

CS5691: Programming Assignment 1

Team 20

Niveath A Shaun Mathew

CS21B060 CS21B076

October 3, 2023

Contents

1 Task 1: Polynomial Curve Fitting for Dataset 1	2
1.1 Training Dataset of size 10	2
1.1.1 Plots of Approximated Functions without Regularization	2
1.1.2 Inferences	3
1.1.3 Plots of Approximated Functions with Regularization	3
1.1.4 Inferences	7
1.1.5 Hyperparameters v/s Training, Validation and Testing Errors	7
1.2 Training Dataset of size 100	8
1.2.1 Plots of Approximated Functions without Regularization	8
1.2.2 Inferences	9
1.2.3 Plots of Approximated Functions with Regularization	9
1.2.4 Inferences	13
1.2.5 Hyperparameters v/s Training, Validation and Testing Errors	14
2 Task 2: Linear Model for Regression using Polynomial Basis Functions for Dataset 2	15
2.1 Training Dataset of size 50	15
2.1.1 Surface Plots of Approximated Functions without Regularization	15
2.1.2 Inferences	16
2.1.3 Surface Plots of Approximated Functions with Regularization	16
2.1.4 Inferences	19
2.1.5 Scatter Plots of the Best Model	20
2.1.6 Hyperparameters v/s Training, Validation and Testing Errors	21
2.2 Training Dataset of size 200	22
2.2.1 Plots of Approximated Functions without Regularization	22
2.2.2 Inferences	23
2.2.3 Plots of Approximated Functions with Regularization	23
2.2.4 Inferences	26
2.2.5 Scatter Plots of the Best Model	27
2.2.6 Hyperparameters v/s Training, Validation and Testing Errors	28
3 Task 3.1: Linear Model for Regression using Gaussian Basis Functions for Dataset 2	29
3.1 Training Dataset of size 50	29
3.1.1 Hyperparameters v/s Training, Validation and Testing Errors	29
3.1.2 Inferences	31
3.1.3 Scatter Plots of the Best Model	31
3.2 Training dataset of size 200	32
3.2.1 Hyperparameters v/s Training, Validation and Testing Errors	32
3.2.2 Inferences	34
3.2.3 Scatter Plots of the Best Model	34
4 Task 3.2: Linear Model for Regression using Gaussian Basis Functions for Dataset 3	35
4.1 Hyperparameters v/s Training, Validation and Testing Errors	35
4.2 Inferences	37
4.3 Scatter Plots of the Best Model	38
5 Task 4: Linear Model for Regression using Polynomial Basis Functions for Dataset 4	39
5.1 Description of Dataset	39
5.2 Procedure	39
5.3 Scatter Plots of Approximated Functions	40
5.4 Inferences	44
5.5 Scatter Plots of the Best Model	45
5.6 Inferences	45
5.7 Hyperparameters v/s Training, Validation and Testing Errors	46

1 Task 1: Polynomial Curve Fitting for Dataset 1

1.1 Training Dataset of size 10

1.1.1 Plots of Approximated Functions without Regularization

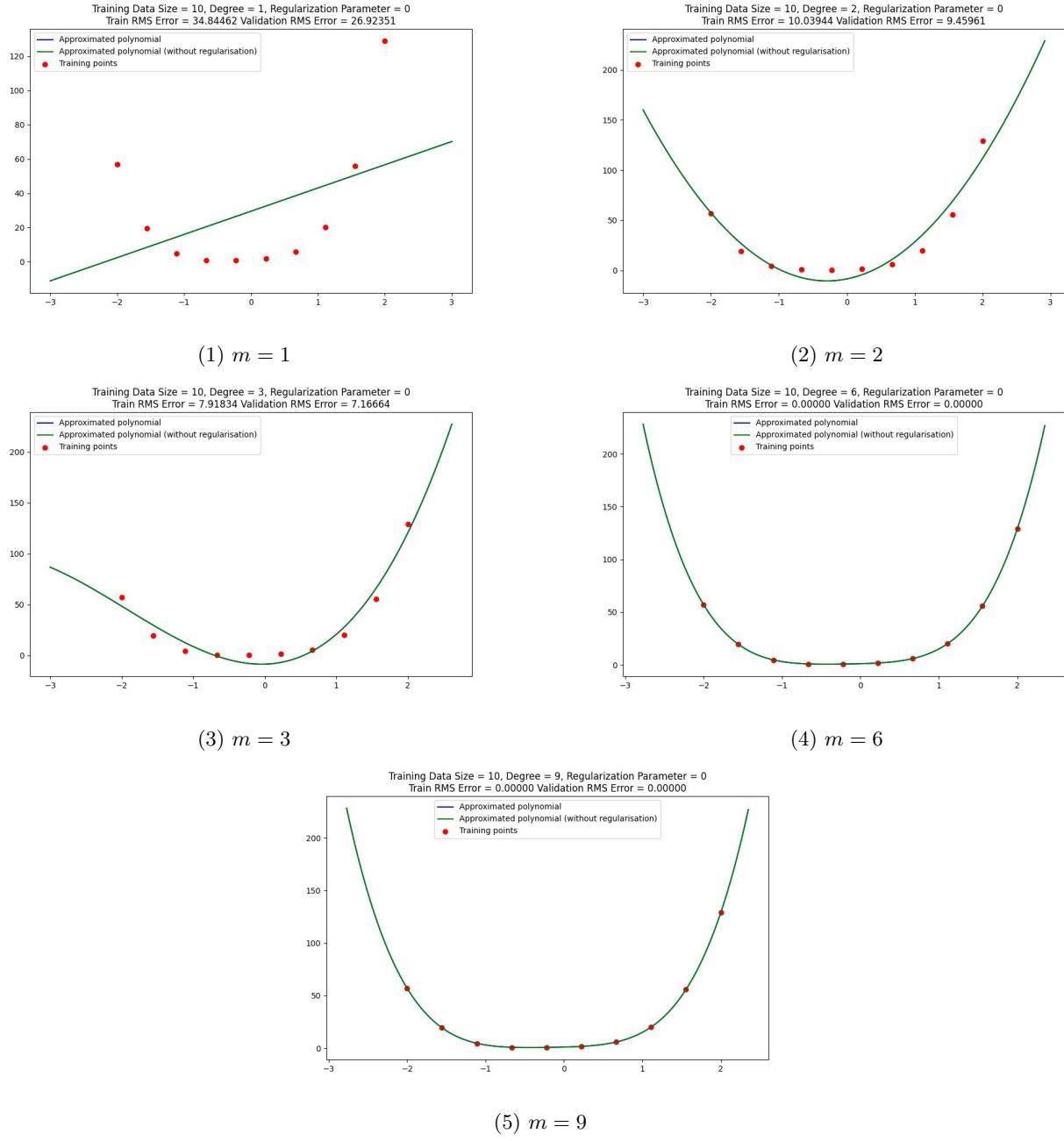


Figure 1: Approximated Functions without Regularization for training dataset of size 10

1.1.2 Inferences

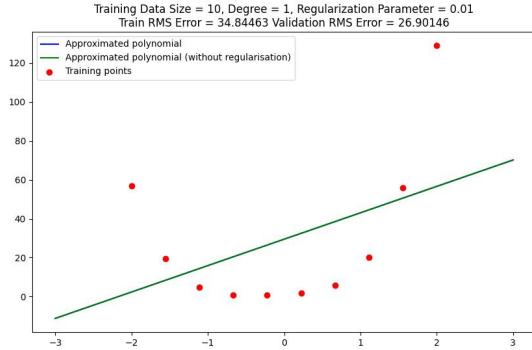
- With small values of degree m such as 1, 2 and 3, the approximated model is unable to fit to the training data accurately. Both the training and validation errors are large. This is a case of underfitting. This indicates that the underlying distribution of the data has a higher complexity than expected.
- With large value of degree $m = 9$, the approximated model fits exactly onto the training dataset. But this can be a case of overfitting as there are only 10 data points available. The number of parameters to estimate is given by

$$D = \frac{(m + d)!}{m! \cdot d!}, m = \text{degree of polynomial}, d = \text{dimensionality of input vector} \quad (1)$$

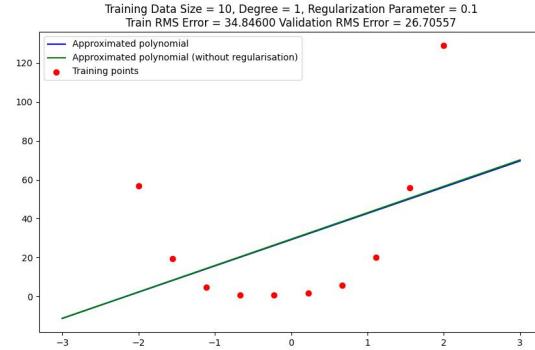
which with $m = 9$ and $d = 1$, gives 10 parameters. So, we need much more than 10 datapoints to train this model effectively. Hence, this is not the best model possible.

- With degree $m = 6$, the approximated model fits exactly onto the training data and the number of datapoints is sufficient to train this model. Hence, this model gives the best result with the given training points. Since this model has nearly zero training and validation error, we don't require the use of regularization to improve the model.

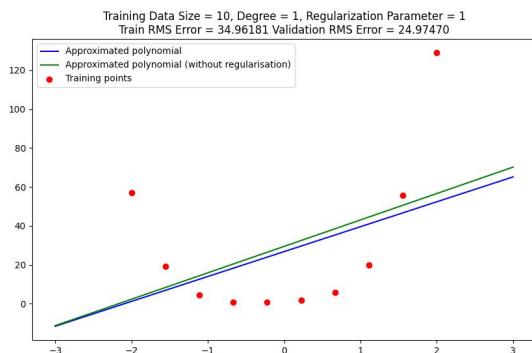
1.1.3 Plots of Approximated Functions with Regularization



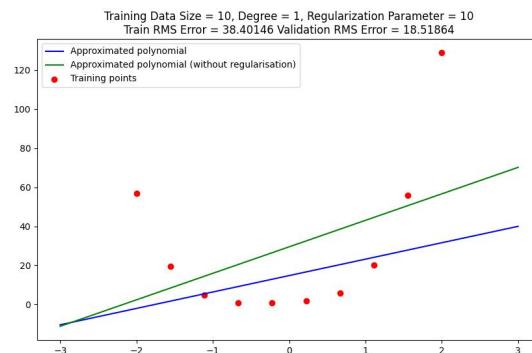
(1) $m = 1, \lambda = 0.01$



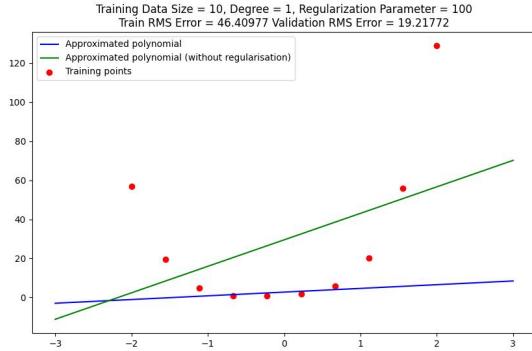
(2) $m = 1, \lambda = 0.1$



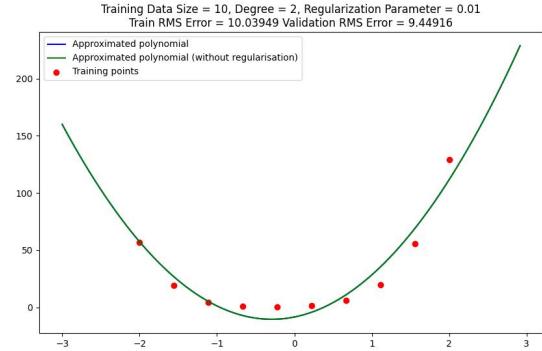
(3) $m = 1, \lambda = 1$



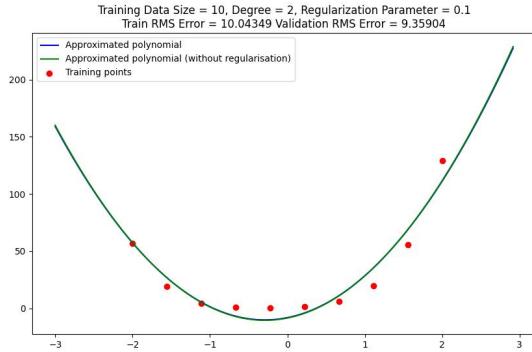
(4) $m = 1, \lambda = 10$



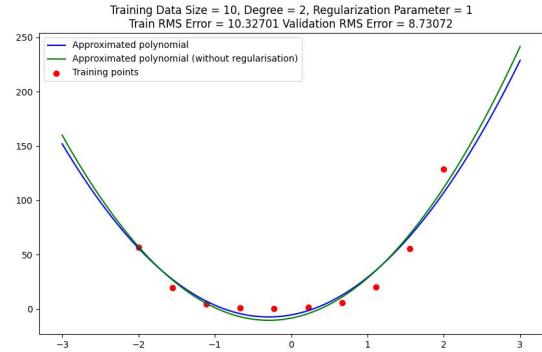
(5) $m = 1, \lambda = 100$



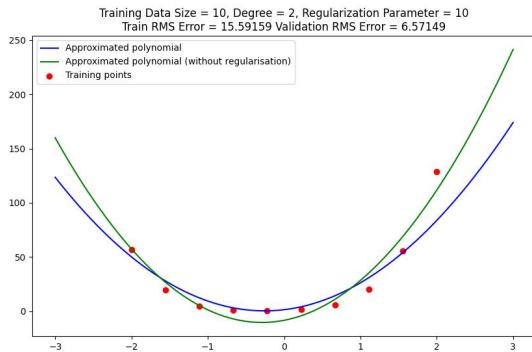
(6) $m = 2, \lambda = 0.01$



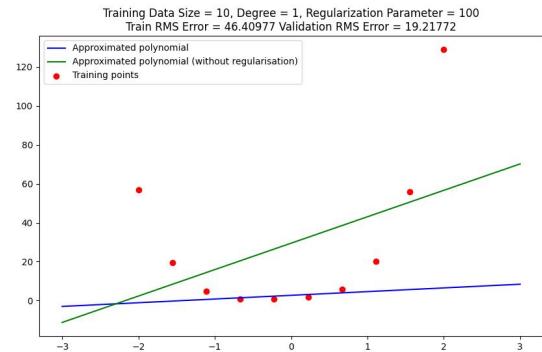
(7) $m = 2, \lambda = 0.1$



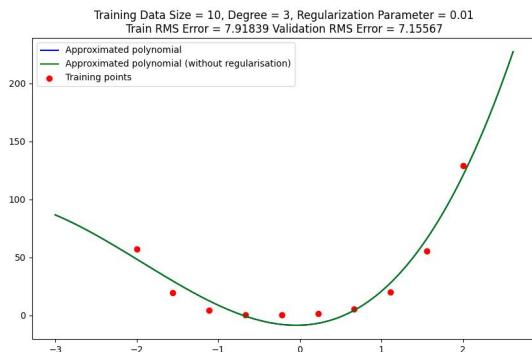
(8) $m = 2, \lambda = 1$



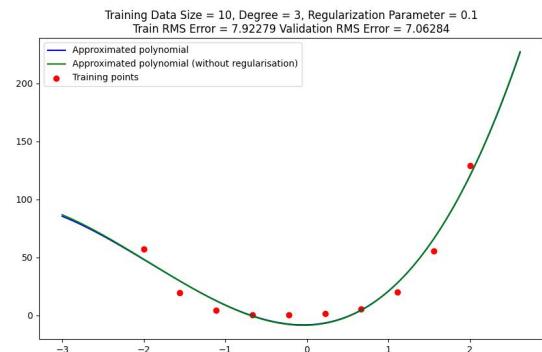
(9) $m = 2, \lambda = 10$



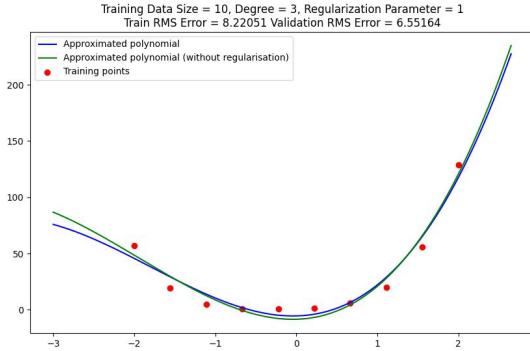
(10) $m = 2, \lambda = 100$



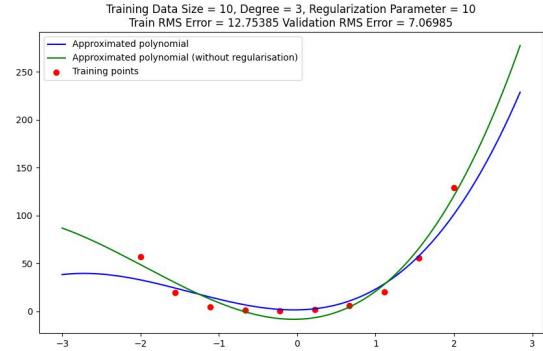
(11) $m = 3, \lambda = 0.01$



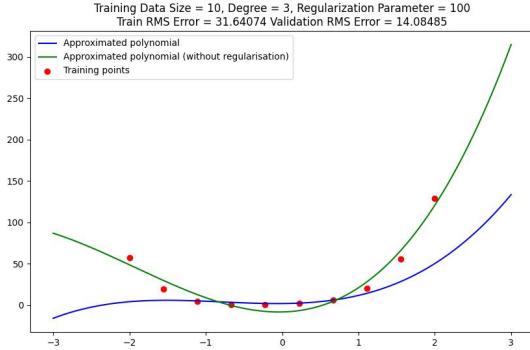
(12) $m = 3, \lambda = 0.1$



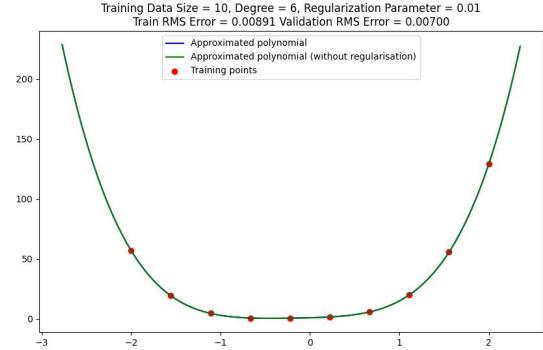
(13) $m = 3, \lambda = 1$



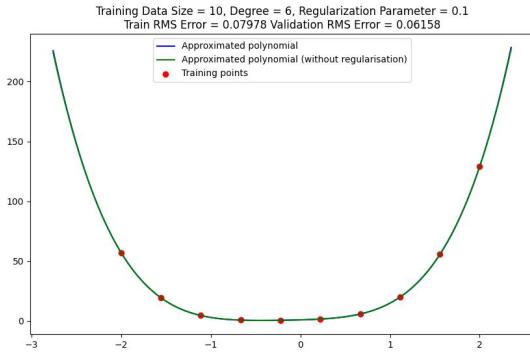
(14) $m = 3, \lambda = 10$



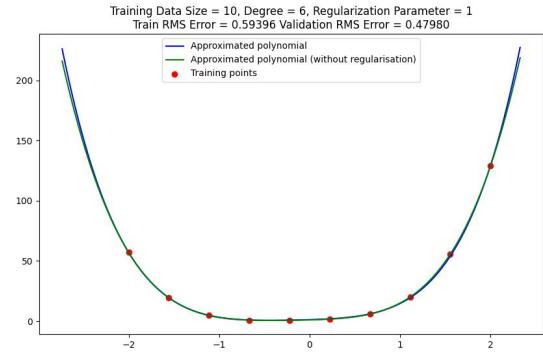
(15) $m = 3, \lambda = 100$



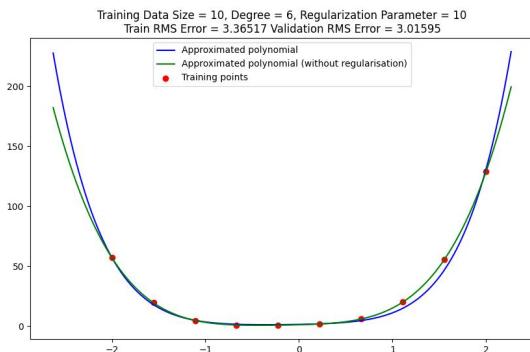
(16) $m = 6, \lambda = 0.01$



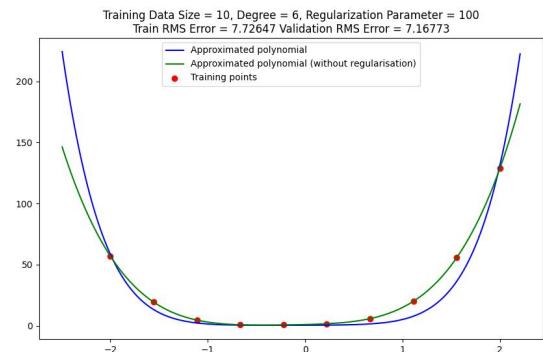
(17) $m = 6, \lambda = 0.1$



(18) $m = 6, \lambda = 1$

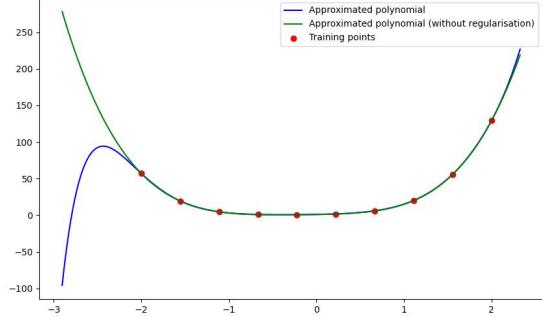


(19) $m = 6, \lambda = 10$



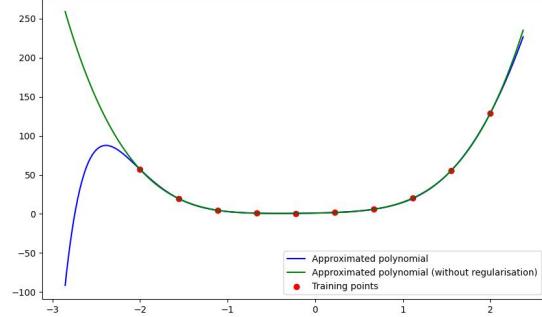
(20) $m = 6, \lambda = 100$

Training Data Size = 10, Degree = 9, Regularization Parameter = 0.01
 Train RMS Error = 0.04611 Validation RMS Error = 0.15320



