

CS5691: Pattern Recognition and Machine Learning
Programming Assignment 1
Project Report

Reetwik Das - CS20D402
Shubhanshu Sharma - EE20D413
Vamsi Sai Krishna Malineni - OE20S302

March 18, 2021

Contents

1	Polynomial curve fitting for dataset	4
1.1	Without Regularization	4
1.1.1	Datapoints in training dataset - 10	4
1.1.2	Datapoints in training dataset - 200	5
1.2	With Regularization	6
1.2.1	Datapoints in training dataset - 10, $\lambda = 1$	6
1.2.2	Datapoints in training dataset - 10, $\lambda = e^{-18}$	7
1.2.3	Datapoints in training dataset - 200, $\lambda = 1$	8
1.2.4	Datapoints in training dataset - 200, $\lambda = e^{-18}$	9
2	Linear model for regression using Polynomial Basis function	10
2.1	Without regularization	10
2.1.1	Datapoints in training dataset - 50	10
2.1.2	Datapoints in training dataset - 200	11
2.1.3	Datapoints in training dataset - 500	12
2.1.4	Scatter plots, training size = 50	13
2.1.5	Scatter plots, training size = 200	13
2.1.6	Scatter plots, training size = 500	13
2.2	With regularization	14
2.2.1	Datapoints in training dataset - 50, $\lambda = e^{-18}$	14
2.2.2	Datapoints in training dataset - 200, $\lambda = e^{-18}$	15
2.2.3	Datapoints in training dataset - 500, $\lambda = e^{-18}$	16
2.2.4	Scatter plots, training size = 50, $\lambda = e^{-18}$	17
2.2.5	Scatter plots, training size = 200, $\lambda = e^{-18}$	17
2.2.6	Scatter plots, training size = 500, $\lambda = e^{-18}$	17
3	Linear model for regression using Gaussian Basis function	18
3.1	Dataset 2 : function2-2d.csv	18
3.1.1	Without regularization	18
3.1.2	With Quadratic regularization	19
3.1.3	With Tikhonov regularization	20
3.2	Dataset 3 : 2-music.txt	21
3.2.1	Without regularization	21
3.2.2	With Quadratic regularization	22
3.2.3	With Tikhonov regularization	23

List of Tables

1.1.1	ERMS for dataset1, training size = 10, $\lambda = 0$	4
1.1.2	ERMS for dataset1, training size = 200, $\lambda = 0$	5
1.2.1	ERMS for dataset1, training size = 10, $\lambda = 1$	6
1.2.2	ERMS for dataset1, training size = 10, $\lambda = e^{-18}$	7
1.2.3	ERMS for dataset1, training size = 200, $\lambda = 1$	8
1.2.4	ERMS for dataset1, training size = 200, $\lambda = e^{-18}$	9

2.1.1 ERMS for dataset2, training size = 50, $\lambda = 0$	10
2.1.2 ERMS for dataset2, training size = 200, $\lambda = 0$	11
2.1.3 ERMS for dataset2, training size = 500, $\lambda = 0$	12
2.2.1 ERMS for dataset2, training size = 50, $\lambda = e^{-18}$	14
2.2.2 ERMS for dataset2, training size = 200, $\lambda = e^{-18}$	15
2.2.3 ERMS for dataset2, training size = 500, $\lambda = e^{-18}$	16
3.1.1 ERMS for dataset2, $\lambda = 0$, $\sigma = 20$	18
3.1.2 ERMS for dataset2, $\lambda = e^{-18}$, $\sigma = 20$	19
3.1.3 ERMS for dataset2, $\lambda = e^{-18}$, $\sigma = 20$	20
3.2.1 ERMS for dataset3, $\lambda = 0$	21
3.2.2 ERMS for dataset3, $\lambda = e^{-18}$	22
3.2.3 ERMS for dataset3, $\lambda = e^{-18}$ using Tikhonov regularization	23

