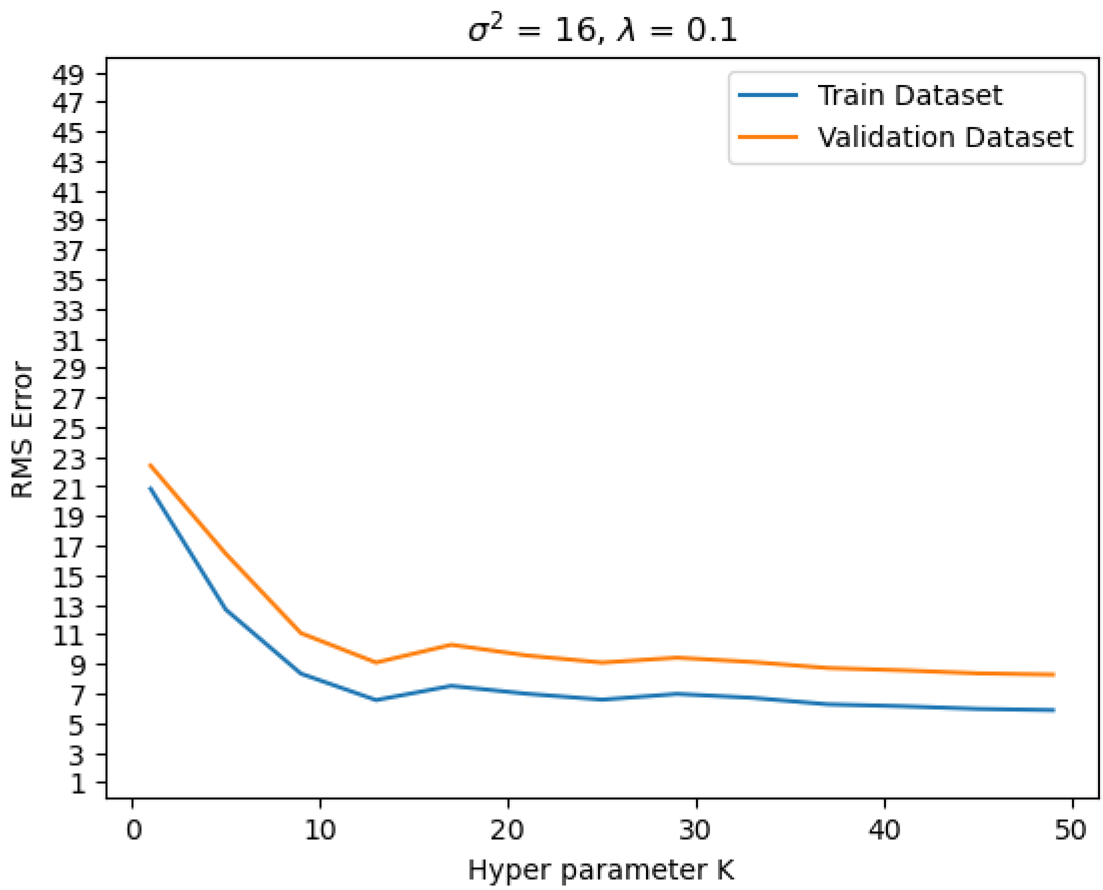
Bivariate Dataset with Regularisation : Variance : 16 and Regularisation : 0.001

```
Regulariation Parameter : 0.001 | Variance : 16
Model RMS Errors for 50 datapoints :

1     | RMS train :    20.85   | RMS validation :    22.41    | RMS test :    24.88
5     | RMS train :    10.46   | RMS validation :    12.21    | RMS test :    12.65
9     | RMS train :     4.87   | RMS validation :     7.37    | RMS test :     8.84
13    | RMS train :     3.45   | RMS validation :     5.88    | RMS test :     7.68
17    | RMS train :     2.33   | RMS validation :     4.35    | RMS test :     5.54
21    | RMS train :     2.56   | RMS validation :     4.58    | RMS test :     6.16
25    | RMS train :     2.22   | RMS validation :     4.17    | RMS test :     5.40
29    | RMS train :     2.13   | RMS validation :     4.09    | RMS test :     5.21
33    | RMS train :     2.14   | RMS validation :     4.10    | RMS test :     5.42
37    | RMS train :     2.39   | RMS validation :     4.25    | RMS test :     5.63
41    | RMS train :     2.05   | RMS validation :     4.02    | RMS test :     5.17
45    | RMS train :     2.08   | RMS validation :     4.02    | RMS test :     5.18
49    | RMS train :     2.05   | RMS validation :     4.00    | RMS test :     5.13
```

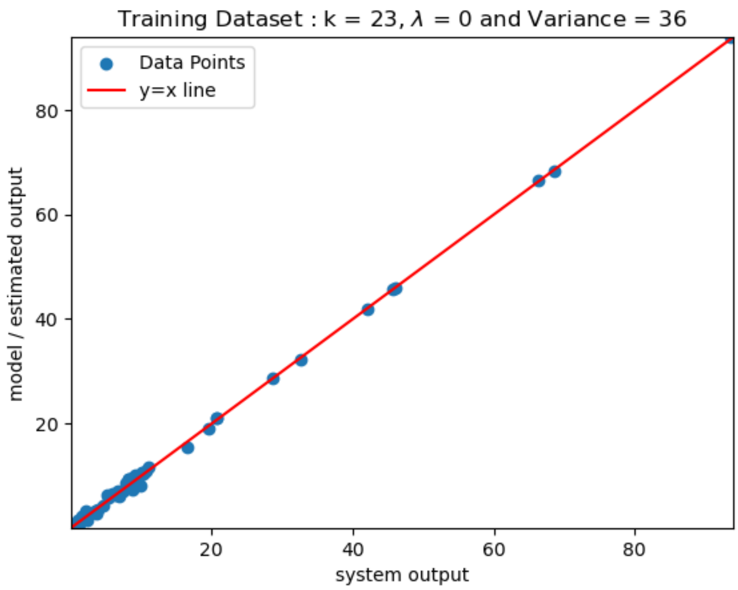


Bivariate Dataset with Regularisation : Variance : 16 and Regularisation : 0.1

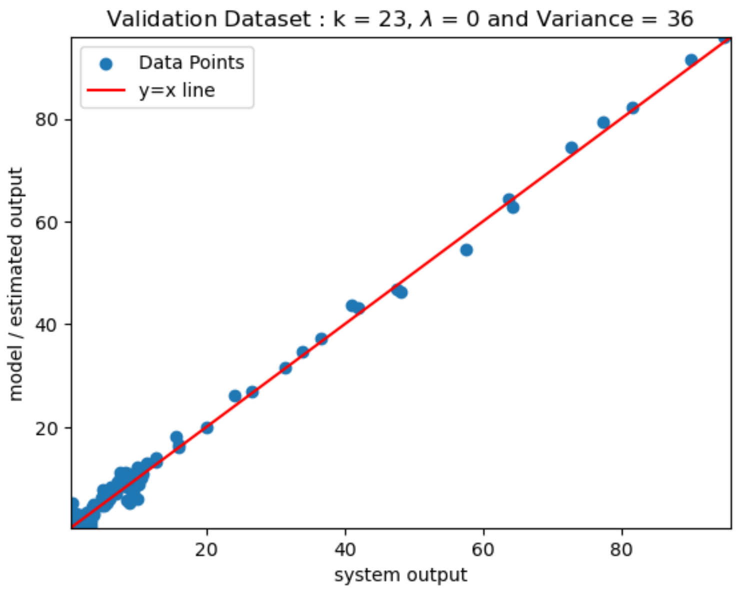
```
Regulariation Parameter : 0.1 | Variance : 16
Model RMS Errors for 50 datapoints :

1     | RMS train :   20.85   | RMS validation :   22.41    | RMS test :   24.88
5     | RMS train :   12.67   | RMS validation :   16.45    | RMS test :   17.47
9     | RMS train :    8.34   | RMS validation :   11.08    | RMS test :   12.67
13    | RMS train :    6.55   | RMS validation :    9.09    | RMS test :   10.46
17    | RMS train :    7.51   | RMS validation :   10.29    | RMS test :   11.63
21    | RMS train :    6.98   | RMS validation :    9.57    | RMS test :   10.70
25    | RMS train :    6.58   | RMS validation :    9.09    | RMS test :   10.26
29    | RMS train :    6.97   | RMS validation :    9.42    | RMS test :   10.44
33    | RMS train :    6.70   | RMS validation :    9.13    | RMS test :   10.28
37    | RMS train :    6.27   | RMS validation :    8.72    | RMS test :    9.96
41    | RMS train :    6.14   | RMS validation :    8.56    | RMS test :    9.80
45    | RMS train :    5.96   | RMS validation :    8.36    | RMS test :    9.63
49    | RMS train :    5.87   | RMS validation :    8.28    | RMS test :    9.53
```



Best Performing Model's Scatter Plot and RMS Error Values [50] :



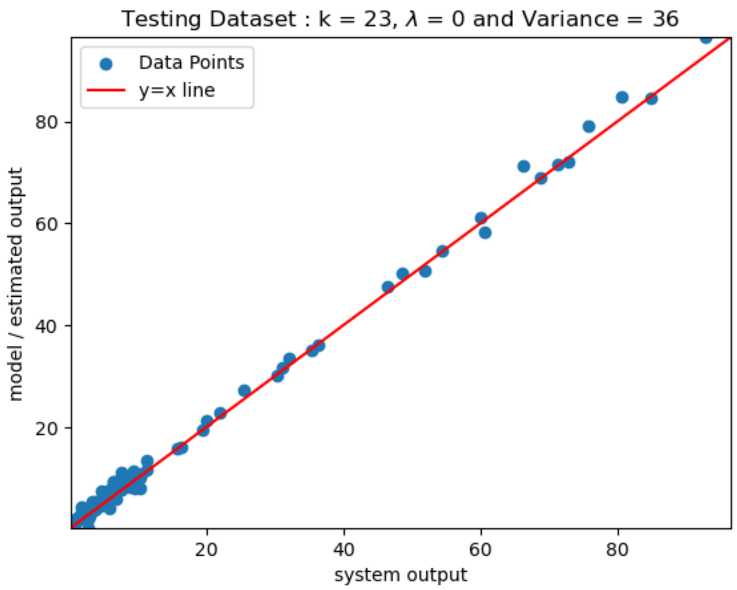
RMS = 0.61



RMS = 1.48

$\lambda = 0$

$\sigma^2 = 9$

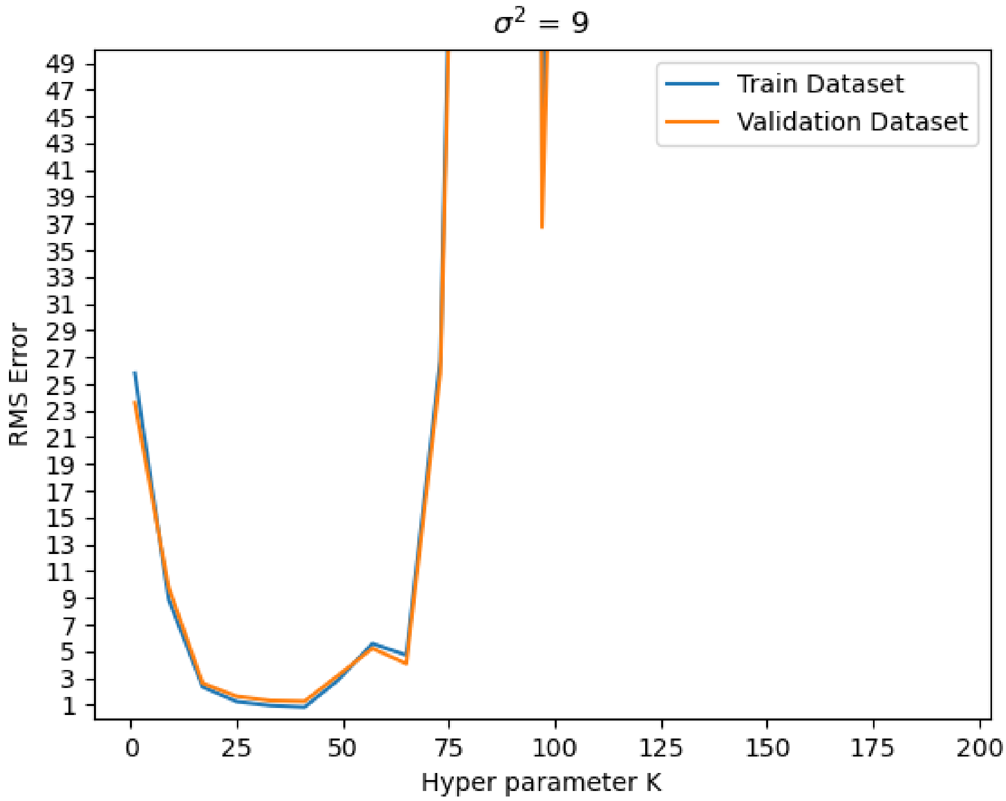


RMS = 1.47

Regression without regularisation for Bivariate Dataset [200 points]