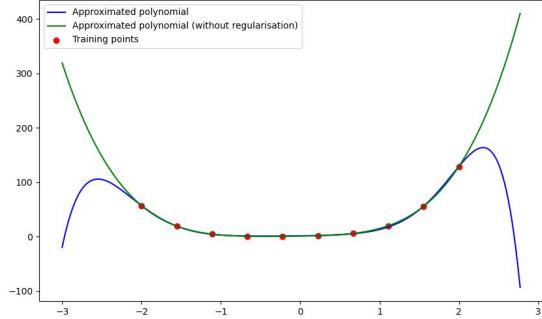
(21) $m = 9, \lambda = 0.01$

Training Data Size = 10, Degree = 9, Regularization Parameter = 0.1
 Train RMS Error = 0.16216 Validation RMS Error = 0.27742



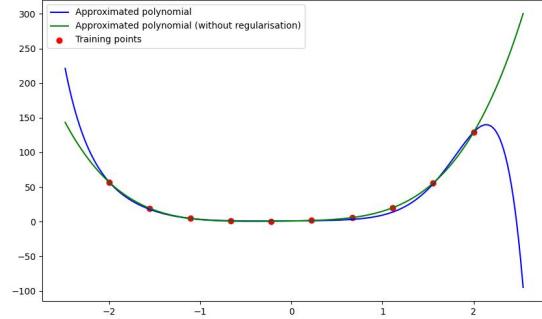
(22) $m = 9, \lambda = 0.1$

Training Data Size = 10, Degree = 9, Regularization Parameter = 1
 Train RMS Error = 0.75962 Validation RMS Error = 0.99700



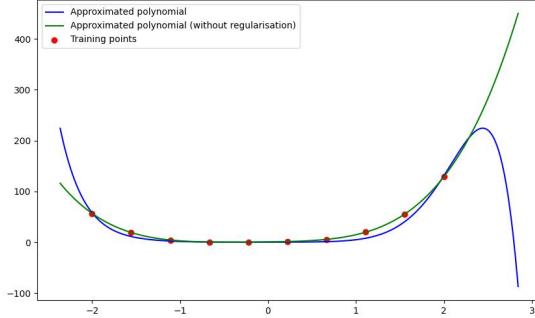
(23) $m = 9, \lambda = 1$

Training Data Size = 10, Degree = 9, Regularization Parameter = 10
 Train RMS Error = 2.24021 Validation RMS Error = 2.90231



(24) $m = 9, \lambda = 10$

Training Data Size = 10, Degree = 9, Regularization Parameter = 100
 Train RMS Error = 6.76901 Validation RMS Error = 6.84056



(25) $m = 9, \lambda = 100$

Figure 2: Approximated Functions with Regularization for training dataset of size 10

1.1.4 Inferences

1. For small values of $m = 1, 2, 3$, the model is still unable to fit to the underlying data. This is a case of underfitting and regularization will not be of help in these models.
2. For large values of $m = 6, 9$, even a small amount of regularization causes the model to deviate a lot from the training data. This can be seen in the marked increase in the training error from $\lambda = 0$ to $\lambda = 0.01$ for $m = 6, 9$. This is likely due to the extremely small size of the training dataset.
3. For all values of the degree m , as the regularization parameter λ increases, the training error increases and the validation error decreases.
4. Regularization forces the model to slightly deviate from the training points, hence reducing possible overfitting. However, with large values of λ such as 100.0, the model deviates too much from the training data and does not capture the underlying distribution. This can be seen by the increase in validation error. Hence, large values of λ can lead to underfitting.
5. We can see that the use of regularization did not improve the model's performance, hence proving validity our initial assertion regarding the same.

1.1.5 Hyperparameters v/s Training, Validation and Testing Errors

Degree	Regularization Parameter	Training Error	Validation Error	Testing Error
1	0.0	34.844 62	26.923 51	24.777 53
1	0.01	34.844 63	26.901 46	24.774 97
1	0.1	34.846 00	26.705 57	24.753 86
1	1.0	34.961 81	24.974 70	24.705 09
1	10.0	38.40146	18.51864	28.38081
1	100.0	46.409 77	19.217 72	37.623 10
2	0.0	10.039 44	9.459 61	8.968 14
2	0.01	10.039 49	9.449 16	8.959 79
2	0.1	10.043 49	9.359 04	8.888 79
2	1.0	10.327 01	8.730 72	8.484 40
2	10.0	15.59159	6.57149	10.95333
2	100.0	34.874 18	11.308 98	27.616 02
3	0.0	7.918 34	7.166 64	8.262 92
3	0.01	7.918 39	7.155 67	8.256 67
3	0.1	7.922 79	7.062 84	8.203 53
3	1.0	8.22051	6.55164	7.88404
3	10.0	12.753 85	7.069 85	8.206 95
3	100.0	31.640 74	14.084 85	23.444 66
6	0.0	8.13774×10^{-12}	5.24059×10^{-12}	7.29628×10^{-12}
6	0.01	0.008 91	0.006 99	0.010 73
6	0.1	0.079 78	0.061 58	0.097 18
6	1.0	0.593 96	0.479 80	0.724 80
6	10.0	3.365 17	3.015 95	3.951 49
6	100.0	7.726 47	7.167 73	9.182 62
9	0.0	3.48263×10^{-9}	3.43536×10^{-9}	3.57539×10^{-9}
9	0.01	0.046 11	0.153 20	0.173 41
9	0.1	0.162 16	0.277 42	0.314 50
9	1.0	0.759 62	0.997 00	1.123 76
9	10.0	2.240 21	2.902 31	2.578 38
9	100.0	6.769 01	6.840 56	7.822 74

Table 1: Training, Validation and Testing Errors for different m, λ for training dataset of size 10. The best model for each m is taken to be the one with the lowest validation error and are shown in boldface.

1.2 Training Dataset of size 100

1.2.1 Plots of Approximated Functions without Regularization

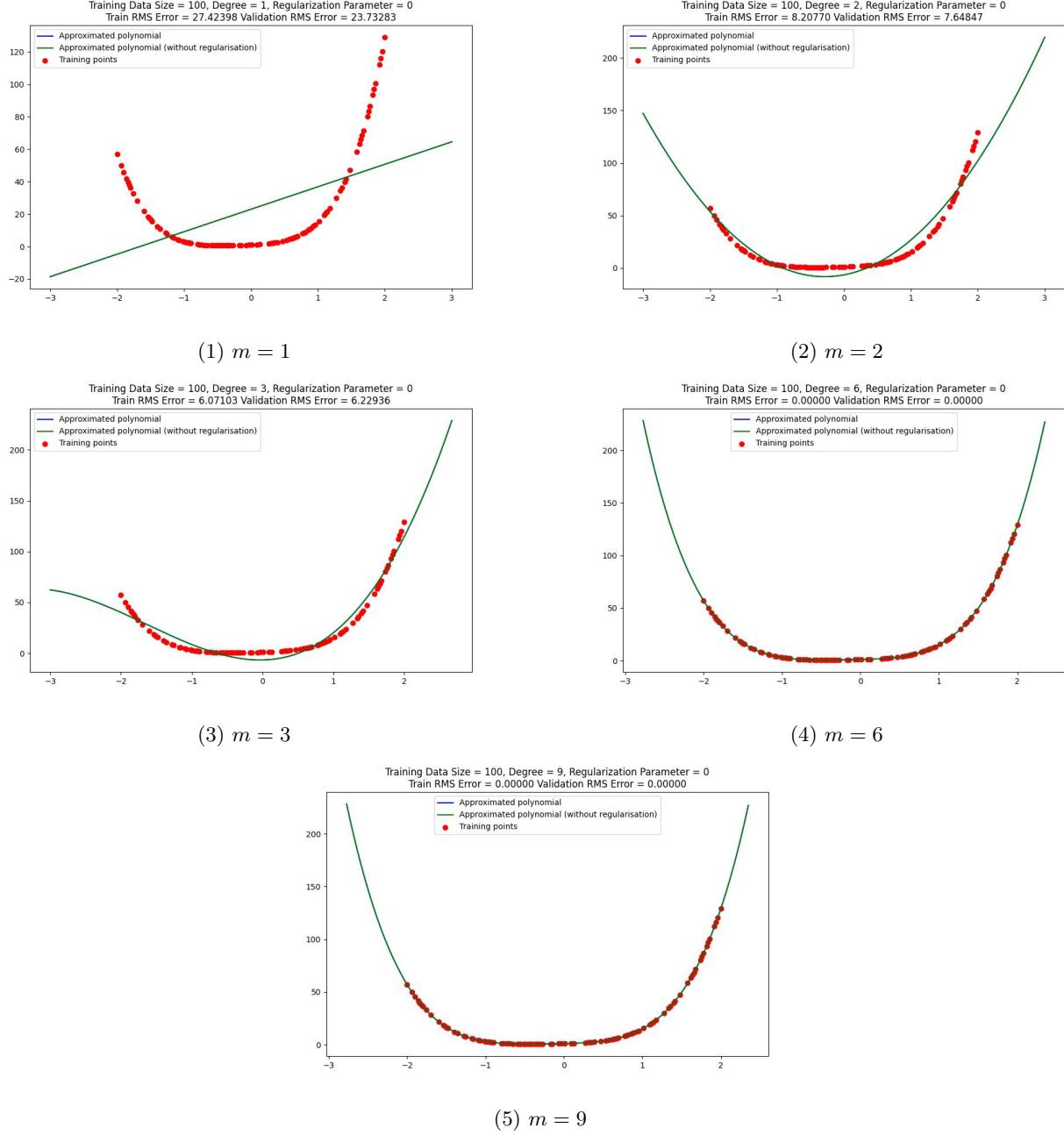
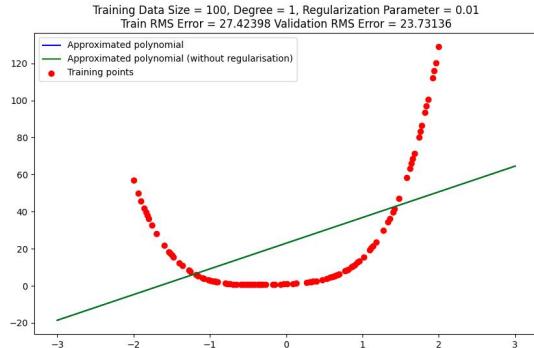


Figure 3: Approximated Functions without Regularization for training dataset of size 100

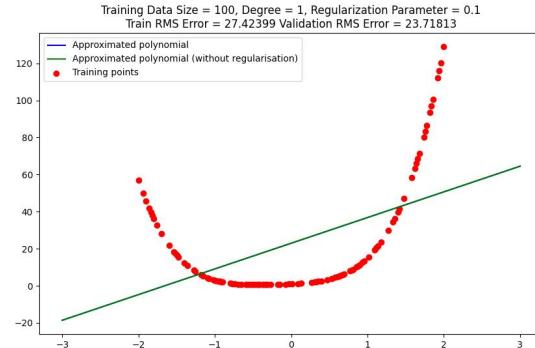
1.2.2 Inferences

1. For small values of m such as 1, 2 and 3, the model does not fit well with the training data and there is underfitting. This is reflected in the high training and validation errors. This indicated that the underlying distribution of data has a much higher complexity than these models.
2. As the model degree increases, the complexity of the model increases and the model fits well with the training data. This is reflected in the lowering of the training and validation errors.
3. However, with higher complexity, the model may overfit with the data. This is reflected in the validation errors for $m = 6$ and $m = 9$, which are 2.4329×10^{-12} and 1.3962×10^{-10} respectively, which is an increase by an order of two.
4. We also find that the performance of the model with $m = 9$, has improved by an order of 1 in comparison to the model trained with just 10 data points. This can be attributed to the increase in the size of the training dataset which gives us a better estimation of the model parameters.
5. The best model among these, will be $m = 6$ as it achieves the least validation error. Since, this model has nearly zero training and testing data, we don't require the use of regularization to improve the model.

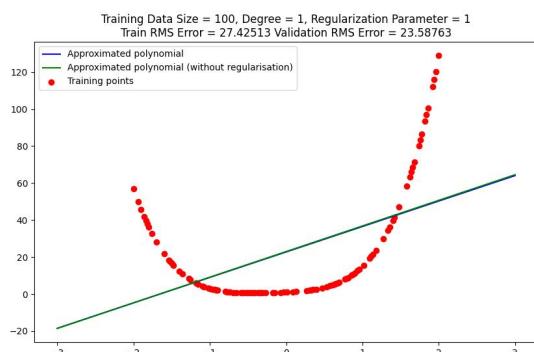
1.2.3 Plots of Approximated Functions with Regularization