List of Figures

1.1.1 Approximated function vs Training points for dataset1, training size = 10, $\lambda = 0$	4
1.1.2 Approximated function vs Training points for dataset1, training size = 100, $\lambda = 0$	5
1.2.1 Approximated function vs Training points for dataset1, training size = 10, $\lambda = 1$	6
1.2.2 Approximated function vs Training points for dataset1, training size = 10, $\lambda = e^{-18}$	7
1.2.3 Approximated function vs Training points for dataset1, training size = 200, $\lambda = 1$	8
1.2.4 Approximated function vs Training points for dataset1, training size = 200, $\lambda = e^{-18}$	9
2.1.1 Approximated surface vs Training points for dataset2, training size = 50, $\lambda = 0$	10
2.1.2 Approximated surface vs Training points for dataset2, training size = 200, $\lambda = 0$	11
2.1.3 Approximated surface vs Training points for dataset2, training size = 500, $\lambda = 0$	12
2.1.4 Model output vs Target output for dataset2, training size = 50, $\lambda = 0$, $M = 2$	13
2.1.5 Model output vs Target output for dataset2, training size = 200, $\lambda = 0$, $M = 2$	13
2.1.6 Model output vs Target output for dataset2, training size = 500, $\lambda = 0$, $M = 3$	13
2.2.1 Approximated surface vs Training points for dataset2, training size = 50, $\lambda = e^{-18}$	14
2.2.2 Approximated surface vs Training points for dataset2, training size = 200, $\lambda = e^{-18}$	15
2.2.3 Approximated surface vs Training points for dataset2, training size = 500, $\lambda = e^{-18}$	16
2.2.4 Model output vs Target output for dataset2, training size = 50, $\lambda = e^{-18}$, $M = 2$	17
2.2.5 Model output vs Target output for dataset2, training size = 200, $\lambda = e^{-18}$, $M = 3$	17
2.2.6 Model output vs Target output for dataset2, training size = 500, $\lambda = e^{-18}$, $M = 3$	17
3.1.1 Model output vs Target output for dataset2, $\lambda = 0$, $\sigma = 20$, $K = 26$	18
3.1.2 Model output vs Target output for dataset2, $\lambda = e^{-18}$, $\sigma = 20$, $K = 44$	19
3.1.3 Model output vs Target output for dataset2, $\lambda = e^{-18}$, $\sigma = 20$, $K = 26$	20
3.2.1 Model output vs Target output for dataset3, $\lambda = 0$	21
3.2.2 Model output vs Target output for dataset3, $\lambda = e^{-18}$	22
3.2.3 Model output vs Target output for dataset3, $\lambda = e^{-18}$ using Tikhonov regularization	23

1 Polynomial curve fitting for dataset

Dataset1 : function2.csv is the dataset used in this section for all the experiments.

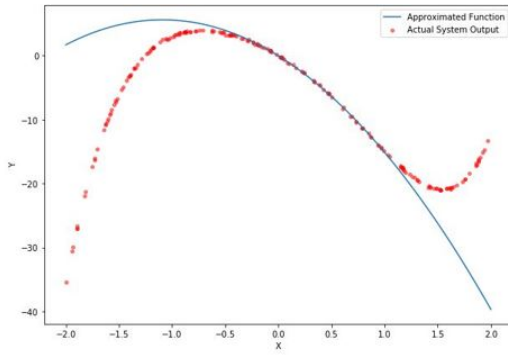
1.1 Without Regularization

1.1.1 Datapoints in training dataset - 10

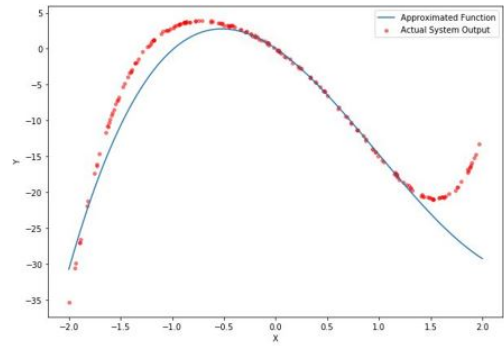
Hyperparameters $M = \{2, 3, 6, 9\}, \lambda = 0$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	0.1123	9.7834	10.5829
3	0.0583	3.9136	4.2917
6	0.0050	1.1671	1.0449
9	$8.81 * 10^{-10}$	0.2006	0.2095

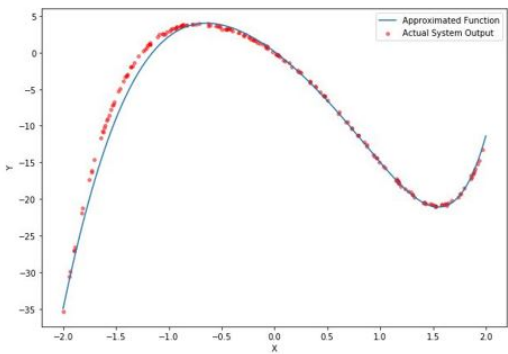
Table 1.1.1: ERMS for dataset1, training size = 10, $\lambda = 0$



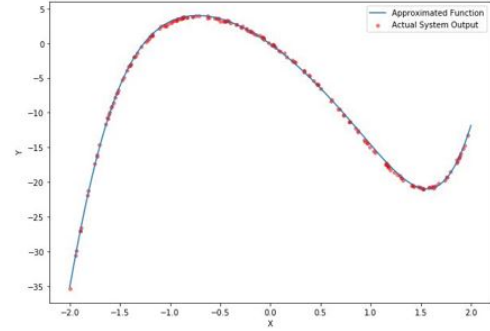
(a) M=2



(b) M=3



(c) M=6



(d) M=9

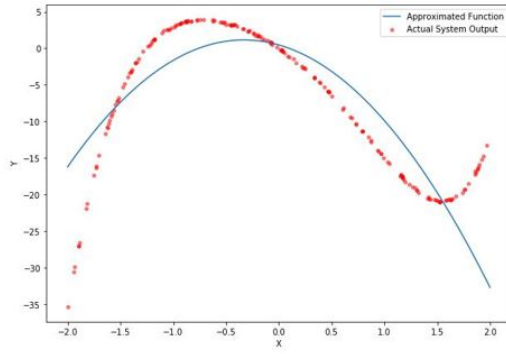
Figure 1.1.1: Approximated function vs Training points for dataset1, training size =10, $\lambda = 0$