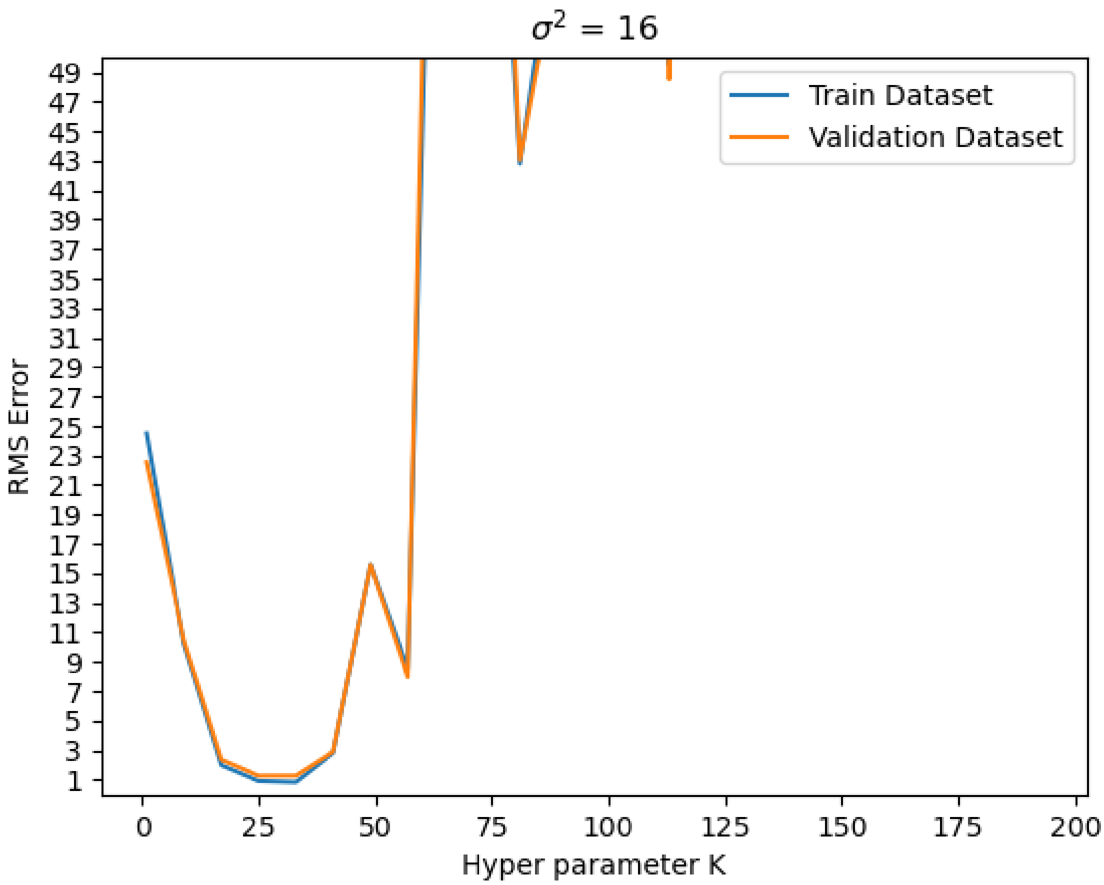
Bivariate Dataset without Regularisation : Variance : 9

1	RMS train :	25.81	RMS validation :	23.60	RMS test :	26.23
9	RMS train :	8.88	RMS validation :	9.77	RMS test :	10.18
17	RMS train :	2.34	RMS validation :	2.61	RMS test :	2.14
25	RMS train :	1.24	RMS validation :	1.63	RMS test :	1.55
33	RMS train :	0.94	RMS validation :	1.31	RMS test :	1.09
41	RMS train :	0.82	RMS validation :	1.26	RMS test :	1.02
49	RMS train :	2.87	RMS validation :	3.22	RMS test :	3.11
57	RMS train :	5.57	RMS validation :	5.22	RMS test :	5.47
65	RMS train :	4.71	RMS validation :	4.07	RMS test :	4.52
73	RMS train :	26.76	RMS validation :	25.76	RMS test :	23.40
81	RMS train :	136.70	RMS validation :	127.41	RMS test :	135.80
89	RMS train :	716.56	RMS validation :	722.17	RMS test :	731.20
97	RMS train :	40.92	RMS validation :	36.71	RMS test :	33.21
105	RMS train :	144.49	RMS validation :	126.09	RMS test :	118.18
113	RMS train :	131.65	RMS validation :	139.82	RMS test :	136.92
121	RMS train :	1689.90	RMS validation :	1875.91	RMS test :	1918.74
129	RMS train :	2665.28	RMS validation :	2689.43	RMS test :	2436.23
137	RMS train :	86.72	RMS validation :	78.60	RMS test :	86.80
145	RMS train :	91.09	RMS validation :	85.71	RMS test :	93.53
153	RMS train :	104.28	RMS validation :	104.98	RMS test :	100.27
161	RMS train :	206.04	RMS validation :	180.10	RMS test :	140.56
169	RMS train :	788.75	RMS validation :	780.58	RMS test :	763.10
177	RMS train :	598.98	RMS validation :	557.60	RMS test :	558.67
185	RMS train :	746.81	RMS validation :	737.04	RMS test :	619.17
193	RMS train :	804.02	RMS validation :	825.47	RMS test :	777.99



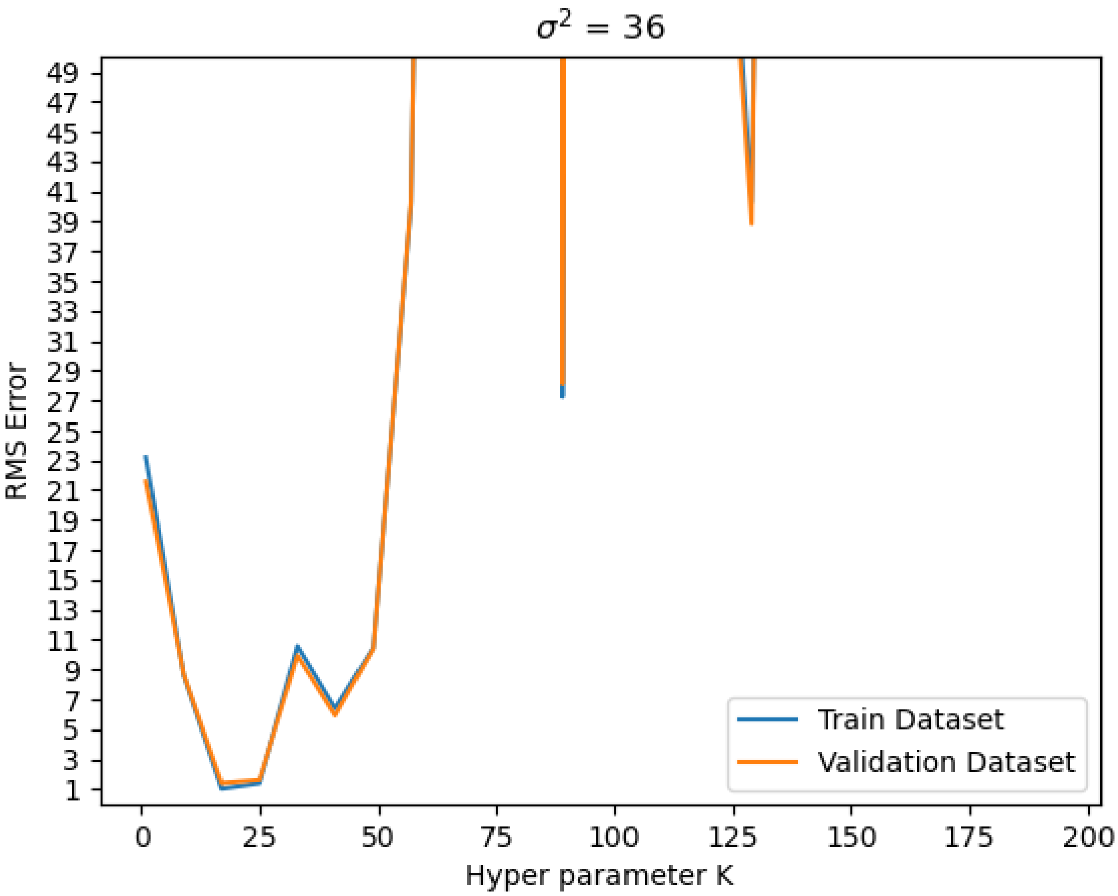
Bivariate Dataset without Regularisation : Variance : 16

```
1    | RMS train :   24.50  | RMS validation :   22.54  | RMS test :   25.00
9    | RMS train :   10.22  | RMS validation :   10.46  | RMS test :    9.75
17   | RMS train :    1.98  | RMS validation :    2.35  | RMS test :    2.07
25   | RMS train :    0.91  | RMS validation :    1.27  | RMS test :    1.03
33   | RMS train :    0.84  | RMS validation :    1.27  | RMS test :    1.04
41   | RMS train :    2.81  | RMS validation :    2.90  | RMS test :    2.65
49   | RMS train :   15.58  | RMS validation :   15.57  | RMS test :   15.20
57   | RMS train :    8.52  | RMS validation :    7.94  | RMS test :    8.72
65   | RMS train :  101.94  | RMS validation :  115.60  | RMS test :  110.98
73   | RMS train :   84.22  | RMS validation :   95.19  | RMS test :   98.33
81   | RMS train :   42.81  | RMS validation :   43.06  | RMS test :   45.69
89   | RMS train :   60.15  | RMS validation :   56.75  | RMS test :   55.11
97   | RMS train :  124.73  | RMS validation :  127.35  | RMS test :  121.16
105  | RMS train :  305.47  | RMS validation :  303.05  | RMS test :  298.68
113  | RMS train :   49.46  | RMS validation :   48.51  | RMS test :   47.08
121  | RMS train :  173.91  | RMS validation :  176.82  | RMS test :  175.55
129  | RMS train :  385.79  | RMS validation :  437.04  | RMS test :  465.27
137  | RMS train :   68.15  | RMS validation :   71.04  | RMS test :   74.49
145  | RMS train :  843.00  | RMS validation :  829.91  | RMS test :  882.38
153  | RMS train :  451.55  | RMS validation :  487.60  | RMS test :  470.72
161  | RMS train :  132.51  | RMS validation :  129.33  | RMS test :  122.68
169  | RMS train :  248.80  | RMS validation :  251.55  | RMS test :  247.34
177  | RMS train :  144.09  | RMS validation :  157.85  | RMS test :  153.32
185  | RMS train :  459.83  | RMS validation :  482.01  | RMS test :  446.68
193  | RMS train :  810.58  | RMS validation :  819.28  | RMS test :  768.14
```



Bivariate Dataset without Regularisation: Variance : 36

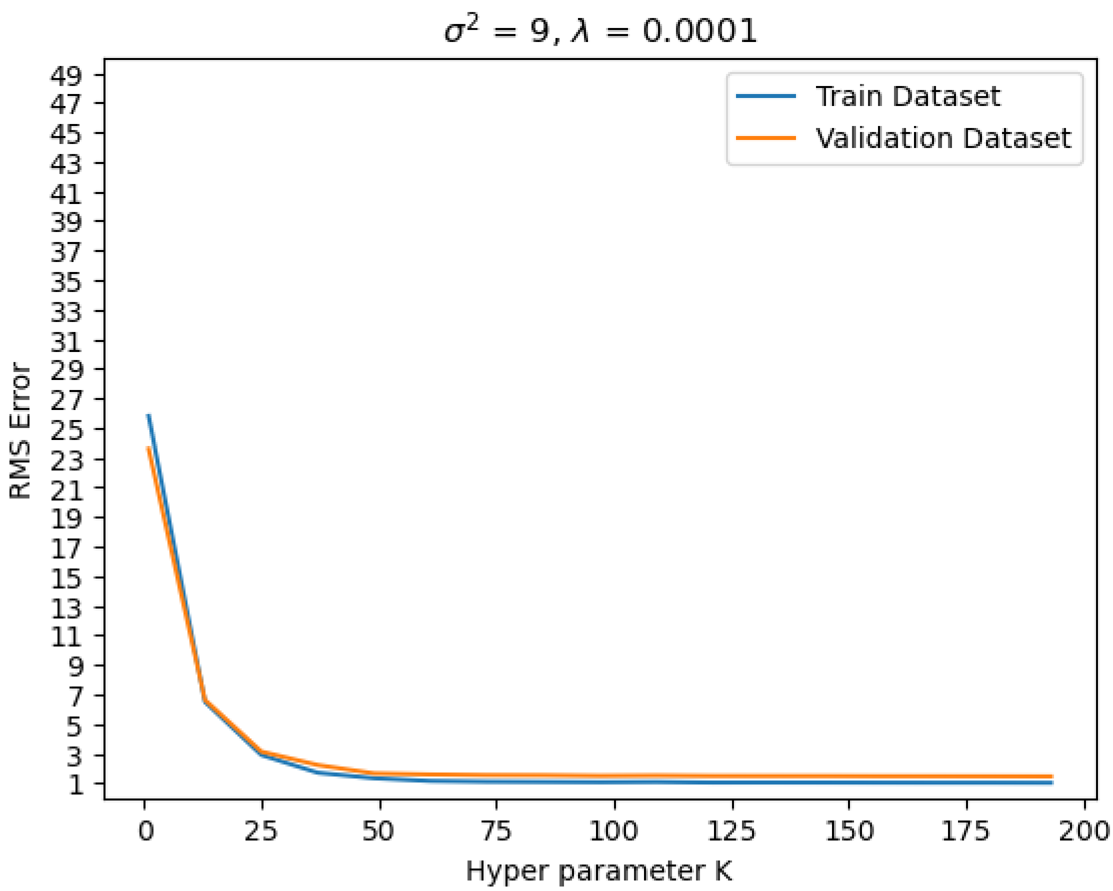
```
1     | RMS train :    23.21   | RMS validation :    21.58   | RMS test :    23.72
9     | RMS train :     8.61   | RMS validation :     8.68   | RMS test :     7.84
17    | RMS train :     1.03   | RMS validation :     1.39   | RMS test :     1.14
25    | RMS train :     1.39   | RMS validation :     1.60   | RMS test :     1.43
33    | RMS train :    10.57   | RMS validation :     9.93   | RMS test :    10.06
41    | RMS train :     6.39   | RMS validation :     5.92   | RMS test :     5.60
49    | RMS train :    10.45   | RMS validation :    10.39   | RMS test :    10.21
57    | RMS train :    40.09   | RMS validation :    40.30   | RMS test :    40.57
65    | RMS train :   180.65   | RMS validation :   179.63   | RMS test :   177.43
73    | RMS train :   223.89   | RMS validation :   222.76   | RMS test :   239.98
81    | RMS train : 16776.69   | RMS validation : 16088.02   | RMS test : 15742.11
89    | RMS train :    27.24   | RMS validation :    28.10   | RMS test :    29.01
97    | RMS train :   638.46   | RMS validation :   618.79   | RMS test :   608.18
105   | RMS train :   136.10   | RMS validation :   140.87   | RMS test :   142.94
113   | RMS train :   165.51   | RMS validation :   188.02   | RMS test :   198.50
121   | RMS train :    79.29   | RMS validation :    75.82   | RMS test :    66.85
129   | RMS train :    40.24   | RMS validation :    38.86   | RMS test :    37.11
137   | RMS train :   201.60   | RMS validation :   207.70   | RMS test :   204.05
145   | RMS train :   592.66   | RMS validation :   588.15   | RMS test :   569.80
153   | RMS train :   688.25   | RMS validation :   727.79   | RMS test :   758.27
161   | RMS train :    54.82   | RMS validation :    53.30   | RMS test :    49.62
169   | RMS train :    59.87   | RMS validation :    59.62   | RMS test :    58.12
177   | RMS train :    90.78   | RMS validation :    85.58   | RMS test :    80.64
185   | RMS train :   164.36   | RMS validation :   159.29   | RMS test :   156.97
193   | RMS train :   450.48   | RMS validation :   434.86   | RMS test :   419.67
```