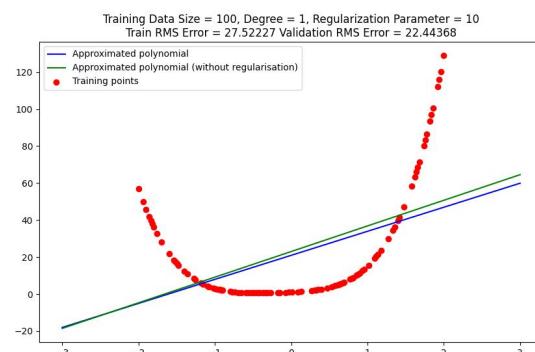
(1) $m = 1, \lambda = 0.01$



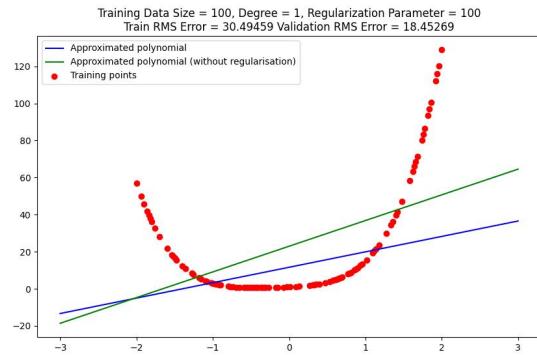
(2) $m = 1, \lambda = 0.1$



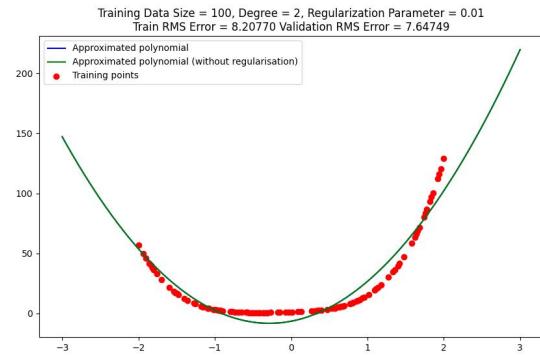
(3) $m = 1, \lambda = 1$



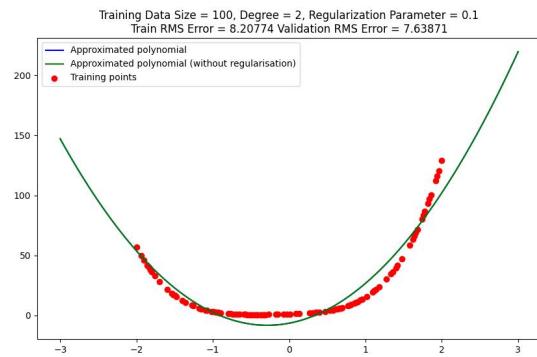
(4) $m = 1, \lambda = 10$



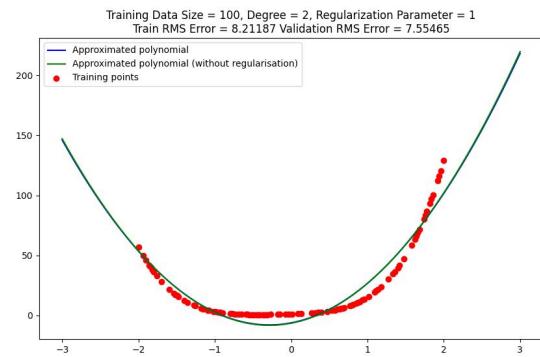
(5) $m = 1, \lambda = 100$



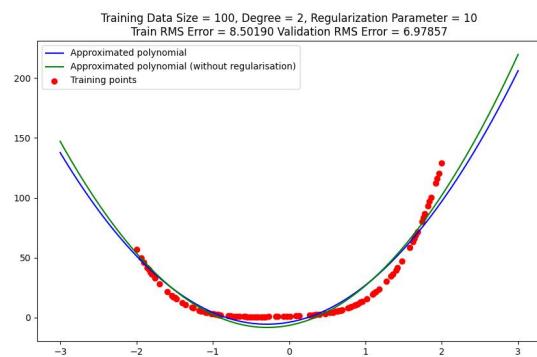
(6) $m = 2, \lambda = 0.01$



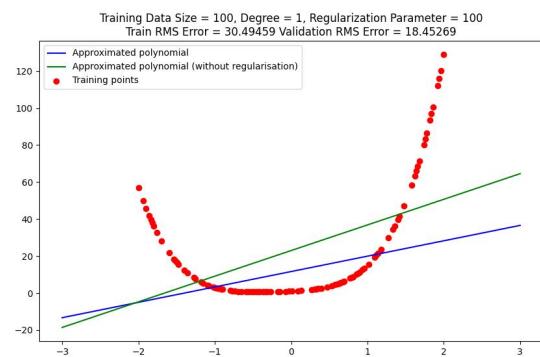
(7) $m = 2, \lambda = 0.1$



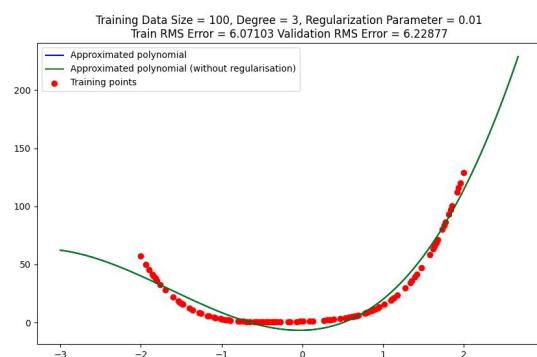
(8) $m = 2, \lambda = 1$



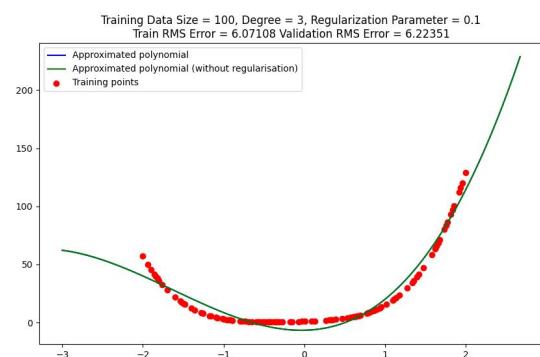
(9) $m = 2, \lambda = 10$



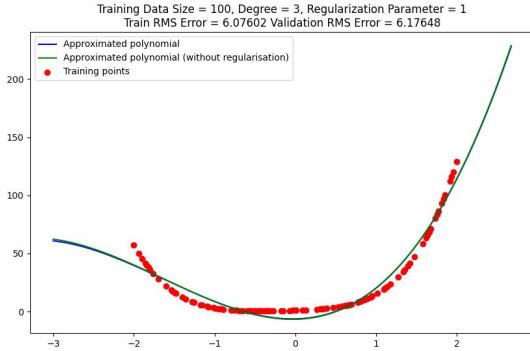
(10) $m = 2, \lambda = 100$



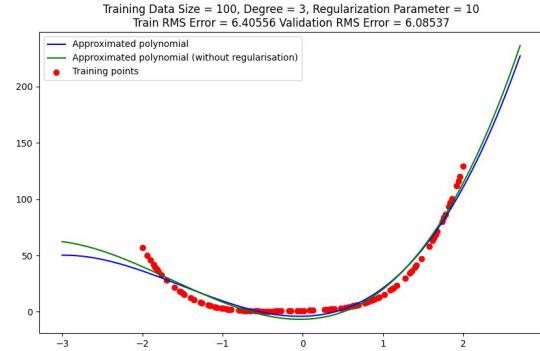
(11) $m = 3, \lambda = 0.01$



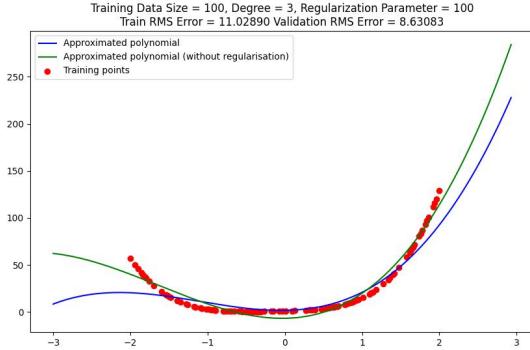
(12) $m = 3, \lambda = 0.1$



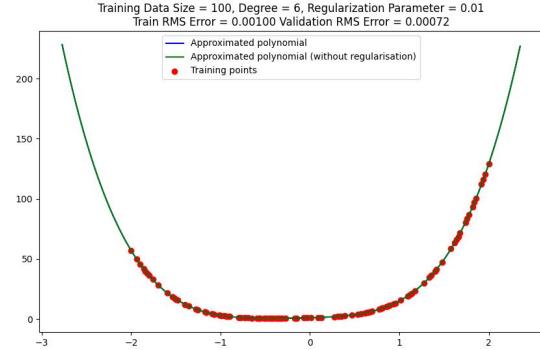
(13) $m = 3, \lambda = 1$



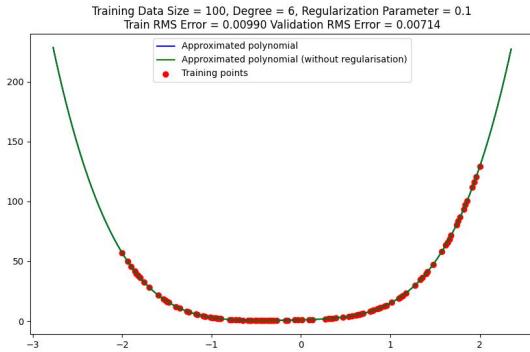
(14) $m = 3, \lambda = 10$



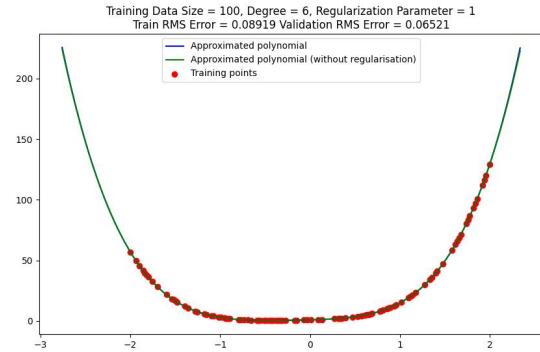
(15) $m = 3, \lambda = 100$



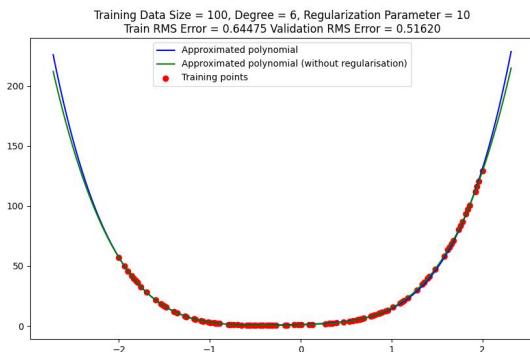
(16) $m = 6, \lambda = 0.01$



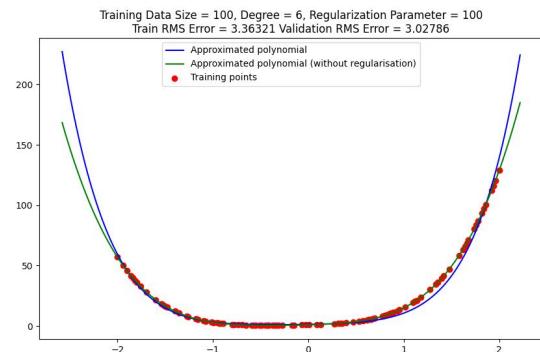
(17) $m = 6, \lambda = 0.1$



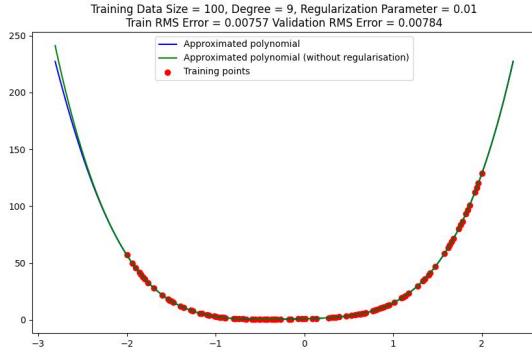
(18) $m = 6, \lambda = 1$



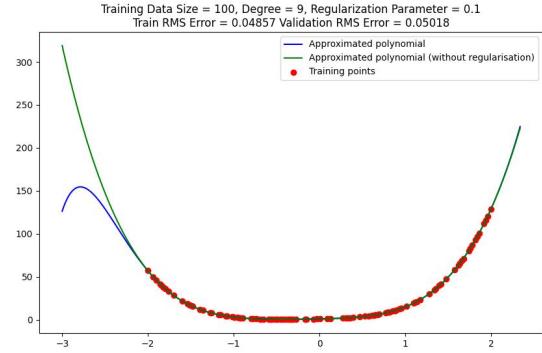
(19) $m = 6, \lambda = 10$



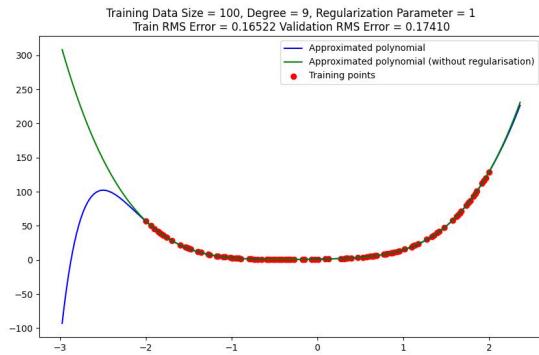
(20) $m = 6, \lambda = 100$



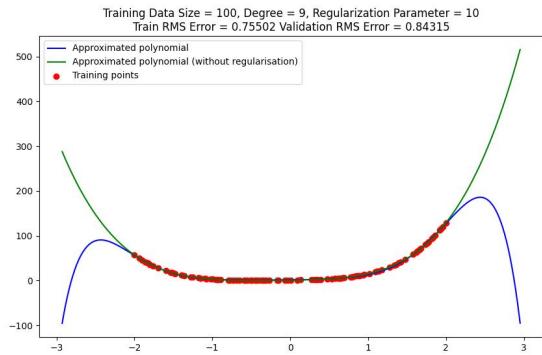
(21) $m = 9, \lambda = 0.01$



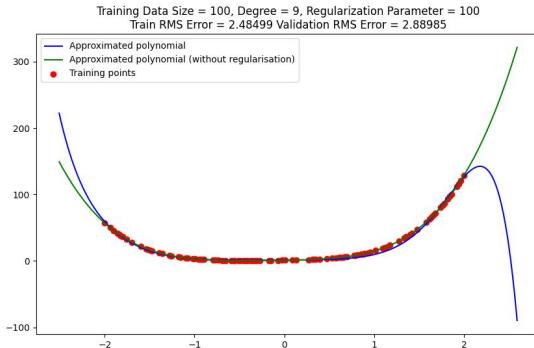
(22) $m = 9, \lambda = 0.1$



(23) $m = 9, \lambda = 1$



(24) $m = 9, \lambda = 10$



(25) $m = 9, \lambda = 100$

Figure 4: Approximated Functions with Regularization for training dataset of size 100

1.2.4 Inferences

1. The models with $m = 1, 2, 3$ continue to inadequately fit to the underlying data. This is a case of underfitting and indicates that the data has a much more complex distribution. Regularization is not of help here, as it enhances validation performance at the expense of training performance.
2. In parallel with the training dataset of size 10, as the regularization parameter λ increases, the training error increases and the validation error decreases.
3. As λ increases, it forces the model to slightly deviate from the training points which reduces overfitting. This can also be seen in increasing training error and reducing validation error as λ increases.
4. However when λ is too large, the model begins to deviate too much from the training data and the training and validation errors increase. This can be seen with $m = 3, 6, 9$ and $\lambda = 100$. Hence, large regularization has led to underfitting of the model in these cases.
5. Once again, we see that the use of regularization has not improved the model's performance, proving the validity of our initial assertion.
6. The dataset appears less sensitive to regularization, as varying regularization parameters doesn't significantly affect the model's performance.
7. Higher polynomial degrees (e.g., 9) don't contribute significantly to model improvement. The dataset seems to have a more linear or lower-degree relationship between features and the target variable.
8. The optimal model complexity seems to lie around degree 6, striking a balance between capturing essential patterns and avoiding unnecessary complexity. This suggests that a polynomial of degree 6 adequately represents the underlying structure of the dataset.

1.2.5 Hyperparameters v/s Training, Validation and Testing Errors

Degree	Regularization Parameter	Training Error	Validation Error	Testing Error
1	0.0	27.423 98	23.732 83	24.773 74
1	0.01	27.423 98	23.731 36	24.774 14
1	0.1	27.423 99	23.718 13	24.777 76
1	1.0	27.425 13	23.587 63	24.814 69
1	10.0	27.522 27	22.443 68	25.240 80
1	100.0	30.49459	18.45269	29.78836
2	0.0	8.207 70	7.648 47	7.958 20
2	0.01	8.207 70	7.647 49	7.958 01
2	0.1	8.207 74	7.638 71	7.956 37
2	1.0	8.211 87	7.554 65	7.943 85
2	10.0	8.501 90	6.978 57	8.098 85
2	100.0	13.65707	5.64613	13.24083
3	0.0	6.071 03	6.229 36	6.707 46
3	0.01	6.071 03	6.228 77	6.707 04
3	0.1	6.071 08	6.223 51	6.703 32
3	1.0	6.076 02	6.176 48	6.669 90
3	10.0	6.40556	6.08537	6.59956
3	100.0	11.028 90	8.630 83	9.782 65
6	0.0	2.19356×10^{-12}	2.43291×10^{-12}	1.86658×10^{-12}
6	0.01	0.001 00	0.000 72	0.001 10
6	0.1	0.009 90	0.007 14	0.010 85
6	1.0	0.089 19	0.065 21	0.100 07
6	10.0	0.644 75	0.516 20	0.743 94
6	100.0	3.363 21	3.027 86	3.755 94
9	0.0	1.60555×10^{-10}	1.39628×10^{-10}	1.88667×10^{-10}
9	0.01	0.007 57	0.007 83	0.006 64
9	0.1	0.048 57	0.050 18	0.043 58
9	1.0	0.165 22	0.174 10	0.163 59
9	10.0	0.755 02	0.843 15	0.741 37
9	100.0	2.484 99	2.889 85	2.388 83

Table 2: Training, Validation and Testing Errors for different m, λ for training dataset of size 100. The best model is taken to be the one with the lowest validation error and are shown in boldface.

2 Task 2: Linear Model for Regression using Polynomial Basis Functions for Dataset 2

2.1 Training Dataset of size 50

2.1.1 Surface Plots of Approximated Functions without Regularization

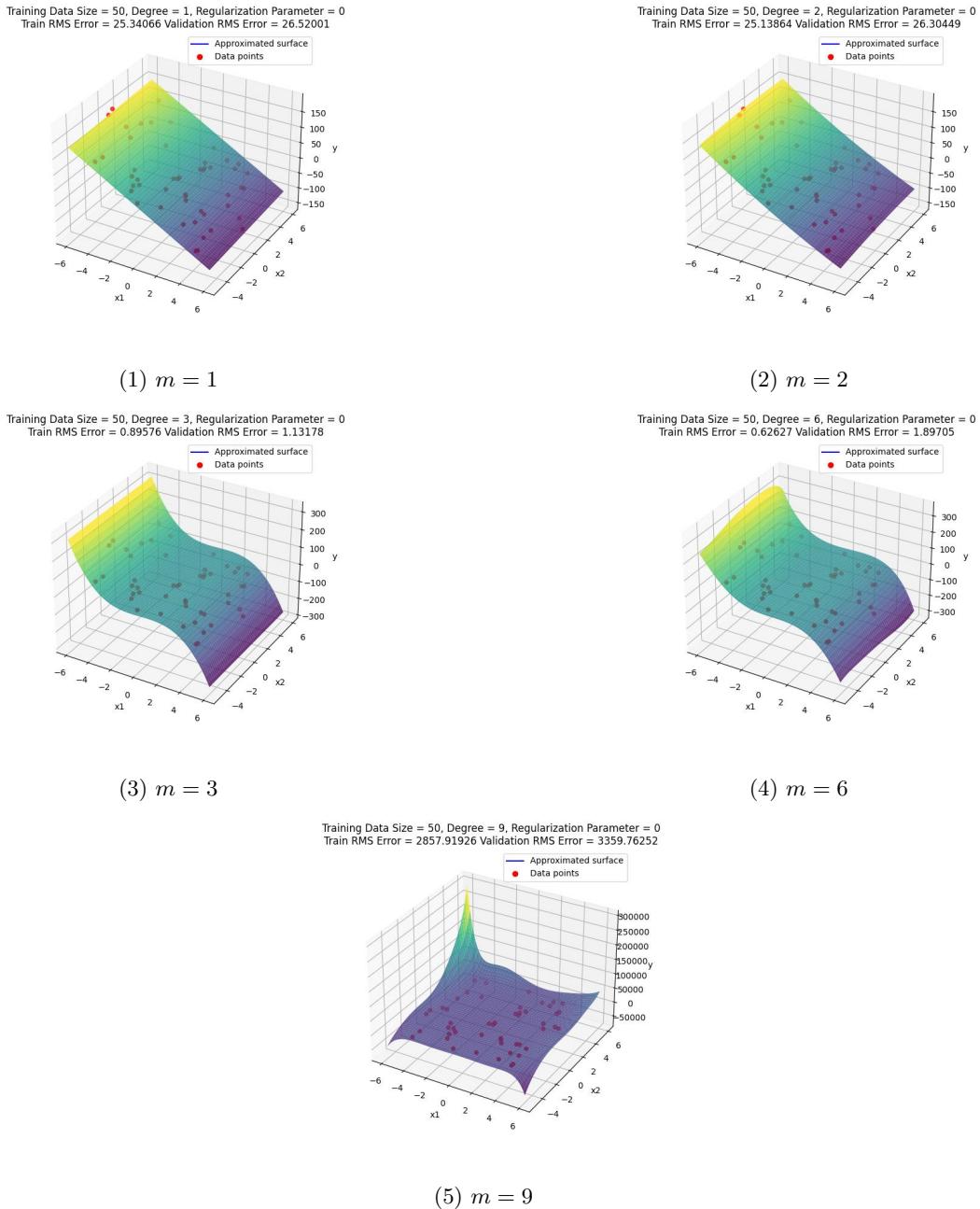


Figure 5: Approximated Functions without Regularization for training dataset of size 50

2.1.2 Inferences

1. **Underfitting (Degree 1-2):** For degrees 1 and 2, the model exhibits underfitting with relatively high training and validation losses. These degrees are too simplistic to capture the complexity of the underlying patterns in the data.

2. **Effective Fitting (Degree 3):** Degree 3 performs well, achieving low training loss (0.89) and validation loss (1.13). This indicates an effective balance between model complexity and fitting the data.

The number of parameters (D) for degree m and input dimensionality d is given by

$$D = \frac{(m+d)!}{m! \cdot d!}.$$

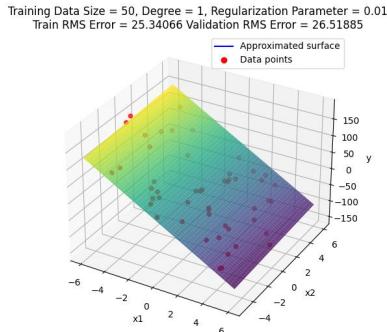
3. **Appropriate Complexity (Degree 6):** Degree 6 shows low training loss (0.62) and a slightly higher validation loss (1.89). The increase in complexity compared to degree 3 is acceptable, and further investigation, such as regularization techniques, could be explored to optimize performance.

4. **Lack of Training Samples (Degree 9):** The extremely high training loss (2857.91) and even higher validation loss (3359.76) for degree 9 indicate that the model is too complex for the limited number of training samples (50). The number of parameters (D) for degree $m = 9$ and input dimensionality $d = 2$ is given by

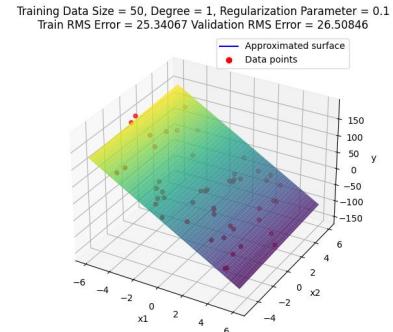
$$D = \frac{(9+2)!}{9! \cdot 2!} = 55.$$

With only 50 samples, the model lacks sufficient data to effectively learn and generalize, resulting in poor performance.