1.1.2 Datapoints in training dataset - 200

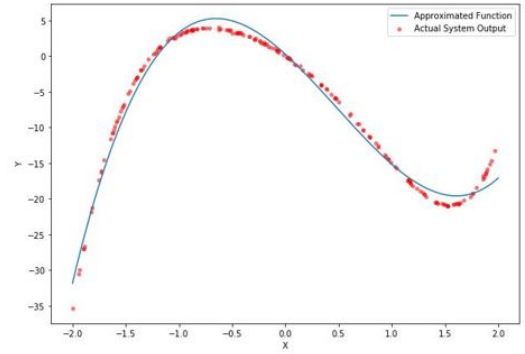
Hyperparameters $M = \{2, 3, 6, 9\}$, $\lambda = 0$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	4.75051	5.4418	6.1629
3	1.0755	1.2381	1.3941
6	0.0884	0.1029	0.0958
9	0.0883	0.1029	0.0959

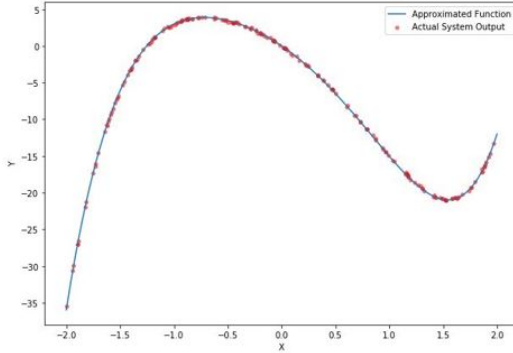
Table 1.1.2: ERMS for dataset1, training size = 200, $\lambda = 0$



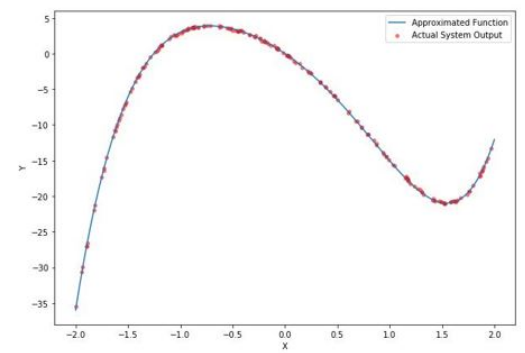
(a) M=2



(b) M=3



(c) M=6



(d) M=9

Figure 1.1.2: Approximated function vs Training points for dataset1, training size = 100, $\lambda = 0$

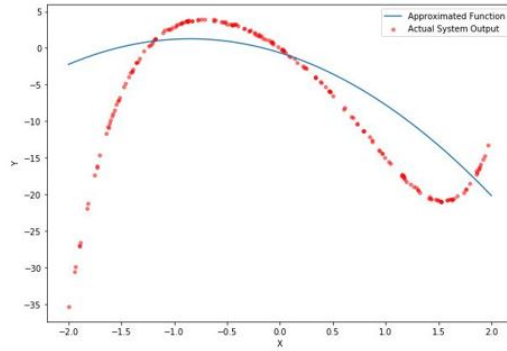
1.2 With Regularization

1.2.1 Datapoints in training dataset - 10, $\lambda = 1$

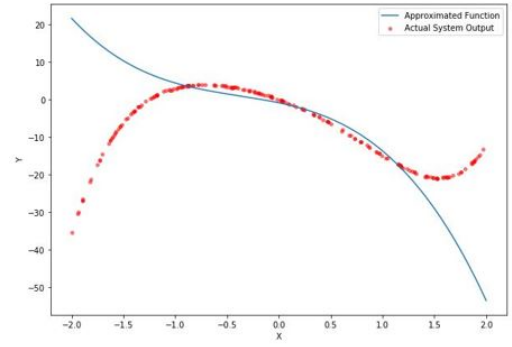
Hyperparameters $M = \{2, 3, 6, 9\}, \lambda = 1$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	1.1111	7.3682	7.9686
3	1.3700	15.0276	16.2461
6	0.4904	12.9961	15.4821
9	0.3985	26.4875	32.9489