Regression with regularisation for Bivariate Dataset [200 points]

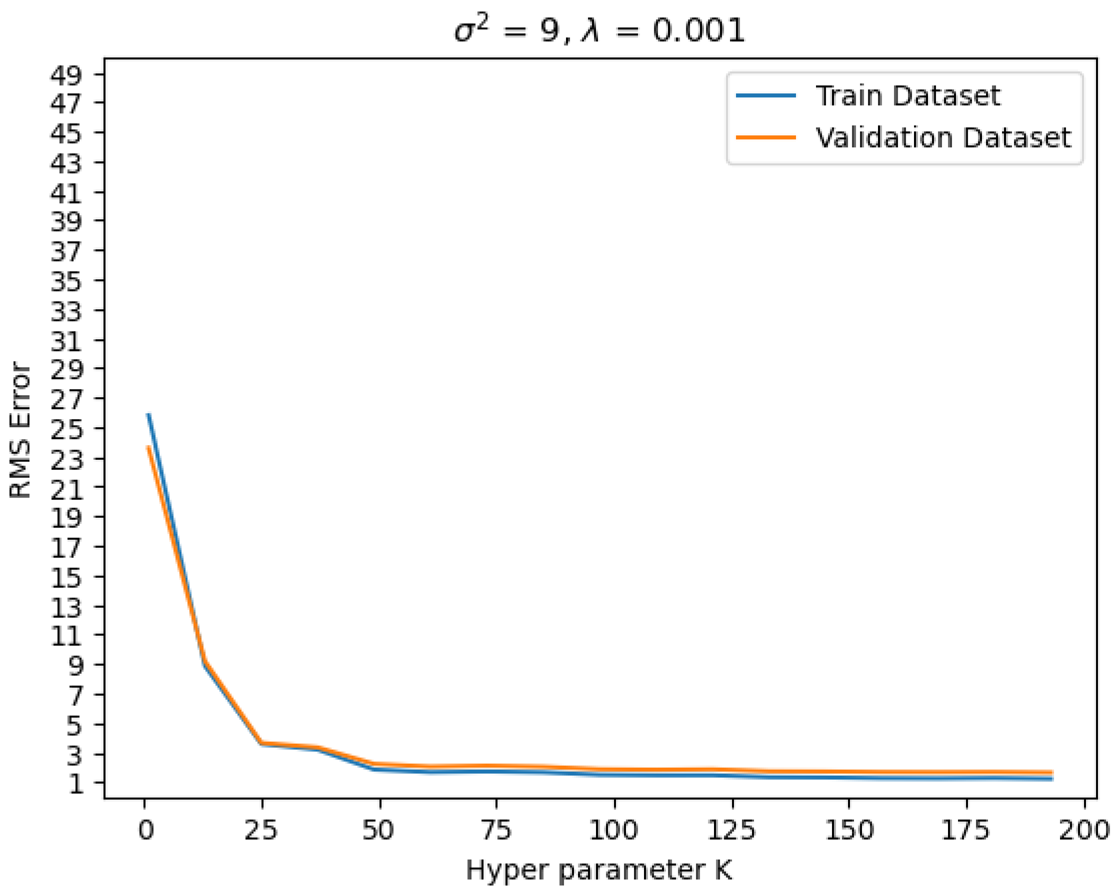
Bivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.0001

```
Regulariation Parameter : 0.0001 | Variance : 9
Model RMS Errors for 200 datapoints :

1     | RMS train :   25.81   | RMS validation :   23.60   | RMS test :   26.23
13    | RMS train :    6.50   | RMS validation :    6.63   | RMS test :    6.48
25    | RMS train :    2.91   | RMS validation :    3.13   | RMS test :    2.87
37    | RMS train :    1.71   | RMS validation :    2.23   | RMS test :    1.90
49    | RMS train :    1.33   | RMS validation :    1.66   | RMS test :    1.50
61    | RMS train :    1.15   | RMS validation :    1.59   | RMS test :    1.37
73    | RMS train :    1.10   | RMS validation :    1.55   | RMS test :    1.34
85    | RMS train :    1.08   | RMS validation :    1.53   | RMS test :    1.33
97    | RMS train :    1.07   | RMS validation :    1.50   | RMS test :    1.29
109   | RMS train :    1.08   | RMS validation :    1.52   | RMS test :    1.31
121   | RMS train :    1.04   | RMS validation :    1.49   | RMS test :    1.28
133   | RMS train :    1.04   | RMS validation :    1.49   | RMS test :    1.28
145   | RMS train :    1.04   | RMS validation :    1.48   | RMS test :    1.28
157   | RMS train :    1.03   | RMS validation :    1.47   | RMS test :    1.27
169   | RMS train :    1.03   | RMS validation :    1.47   | RMS test :    1.26
181   | RMS train :    1.03   | RMS validation :    1.47   | RMS test :    1.26
193   | RMS train :    1.02   | RMS validation :    1.46   | RMS test :    1.25
```



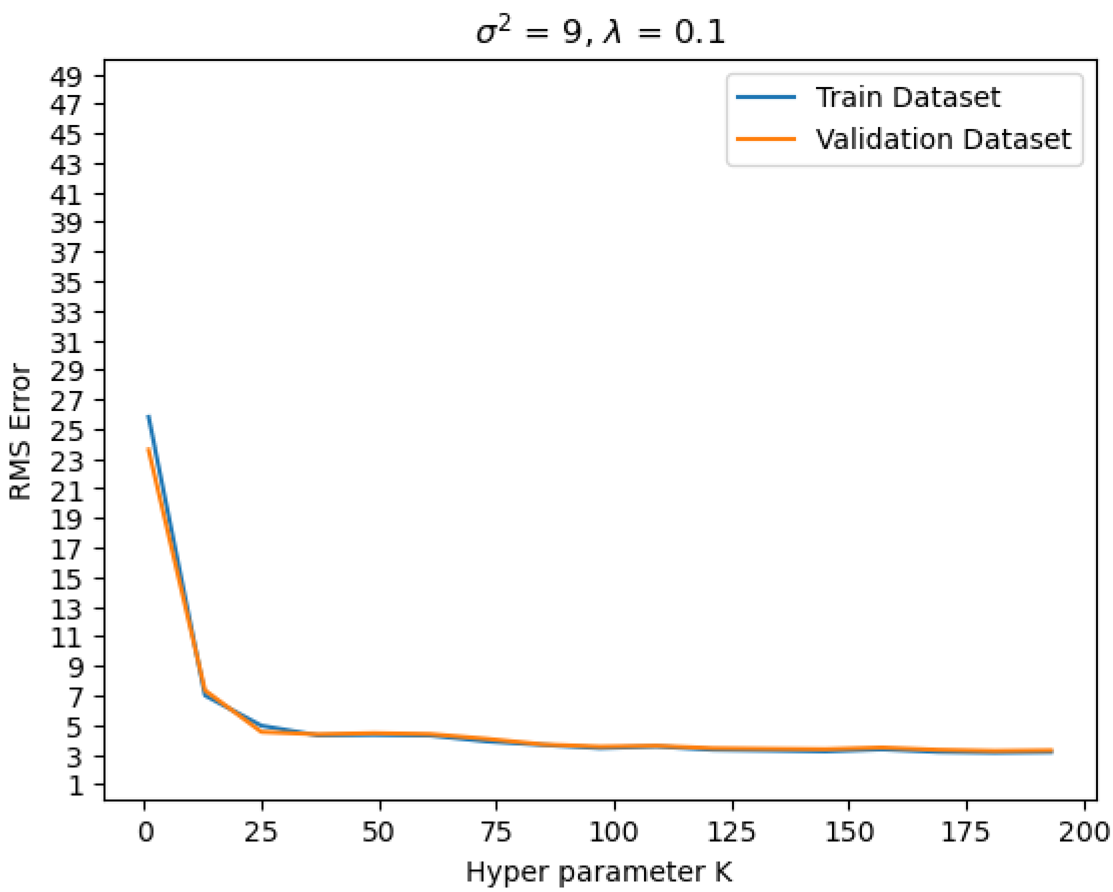
Bivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.001

```
Regulariation Parameter : 0.001 | Variance : 9
Model RMS Errors for 200 datapoints :

1      | RMS train :   25.81   | RMS validation :   23.60    | RMS test :   26.23
13     | RMS train :    8.86   | RMS validation :    9.19    | RMS test :   10.20
25     | RMS train :    3.58   | RMS validation :    3.66    | RMS test :    3.94
37     | RMS train :    3.20   | RMS validation :    3.36    | RMS test :    3.25
49     | RMS train :    1.85   | RMS validation :    2.24    | RMS test :    1.98
61     | RMS train :    1.68   | RMS validation :    2.06    | RMS test :    1.82
73     | RMS train :    1.73   | RMS validation :    2.12    | RMS test :    1.84
85     | RMS train :    1.67   | RMS validation :    2.05    | RMS test :    1.80
97     | RMS train :    1.50   | RMS validation :    1.88    | RMS test :    1.65
109    | RMS train :    1.47   | RMS validation :    1.85    | RMS test :    1.64
121    | RMS train :    1.47   | RMS validation :    1.88    | RMS test :    1.64
133    | RMS train :    1.34   | RMS validation :    1.74    | RMS test :    1.51
145    | RMS train :    1.32   | RMS validation :    1.73    | RMS test :    1.51
157    | RMS train :    1.27   | RMS validation :    1.69    | RMS test :    1.47
169    | RMS train :    1.26   | RMS validation :    1.68    | RMS test :    1.46
181    | RMS train :    1.28   | RMS validation :    1.69    | RMS test :    1.47
193    | RMS train :    1.23   | RMS validation :    1.66    | RMS test :    1.44
```



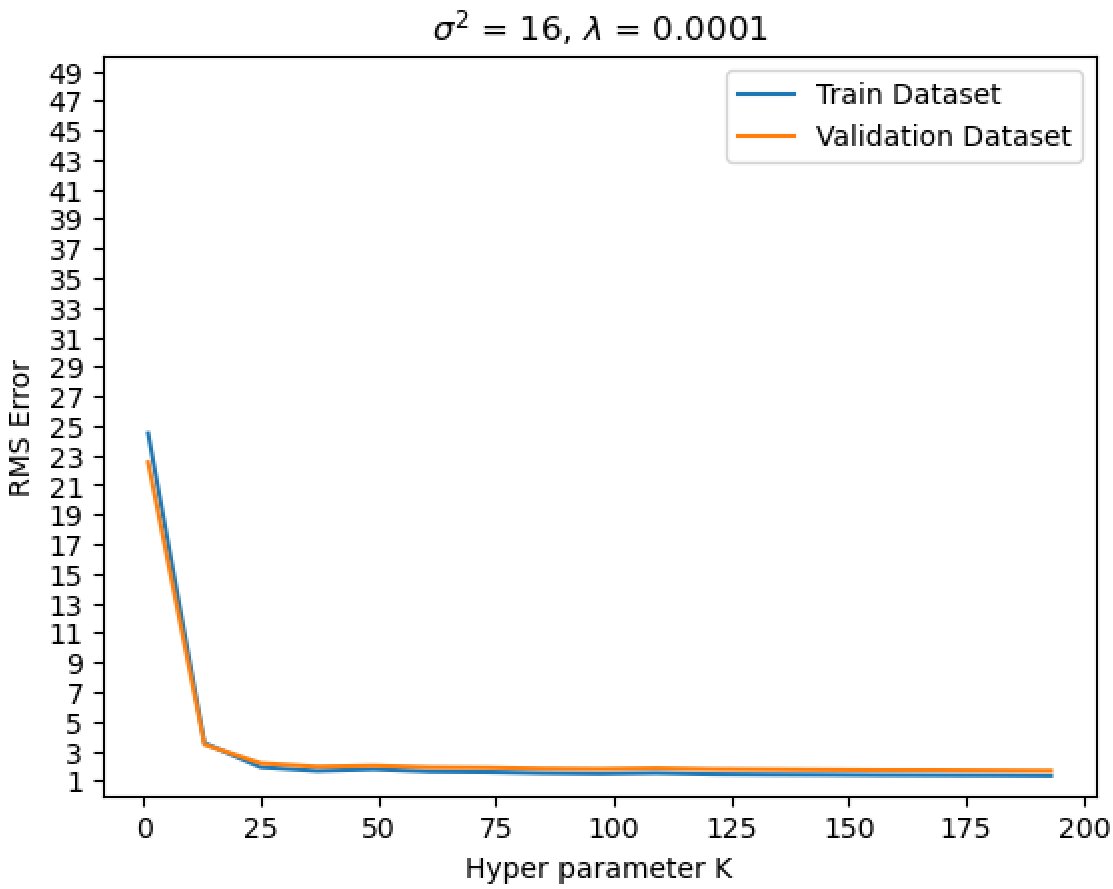
Bivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.1

```
Regulariation Parameter : 0.1 | Variance : 9
Model RMS Errors for 200 datapoints :

1     | RMS train :   25.81   | RMS validation :   23.60   | RMS test :   26.23
13    | RMS train :    7.00   | RMS validation :    7.35   | RMS test :    7.57
25    | RMS train :    4.95   | RMS validation :    4.53   | RMS test :    4.73
37    | RMS train :    4.30   | RMS validation :    4.37   | RMS test :    4.11
49    | RMS train :    4.33   | RMS validation :    4.45   | RMS test :    4.24
61    | RMS train :    4.29   | RMS validation :    4.39   | RMS test :    4.22
73    | RMS train :    3.89   | RMS validation :    4.07   | RMS test :    3.78
85    | RMS train :    3.65   | RMS validation :    3.69   | RMS test :    3.56
97    | RMS train :    3.44   | RMS validation :    3.54   | RMS test :    3.33
109   | RMS train :    3.55   | RMS validation :    3.59   | RMS test :    3.48
121   | RMS train :    3.33   | RMS validation :    3.43   | RMS test :    3.23
133   | RMS train :    3.28   | RMS validation :    3.39   | RMS test :    3.19
145   | RMS train :    3.22   | RMS validation :    3.37   | RMS test :    3.15
157   | RMS train :    3.36   | RMS validation :    3.48   | RMS test :    3.29
169   | RMS train :    3.18   | RMS validation :    3.32   | RMS test :    3.12
181   | RMS train :    3.13   | RMS validation :    3.25   | RMS test :    3.08
193   | RMS train :    3.17   | RMS validation :    3.29   | RMS test :    3.13
```