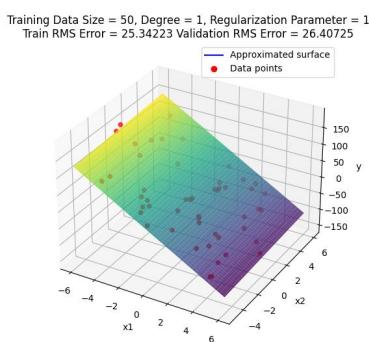
2.1.3 Surface Plots of Approximated Functions with Regularization



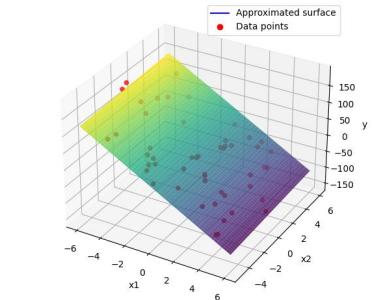
(1) $m = 1, \lambda = 0.01$



(2) $m = 1, \lambda = 0.1$

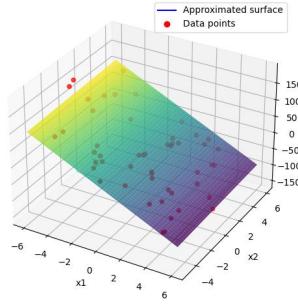


(3) $m = 1, \lambda = 1$



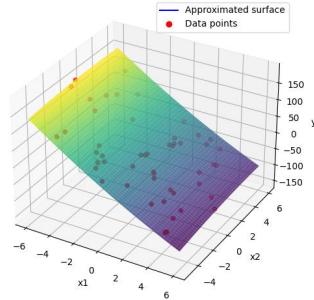
(4) $m = 1, \lambda = 10$

Training Data Size = 50, Degree = 1, Regularization Parameter = 100
 Train RMS Error = 28.81333 Validation RMS Error = 24.40382



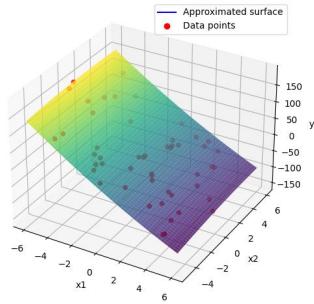
(5) $m = 1, \lambda = 100$

Training Data Size = 50, Degree = 2, Regularization Parameter = 0.01
 Train RMS Error = 25.13864 Validation RMS Error = 26.30349



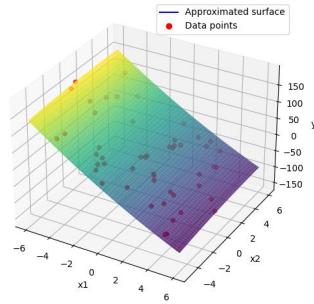
(6) $m = 2, \lambda = 0.01$

Training Data Size = 50, Degree = 2, Regularization Parameter = 0.1
 Train RMS Error = 25.13867 Validation RMS Error = 26.29456



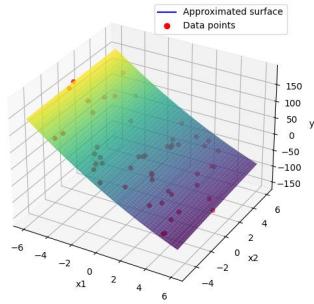
(7) $m = 2, \lambda = 0.1$

Training Data Size = 50, Degree = 2, Regularization Parameter = 1
 Train RMS Error = 25.14150 Validation RMS Error = 26.21237



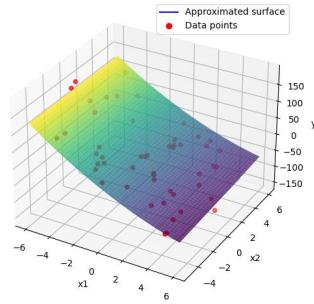
(8) $m = 2, \lambda = 1$

Training Data Size = 50, Degree = 2, Regularization Parameter = 10
 Train RMS Error = 25.26194 Validation RMS Error = 25.73134



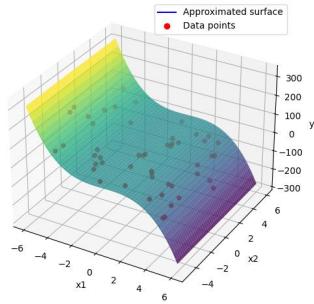
(9) $m = 2, \lambda = 10$

Training Data Size = 50, Degree = 2, Regularization Parameter = 100
 Train RMS Error = 27.84956 Validation RMS Error = 25.14181



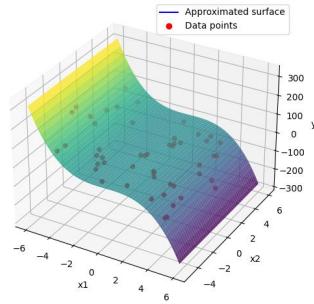
(10) $m = 2, \lambda = 100$

Training Data Size = 50, Degree = 3, Regularization Parameter = 0.01
 Train RMS Error = 0.89576 Validation RMS Error = 1.13174



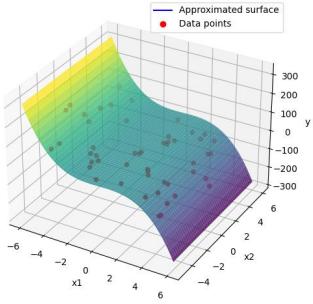
(11) $m = 3, \lambda = 0.01$

Training Data Size = 50, Degree = 3, Regularization Parameter = 0.1
 Train RMS Error = 0.89578 Validation RMS Error = 1.13140



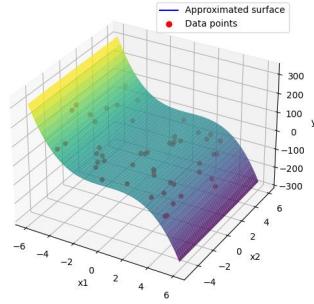
(12) $m = 3, \lambda = 0.1$

Training Data Size = 50, Degree = 3, Regularization Parameter = 1
 Train RMS Error = 0.89773 Validation RMS Error = 1.12922



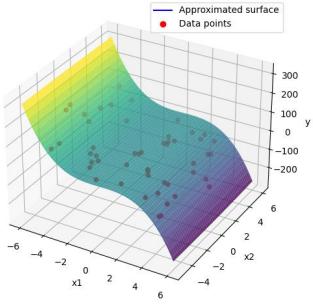
(13) $m = 3, \lambda = 1$

Training Data Size = 50, Degree = 3, Regularization Parameter = 10
 Train RMS Error = 0.97508 Validation RMS Error = 1.14907



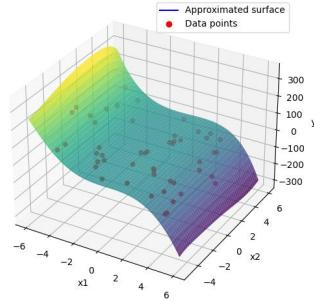
(14) $m = 3, \lambda = 10$

Training Data Size = 50, Degree = 3, Regularization Parameter = 100
 Train RMS Error = 1.39033 Validation RMS Error = 1.35911



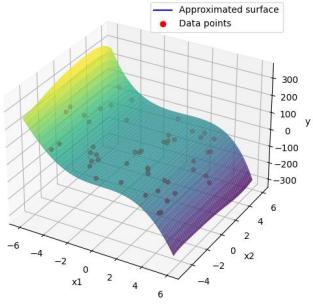
(15) $m = 3, \lambda = 100$

Training Data Size = 50, Degree = 6, Regularization Parameter = 0.01
 Train RMS Error = 0.62627 Validation RMS Error = 1.89561



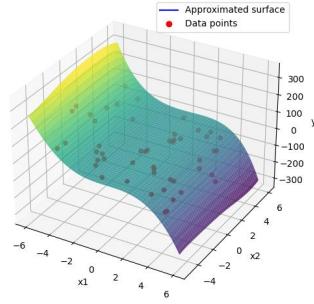
(16) $m = 6, \lambda = 0.01$

Training Data Size = 50, Degree = 6, Regularization Parameter = 0.1
 Train RMS Error = 0.62653 Validation RMS Error = 1.88820



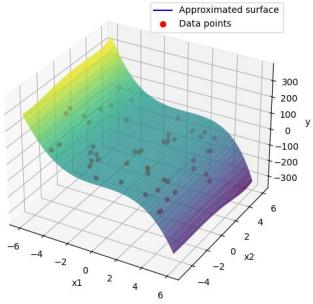
(17) $m = 6, \lambda = 0.1$

Training Data Size = 50, Degree = 6, Regularization Parameter = 1
 Train RMS Error = 0.64222 Validation RMS Error = 1.85420



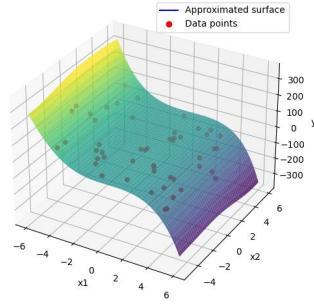
(18) $m = 6, \lambda = 1$

Training Data Size = 50, Degree = 6, Regularization Parameter = 10
 Train RMS Error = 0.79505 Validation RMS Error = 2.08828



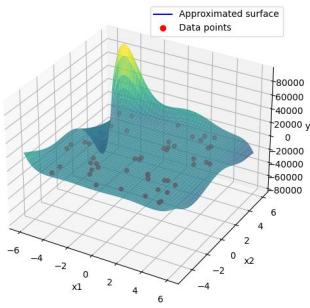
(19) $m = 6, \lambda = 10$

Training Data Size = 50, Degree = 6, Regularization Parameter = 100
 Train RMS Error = 1.17847 Validation RMS Error = 2.63916



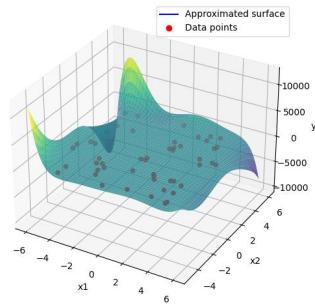
(20) $m = 6, \lambda = 100$

Training Data Size = 50, Degree = 9, Regularization Parameter = 0.01
 Train RMS Error = 0.14174 Validation RMS Error = 733.94670



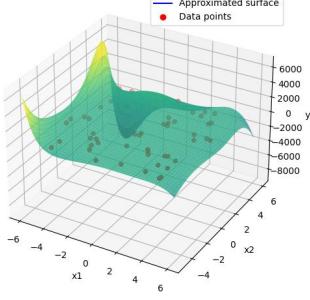
(21) $m = 9, \lambda = 0.01$

Training Data Size = 50, Degree = 9, Regularization Parameter = 0.1
 Train RMS Error = 0.32412 Validation RMS Error = 96.52342



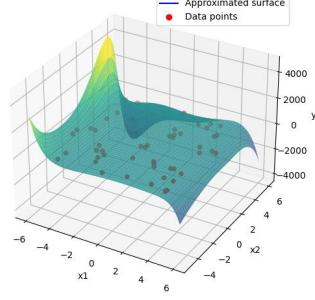
(22) $m = 9, \lambda = 0.1$

Training Data Size = 50, Degree = 9, Regularization Parameter = 1
 Train RMS Error = 0.39136 Validation RMS Error = 77.96556



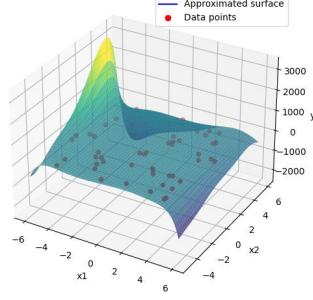
(23) $m = 9, \lambda = 1$

Training Data Size = 50, Degree = 9, Regularization Parameter = 10
 Train RMS Error = 0.55262 Validation RMS Error = 37.53152



(24) $m = 9, \lambda = 10$

Training Data Size = 50, Degree = 9, Regularization Parameter = 100
 Train RMS Error = 0.69083 Validation RMS Error = 21.17746



(25) $m = 9, \lambda = 100$

Figure 6: Approximated Functions with Regularization for training dataset of size 50

2.1.4 Inferences

- Degree 1:** As the regularization parameter λ increases, both training and validation errors increase consistently. This aligns with the expected behavior of regularization, which tends to prevent overfitting. The increase in errors indicates a constraint on the model's complexity, preventing it from fitting the training data too closely.
- Degree 2:** Similar to degree 1, increasing λ leads to higher training and validation errors, indicating consistent regularization effects. The model, being of higher degree, exhibits a more pronounced sensitivity to regularization, emphasizing the need for appropriate parameter tuning.
- Degree 3:** For degree 3, the pattern of increasing errors with higher λ continues, but the impact is less severe compared to degrees 1 and 2. This suggests that the model with degree 3 is more robust to regularization, potentially due to its moderate complexity.

4. **Degree 6:** With degree 6, the trend of increasing errors with higher λ persists. The model appears sensitive to regularization, particularly with $\lambda = 100.0$. This indicates that, at this level of complexity, too much regularization can negatively impact performance.
5. **Degree 9:** Degree 9 exhibits a distinct behavior. With $\lambda = 0.0$, the model has an extremely high training error, suggesting potential underfitting. However, as λ increases, both training and validation errors decrease. This unexpected behaviour could be a result of numerical instability and a lack of input samples to properly train the model.
6. Generally, increasing λ has a regularizing effect, helping to prevent overfitting. However, the impact of regularization varies with the degree of the polynomial and the sample size used to train the model. Polynomial basis functions of degree 3 and 6, tend to have a better overall performance in the train, test and validation dataset.

2.1.5 Scatter Plots of the Best Model

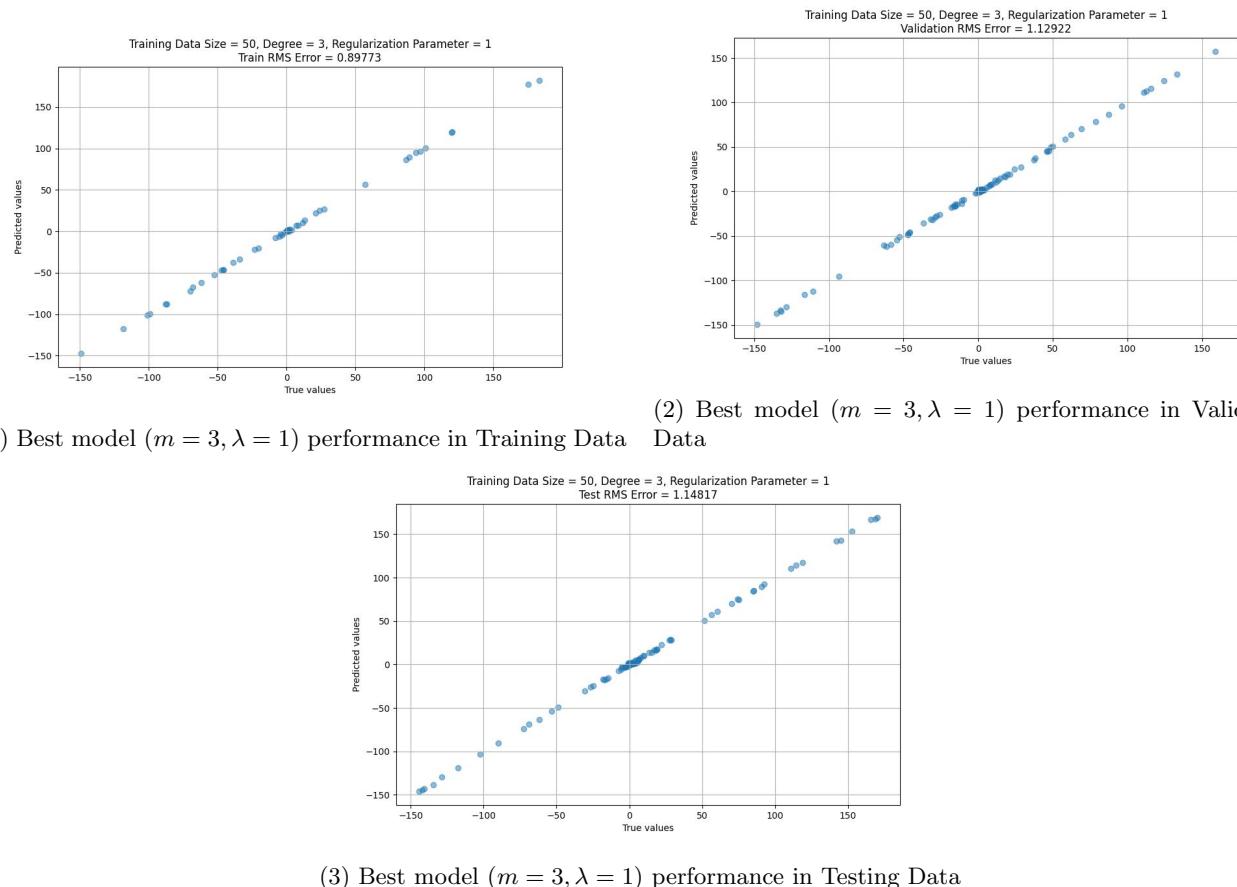


Figure 7: Best model performance with Training, Validation and Testing Data

2.1.6 Hyperparameters v/s Training, Validation and Testing Errors

Degree	Regularization Parameter	Training Error	Validation Error	Testing Error
1	0.0	25.340 66	26.520 01	27.886 23
1	0.01	25.340 66	26.518 85	27.885 66
1	0.1	25.340 67	26.508 46	27.880 51
1	1.0	25.342 23	26.407 25	27.831 35
1	10.0	25.461 38	25.614 48	27.519 58
1	100.0	28.81333	24.40382	29.12246
2	0.0	25.138 64	26.304 49	27.308 84
2	0.01	25.138 64	26.303 49	27.308 37
2	0.1	25.138 67	26.294 56	27.304 11
2	1.0	25.141 50	26.212 37	27.266 80
2	10.0	25.261 94	25.731 34	27.123 93
2	100.0	27.84956	25.14181	28.27266
3	0.0	0.895 76	1.131 78	1.153 01
3	0.01	0.895 76	1.131 74	1.152 94
3	0.1	0.895 78	1.131 40	1.152 37
3	1.0	0.89773	1.12922	1.14817
3	10.0	0.975 08	1.149 07	1.163 26
3	100.0	1.390 33	1.359 11	1.438 95
6	0.0	0.626 27	1.897 05	2.257 23
6	0.01	0.626 27	1.895 61	2.252 13
6	0.1	0.626 53	1.884 07	2.208 32
6	1.0	0.64222	1.85420	1.91613
6	10.0	0.795 05	2.088 28	1.607 29
6	100.0	1.178 47	2.639 16	2.585 78
9	0.0	2857.919 26	3359.762 52	3154.958 08
9	0.01	0.141 74	733.946 70	691.732 34
9	0.1	0.324 12	96.523 42	99.826 46
9	1.0	0.391 36	77.965 56	89.146 19
9	10.0	0.552 62	37.531 52	45.308 10
9	100.0	0.69083	21.17746	8.48779

Table 3: Training, Validation and Testing Errors for different m, λ for training dataset of size 50. The best model is taken to be the one with the lowest validation error and are shown in boldface.

2.2 Training Dataset of size 200

2.2.1 Plots of Approximated Functions without Regularization

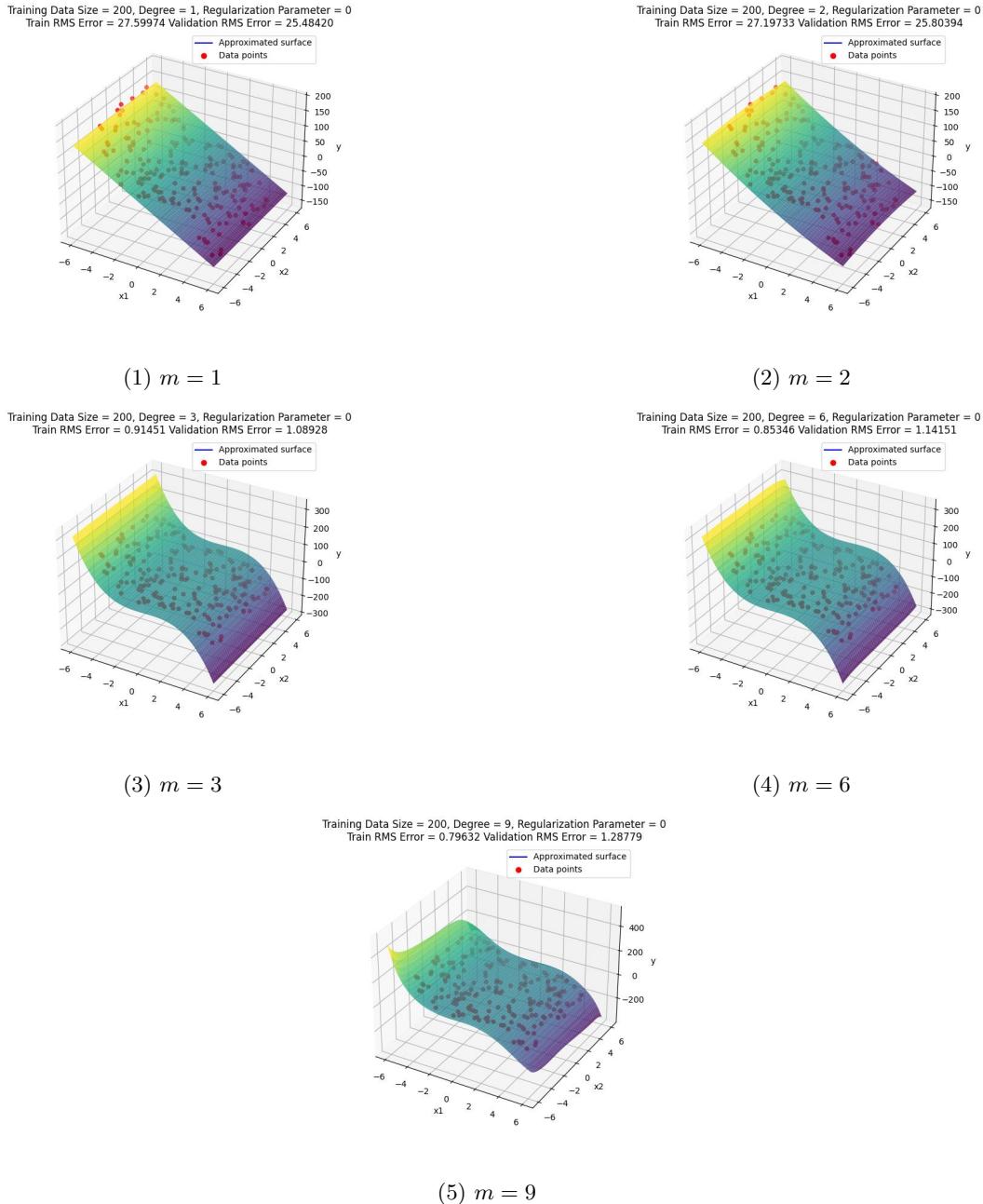


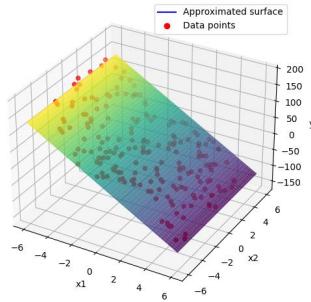
Figure 8: Approximated Functions without Regularization for training dataset of size 200

2.2.2 Inferences

1. For Degree 1 and Degree 2, errors are relatively high across all datasets, indicating potential underfitting and suggesting the need for regularization.
2. Degree 3 exhibits a substantial decrease in errors, indicating improved model fit, especially in the testing dataset.
3. Degrees 6 and 9 continue the trend of decreasing errors, showcasing the benefits of higher degrees. Though with higher degree, we face the issue of overfitting.
4. Generally, there is a need for regularization across all degrees to avoid overfitting. Higher degrees (6 and 9) show a risk of overfitting, particularly when regularization is absent.
5. With 200 data points and no regularization, underfitting is not as pronounced, especially in the case of higher degrees like Degree 6 and 9.
6. The larger number of training samples seems to mitigate the underfitting effect, emphasizing the impact of dataset size on regularization outcomes.
7. The models with higher degrees such as $m = 6, 9$ have much lower training error than validation error. Hence, this may be a case where regularization might improve the generalizability of the model.