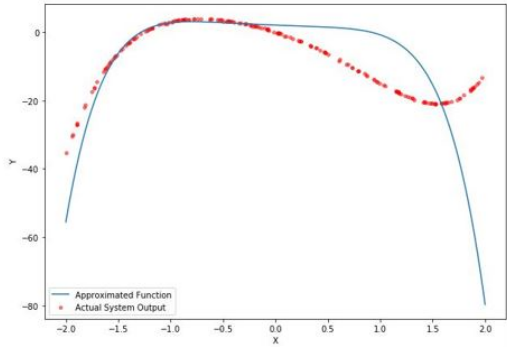
Table 1.2.1: ERMS for dataset1, training size =10, $\lambda = 1$



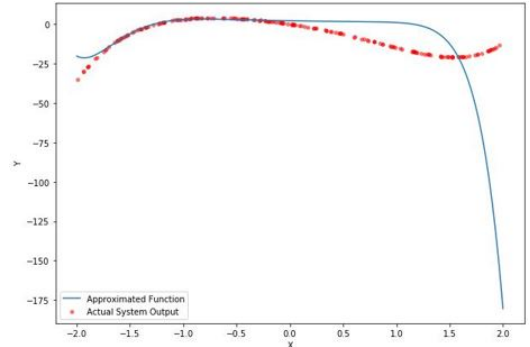
(a) M=2



(b) M=3



(c) M=6



(d) M=9

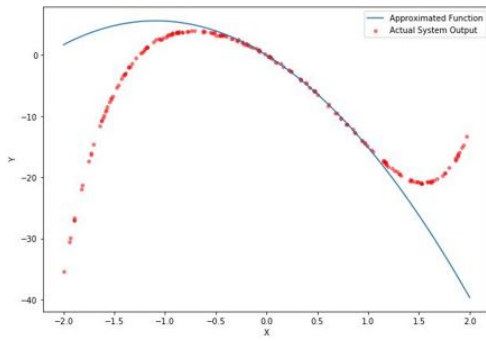
Figure 1.2.1: Approximated function vs Training points for dataset1, training size =10, $\lambda = 1$

1.2.2 Datapoints in training dataset - 10, $\lambda = e^{-18}$

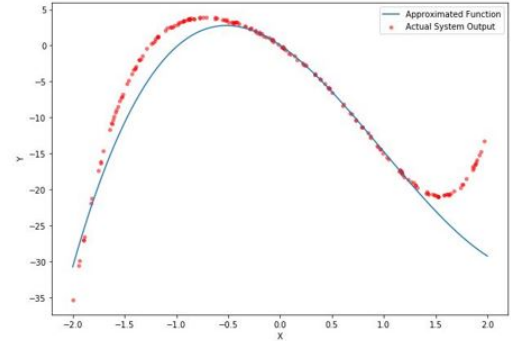
Hyperparameters $M = \{2, 3, 6, 9\}$, $\lambda = 10^{-18}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	0.1123	9.7834	10.5829
3	0.0582	3.9136	4.2917
6	0.0050	1.1671	1.0444
9	$3.61 * 10^{-7}$	4.5676	5.3469

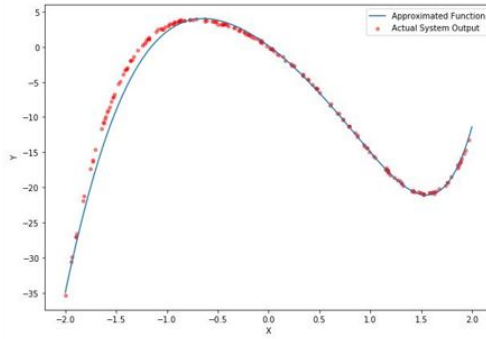
Table 1.2.2: ERMS for dataset1, training size =10, $\lambda = e^{-18}$



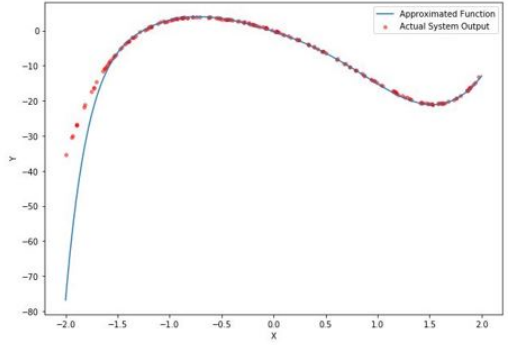
(a) M=2



(b) M=3



(c) M=6



(d) M=9

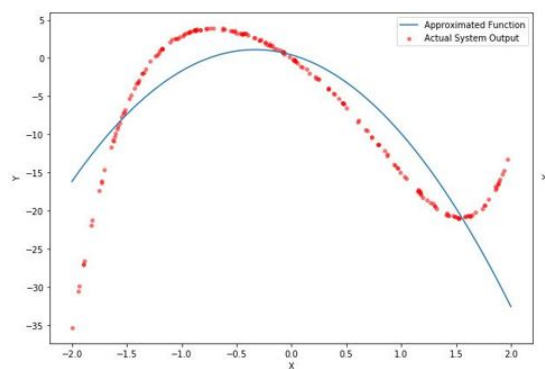
Figure 1.2.2: Approximated function vs Training points for dataset1, training size = 10, $\lambda = e^{-18}$