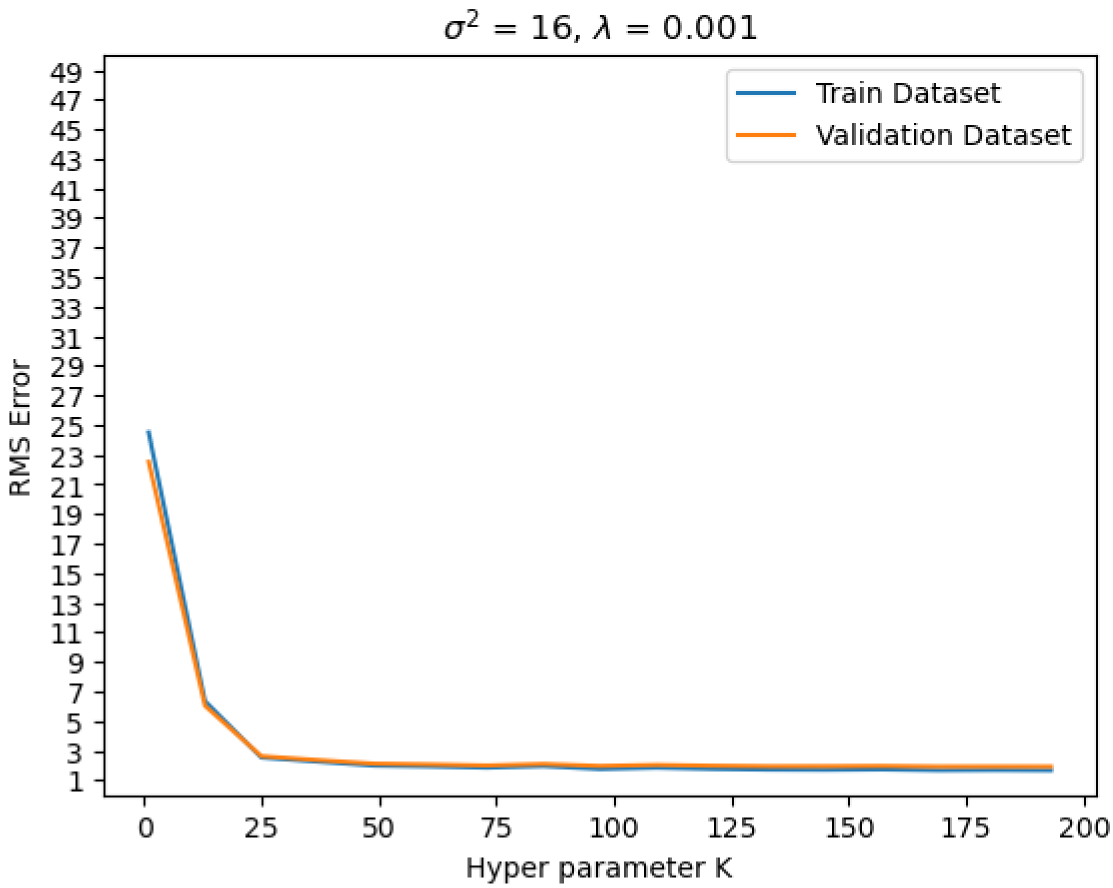
Bivariate Dataset with Regularisation : Variance : 16 and Regularisation : 0.0001

```
Regulariation Parameter : 0.0001 | Variance : 16
Model RMS Errors for 200 datapoints :

1      | RMS train :    24.50    | RMS validation :    22.54     | RMS test :    25.00
13     | RMS train :     3.53    | RMS validation :     3.43     | RMS test :     3.48
25     | RMS train :     1.91    | RMS validation :     2.18     | RMS test :     2.01
37     | RMS train :     1.66    | RMS validation :     1.95     | RMS test :     1.79
49     | RMS train :     1.77    | RMS validation :     2.02     | RMS test :     1.89
61     | RMS train :     1.62    | RMS validation :     1.91     | RMS test :     1.75
73     | RMS train :     1.59    | RMS validation :     1.89     | RMS test :     1.74
85     | RMS train :     1.51    | RMS validation :     1.82     | RMS test :     1.66
97     | RMS train :     1.47    | RMS validation :     1.79     | RMS test :     1.63
109    | RMS train :     1.52    | RMS validation :     1.84     | RMS test :     1.67
121    | RMS train :     1.44    | RMS validation :     1.78     | RMS test :     1.59
133    | RMS train :     1.42    | RMS validation :     1.76     | RMS test :     1.58
145    | RMS train :     1.40    | RMS validation :     1.74     | RMS test :     1.56
157    | RMS train :     1.38    | RMS validation :     1.72     | RMS test :     1.54
169    | RMS train :     1.36    | RMS validation :     1.72     | RMS test :     1.53
181    | RMS train :     1.35    | RMS validation :     1.70     | RMS test :     1.52
193    | RMS train :     1.33    | RMS validation :     1.68     | RMS test :     1.50
```



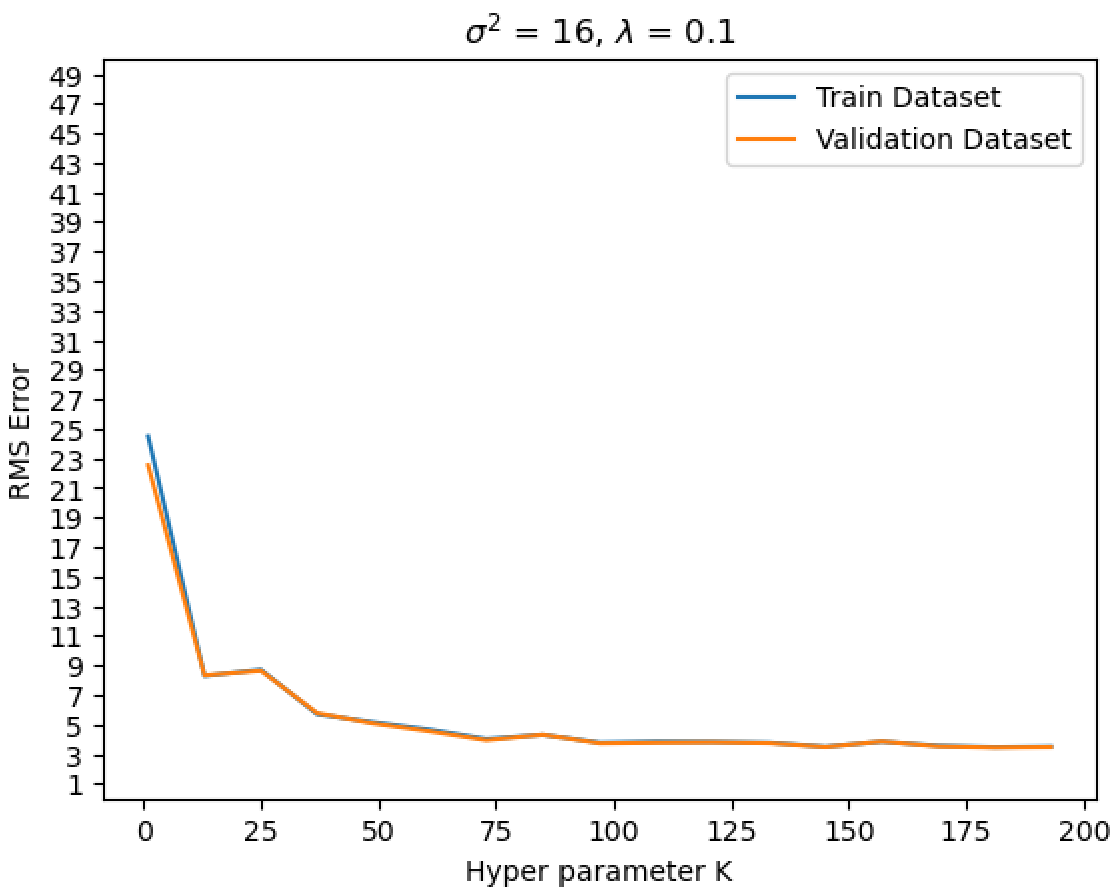
Bivariate Dataset with Regularisation : Variance : 16 and Regularisation : 0.001

```
Regulariation Parameter : 0.001 | Variance : 16
Model RMS Errors for 200 datapoints :

1     | RMS train :   24.50   | RMS validation :   22.54     | RMS test :   25.00
13    | RMS train :    6.35   | RMS validation :    6.03     | RMS test :    6.16
25    | RMS train :    2.53   | RMS validation :    2.62     | RMS test :    2.47
37    | RMS train :    2.25   | RMS validation :    2.36     | RMS test :    2.25
49    | RMS train :    1.98   | RMS validation :    2.12     | RMS test :    2.02
61    | RMS train :    1.93   | RMS validation :    2.06     | RMS test :    1.98
73    | RMS train :    1.85   | RMS validation :    1.99     | RMS test :    1.92
85    | RMS train :    1.95   | RMS validation :    2.09     | RMS test :    2.01
97    | RMS train :    1.75   | RMS validation :    1.96     | RMS test :    1.85
109   | RMS train :    1.84   | RMS validation :    2.03     | RMS test :    1.92
121   | RMS train :    1.76   | RMS validation :    1.97     | RMS test :    1.86
133   | RMS train :    1.71   | RMS validation :    1.94     | RMS test :    1.82
145   | RMS train :    1.70   | RMS validation :    1.94     | RMS test :    1.81
157   | RMS train :    1.72   | RMS validation :    1.96     | RMS test :    1.83
169   | RMS train :    1.67   | RMS validation :    1.92     | RMS test :    1.79
181   | RMS train :    1.68   | RMS validation :    1.92     | RMS test :    1.80
193   | RMS train :    1.67   | RMS validation :    1.92     | RMS test :    1.79
```