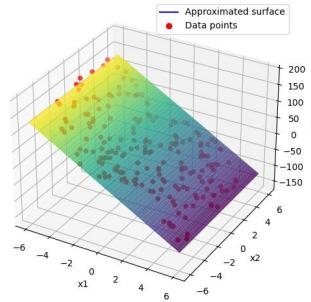
2.2.3 Plots of Approximated Functions with Regularization

Training Data Size = 200, Degree = 1, Regularization Parameter = 0.01
Train RMS Error = 27.59974 Validation RMS Error = 25.48407



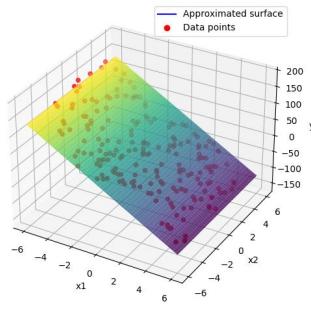
(1) $m = 1, \lambda = 0.01$

Training Data Size = 200, Degree = 1, Regularization Parameter = 0.1
Train RMS Error = 27.59974 Validation RMS Error = 25.48293



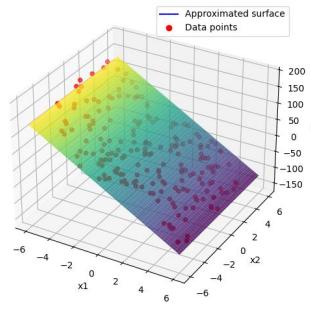
(2) $m = 1, \lambda = 0.1$

Training Data Size = 200, Degree = 1, Regularization Parameter = 1
Train RMS Error = 27.59978 Validation RMS Error = 25.47161



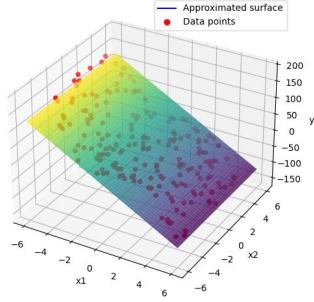
(3) $m = 1, \lambda = 1$

Training Data Size = 200, Degree = 1, Regularization Parameter = 10
Train RMS Error = 27.60367 Validation RMS Error = 25.36355



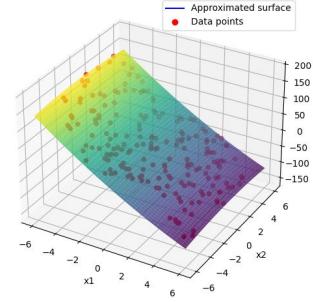
(4) $m = 1, \lambda = 10$

Training Data Size = 200, Degree = 1, Regularization Parameter = 100
 Train RMS Error = 27.87339 Validation RMS Error = 24.63102



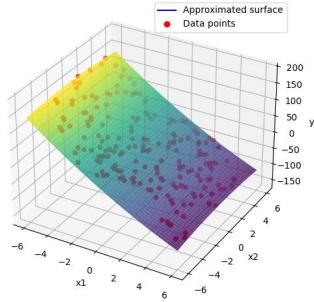
(5) $m = 1, \lambda = 100$

Training Data Size = 200, Degree = 2, Regularization Parameter = 0.01
 Train RMS Error = 27.19733 Validation RMS Error = 25.80385



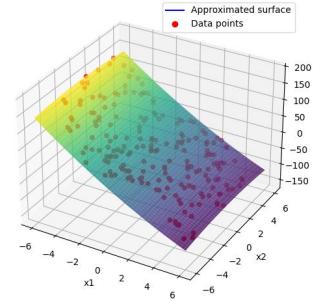
(6) $m = 2, \lambda = 0.01$

Training Data Size = 200, Degree = 2, Regularization Parameter = 0.1
 Train RMS Error = 27.19733 Validation RMS Error = 25.80303



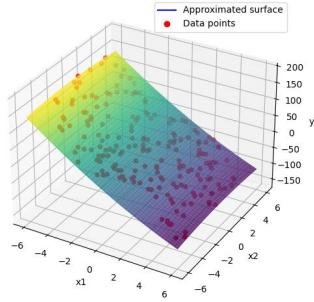
(7) $m = 2, \lambda = 0.1$

Training Data Size = 200, Degree = 2, Regularization Parameter = 1
 Train RMS Error = 27.19738 Validation RMS Error = 25.79493



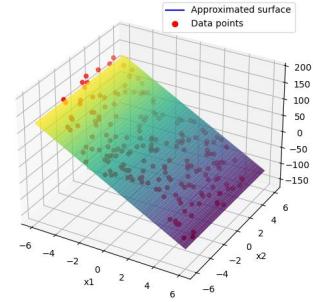
(8) $m = 2, \lambda = 1$

Training Data Size = 200, Degree = 2, Regularization Parameter = 10
 Train RMS Error = 27.20140 Validation RMS Error = 25.71845



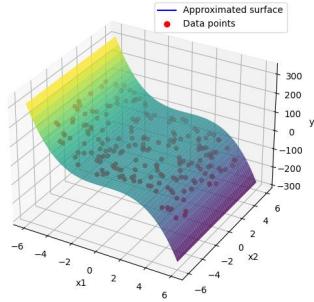
(9) $m = 2, \lambda = 10$

Training Data Size = 200, Degree = 1, Regularization Parameter = 100
 Train RMS Error = 27.87339 Validation RMS Error = 24.63102



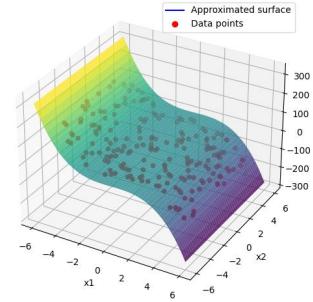
(10) $m = 2, \lambda = 100$

Training Data Size = 200, Degree = 3, Regularization Parameter = 0.01
 Train RMS Error = 0.91451 Validation RMS Error = 1.08889



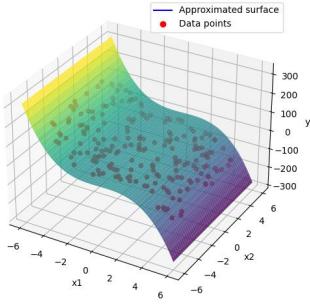
(11) $m = 3, \lambda = 0.01$

Training Data Size = 200, Degree = 3, Regularization Parameter = 0.1
 Train RMS Error = 0.91451 Validation RMS Error = 1.08889



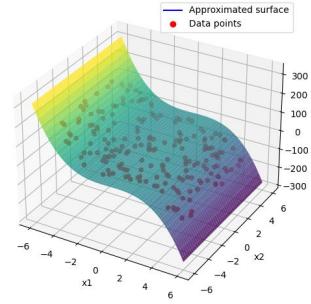
(12) $m = 3, \lambda = 0.1$

Training Data Size = 200, Degree = 3, Regularization Parameter = 1
 Train RMS Error = 0.91469 Validation RMS Error = 1.08557



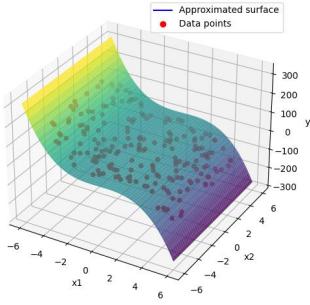
(13) $m = 3, \lambda = 1$

Training Data Size = 200, Degree = 3, Regularization Parameter = 10
 Train RMS Error = 0.92767 Validation RMS Error = 1.06763



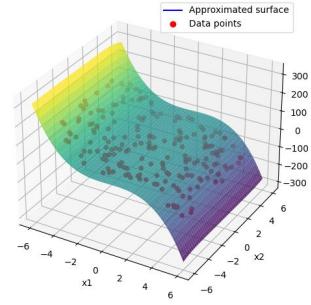
(14) $m = 3, \lambda = 10$

Training Data Size = 200, Degree = 3, Regularization Parameter = 100
 Train RMS Error = 1.14335 Validation RMS Error = 1.14774



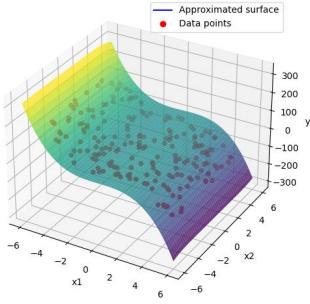
(15) $m = 3, \lambda = 100$

Training Data Size = 200, Degree = 6, Regularization Parameter = 0.01
 Train RMS Error = 0.85346 Validation RMS Error = 1.14144



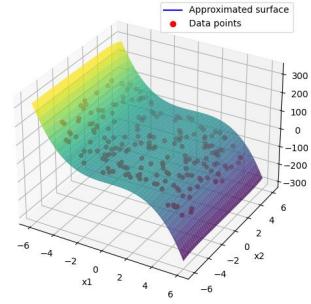
(16) $m = 6, \lambda = 0.01$

Training Data Size = 200, Degree = 6, Regularization Parameter = 0.1
 Train RMS Error = 0.85347 Validation RMS Error = 1.14082



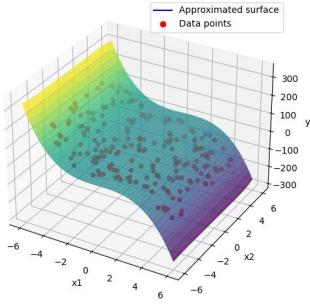
(17) $m = 6, \lambda = 0.1$

Training Data Size = 200, Degree = 6, Regularization Parameter = 1
 Train RMS Error = 0.85452 Validation RMS Error = 1.13587



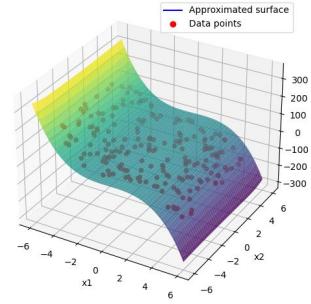
(18) $m = 6, \lambda = 1$

Training Data Size = 200, Degree = 6, Regularization Parameter = 10
 Train RMS Error = 0.88953 Validation RMS Error = 1.13379



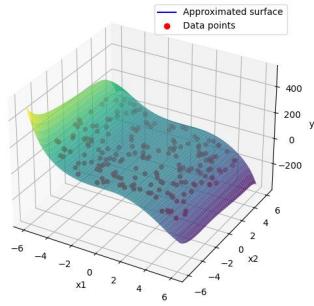
(19) $m = 6, \lambda = 10$

Training Data Size = 200, Degree = 6, Regularization Parameter = 100
 Train RMS Error = 1.05975 Validation RMS Error = 1.22639



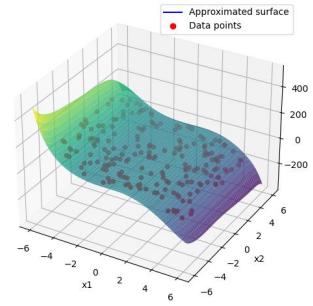
(20) $m = 6, \lambda = 100$

Training Data Size = 200, Degree = 9, Regularization Parameter = 0.01
 Train RMS Error = 0.79632 Validation RMS Error = 1.28793



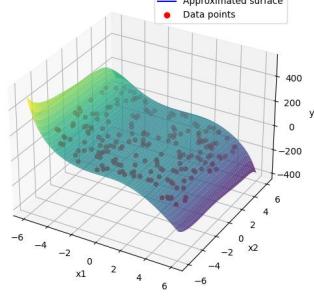
(21) $m = 9, \lambda = 0.01$

Training Data Size = 200, Degree = 9, Regularization Parameter = 0.1
 Train RMS Error = 0.79635 Validation RMS Error = 1.28922



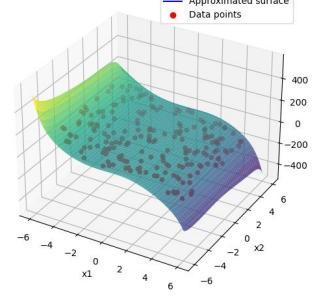
(22) $m = 9, \lambda = 0.1$

Training Data Size = 200, Degree = 9, Regularization Parameter = 1
 Train RMS Error = 0.79839 Validation RMS Error = 1.30320



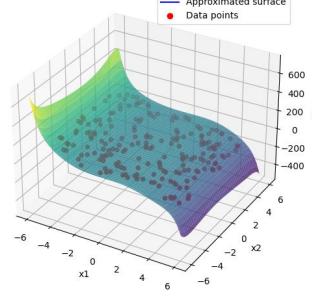
(23) $m = 9, \lambda = 1$

Training Data Size = 200, Degree = 9, Regularization Parameter = 10
 Train RMS Error = 0.83213 Validation RMS Error = 1.38413



(24) $m = 9, \lambda = 10$

Training Data Size = 200, Degree = 9, Regularization Parameter = 100
 Train RMS Error = 0.99531 Validation RMS Error = 1.55550



(25) $m = 9, \lambda = 100$

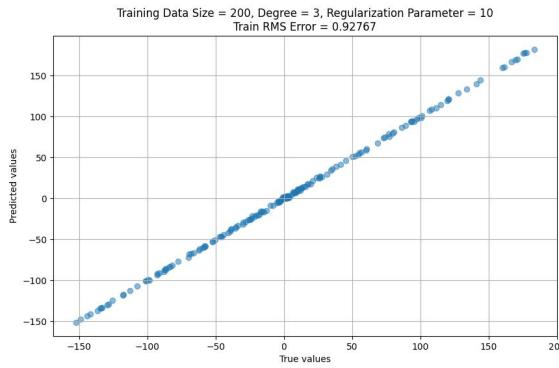
Figure 9: Approximated Functions with Regularization for training dataset of size 200

2.2.4 Inferences

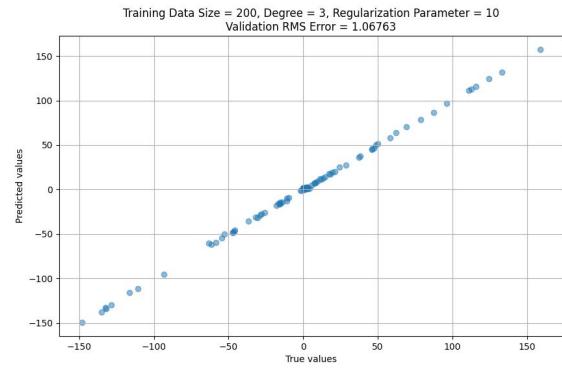
1. Higher degrees generally reduce the training error. But similar to the 50 training sample case, we observe overfitting in these cases.
2. Low degrees $m = 1, 2$ exhibit underfitting, with validation errors decreasing as regularization increases.
3. For Degree 3, a moderate regularization parameter $\lambda = 1$ yields good results, but higher degrees show overfitting tendencies.
4. Degrees 6 and 9 follow similar trends, with higher regularization improving performance, but overfitting is more pronounced.
5. Low regularization $\lambda = 0, 0.01$ leads to high errors due to overfitting in case of higher degrees $m = 6, 9$.

6. Moderate regularization $\lambda = 0.1$, 1 strikes a balance, preventing overfitting and providing good generalization.
7. High regularization $\lambda = 10$ is effective, but very high $\lambda = 100$ leads to underfitting.
8. Underfitting is observed with simple models i.e, low degrees and high regularization.
9. Overfitting is evident with higher degrees and low regularization, leading to a decrease in testing performance.
10. Beyond polynomial degree 6, the performance improvement becomes marginal, suggesting diminishing returns with increasing model complexity.
11. Polynomial degrees 6 and 3 outperform degree 9, indicating an optimal complexity level that balances bias and variance for this dataset.
12. Regularization has improved the performance of the model, thus validating our original assertion about the same.

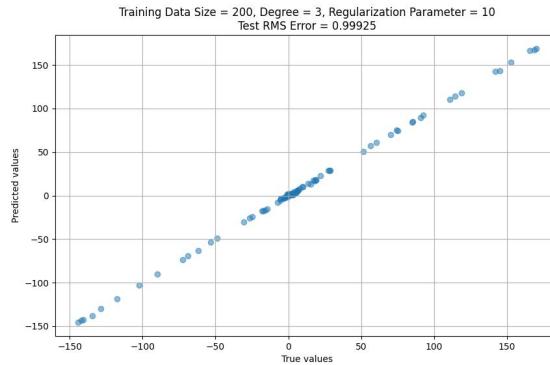
2.2.5 Scatter Plots of the Best Model



(1) Best model ($m = 3, \lambda = 10$) performance in Training Data



(2) Best model ($m = 3, \lambda = 10$) performance in Validation Data



(3) Best model ($m = 3, \lambda = 10$) performance in Testing Data

Figure 10: Best model performance with Training, Validation and Testing Data

2.2.6 Hyperparameters v/s Training, Validation and Testing Errors

Degree	Regularization Parameter	Training Error	Validation Error	Testing Error
1	0.0	27.599 74	25.484 20	27.372 36
1	0.01	27.599 74	25.484 07	27.372 33
1	0.1	27.599 74	25.482 93	27.372 12
1	1.0	27.599 78	25.471 61	27.370 05
1	10.0	27.603 67	25.363 55	27.353 14
1	100.0	27.87339	24.63102	27.42702
2	0.0	27.197 33	25.803 94	26.946 46
2	0.01	27.197 33	25.803 85	26.946 49
2	0.1	27.197 33	25.803 03	26.946 75
2	1.0	27.197 38	25.794 93	26.949 24
2	10.0	27.201 40	25.718 45	26.970 62
2	100.0	27.41277	25.14321	27.09778
3	0.0	0.914 51	1.089 28	1.029 57
3	0.01	0.914 51	1.089 24	1.029 52
3	0.1	0.914 51	1.088 89	1.029 04
3	1.0	0.914 69	1.085 57	1.024 50
3	10.0	0.92767	1.06763	0.99925
3	100.0	1.143 35	1.147 74	1.119 98
6	0.0	0.853 46	1.141 51	1.185 74
6	0.01	0.853 46	1.141 44	1.185 52
6	0.1	0.853 47	1.140 82	1.183 55
6	1.0	0.854 52	1.135 87	1.166 50
6	10.0	0.88953	1.13379	1.11094
6	100.0	1.059 75	1.226 39	1.255 13
9	0.0	0.796 32	1.287 79	1.217 48
9	0.01	0.79632	1.28793	1.21703
9	0.1	0.796 35	1.289 22	1.213 25
9	1.0	0.798 39	1.303 20	1.189 97
9	10.0	0.832 13	1.384 13	1.191 77
9	100.0	0.995 31	1.555 50	1.460 03

Table 4: Training, Validation and Testing Errors for different m, λ for training dataset of size 200. The best model is taken to be the one with the lowest validation error and are shown in boldface.