1.2.3 Datapoints in training dataset - 200, $\lambda = 1$

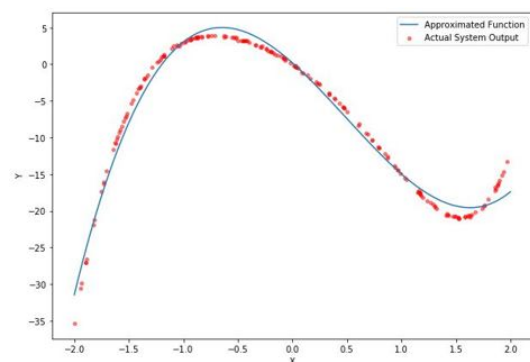
Hyperparameters $M = \{2, 3, 6, 9\}, \lambda = 1$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	4.7052	5.4348	6.1529
3	1.0877	1.2572	1.4274
6	0.2360	0.2285	0.2402
9	0.2390	0.2247	0.2365

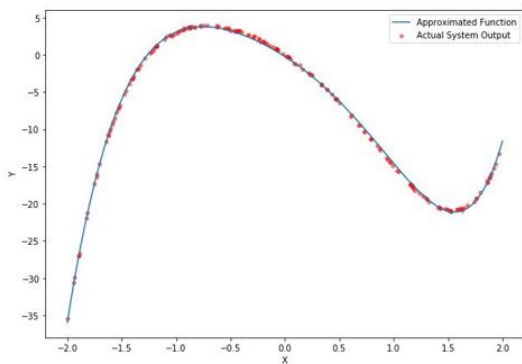
Table 1.2.3: ERMS for dataset1, training size = 200, $\lambda = 1$



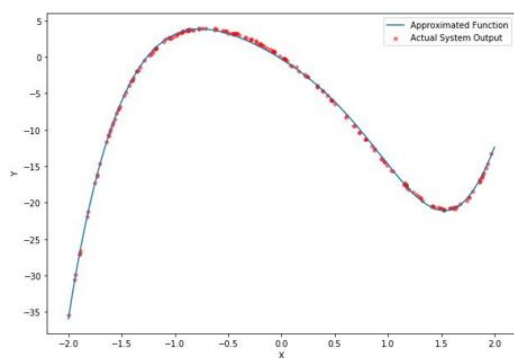
(a) M=2



(b) M=3



(c) M=6



(d) M=9

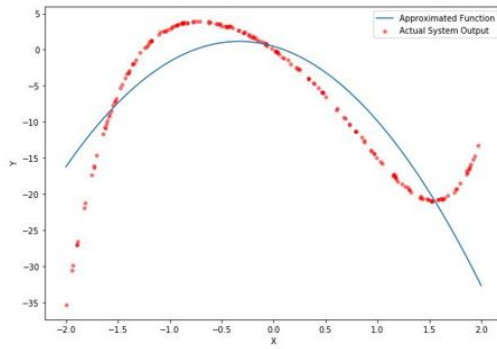
Figure 1.2.3: Approximated function vs Training points for dataset1, training size =200, $\lambda = 1$

1.2.4 Datapoints in training dataset - 200, $\lambda = e^{-18}$

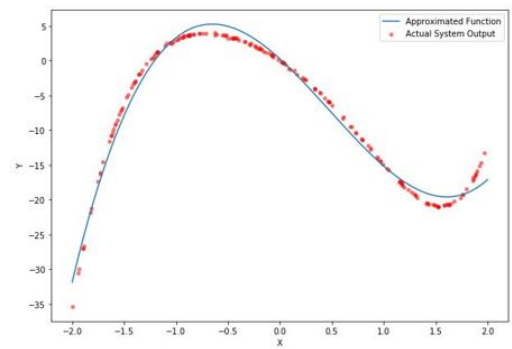
Hyperparameters $M = \{2, 3, 6, 9\}$, $\lambda = e^{-18}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	4.7051	5.4418	6.1629
3	1.0755	1.2381	1.3941
6	0.0884	0.1029	0.0958
9	0.0883	0.1029	0.0959

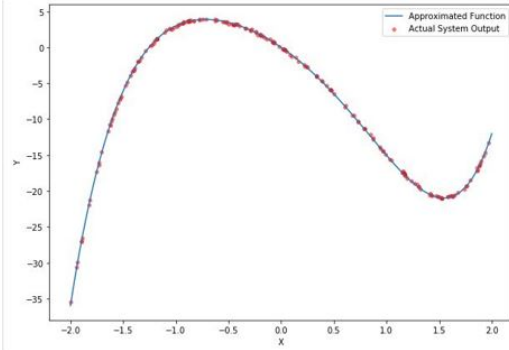
Table 1.2.4: ERMS for dataset1, training size =200, $\lambda = e^{-18}$