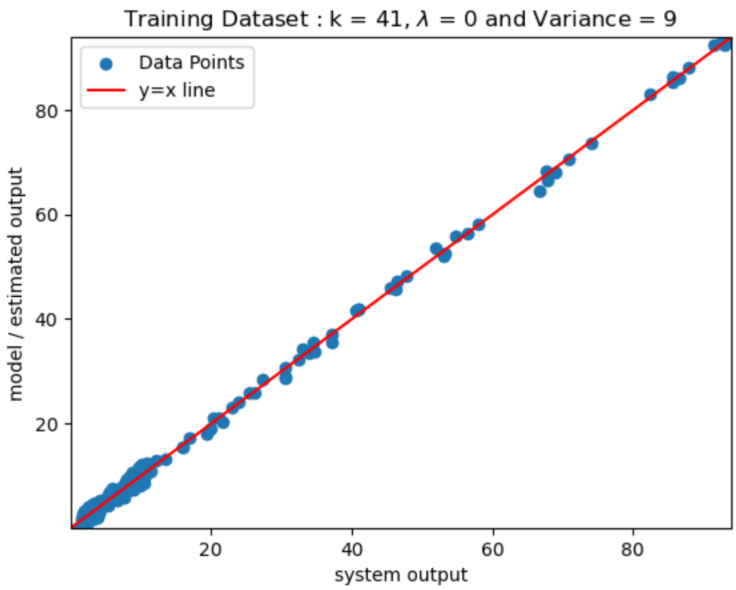


Bivariate Dataset with Regularisation : Variance : 16 and Regularisation : 0.1

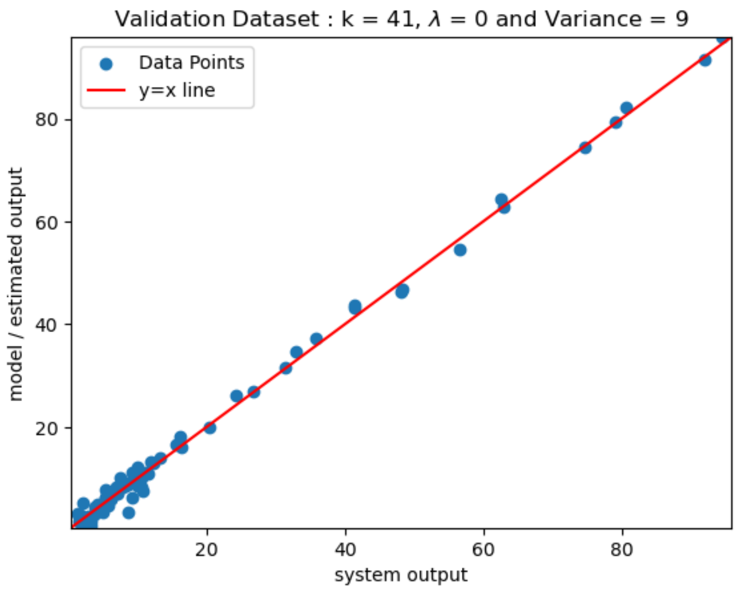
```
Regulariation Parameter : 0.1 | Variance : 16
Model RMS Errors for 200 datapoints :

1     | RMS train :   24.50   | RMS validation :   22.54   | RMS test :   25.00
13    | RMS train :    8.29   | RMS validation :    8.33   | RMS test :    8.21
25    | RMS train :    8.71   | RMS validation :    8.65   | RMS test :    8.76
37    | RMS train :    5.70   | RMS validation :    5.77   | RMS test :    6.12
49    | RMS train :    5.15   | RMS validation :    5.07   | RMS test :    5.18
61    | RMS train :    4.65   | RMS validation :    4.54   | RMS test :    4.69
73    | RMS train :    4.02   | RMS validation :    3.95   | RMS test :    4.06
85    | RMS train :    4.31   | RMS validation :    4.31   | RMS test :    4.44
97    | RMS train :    3.78   | RMS validation :    3.73   | RMS test :    3.86
109   | RMS train :    3.84   | RMS validation :    3.77   | RMS test :    3.90
121   | RMS train :    3.82   | RMS validation :    3.80   | RMS test :    3.87
133   | RMS train :    3.77   | RMS validation :    3.75   | RMS test :    3.87
145   | RMS train :    3.50   | RMS validation :    3.48   | RMS test :    3.53
157   | RMS train :    3.84   | RMS validation :    3.86   | RMS test :    3.98
169   | RMS train :    3.56   | RMS validation :    3.52   | RMS test :    3.61
181   | RMS train :    3.47   | RMS validation :    3.45   | RMS test :    3.50
193   | RMS train :    3.51   | RMS validation :    3.48   | RMS test :    3.52
```



Best Performing Model's Scatter Plot and RMS Error Values [200] :



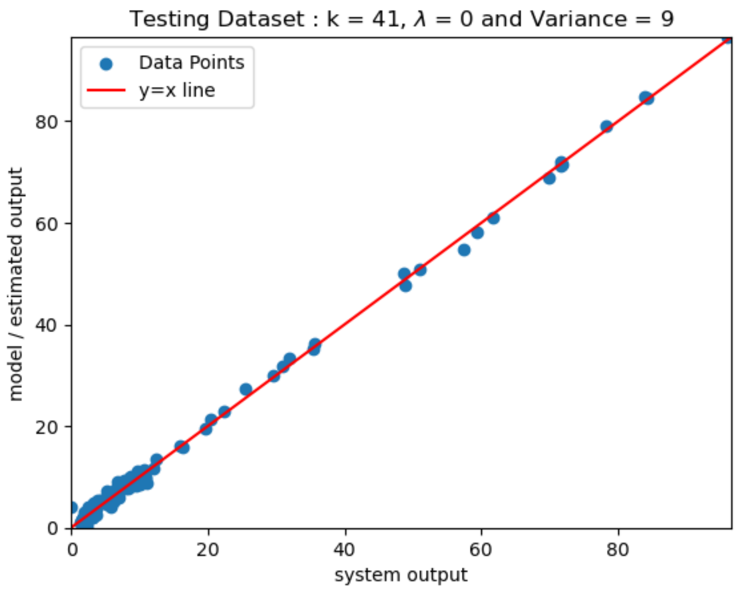
RMS = 0.82



RMS = 1.26

λ = 0

σ^2 = 9

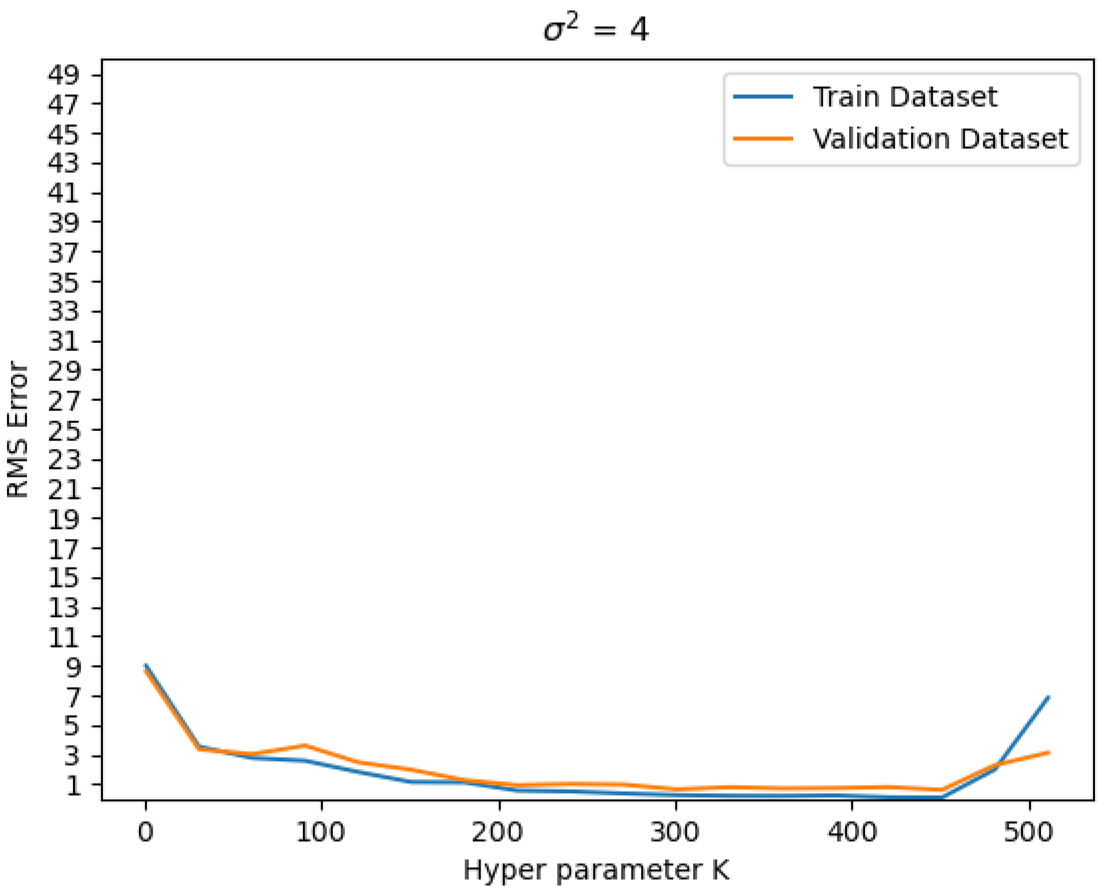


RMS = 1.02

Linear Regression without Regularisation for Multivariate Dataset

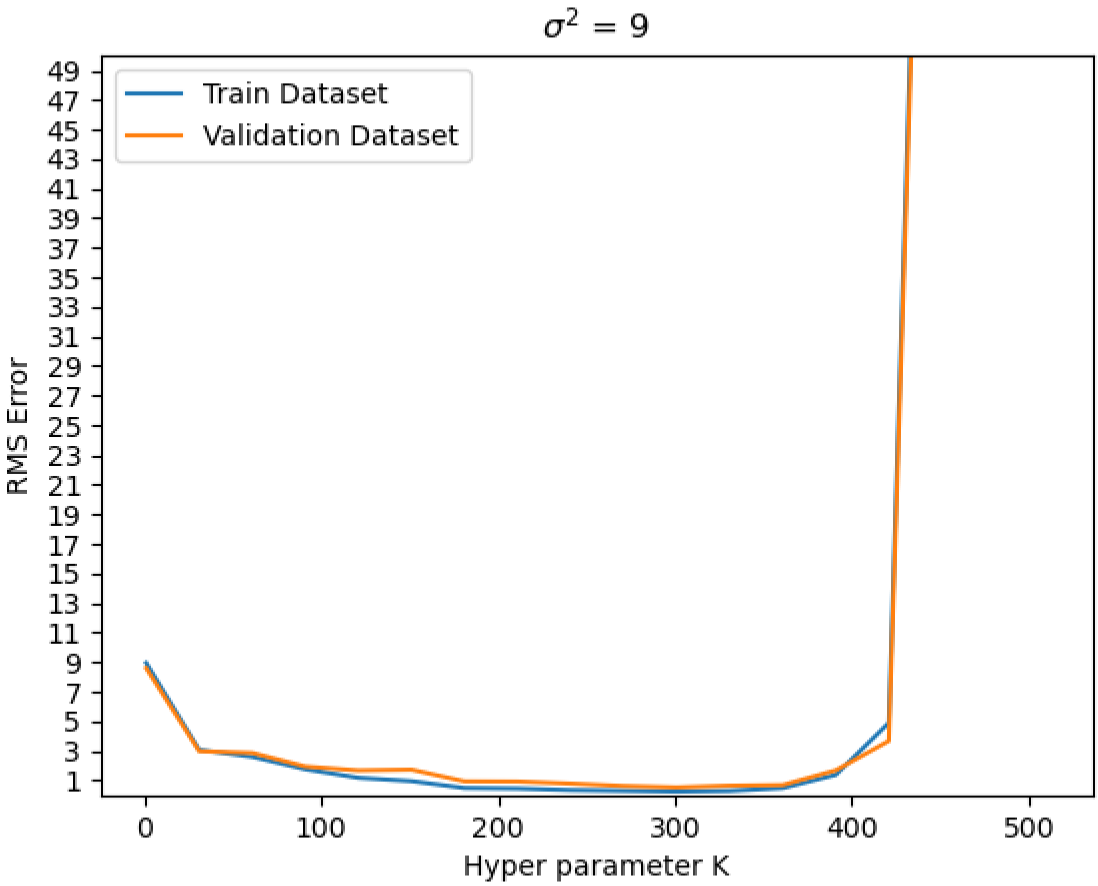
Multivariate Dataset without Regularisation : Variance : 4

```
k : 1     | RMS train :   9.00   | RMS validation :   8.63   | RMS test :   9.32
k : 31    | RMS train :   3.53   | RMS validation :   3.35   | RMS test :   4.07
k : 61    | RMS train :   2.77   | RMS validation :   3.03   | RMS test :   3.73
k : 91    | RMS train :   2.58   | RMS validation :   3.60   | RMS test :   3.14
k : 121   | RMS train :   1.82   | RMS validation :   2.48   | RMS test :   2.79
k : 151   | RMS train :   1.14   | RMS validation :   1.97   | RMS test :   1.97
k : 181   | RMS train :   1.12   | RMS validation :   1.27   | RMS test :   1.60
k : 211   | RMS train :   0.57   | RMS validation :   0.93   | RMS test :   1.16
k : 241   | RMS train :   0.50   | RMS validation :   1.01   | RMS test :   1.38
k : 271   | RMS train :   0.37   | RMS validation :   0.97   | RMS test :   0.79
k : 301   | RMS train :   0.26   | RMS validation :   0.65   | RMS test :   0.83
k : 331   | RMS train :   0.21   | RMS validation :   0.79   | RMS test :   0.89
k : 361   | RMS train :   0.20   | RMS validation :   0.71   | RMS test :   0.78
k : 391   | RMS train :   0.23   | RMS validation :   0.73   | RMS test :   0.80
k : 421   | RMS train :   0.11   | RMS validation :   0.80   | RMS test :   0.97
k : 451   | RMS train :   0.09   | RMS validation :   0.61   | RMS test :   0.96
k : 481   | RMS train :   1.99   | RMS validation :   2.28   | RMS test :   2.02
k : 511   | RMS train :   6.84   | RMS validation :   3.12   | RMS test :   4.90
```