3 Task 3.1: Linear Model for Regression using Gaussian Basis Functions for Dataset 2

3.1 Training Dataset of size 50

3.1.1 Hyperparameters v/s Training, Validation and Testing Errors

Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
5	5	0	35.273 06	36.402 88	39.326 88
5	5	0.01	35.662 19	37.195 47	40.757 77
5	5	0.1	36.017 30	37.429 22	41.299 28
5	5	1	35.51236	34.34175	38.17339
5	5	10	47.830 11	38.069 93	45.261 82
5	5	100	65.215 37	52.063 07	60.162 15
5	10	0	28.086 64	29.910 68	31.782 25
5	10	0.01	28.126 03	29.780 30	31.619 11
5	10	0.1	28.281 30	29.452 46	31.702 75
5	10	1	31.66447	28.41727	33.07450
5	10	10	53.795 90	42.487 39	49.947 23
5	10	100	67.416 66	53.995 61	61.701 13
5	15	0	26.530 99	27.961 72	29.473 38
5	15	0.01	26.581 31	27.901 42	29.571 23
5	15	0.1	27.10568	27.00792	29.54564
5	15	1	44.118 60	34.592 18	41.576 21
5	15	10	63.118 73	50.335 65	57.885 77
5	15	100	68.664 24	55.106 72	62.638 23
5	20	0	25.896 91	27.217 46	28.440 12
5	20	0.01	26.010 19	27.036 01	28.735 01
5	20	0.1	29.32086	26.02244	30.20199
5	20	1	53.103 54	41.863 07	49.099 26
5	20	10	66.863 01	53.571 28	61.071 60
5	20	100	69.045 22	55.419 98	63.084 28
10	5	0	9.78331	14.81851	15.41025
10	5	0.01	24.348 73	30.847 36	34.439 38
10	5	0.1	29.235 12	33.097 72	36.741 51
10	5	1	33.389 20	33.764 48	37.482 36
10	5	10	42.443 71	34.397 77	41.354 92
10	5	100	61.767 64	49.007 86	56.615 64
10	10	0	6.33656	9.21987	6.98774
10	10	0.01	27.818 66	29.621 02	31.339 42
10	10	0.1	27.979 84	29.462 16	31.365 40
10	10	1	29.418 51	27.811 33	31.142 77
10	10	10	44.327 61	35.476 66	42.032 02
10	10	100	66.015 04	52.743 45	60.725 92
10	15	0	2.69528	2.53537	2.89432
10	15	0.01	26.502 52	27.868 94	29.431 10
10	15	0.1	26.671 73	27.449 14	29.476 11
10	15	1	33.664 20	27.966 71	33.317 65
10	15	10	59.107 91	46.866 90	54.368 10
10	15	100	67.953 58	54.493 56	62.038 27
10	20	0	2.03270	3.00227	3.01889
10	20	0.01	26.087 29	27.199 46	28.956 31
10	20	0.1	26.777 83	26.230 96	28.822 34
10	20	1	44.197 11	34.753 17	41.573 78

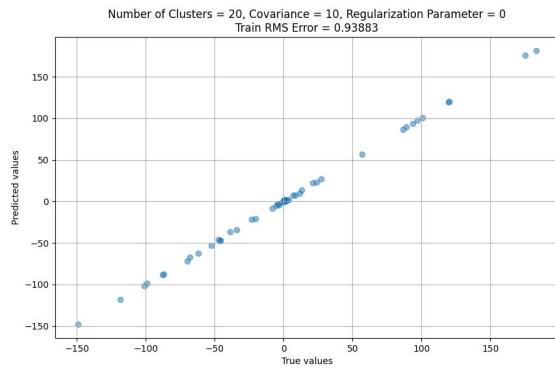
Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
10	20	10	64.87775	51.79070	59.34001
10	20	100	68.78092	55.19547	62.79604
15	5	0	8.01771	10.01373	11.22802
15	5	0.01	13.32150	17.55516	18.77483
15	5	0.1	25.07280	28.79990	31.60346
15	5	1	33.40876	34.67799	38.30222
15	5	10	39.45771	32.72374	39.27926
15	5	100	58.91456	46.50618	54.46583
15	10	0	1.70331	3.32245	3.24209
15	10	0.01	27.71668	29.55517	31.31712
15	10	0.1	27.97338	29.47954	31.44220
15	10	1	28.74118	28.30975	31.02795
15	10	10	41.31934	33.01606	39.38697
15	10	100	63.66620	50.74322	58.41065
15	15	0	1.66608	2.65128	2.83373
15	15	0.01	26.52279	27.90670	29.46858
15	15	0.1	26.65168	27.65500	29.63111
15	15	1	31.54201	27.09625	31.88538
15	15	10	54.59939	43.18430	50.31619
15	15	100	67.52468	54.07945	61.75183
15	20	0	5.80971	5.51535	5.51729
15	20	0.01	26.01961	27.25059	28.79771
15	20	0.1	26.46220	26.21044	28.57625
15	20	1	40.48936	32.05142	38.56713
15	20	10	62.77663	50.04402	57.47787
15	20	100	68.31596	54.80567	62.34214
20	5	0	3.10027	7.55743	7.22285
20	5	0.01	11.39858	15.91433	16.03120
20	5	0.1	23.67399	28.15636	30.68577
20	5	1	32.32732	34.24839	37.94277
20	5	10	39.56077	32.68318	38.77947
20	5	100	56.83247	44.43328	52.13406
20	10	0	0.93883	1.79379	1.78864
20	10	0.01	27.50591	29.42691	31.23208
20	10	0.1	28.03943	29.42641	31.23309
20	10	1	28.48698	28.24455	30.74272
20	10	10	38.11847	30.88401	37.00021
20	10	100	63.62359	50.63896	58.44636
20	15	0	13.77664	13.54697	13.53529
20	15	0.01	26.44217	27.87578	29.35038
20	15	0.1	26.68331	27.39421	29.13552
20	15	1	29.19267	26.42054	30.38021
20	15	10	53.62104	42.15080	49.63501
20	15	100	66.62654	53.31717	60.87135
20	20	0	37.77226	38.31106	38.36692
20	20	0.01	26.03850	27.33075	28.93332
20	20	0.1	26.25120	26.59861	28.82540
20	20	1	34.59378	28.77672	34.03547
20	20	10	62.24702	49.65485	57.09016
20	20	100	68.07175	54.59634	62.15012

Table 5: Training, Validation and Testing Errors for different K, σ, λ for training dataset of size 50. The best model is taken to be the one with the lowest validation error and are shown in boldface.

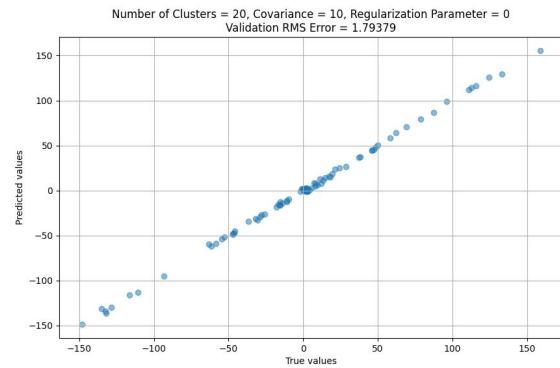
3.1.2 Inferences

1. Generally, increasing clusters tends to decrease training errors, possibly leading to overfitting.
2. Configuration with 20 clusters, $\lambda = 0$, and covariance of 10 exhibits the least training error at 0.9388. It also performs well on the validation and test data.
3. Increasing regularization parameter λ raises the training and reduces the validation errors. However we observe an increase in the validation error for large values of λ (10, 100).
4. Non-zero $\lambda = 0.01, 0.1$ values show a decrease in validation errors, suggesting a reduction in overfitting.
5. The number of clusters influences training errors. An increase in clusters generally reduces the errors.
6. Configuration with $\lambda = 100$ consistently shows higher errors across all datasets, indicating potential oversensitivity to regularization.
7. The error is also dependent on covariance. A general observed trend is that the error drops as the covariance increases. But above a certain threshold, the error starts rising with a rise in covariance. The optimal value of covariance depends on number of clusters chosen for the data. Typically, we can observe a moderate value of covariance $\sigma = 10, 15$, which seems to be the optimal value.
8. The model has roughly similar performance on train, test and validation data, given that the optimal parameters are chosen.
9. The model performs optimally with $\lambda = 0$, covariance = 10, and 20 clusters.

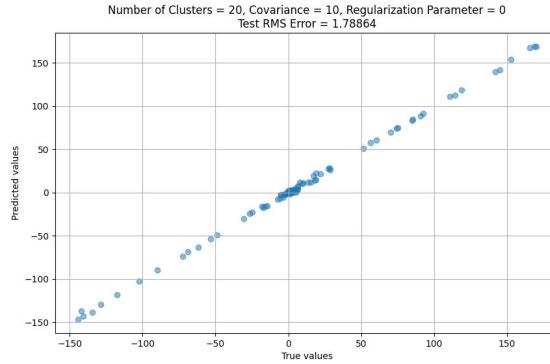
3.1.3 Scatter Plots of the Best Model



(1) Best model ($K = 20, \sigma = 10, \lambda = 0$) performance in Training Data



(2) Best model ($K = 20, \sigma = 10, \lambda = 0$) performance in Validation Data



(3) Best model ($K = 20, \sigma = 10, \lambda = 0$) performance in Testing Data

Figure 11: Best model performance with Training, Validation and Testing Data

3.2 Training dataset of size 200

3.2.1 Hyperparameters v/s Training, Validation and Testing Errors

Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
5	5	0	40.640 18	35.862 42	39.960 56
5	5	0.01	39.063 53	35.150 61	39.696 44
5	5	0.1	39.066 89	35.045 01	39.600 92
5	5	1	40.145 58	34.767 84	39.609 70
5	5	10	41.42510	32.88108	38.93777
5	5	100	59.090 50	44.155 42	52.214 15
5	10	0	31.468 82	28.752 14	30.895 21
5	10	0.01	31.603 62	28.571 43	31.269 16
5	10	0.1	31.662 34	28.552 36	31.212 75
5	10	1	31.84037	27.79075	31.25499
5	10	10	41.112 52	30.708 98	37.283 74
5	10	100	64.658 12	48.857 58	56.357 48
5	15	0	29.268 93	26.998 04	28.730 17
5	15	0.01	29.328 68	26.872 98	29.005 99
5	15	0.1	29.398 36	26.578 80	29.095 30
5	15	1	31.91692	25.74882	30.47151
5	15	10	54.532 52	40.662 65	47.942 13
5	15	100	70.135 44	53.516 17	61.060 98
5	20	0	28.582 94	26.240 94	28.205 88
5	20	0.01	28.547 26	26.169 77	28.243 77
5	20	0.1	28.89758	25.43993	28.38748
5	20	1	38.366 52	28.673 93	34.989 81
5	20	10	64.494 18	48.826 59	56.232 59
5	20	100	71.691 89	54.838 27	62.354 84
10	5	0	24.491 47	22.512 41	24.501 72
10	5	0.01	19.87771	17.69592	17.98640
10	5	0.1	35.333 69	31.691 05	35.025 74
10	5	1	37.262 93	32.703 40	37.243 84
10	5	10	39.950 33	32.777 68	38.162 86
10	5	100	52.665 87	38.970 96	47.058 18
10	10	0	4.71791	4.30419	3.74850
10	10	0.01	30.678 25	28.534 03	30.699 82
10	10	0.1	31.118 71	28.545 52	30.844 56
10	10	1	31.420 10	27.994 70	30.831 35
10	10	10	35.251 07	27.656 55	32.992 69
10	10	100	58.586 30	43.957 51	51.250 08
10	15	0	2.20590	1.84296	2.07088
10	15	0.01	29.087 43	26.909 81	28.808 13
10	15	0.1	29.274 59	26.670 68	28.867 35
10	15	1	30.015 63	25.706 83	29.185 21
10	15	10	44.223 69	32.677 04	39.594 59
10	15	100	67.194 67	51.078 77	58.564 54
10	20	0	4.81130	5.03836	4.42378
10	20	0.01	28.496 44	26.241 69	28.247 20
10	20	0.1	28.615 20	25.820 12	28.201 92
10	20	1	32.043 04	25.369 08	30.244 51
10	20	10	57.172 37	42.786 35	49.930 28
10	20	100	70.509 48	53.866 52	61.307 50

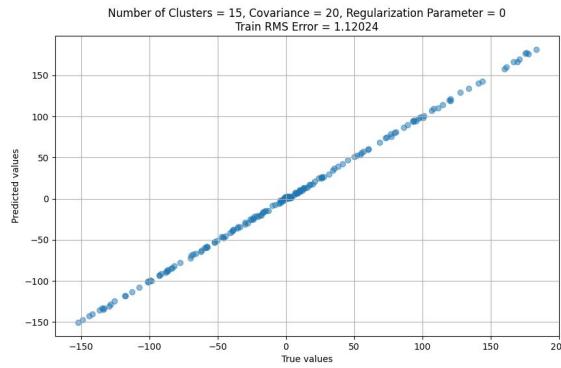
Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
15	5	0	8.37685	7.66738	8.59318
15	5	0.01	10.64133	9.13697	10.13170
15	5	0.1	21.21437	19.04304	21.29036
15	5	1	34.37882	30.66566	34.08519
15	5	10	39.26076	32.82260	38.06710
15	5	100	49.17241	36.72589	44.34394
15	10	0	2.70458	2.68651	3.33026
15	10	0.01	30.30748	28.25943	30.43547
15	10	0.1	31.04407	28.47048	30.81257
15	10	1	31.25951	28.20297	30.84338
15	10	10	33.76607	27.30418	32.37265
15	10	100	54.93497	40.75410	48.46101
15	15	0	1.45429	1.46338	1.65193
15	15	0.01	28.97072	26.95058	28.69494
15	15	0.1	29.19181	26.76684	28.88367
15	15	1	29.63392	25.92728	28.98601
15	15	10	39.63551	29.46910	35.75267
15	15	100	64.54202	48.82060	56.18210
15	20	0	1.12024	1.20874	1.27879
15	20	0.01	28.48955	26.23874	28.21082
15	20	0.1	28.59211	26.01817	28.33161
15	20	1	30.78269	25.01692	29.40824
15	20	10	51.21829	38.08195	45.38690
15	20	100	69.49993	52.97435	60.55205
20	5	0	6.69271	6.65142	7.27503
20	5	0.01	8.53701	7.32387	8.49601
20	5	0.1	19.27048	17.51689	18.18428
20	5	1	33.35760	29.61330	33.02124
20	5	10	38.71246	32.48889	37.25133
20	5	100	46.22424	35.02603	42.57546
20	10	0	2.03519	2.17390	2.44225
20	10	0.01	30.10714	28.05197	30.12822
20	10	0.1	30.87145	28.52602	30.69710
20	10	1	31.32284	28.23347	30.89666
20	10	10	32.53992	27.29130	31.54649
20	10	100	51.74868	38.30528	45.72373
20	15	0	1.83906	2.24362	2.42230
20	15	0.01	28.93564	26.97073	28.68273
20	15	0.1	29.20468	26.79478	28.90353
20	15	1	29.52257	26.08519	28.91251
20	15	10	36.74244	27.83536	33.74877
20	15	100	62.90249	47.40523	55.10658
20	20	0	14.65931	14.82575	14.77602
20	20	0.01	28.43902	26.23238	28.13948
20	20	0.1	28.57121	26.02946	28.21959
20	20	1	29.63908	25.16323	28.85796
20	20	10	46.70780	34.47975	41.39973
20	20	100	68.24246	51.96459	59.45732

Table 6: Training, Validation and Testing Errors for different K, σ, λ for training dataset 2 of size 200. The best model is taken to be the one with the lowest validation error and are shown in boldface.

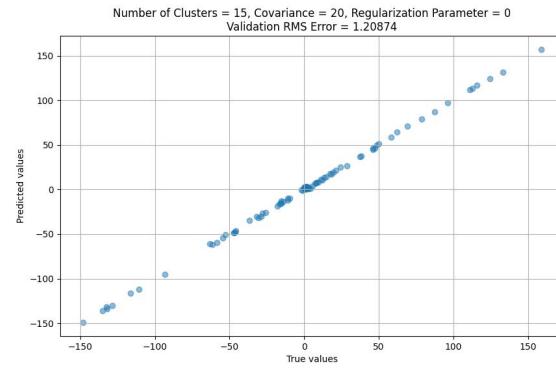
3.2.2 Inferences

1. Increasing the number of clusters generally reduces training errors but can lead to overfitting.
2. Effect of regularization on the errors is similar to what was observed in case of 50 data points.
3. With 200 samples, the model consistently exhibits a similar magnitude of the training and validation errors, indicating less overfitting and generalization.
4. The relation between covariance and model performance is similar to the case of 50 sample test inputs. Here a moderate covariance value of $\sigma = 15$, seems to be the optimal value for different cluster sizes.
5. The model has a similar performance on the training, validation and testing data when optimal parameters are chosen.
6. The model performs optimally with $\lambda = 0$, covariance = 20, and 15 clusters.
7. The dataset exhibits resilience against overfitting, as evidenced by the optimal configuration's preference for a regularization parameter of 0. This suggests that the model generalizes well without the need for strong regularization
8. The model's performance appears to be sensitive to changes in covariance values. The fact that the optimal configuration includes a covariance of 10 indicates that the dataset responds favorably to a moderate level of covariance.