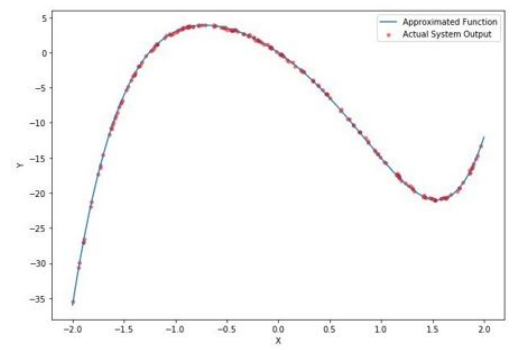
(a) M=2



(b) M=3



(c) M=6



(d) M=9

Figure 1.2.4: Approximated function vs Training points for dataset1, training size = 200, $\lambda = e^{-18}$

Observations :

We observed in the section that when we don't have enough data for training i.e when we had only 10 points for training and the model complexity is high, the ERMS for training data is pretty low although the ERMS for test and validation is higher. This is because of overfitting and when we have more data i.e 200 points for training and the model complexity is high this problem is resolved.

2 Linear model for regression using Polynomial Basis function

Dataset2 : function2-2d.csv is used for all the experiments in the given section.

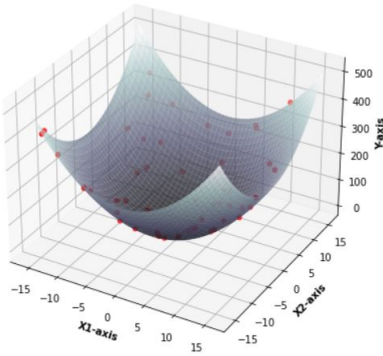
2.1 Without regularization

2.1.1 Datapoints in training dataset - 50

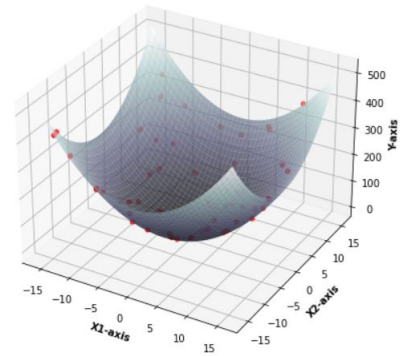
Hyperparameters $M = \{2, 3, 6\}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	$4.562 * 10^{-14}$	$5.6838 * 10^{-14}$	$5.7531 * 10^{-14}$
3	$1.5679 * 10^{-13}$	$2.1288 * 10^{-13}$	$2.2070 * 10^{-13}$
6	$7.3818 * 10^{-11}$	$3.5131 * 10^{-10}$	$3.5751 * 10^{-10}$

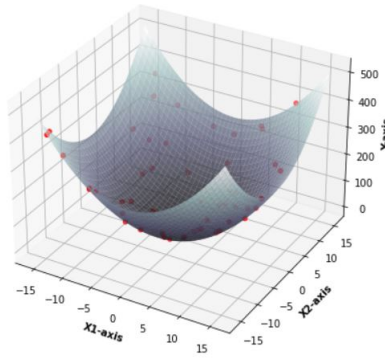
Table 2.1.1: ERMS for dataset2, training size = 50, $\lambda = 0$



(a) M=2



(b) M=3



(c) M=6

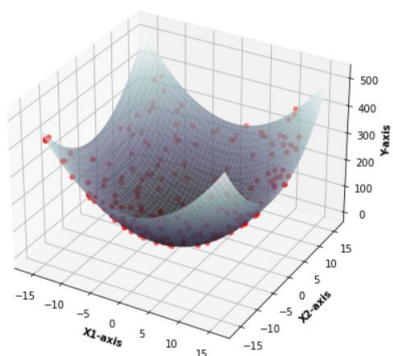
Figure 2.1.1: Approximated surface vs Training points for dataset2, training size = 50, $\lambda = 0$

2.1.2 Datapoints in training dataset - 200

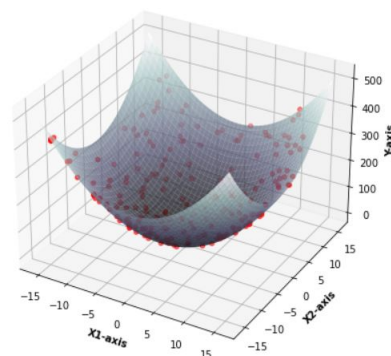
Hyperparameters $M = \{2, 3, 6\}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	$6.0827 * 10^{-14}$	$5.9075 * 10^{-14}$	$6.0578 * 10^{-14}$
3	$8.4333 * 10^{-14}$	$9.3186 * 10^{-14}$	$9.9068 * 10^{-14}$
6	$9.2418 * 10^{-11}$	$9.7743 * 10^{-11}$	$9.6217 * 10^{-11}$