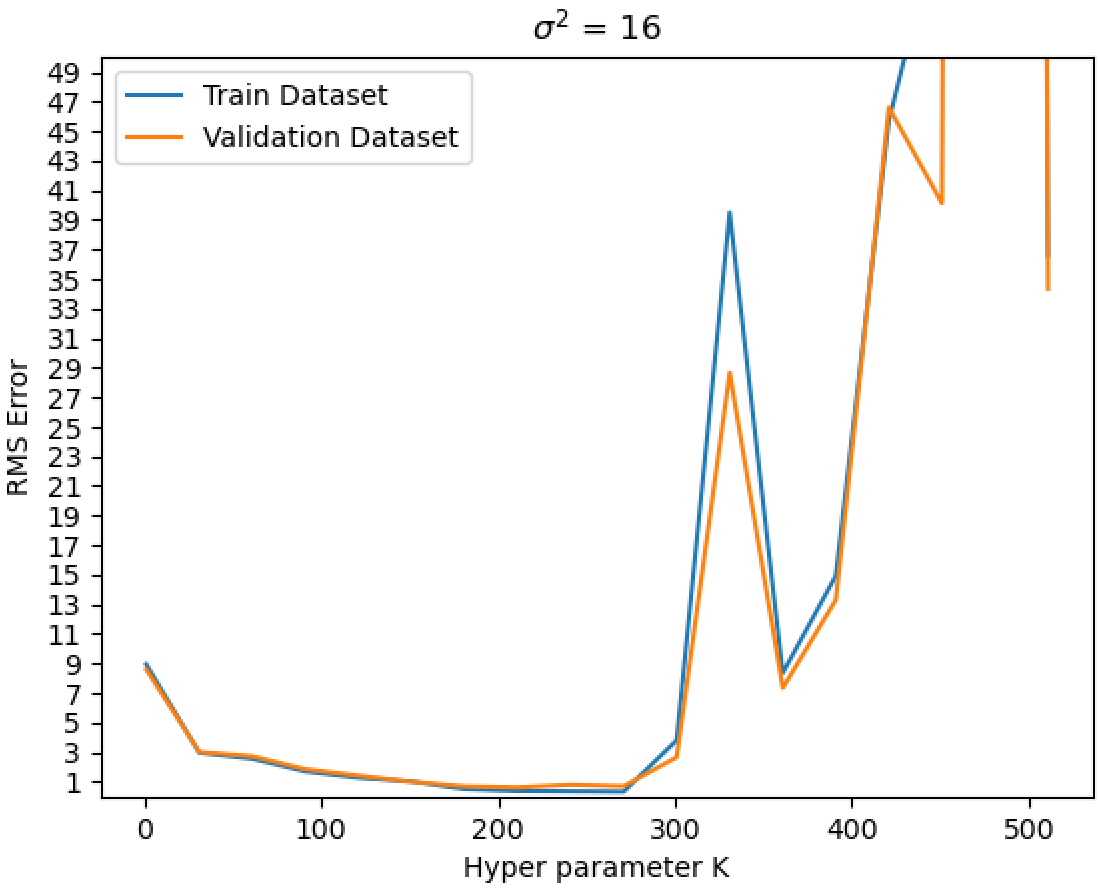
Multivariate Dataset without Regularisation : Variance : 9

```
k : 1     | RMS train :    8.95     | RMS validation :    8.60    | RMS test :    9.27
k : 31    | RMS train :    3.06     | RMS validation :    2.97    | RMS test :    3.67
k : 61    | RMS train :    2.59     | RMS validation :    2.85    | RMS test :    2.81
k : 91    | RMS train :    1.74     | RMS validation :    1.93    | RMS test :    1.95
k : 121   | RMS train :    1.16     | RMS validation :    1.66    | RMS test :    1.62
k : 151   | RMS train :    0.94     | RMS validation :    1.71    | RMS test :    1.52
k : 181   | RMS train :    0.47     | RMS validation :    0.92    | RMS test :    1.12
k : 211   | RMS train :    0.43     | RMS validation :    0.91    | RMS test :    0.95
k : 241   | RMS train :    0.33     | RMS validation :    0.77    | RMS test :    0.61
k : 271   | RMS train :    0.28     | RMS validation :    0.60    | RMS test :    0.80
k : 301   | RMS train :    0.23     | RMS validation :    0.51    | RMS test :    0.68
k : 331   | RMS train :    0.28     | RMS validation :    0.63    | RMS test :    0.75
k : 361   | RMS train :    0.46     | RMS validation :    0.67    | RMS test :    0.83
k : 391   | RMS train :    1.36     | RMS validation :    1.67    | RMS test :    1.41
k : 421   | RMS train :    4.89     | RMS validation :    3.66    | RMS test :    5.33
k : 451   | RMS train :  117.40     | RMS validation :  114.05    | RMS test :  110.14
k : 481   | RMS train :   84.50     | RMS validation :   85.86    | RMS test :   85.71
k : 511   | RMS train : 1112.61     | RMS validation : 1139.95    | RMS test : 1151.19
```



Multivariate Dataset without Regularisation : Variance : 16

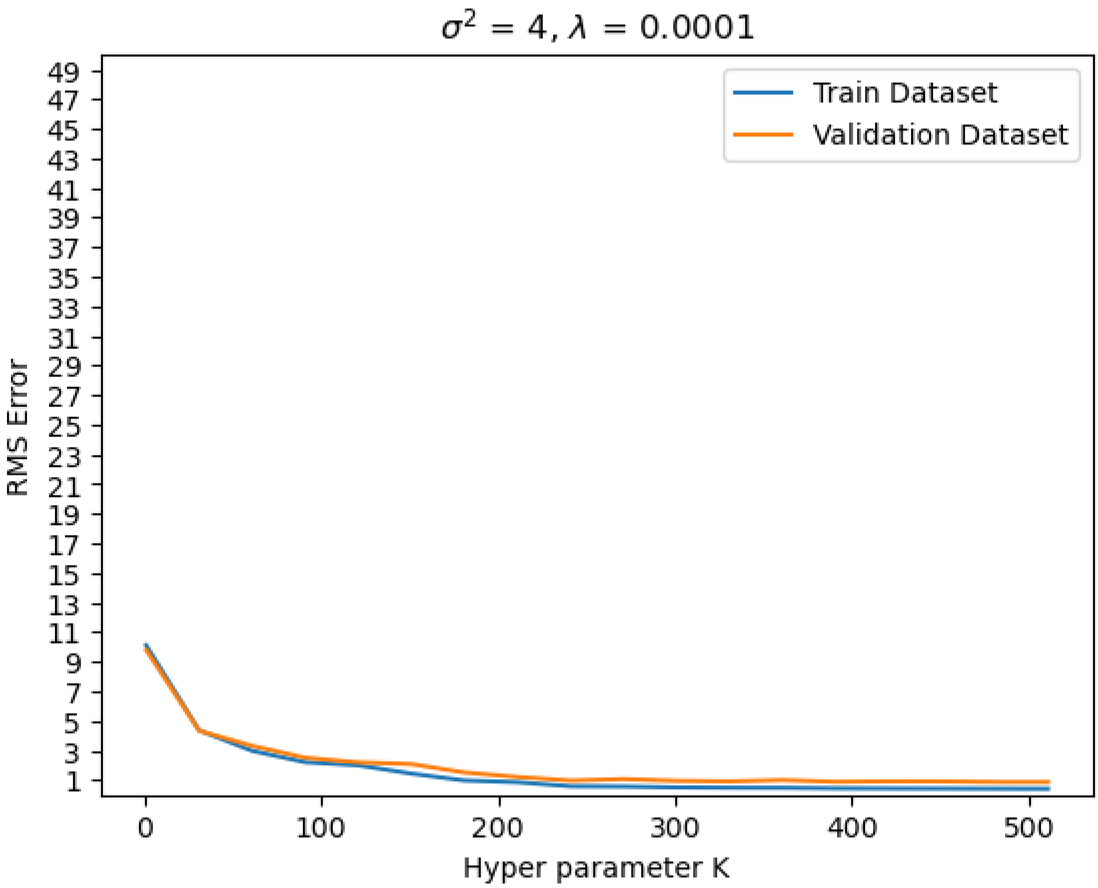
```
k : 1     | RMS train :    8.94     | RMS validation :    8.59    | RMS test :    9.26
k : 31    | RMS train :    2.96     | RMS validation :    3.00    | RMS test :    3.24
k : 61    | RMS train :    2.55     | RMS validation :    2.72    | RMS test :    2.98
k : 91    | RMS train :    1.70     | RMS validation :    1.84    | RMS test :    1.66
k : 121   | RMS train :    1.27     | RMS validation :    1.41    | RMS test :    2.18
k : 151   | RMS train :    1.01     | RMS validation :    0.99    | RMS test :    1.19
k : 181   | RMS train :    0.51     | RMS validation :    0.68    | RMS test :    0.88
k : 211   | RMS train :    0.37     | RMS validation :    0.61    | RMS test :    0.71
k : 241   | RMS train :    0.35     | RMS validation :    0.79    | RMS test :    0.84
k : 271   | RMS train :    0.32     | RMS validation :    0.69    | RMS test :    0.92
k : 301   | RMS train :    3.80     | RMS validation :    2.66    | RMS test :    3.70
k : 331   | RMS train :   39.51     | RMS validation :   28.67    | RMS test :   40.52
k : 361   | RMS train :    8.35     | RMS validation :    7.34    | RMS test :    7.57
k : 391   | RMS train :   14.92     | RMS validation :   13.30    | RMS test :   15.05
k : 421   | RMS train :   45.70     | RMS validation :   46.60    | RMS test :   42.04
k : 451   | RMS train :   60.10     | RMS validation :   40.12    | RMS test :   58.37
k : 481   | RMS train :  776.67     | RMS validation :  738.13    | RMS test :  923.11
k : 511   | RMS train :   36.57     | RMS validation :   34.34    | RMS test :   32.05
```



Linear Regression with Regularisation for Multivariate Dataset

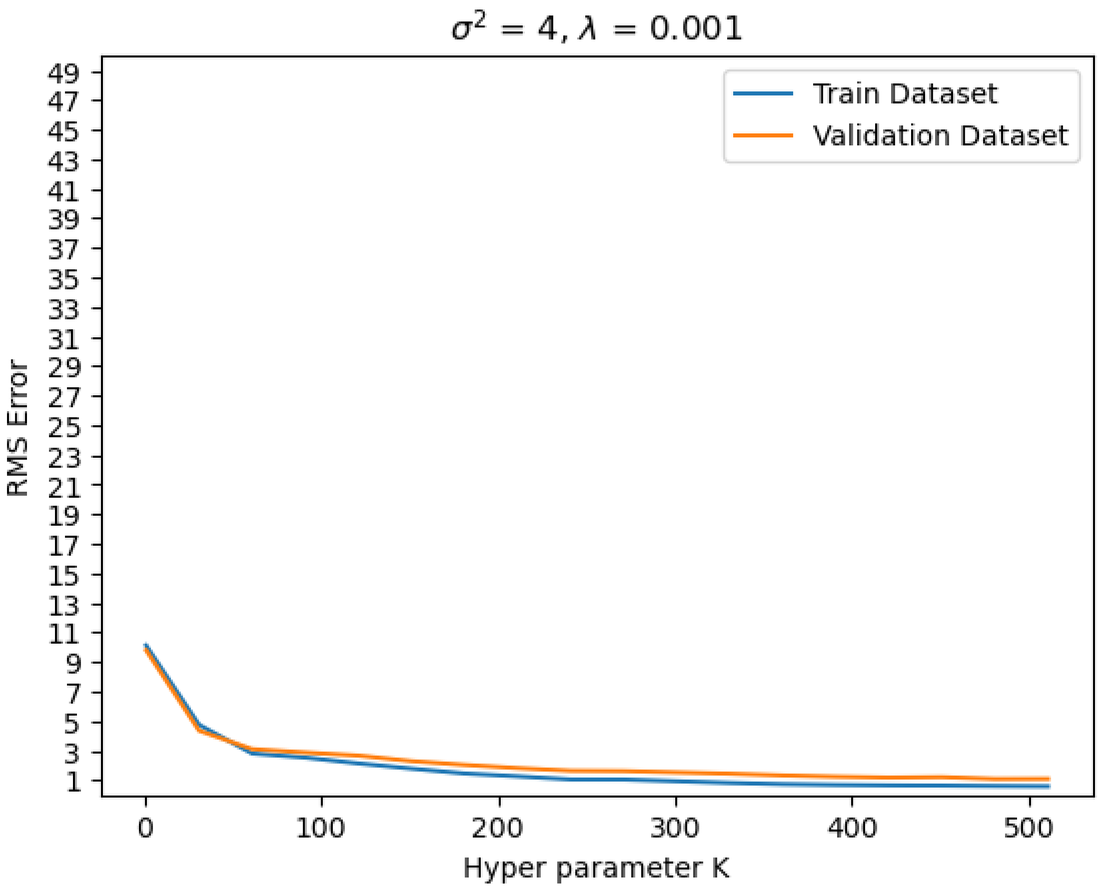
Multivariate Dataset with Regularisation : Variance : 4 and Regularisation : 0.0001

k : 1	RMS train : 10.12	RMS validation : 9.80	RMS test : 10.34
k : 31	RMS train : 4.37	RMS validation : 4.33	RMS test : 5.14
k : 61	RMS train : 2.98	RMS validation : 3.31	RMS test : 3.09
k : 91	RMS train : 2.21	RMS validation : 2.51	RMS test : 2.88
k : 121	RMS train : 2.00	RMS validation : 2.20	RMS test : 2.31
k : 151	RMS train : 1.44	RMS validation : 2.09	RMS test : 2.11
k : 181	RMS train : 0.98	RMS validation : 1.52	RMS test : 1.67
k : 211	RMS train : 0.86	RMS validation : 1.21	RMS test : 1.33
k : 241	RMS train : 0.59	RMS validation : 0.97	RMS test : 1.15
k : 271	RMS train : 0.57	RMS validation : 1.07	RMS test : 1.20
k : 301	RMS train : 0.51	RMS validation : 0.95	RMS test : 1.11
k : 331	RMS train : 0.48	RMS validation : 0.91	RMS test : 1.07
k : 361	RMS train : 0.48	RMS validation : 1.00	RMS test : 1.16
k : 391	RMS train : 0.45	RMS validation : 0.88	RMS test : 1.03
k : 421	RMS train : 0.44	RMS validation : 0.91	RMS test : 1.00
k : 451	RMS train : 0.43	RMS validation : 0.91	RMS test : 1.01
k : 481	RMS train : 0.42	RMS validation : 0.87	RMS test : 0.95
k : 511	RMS train : 0.41	RMS validation : 0.87	RMS test : 0.97