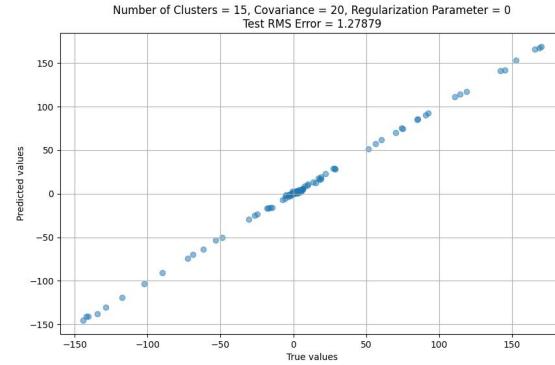
3.2.3 Scatter Plots of the Best Model



(1) Best model ($K = 15, \sigma = 20, \lambda = 0$) performance in Training Data



(2) Best model ($K = 15, \sigma = 20, \lambda = 0$) performance in Validation Data



(3) Best model ($K = 15, \sigma = 20, \lambda = 0$) performance in Testing Data

Figure 12: Best model performance with Training, Validation and Testing Data

4 Task 3.2: Linear Model for Regression using Gaussian Basis Functions for Dataset 3

4.1 Hyperparameters v/s Training, Validation and Testing Errors

Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
70	30	0	4.859 90	5.860 90	4.884 40
70	30	0.01	4.547 01	5.758 11	4.375 50
70	30	0.1	4.97758	5.69835	4.97815
70	30	1	5.454 79	6.044 21	5.148 35
70	30	10	6.604 19	6.780 48	5.832 49
70	30	100	7.588 41	7.694 55	6.444 55
70	35	0	4.492 46	5.550 18	4.308 94
70	35	0.01	4.44613	5.13883	4.13445
70	35	0.1	4.642 19	5.534 98	4.565 31
70	35	1	5.195 98	5.673 02	4.623 87
70	35	10	6.545 91	6.694 33	5.666 87
70	35	100	7.467 05	7.594 60	6.411 17
70	40	0	4.179 18	7.960 01	10.644 99
70	40	0.01	4.26652	4.94447	4.23797
70	40	0.1	4.492 84	5.031 10	4.224 03
70	40	1	4.919 02	5.403 96	4.316 62
70	40	10	6.292 91	6.459 31	5.217 22
70	40	100	7.371 95	7.530 51	6.353 26
70	45	0	4.057 28	5.010 81	3.630 40
70	45	0.01	4.03433	4.83836	4.55810
70	45	0.1	4.490 24	4.948 23	3.766 78
70	45	1	4.799 52	5.357 22	4.113 64
70	45	10	6.152 42	6.306 33	5.284 58
70	45	100	7.379 11	7.518 80	6.305 01
80	30	0	4.790 50	6.589 12	5.314 81
80	30	0.01	4.529 92	5.563 67	4.749 55
80	30	0.1	4.82768	5.44171	4.94848
80	30	1	5.441 19	6.019 45	5.330 29
80	30	10	6.551 03	6.703 06	5.807 79
80	30	100	7.507 56	7.590 85	6.401 70
80	35	0	3.993 17	5.274 19	3.811 21
80	35	0.01	4.166 27	5.014 27	4.186 11
80	35	0.1	4.29566	4.98269	4.42704
80	35	1	5.179 35	5.852 10	4.820 69
80	35	10	6.424 90	6.562 28	5.509 46
80	35	100	7.354 18	7.475 00	6.302 53
80	40	0	4.286 54	5.161 95	4.546 75
80	40	0.01	3.96875	4.80560	4.07899
80	40	0.1	4.314 02	5.064 04	4.439 76
80	40	1	4.808 23	5.548 80	4.428 92
80	40	10	6.157 46	6.323 72	5.217 49
80	40	100	7.391 10	7.505 59	6.308 61
80	45	0	3.692 79	5.050 54	4.402 99
80	45	0.01	4.06057	4.61357	3.90831
80	45	0.1	3.778 29	4.822 18	4.141 81
80	45	1	4.771 64	5.274 37	3.949 63
80	45	10	6.031 36	6.213 91	4.999 07

Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
80	45	100	7.200 76	7.281 94	6.107 75
90	30	0	3.941 69	6.145 66	4.813 93
90	30	0.01	4.37316	6.13613	5.00570
90	30	0.1	4.360 28	5.680 26	4.620 16
90	30	1	5.147 67	5.780 47	4.841 31
90	30	10	6.550 92	6.751 44	5.688 54
90	30	100	7.446 56	7.546 54	6.407 53
90	35	0	4.019 00	5.367 29	4.204 06
90	35	0.01	4.06997	5.04099	4.85758
90	35	0.1	4.176 59	5.192 02	4.935 16
90	35	1	4.886 31	5.532 47	4.488 77
90	35	10	6.258 47	6.437 67	5.486 50
90	35	100	7.350 40	7.502 49	6.294 98
90	40	0	4.118 35	5.410 62	4.360 57
90	40	0.01	4.167 99	4.936 26	4.369 52
90	40	0.1	4.17786	4.85045	3.76992
90	40	1	4.582 11	5.356 71	4.291 08
90	40	10	6.221 07	6.364 06	5.335 81
90	40	100	7.261 67	7.410 75	6.284 28
90	45	0	3.629 36	5.263 19	4.299 99
90	45	0.01	4.00408	4.92861	4.12769
90	45	0.1	4.158 19	5.007 33	3.903 64
90	45	1	4.608 48	5.221 12	4.268 25
90	45	10	6.003 92	6.120 66	5.076 66
90	45	100	7.178 24	7.285 04	6.138 11
100	30	0	4.029 23	5.997 53	4.539 48
100	30	0.01	3.969 63	6.088 81	4.668 52
100	30	0.1	3.93209	5.07671	4.91263
100	30	1	5.016 64	5.691 45	5.089 15
100	30	10	6.437 87	6.666 13	5.632 27
100	30	100	7.450 83	7.567 68	6.381 36
100	35	0	4.074 09	5.849 14	4.217 43
100	35	0.01	3.972 32	5.689 57	4.741 08
100	35	0.1	3.84848	5.06922	4.58001
100	35	1	4.778 13	5.361 12	4.572 75
100	35	10	6.262 18	6.385 83	5.540 69
100	35	100	7.301 44	7.365 32	6.238 57
100	40	0	3.672 63	4.945 74	4.373 85
100	40	0.01	3.62957	4.90307	4.01890
100	40	0.1	3.612 08	4.958 22	3.737 03
100	40	1	4.708 01	5.573 62	4.623 02
100	40	10	6.065 52	6.266 21	5.000 80
100	40	100	7.268 61	7.387 39	6.231 37
100	45	0	3.24214	4.61831	3.62015
100	45	0.01	3.776 04	4.761 22	4.05799
100	45	0.1	3.701 40	4.771 69	4.152 95
100	45	1	4.448 08	5.083 10	3.840 32
100	45	10	5.934 19	6.145 46	4.888 25
100	45	100	7.111 78	7.219 86	6.085 91

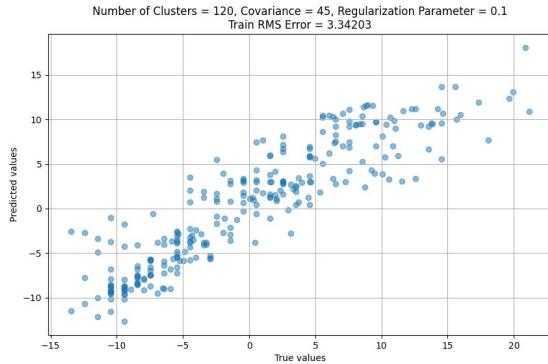
Clusters	Covariance	Regularization Parameter	Training Error	Validation Error	Testing Error
120	30	0	3.72769	5.17559	4.66918
120	30	0.01	3.80874	5.39389	4.50766
120	30	0.1	3.87249	5.11061	4.68030
120	30	1	4.81172	5.64988	4.96879
120	30	10	6.35181	6.61774	5.60494
120	30	100	7.37618	7.45249	6.30966
120	35	0	3.57649	4.73260	3.53759
120	35	0.01	3.34096	4.95357	3.94892
120	35	0.1	3.64543	5.07657	4.42674
120	35	1	4.63930	5.51431	4.70962
120	35	10	6.19591	6.39044	5.39481
120	35	100	7.25015	7.35105	6.22525
120	40	0	3.24904	4.72630	3.76178
120	40	0.01	3.48987	5.11746	4.40299
120	40	0.1	3.62117	4.80867	3.66593
120	40	1	4.33636	5.08381	4.08022
120	40	10	5.93880	6.16045	4.97875
120	40	100	7.14989	7.26303	6.10004
120	45	0	3.55332	4.58758	3.99585
120	45	0.01	3.24345	4.70685	3.81802
120	45	0.1	3.34203	4.57870	3.97437
120	45	1	4.50599	4.93747	4.03946
120	45	10	5.85053	6.02741	4.73581
120	45	100	6.98048	7.01748	6.00187

Table 7: Training, Validation and Testing Errors for different K, σ, λ for training dataset 2 of size 274. The best model is taken to be the one with the lowest validation error and are shown in boldface.

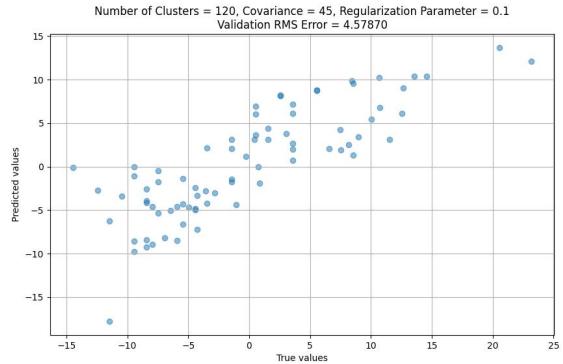
4.2 Inferences

1. Higher λ values lead to increased training errors. Validation and testing errors similarly rise with higher λ .
2. Increasing regularization leads to higher errors ($\lambda = 100$, generally has large errors for all the datasets).
3. Changes in covariance can affect errors. We can roughly observe that increasing covariance reduces the errors.
4. Higher cluster counts tend to reduce training errors; e.g., higher counts like 70 or 80 clusters exhibit lower training errors. However, higher cluster counts, as seen in the 120 clusters, 40 covariance, and regularization parameter 0 configuration, might lead to overfitting.
5. The model has roughly similar performance on the training, validation and testing data when optimal hyperparameters are chosen.
6. Model with 120 clusters, 45 covariance, and regularization parameter 0.1 performs optimally with both training and validation datasets.
7. The dataset exhibits sensitivity to regularization, emphasizing the importance of controlling model complexity to avoid overfitting.
8. Higher cluster counts (e.g., 120 clusters) contribute to improved model performance, suggesting that the underlying patterns in the dataset benefit from a more intricate representation.
9. The dataset benefits from a higher covariance (45), indicating that a broader spread of data is essential for capturing the nuanced relationships within the dataset.

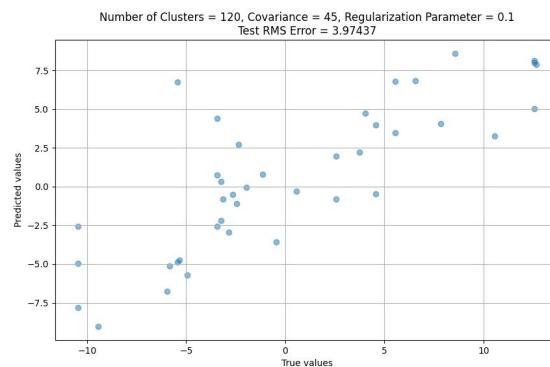
4.3 Scatter Plots of the Best Model



(1) Best model ($K = 120, \sigma = 45, \lambda = 0.1$) performance in Training Data



(2) Best model ($K = 120, \sigma = 45, \lambda = 0.1$) performance in Validation Data



(3) Best model ($K = 120, \sigma = 45, \lambda = 0.1$) performance in Testing Data

Figure 13: Best model performance with Training, Validation and Testing Data

5 Task 4: Linear Model for Regression using Polynomial Basis Functions for Dataset 4

5.1 Description of Dataset

We have chosen the **California Real Estate Dataset**. The input features pertain to several features of the a typical house in a Californian district, which are listed below and an output feature that is the median housing value in that Californian district. The entire dataset has a total of 17,000 data points. The random sample of 500 data points were chosen for the purpose of this assignment. Out of this, 60% of the data was taken for the training split, 20% for the validation split and 20% for the testing split.

Input Features:

1. Latitude
2. Longitude
3. Median Housing Age
4. Total Rooms
5. Total Bedrooms
6. Population
7. Number of households
8. Median Income

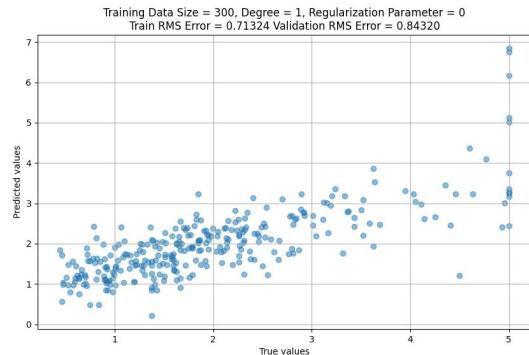
Output Feature:

1. Median House Value

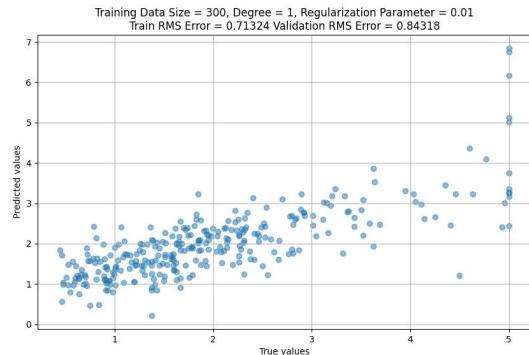
5.2 Procedure

- We performed a linear regression using polynomial basis functions for the dataset.
- The input features have varying scales and hence each feature was divided by a factor to make all the data similar in scale.
- Several different combinations of the degree of the polynomial, $m = 1, 2, 3, 4, 5, 8$, and the regularization parameter, $\lambda = 0, 0.01, 0.1, 1, 10, 100$ were tested.
- The model was trained on the training data and tested using validation data.
- The best model was chosen using the Cross-Validation method and the final model was tested on the testing data.
- All findings are detailed below.

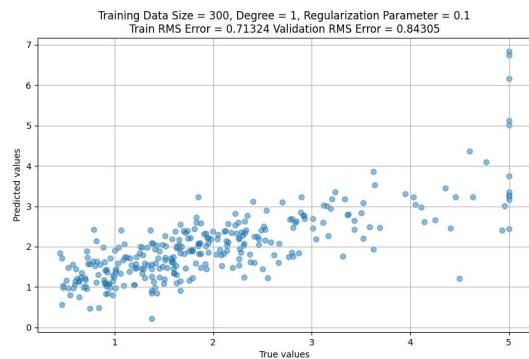
5.3 Scatter Plots of Approximated Functions