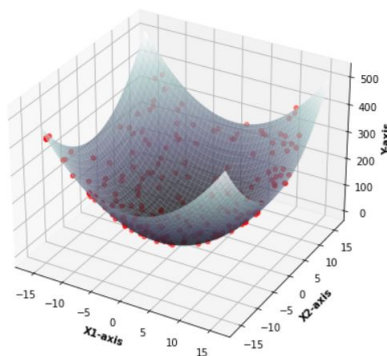
Table 2.1.2: ERMS for dataset2, training size = 200, $\lambda = 0$



(a) M=2



(b) M=3



(c) M=6

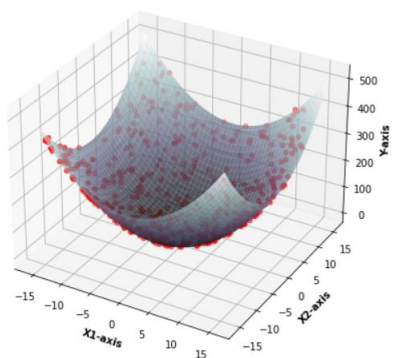
Figure 2.1.2: Approximated surface vs Training points for dataset2, training size = 200, $\lambda = 0$

2.1.3 Datapoints in training dataset - 500

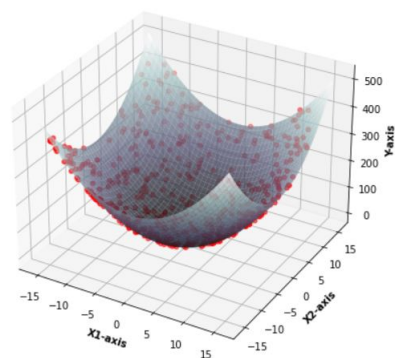
Hyperparameters $M = \{2, 3, 6\}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	$9.2728 * 10^{-14}$	$8.8556 * 10^{-14}$	$9.3566 * 10^{-14}$
3	$5.5874 * 10^{-14}$	$5.3009 * 10^{-14}$	$5.2778 * 10^{-14}$
6	$9.0462 * 10^{-11}$	$9.5982 * 10^{-11}$	$9.1015 * 10^{-11}$

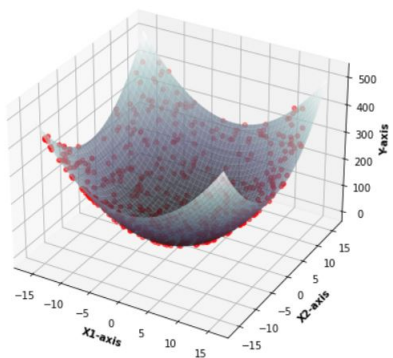
Table 2.1.3: ERMS for dataset2, training size = 500, $\lambda = 0$



(a) M=2



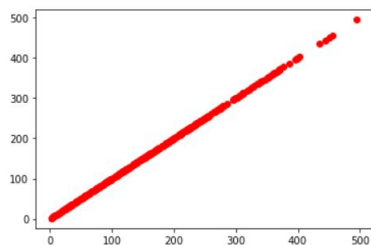
(b) M=3



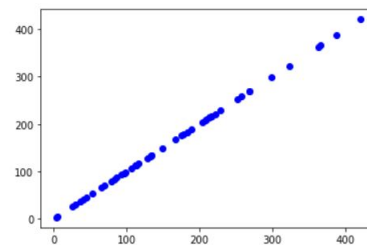
(c) M=6

Figure 2.1.3: Approximated surface vs Training points for dataset2, training size = 500, $\lambda = 0$

2.1.4 Scatter plots, training size = 50



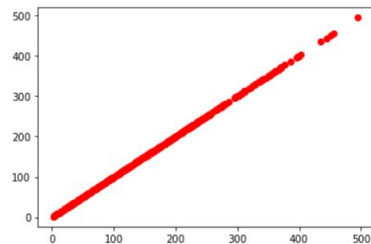
(a) Testing Data



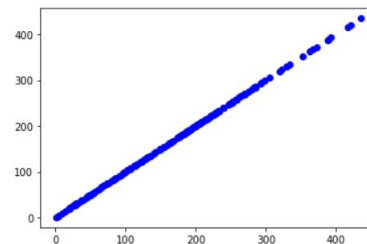
(b) Training Data

Figure 2.1.4: Model output vs Target output for dataset2, training size = 50, $\lambda = 0$, $M = 2$

2.1.5 Scatter plots, training size = 200



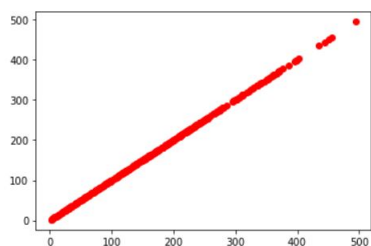
(a) Testing Data



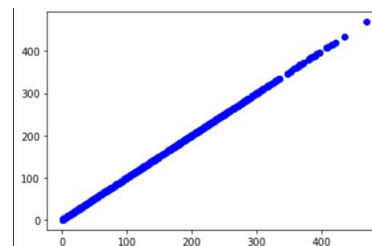
(b) Training Data

Figure 2.1.5: Model output vs Target output for dataset2, training size = 200, $\lambda = 0$, $M = 2$