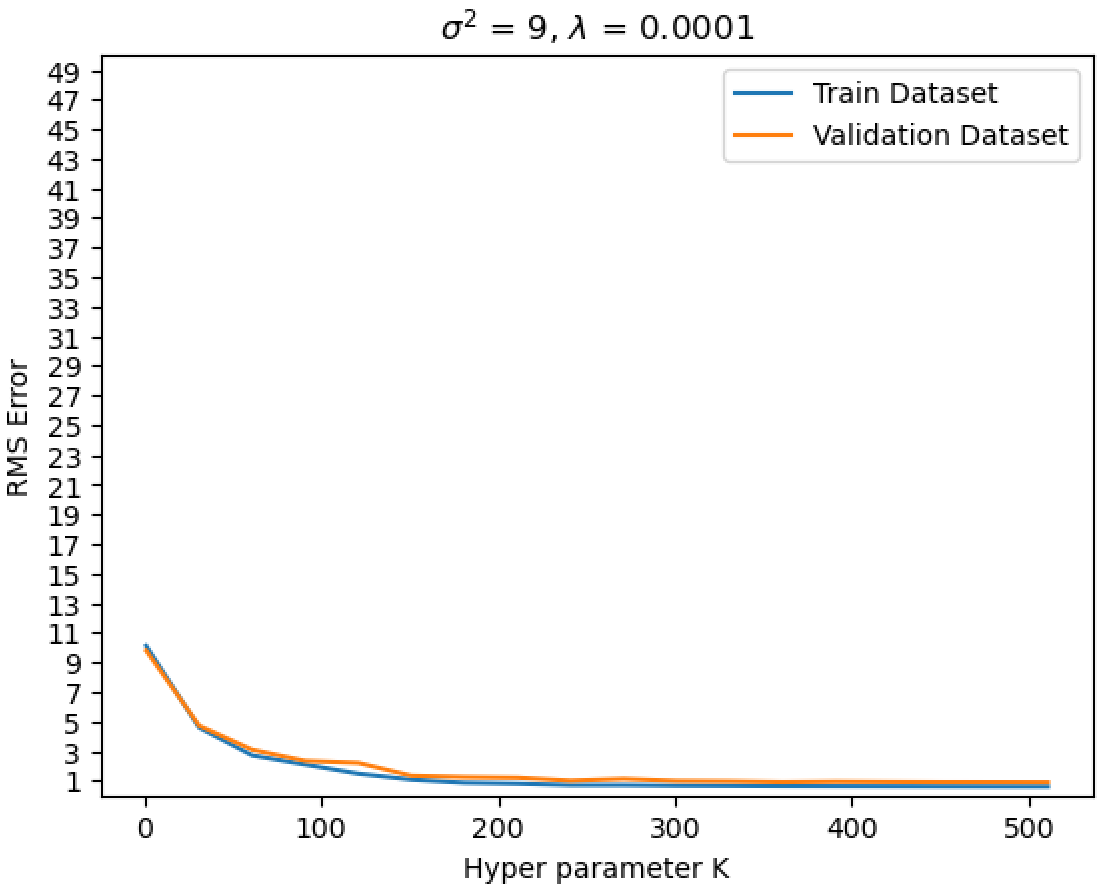
Multivariate Dataset with Regularisation : Variance : 4 and Regularisation : 0.001

```
k : 1     | RMS train :   10.12       | RMS validation :     9.80     | RMS test :    10.34
k : 31    | RMS train :    4.75       | RMS validation :     4.35     | RMS test :     5.01
k : 61    | RMS train :    2.81       | RMS validation :     3.10     | RMS test :     3.42
k : 91    | RMS train :    2.53       | RMS validation :     2.87     | RMS test :     3.08
k : 121   | RMS train :    2.14       | RMS validation :     2.67     | RMS test :     2.67
k : 151   | RMS train :    1.78       | RMS validation :     2.29     | RMS test :     2.28
k : 181   | RMS train :    1.45       | RMS validation :     2.04     | RMS test :     2.08
k : 211   | RMS train :    1.26       | RMS validation :     1.82     | RMS test :     1.95
k : 241   | RMS train :    1.05       | RMS validation :     1.64     | RMS test :     1.81
k : 271   | RMS train :    1.04       | RMS validation :     1.62     | RMS test :     1.59
k : 301   | RMS train :    0.93       | RMS validation :     1.52     | RMS test :     1.62
k : 331   | RMS train :    0.81       | RMS validation :     1.44     | RMS test :     1.52
k : 361   | RMS train :    0.73       | RMS validation :     1.32     | RMS test :     1.42
k : 391   | RMS train :    0.68       | RMS validation :     1.24     | RMS test :     1.35
k : 421   | RMS train :    0.65       | RMS validation :     1.18     | RMS test :     1.34
k : 451   | RMS train :    0.63       | RMS validation :     1.21     | RMS test :     1.29
k : 481   | RMS train :    0.60       | RMS validation :     1.08     | RMS test :     1.28
k : 511   | RMS train :    0.58       | RMS validation :     1.09     | RMS test :     1.22
```



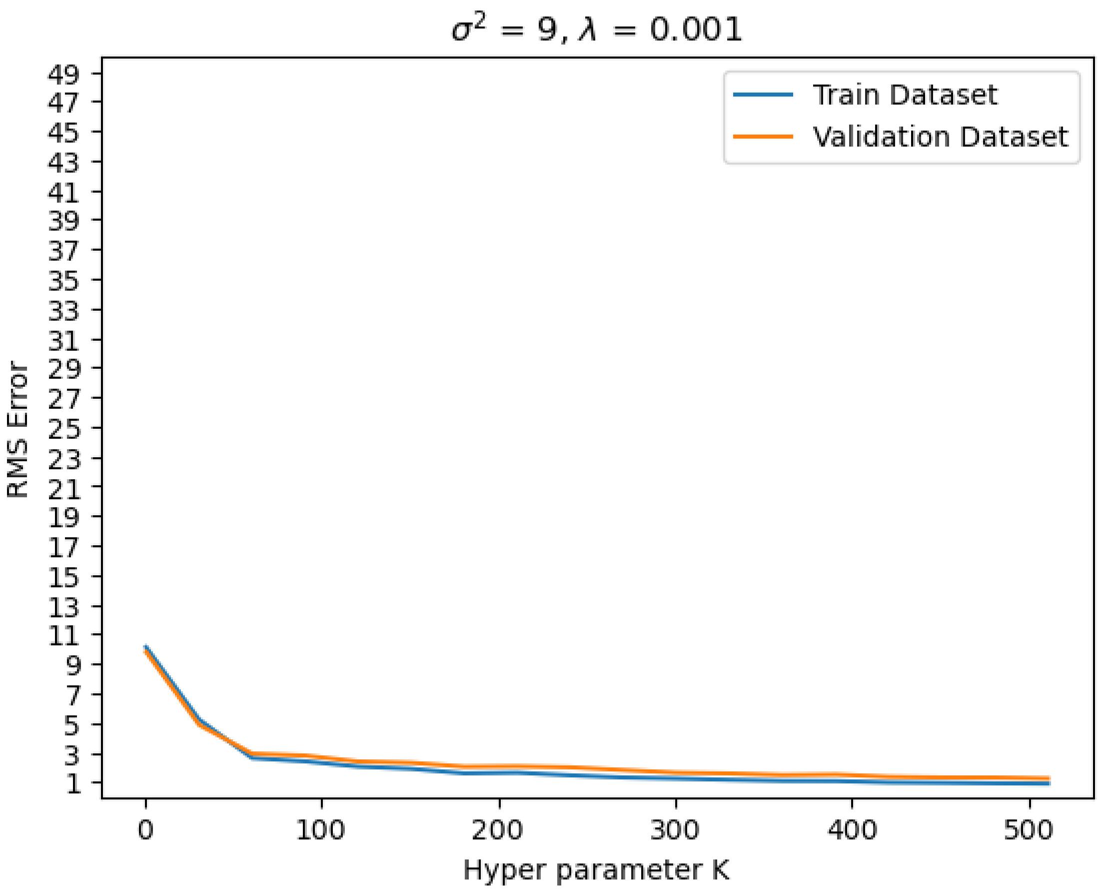
Multivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.0001

```
k : 1     | RMS train :   10.12     | RMS validation :    9.80    | RMS test :   10.34
k : 31    | RMS train :    4.58     | RMS validation :    4.72    | RMS test :    5.10
k : 61    | RMS train :    2.71     | RMS validation :    3.09    | RMS test :    3.31
k : 91    | RMS train :    2.09     | RMS validation :    2.34    | RMS test :    2.50
k : 121   | RMS train :    1.47     | RMS validation :    2.20    | RMS test :    1.88
k : 151   | RMS train :    1.07     | RMS validation :    1.33    | RMS test :    1.38
k : 181   | RMS train :    0.86     | RMS validation :    1.24    | RMS test :    1.17
k : 211   | RMS train :    0.80     | RMS validation :    1.20    | RMS test :    1.14
k : 241   | RMS train :    0.70     | RMS validation :    1.00    | RMS test :    1.04
k : 271   | RMS train :    0.70     | RMS validation :    1.12    | RMS test :    1.05
k : 301   | RMS train :    0.67     | RMS validation :    0.98    | RMS test :    0.98
k : 331   | RMS train :    0.66     | RMS validation :    0.96    | RMS test :    0.99
k : 361   | RMS train :    0.64     | RMS validation :    0.91    | RMS test :    0.97
k : 391   | RMS train :    0.63     | RMS validation :    0.94    | RMS test :    0.97
k : 421   | RMS train :    0.62     | RMS validation :    0.92    | RMS test :    0.96
k : 451   | RMS train :    0.61     | RMS validation :    0.91    | RMS test :    0.98
k : 481   | RMS train :    0.60     | RMS validation :    0.90    | RMS test :    0.97
k : 511   | RMS train :    0.60     | RMS validation :    0.90    | RMS test :    0.96
```

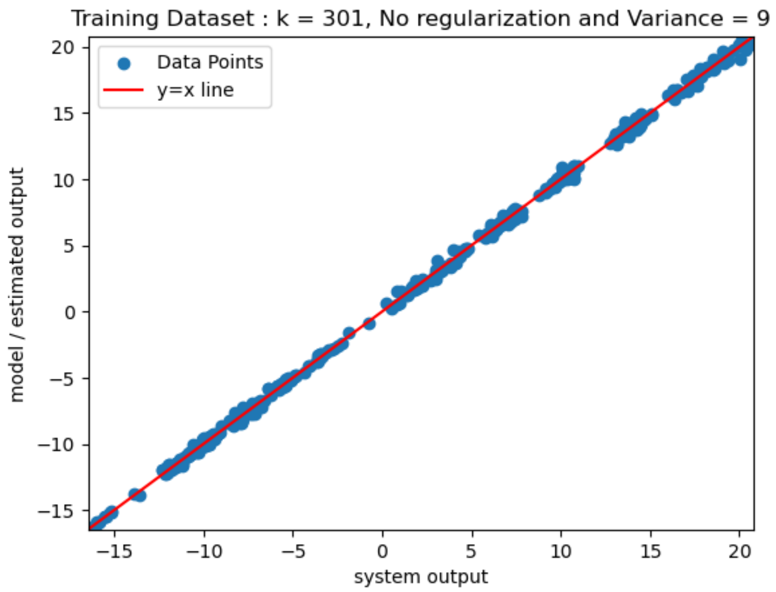


Multivariate Dataset with Regularisation : Variance : 9 and Regularisation : 0.001

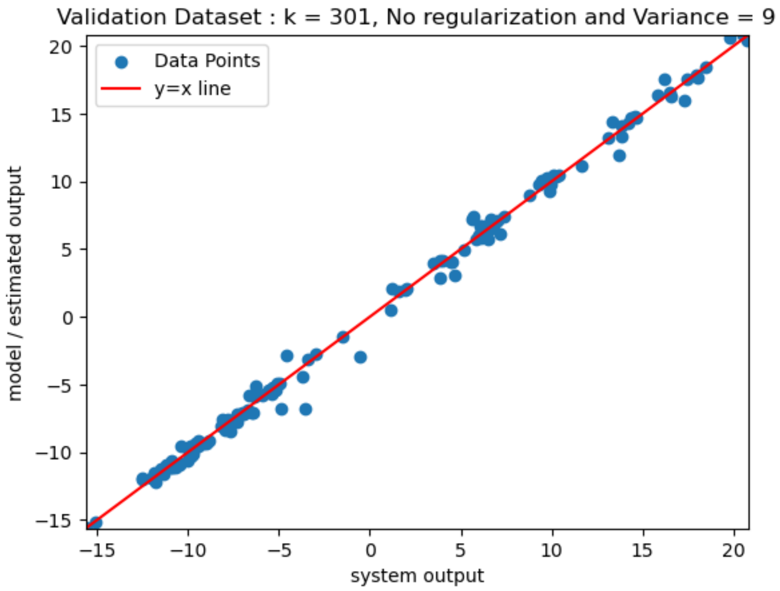
k : 1	RMS train : 10.12	RMS validation : 9.80	RMS test : 10.34
k : 31	RMS train : 5.23	RMS validation : 4.87	RMS test : 5.22
k : 61	RMS train : 2.63	RMS validation : 2.92	RMS test : 3.10
k : 91	RMS train : 2.40	RMS validation : 2.79	RMS test : 2.80
k : 121	RMS train : 2.05	RMS validation : 2.40	RMS test : 2.58
k : 151	RMS train : 1.90	RMS validation : 2.31	RMS test : 2.32
k : 181	RMS train : 1.59	RMS validation : 2.04	RMS test : 2.10
k : 211	RMS train : 1.64	RMS validation : 2.07	RMS test : 2.12
k : 241	RMS train : 1.46	RMS validation : 2.00	RMS test : 1.82
k : 271	RMS train : 1.31	RMS validation : 1.81	RMS test : 1.69
k : 301	RMS train : 1.24	RMS validation : 1.64	RMS test : 1.57
k : 331	RMS train : 1.15	RMS validation : 1.58	RMS test : 1.56
k : 361	RMS train : 1.08	RMS validation : 1.48	RMS test : 1.47
k : 391	RMS train : 1.07	RMS validation : 1.51	RMS test : 1.40
k : 421	RMS train : 0.98	RMS validation : 1.36	RMS test : 1.34
k : 451	RMS train : 0.95	RMS validation : 1.31	RMS test : 1.31
k : 481	RMS train : 0.92	RMS validation : 1.31	RMS test : 1.28
k : 511	RMS train : 0.89	RMS validation : 1.25	RMS test : 1.24