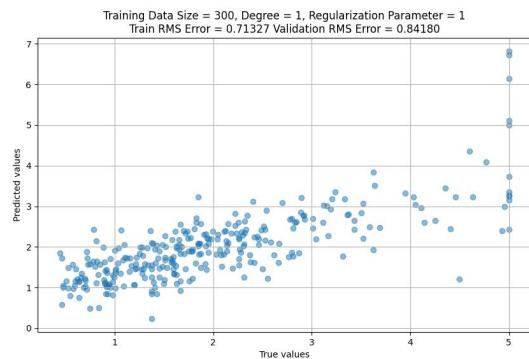
(1) $m = 1, \lambda = 0$



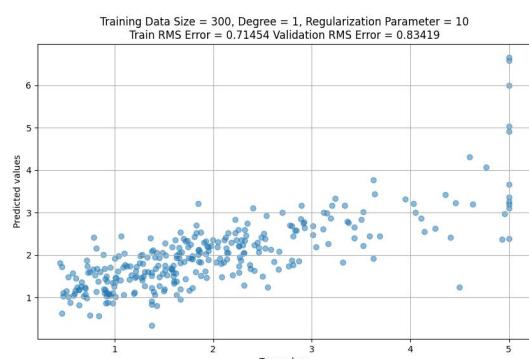
(2) $m = 1, \lambda = 0.01$



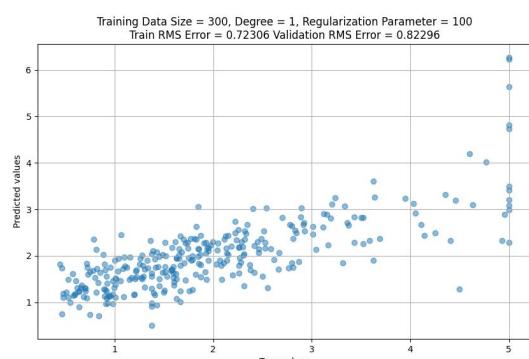
(3) $m = 1, \lambda = 0.1$



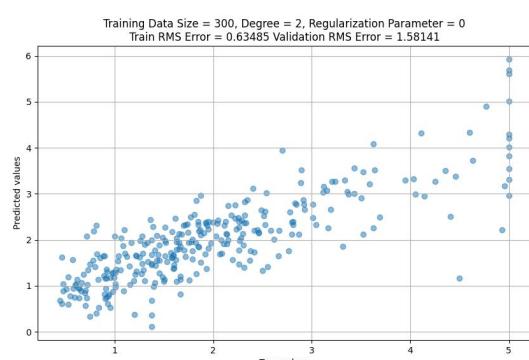
(4) $m = 1, \lambda = 1$



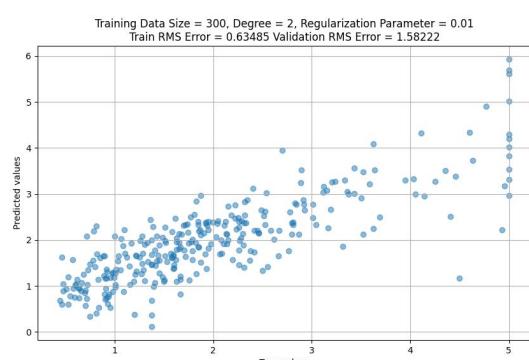
(5) $m = 1, \lambda = 10$



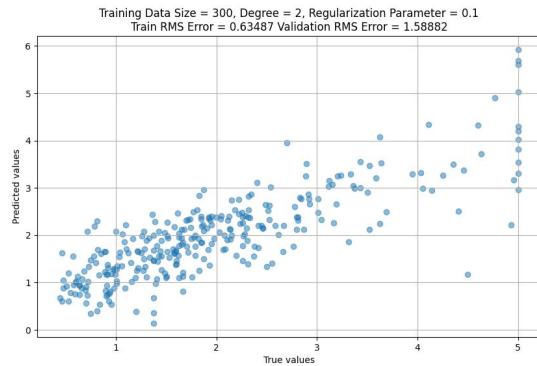
(6) $m = 1, \lambda = 100$



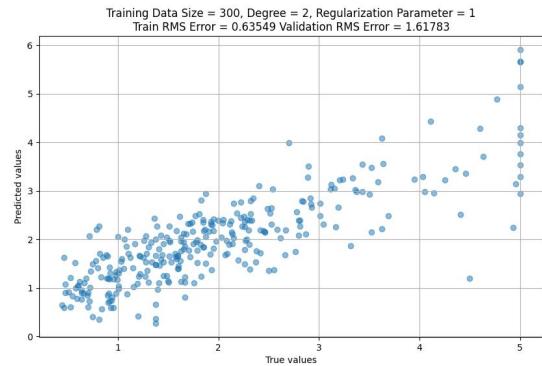
(7) $m = 2, \lambda = 0$



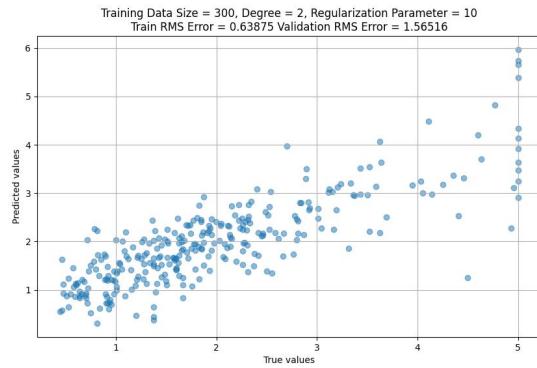
(8) $m = 2, \lambda = 0.01$



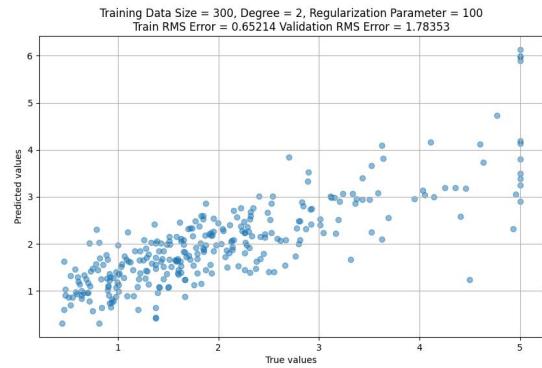
(9) $m = 2, \lambda = 0.1$



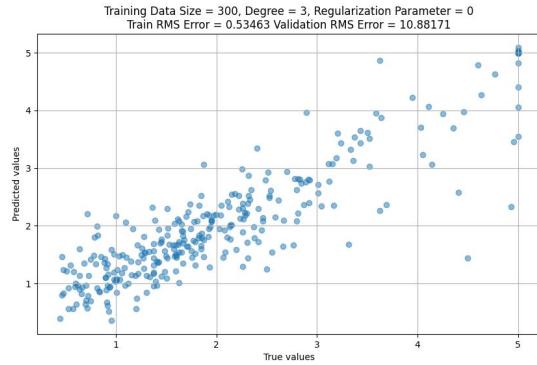
(10) $m = 2, \lambda = 1$



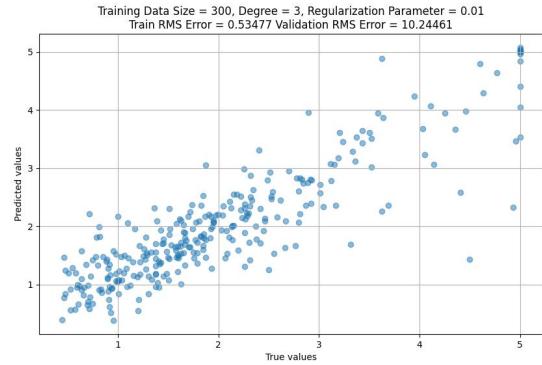
(11) $m = 2, \lambda = 10$



(12) $m = 2, \lambda = 100$



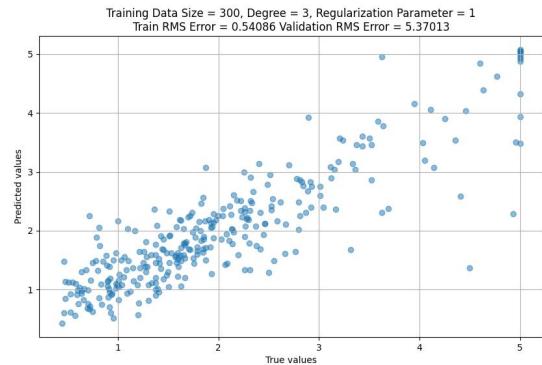
(13) $m = 3, \lambda = 0$



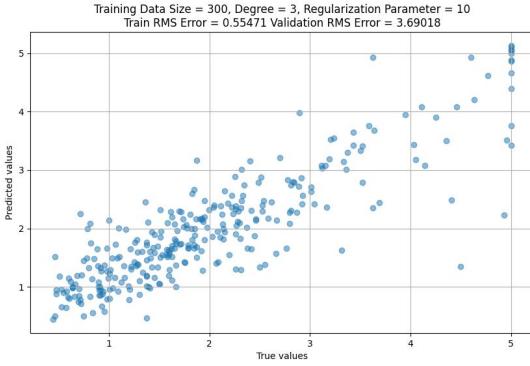
(14) $m = 3, \lambda = 0.01$



(15) $m = 3, \lambda = 0.1$



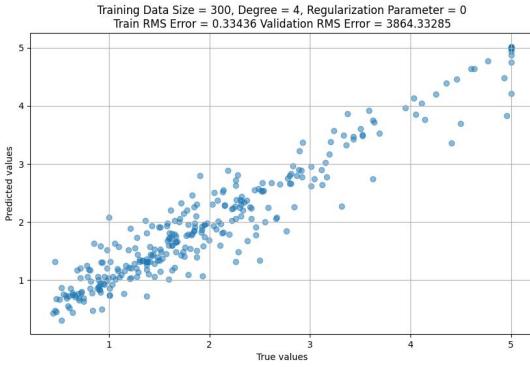
(16) $m = 3, \lambda = 1$



(17) $m = 3, \lambda = 10$



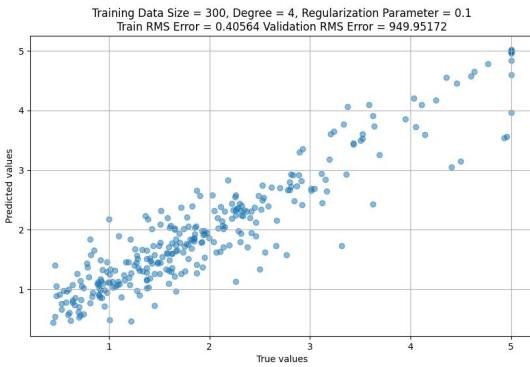
(18) $m = 3, \lambda = 100$



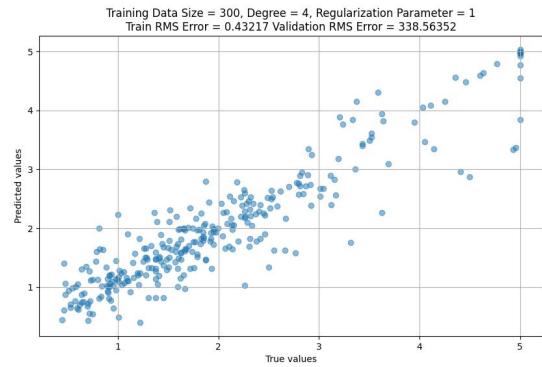
(19) $m = 4, \lambda = 0$



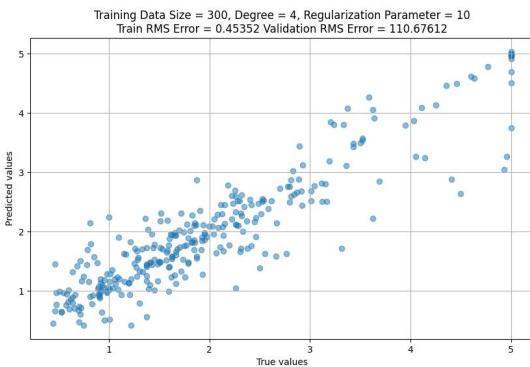
(20) $m = 4, \lambda = 0.01$



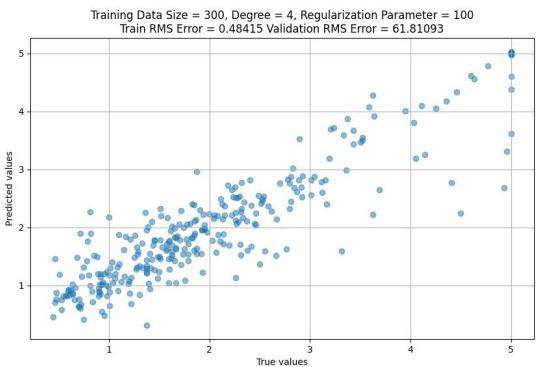
(21) $m = 4, \lambda = 0.1$



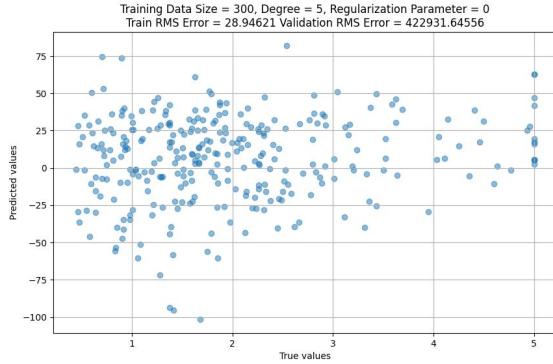
(22) $m = 4, \lambda = 1$



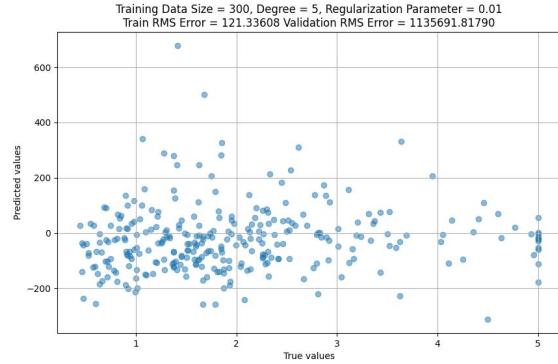
(23) $m = 4, \lambda = 10$



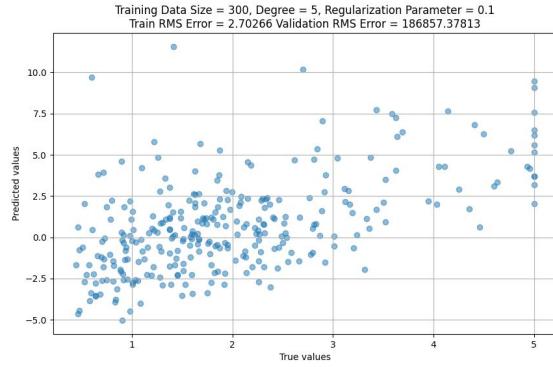
(24) $m = 4, \lambda = 100$



(25) $m = 5, \lambda = 0$



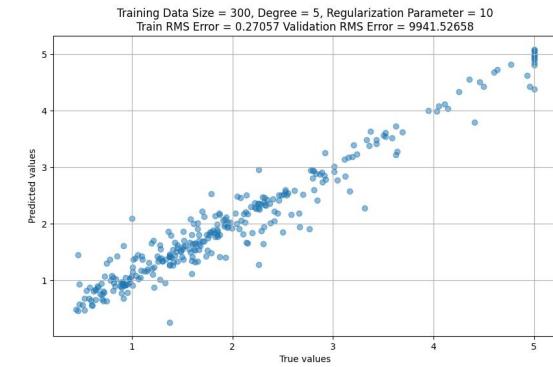
(26) $m = 5, \lambda = 0.01$



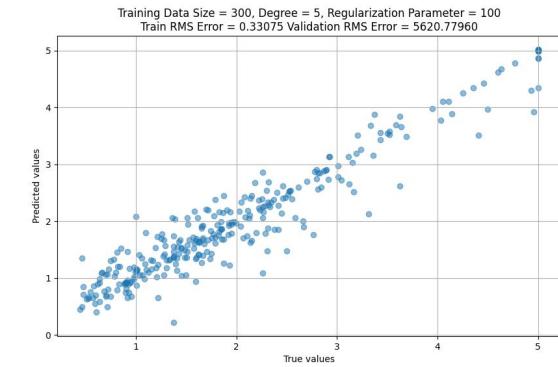
(27) $m = 5, \lambda = 0.1$



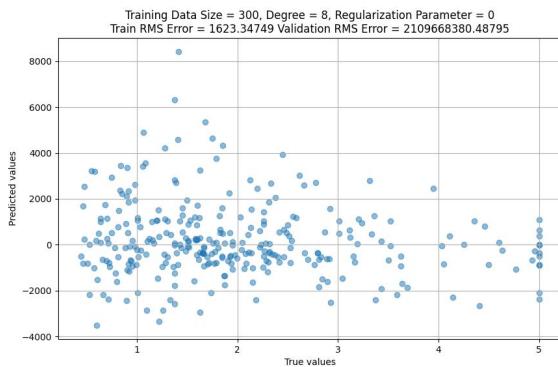
(28) $m = 5, \lambda = 1$



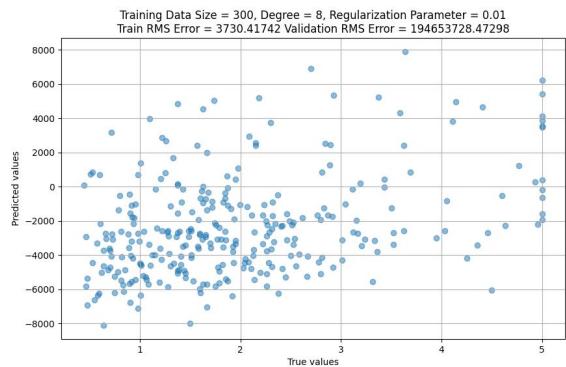
(29) $m = 5, \lambda = 10$



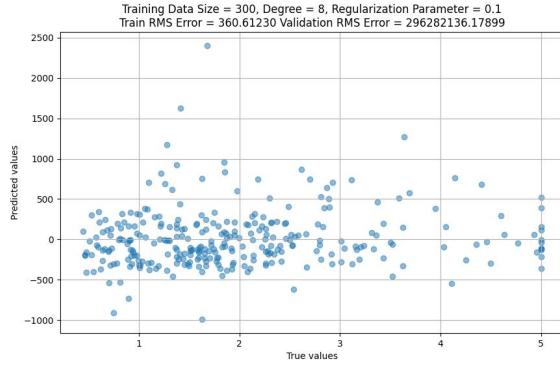
(30) $m = 5, \lambda = 100$



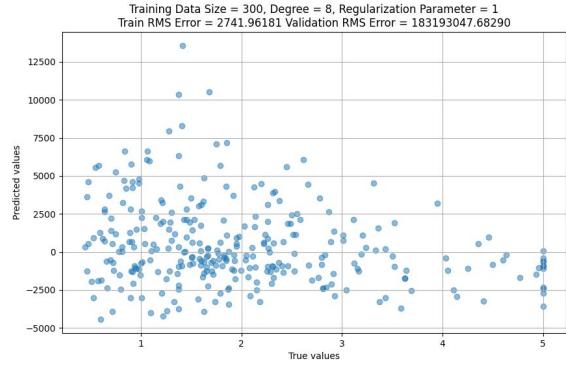
(31) $m = 8, \lambda = 0$



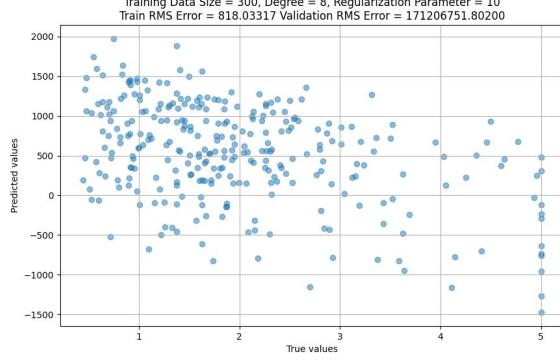
(32) $m = 8, \lambda = 0.01$



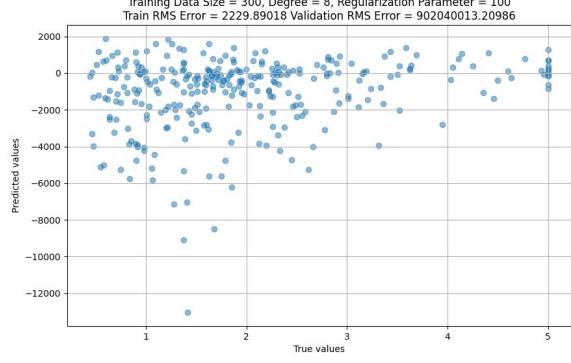
(33) $m = 8, \lambda = 0.1$



(34) $m = 8, \lambda = 1$



(35) $m = 8, \lambda = 10$



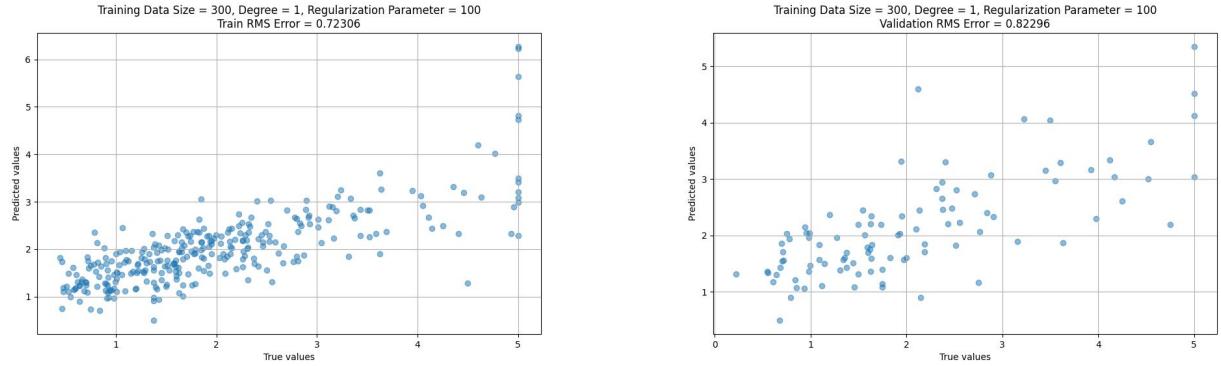
(36) $m = 8, \lambda = 100$

Figure 14: Scatter Plots of Approximated Functions for different m, λ

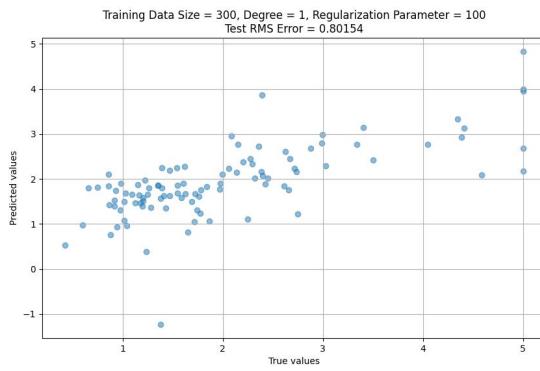
5.4 Inferences

- For $m = 1$, the model fits well with the training data and also performs well with the validation data.
- With regularization, the model performance on the validation data further increases with minimal drop in performance in the training data.
- As the degree increases, the model begins to overfit on the training data. This can be seen with the decrease in the training error and increase in the validation error.
- At degree $m = 4$, with no regularization the model has the lowest training error, but the validation error is huge.
- The validation error improves with the usage of regularization but the model, still, is highly overfitted on the training data.
- For degrees, $m = 5, 8$ the model begins to fall apart and both the training and validation errors shoot up. This is due to the insufficient number of samples for the training of the model. For $m = 5$, we require atleast 462 samples and for $m = 8$, we require atleast 3003 samples for a good approximation. These values are given by equation (1).
- The best model is identified to be one with parameters of $m = 1, \lambda = 100$.

5.5 Scatter Plots of the Best Model



(1) Best model ($m = 1, \lambda = 100$) performance in Training Data (2) Best model ($m = 1, \lambda = 100$) performance in Validation Data



(1) Best model ($m = 1, \lambda = 100$) performance in Testing Data

Figure 15: Best model performance with Training, Validation and Testing Data

5.6 Inferences

1. The model has similar errors for training, validation and testing data. This shows that the model has not overfitted with the training data and is generalizable.
2. The usage of regularization has improved the performance of the model marginally. The training error has increased by 0.00982 while the validation error has reduced by 0.02024.
3. So, the usage of regularization is not necessary in this situation.
4. High errors on validation and testing datasets when higher degrees are used, even with regularization, suggest a substantial overfitting problem. The model struggles to generalize to unseen data when larger degree polynomials are chosen $m = 4, 5, 8$.
5. The observed increase in errors with higher polynomial degrees indicates that the dataset might not have complex relationships that require a polynomial of higher degree. Overfitting becomes more pronounced as the model tries to fit noise in the data.
6. The fact that the best performance is observed with a high regularization parameter (100) suggests that the dataset is particularly sensitive to overfitting. Strong regularization is required to control model complexity and prevent fitting noise.
7. The observed trend of increasing errors with higher degrees and the necessity for strong regularization implies that the dataset may have inherent noise or is inherently less complex.

5.7 Hyperparameters v/s Training, Validation and Testing Errors

Degree	Regularization Parameter	Training Error	Validation Error	Testing Error
1	0.0	0.713 24	0.843 20	0.815 70
1	0.01	0.713 24	0.843 18	0.815 70
1	0.1	0.713 24	0.843 05	0.815 70
1	1.0	0.713 27	0.841 80	0.815 73
1	10.0	0.714 54	0.834 19	0.815 56
1	100.0	0.72306	0.82296	0.80154
2	0.0	0.634 85	1.581 41	1.074 04
2	0.01	0.634 85	1.582 22	1.073 27
2	0.1	0.634 87	1.588 82	1.066 79
2	1.0	0.635 49	1.617 83	1.027 31
2	10.0	0.63875	1.56516	0.93007
2	100.0	0.652 14	1.783 53	0.933 89
3	0.0	0.534 63	10.881 71	8.562 20
3	0.01	0.534 77	10.244 61	7.198 08
3	0.1	0.536 41	8.241 80	3.448 51
3	1.0	0.540 86	5.370 13	1.172 56
3	10.0	0.55471	3.69018	1.05909
3	100.0	0.571 22	4.633 24	1.239 05
4	0.0	0.334 36	3864.33285	3628.20562
4	0.01	0.378 02	1664.17750	2366.81740
4	0.1	0.405 64	949.95172	1054.38088
4	1.0	0.432 17	338.56352	164.14853
4	10.0	0.453 52	110.67612	313.24751
4	100.0	0.48415	61.81093	347.27459
5	0.0	28.946 21	422 931.645 56	268 060.953 15
5	0.01	121.336 08	1 135 691.817 90	856 210.757 88
5	0.1	2.702 66	186 857.378 13	148 532.377 73
5	1.0	0.621 66	44 934.855 87	60 520.811 26
5	10.0	0.270 57	9941.526 58	16 778.371 84
5	100.0	0.33075	5620.77960	5049.82310
8	0.0	1623.347 49	2 109 668 380.487 95	2 269 890 415.068 18
8	0.01	3730.417 42	194 653 728.472 98	15 490 863.695 70
8	0.1	360.612 30	296 282 136.178 99	29 393 537.967 09
8	1.0	2741.961 81	183 193 047.682 90	58 460 107.929 91
8	10.0	818.03317	171206751.80200	44590205.69076
8	100.0	2229.890 18	902 040 013.209 86	49 619 018.591 18

Table 8: Training, Validation and Testing Errors for different m, λ for training dataset of size 200. The best model is taken to be the one with the lowest validation error and are shown in boldface.