Best Performing Model's Scatter Plot and RMS Error Values :



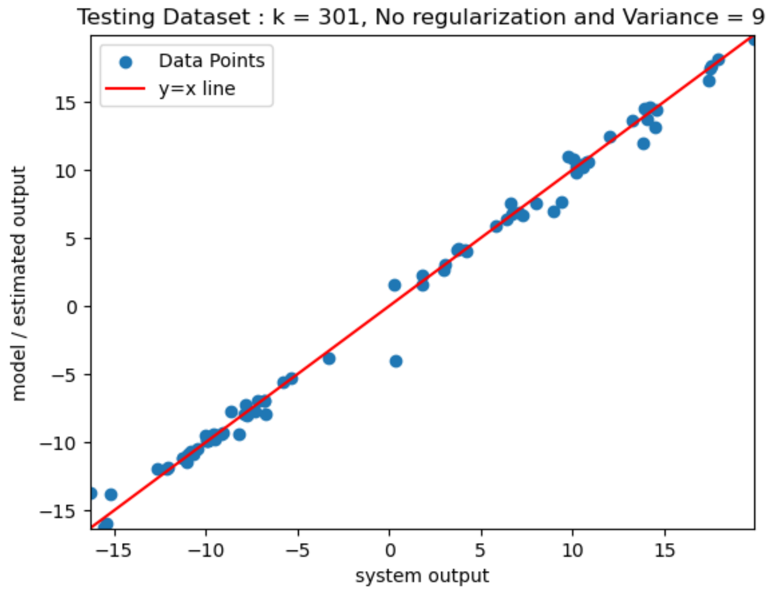
RMS = 0.23



RMS = 0.51

λ = 0

σ^2 = 9



RMS = 0.68

Observations & Conclusions :

- Bivariate Dataset :
  - As the size of the dataset increases from 50 to 200 we can see the RMSE for the test data reduces from **1.47** to **1.02**
  - From the various plots we can see that as the variance hyperparameter increases models will less clusters perform better, if the variance decreases then models with more clusters will perform better
  - Quadratic Regularisation had increased RMSE as compared to models without Regularisation, but it had prevented the RMSE from spiralling upwards for larger clusters
  - This could be due to the Regularisation parameter shifting the eigenvalues of the square matrix in the Pseudo inverse of Regression term, thereby moving the square matrix from near singularity conditions
  - This would prevent regression parameters from taking huge values, effectively reducing error for models with large number of clusters and higher variance

- Multivariate Dataset :
  - Similar results are obtained with this dataset as well

Task 4:

Linear model for regression using Polynomial basis functions for Datasets 4
Dataset 4 : Multivariate Dataset - 8d data

File Name	Data Points
test_14.csv	206
train_14.csv	618
val_14.csv	206

The real world dataset used in Task 4 was to model concrete strength which is a nonlinear function of its ingredients.

Regression using polynomial basis functions was carried out and the performance of it is as follows:

Best Performing Model's Scatter Plot(Using cross validation):

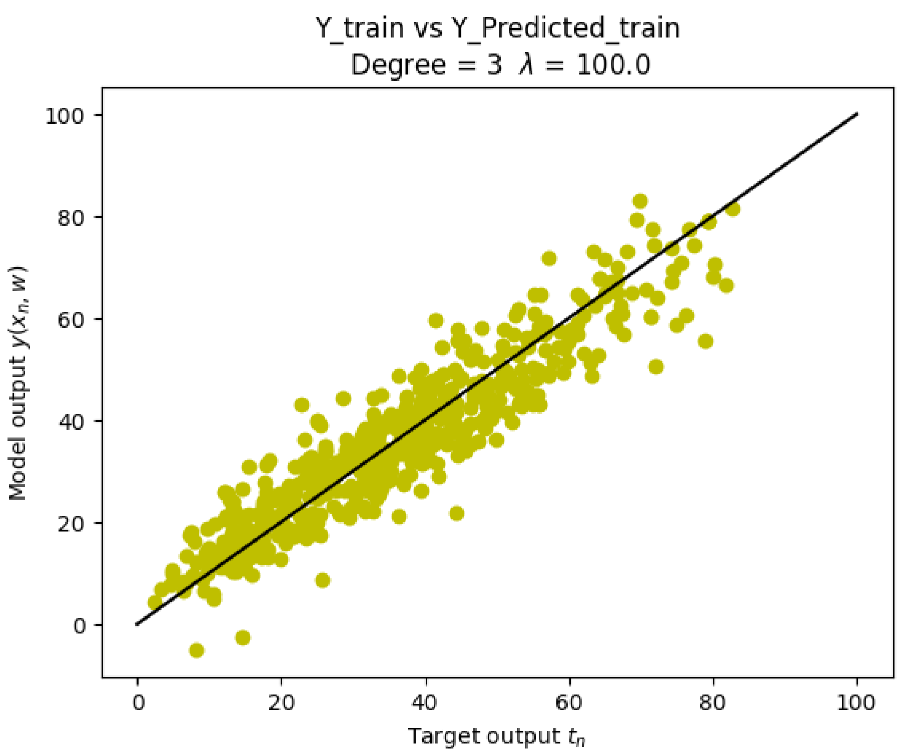


Fig 4.1: Scatter plot for training data

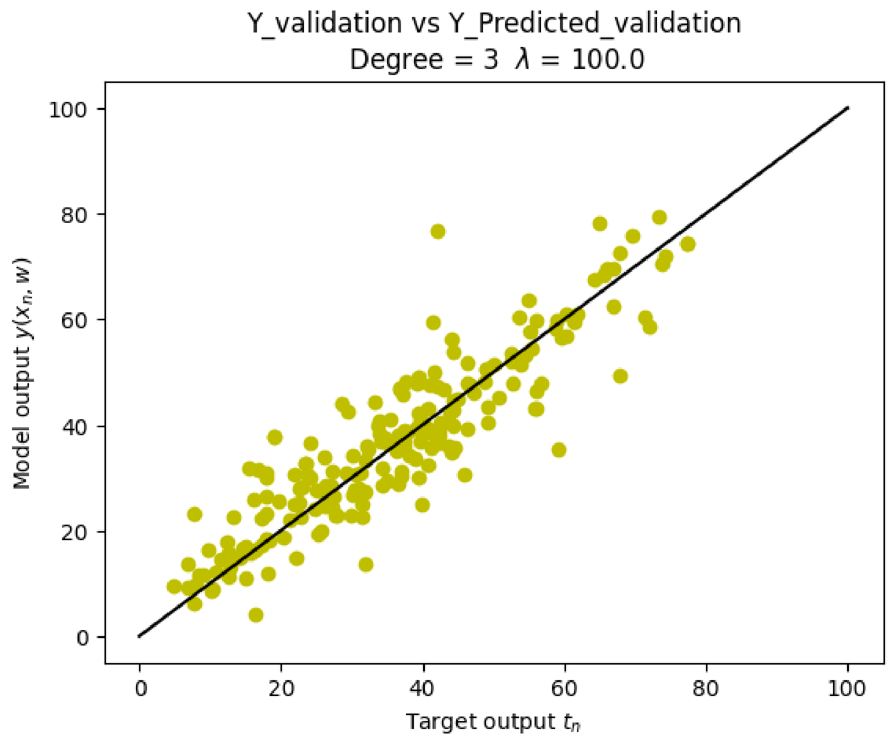


Fig 4.2: Scatter plot for validation data

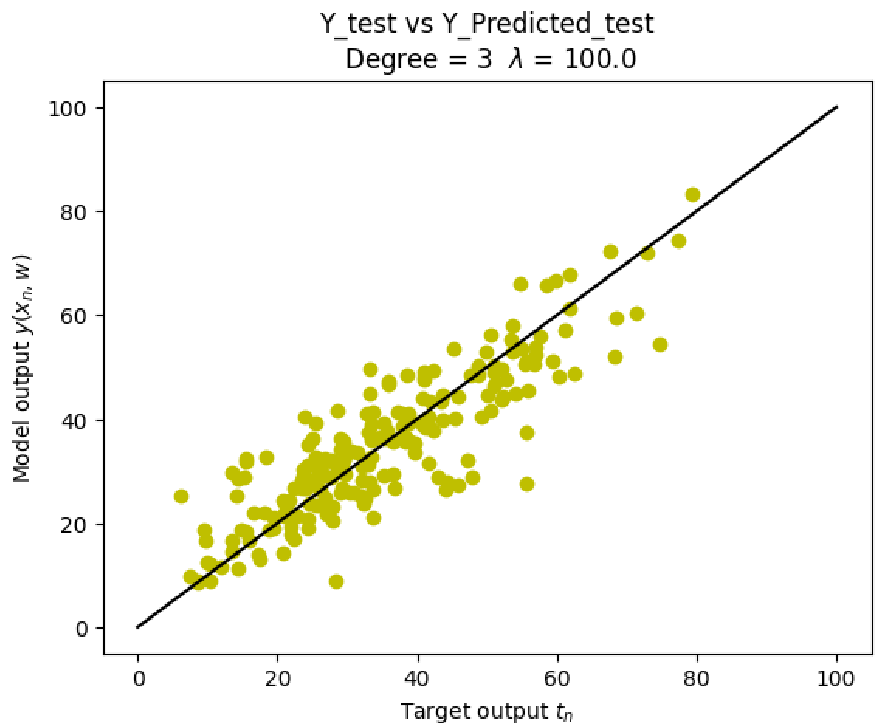


Fig 4.3: Scatter plot for test data

From the above scatter plot results, it's evident that the best performing model fits the test data well which was selected through the cross validation method.

RMS Error Values:

Degree	Size	λ	ERMS_train	ERMS_val	ERMS_test
1	618	0	11.58870854	11.2604337	10.86354431
		1.00E-06	11.58870854	11.26043038	10.86354144
		1.00E-02	11.59070889	11.23057882	10.83793879
		1	12.91950121	12.10750657	11.73698863
		100	13.89601466	13.09798986	12.69232879
2	618	0	8.518829132	8.362130089	8.505729546
		1.00E-06	8.51882913	8.362129648	8.505729043
		1.00E-02	8.518825992	8.360924483	8.504346876
		1	8.519372114	8.379647557	8.524829744
		100	8.533191842	8.44029723	8.541752448
3	618	0	6.279970126	7.308509141	7.675958826
		1.00E-06	6.279970126	7.308509142	7.675958823
		1.00E-02	6.279960158	7.308421185	7.675890155
		1	6.280459293	7.308939671	7.675300332
		100	6.260615818	7.266586319	7.553996321
4	618	0	4.571177734	9.540639708	11.7051196
		1.00E-06	4.571177734	9.540640065	11.705121
		1.00E-02	4.571177734	9.540639356	11.70511943
		1	4.571059099	9.545614252	11.70742702
		100	4.570803882	9.540604654	11.70563277
5	618	0	3.766445732	23.51914609	20.10714394
		1.00E-06	3.766445732	23.51914609	20.10714394
		1.00E-02	3.766444156	23.51880214	20.10696674
		1	3.766444749	23.51900816	20.10706813
		100	3.767116694	23.5165702	20.10489824

Table 4.1: E_{RMS} data for Task 4

From the above error table, it's clear that the model with **degree 3** and regularisation parameter **λ=100** is the best performing model as it has the minimum $\mathbf{E_{RMS}}$ = 7.2666.

For models with higher complexity , i.e, degree > 3, the training overfits and even with regularization, it doesn't perform as good as the degree 3 model.

Observation & Conclusion:

In this analysis, polynomial regression was employed with varying degrees (**1 to 10**) and applied regularization with a strength of 100.0 to predict concrete strength in a multivariate dataset.
After a thorough evaluation, it became evident that the **degree 3** polynomial regression model with strong regularization (**λ = 100.0**) outperformed models with degrees ranging from **1 to 10.**

- **Overfitting Mitigation:** The **degree 3** model with high regularization **(λ = 100.0)** effectively mitigated the risk of overfitting. Higher-degree polynomial models (eg., degree = [6, 7, 8, 9, 10]) tend to be more susceptible to overfitting when the dataset is limited.
- **Bias-Variance Tradeoff**: A **degree 1** model would have been too simplistic to capture the underlying complexity of the concrete strength prediction problem, resulting in high bias (**underfitting**). In contrast, the degree 3 model found an optimal balance between bias and variance, resulting in a model that better captures the nonlinear relationships in the data.
- **Multivariate Dataset:** A degree 3 polynomial basis allowed the model to capture the complex interactions between different features, more effectively than lower-degree models, which might miss important nonlinear relationships.

In conclusion, based on the multivariate dataset and the performance evaluation criteria, the **degree 3** polynomial regression model with strong regularization (**λ = 100.0**) emerged as the top-performing model for predicting concrete strength.
It struck a balance between capturing the nonlinear relationships in the data and preventing overfitting, making it the preferred choice among the considered degrees (1 to 10) which is backed by the scatter plot.

This conclusion highlights the importance of both polynomial degree selection and regularization strength in achieving accurate predictions in regression tasks.