2.1.6 Scatter plots, training size = 500



(a) Testing Data



(b) Training Data

Figure 2.1.6: Model output vs Target output for dataset2, training size = 500, $\lambda = 0$, $M = 3$

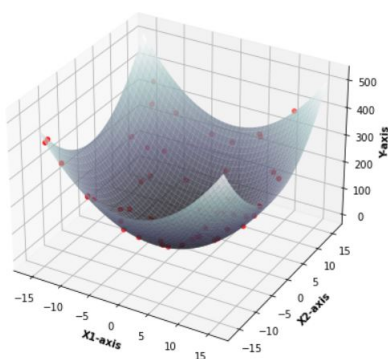
2.2 With regularization

2.2.1 Datapoints in training dataset - 50, $\lambda = e^{-18}$

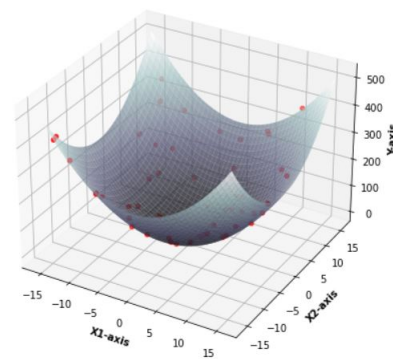
Hyperparameters $M = \{2, 3, 6\}$, $\lambda = e^{-18}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	$4.8206 * 10^{-12}$	$3.9873 * 10^{-12}$	$3.8850 * 10^{-12}$
3	$5.1383 * 10^{-12}$	$5.2417 * 10^{-12}$	$5.0842 * 10^{-12}$
6	$1.1819 * 10^{-10}$	$3.0351 * 10^{-10}$	$3.6636 * 10^{-10}$

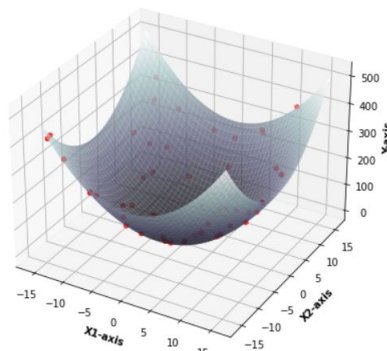
Table 2.2.1: ERMS for dataset2, training size = 50, $\lambda = e^{-18}$



(a) M=2



(b) M=3



(c) M=6

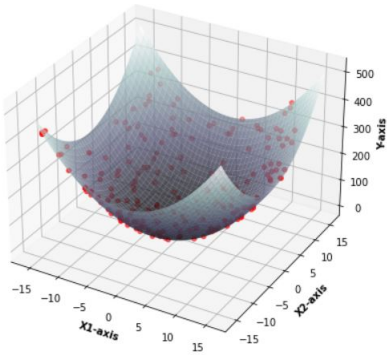
Figure 2.2.1: Approximated surface vs Training points for dataset2, training size = 50, $\lambda = e^{-18}$

2.2.2 Datapoints in training dataset - 200, $\lambda = e^{-18}$

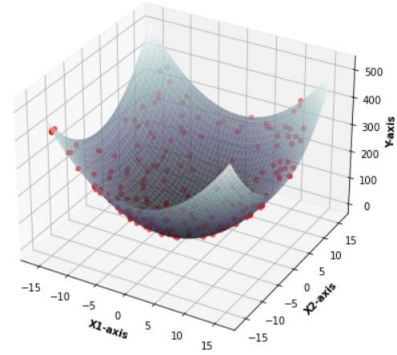
Hyperparameters $M = \{2, 3, 6\}$, $\lambda = e^{-18}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	$1.1856 * 10^{-12}$	$1.0939 * 10^{-12}$	$1.1856 * 10^{-12}$
3	$1.0120 * 10^{-12}$	$9.7215 * 10^{-13}$	$1.0120 * 10^{-12}$
6	$8.2470 * 10^{-11}$	$9.9772 * 10^{-11}$	$1.3023 * 10^{-10}$

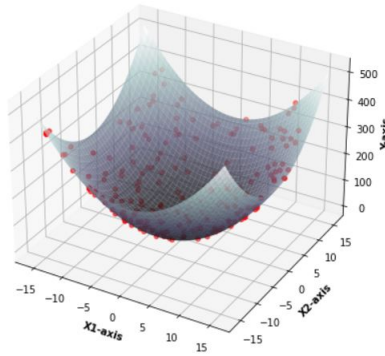
Table 2.2.2: ERMS for dataset2, training size = 200, $\lambda = e^{-18}$



(a) M=2



(b) M=3



(c) M=6

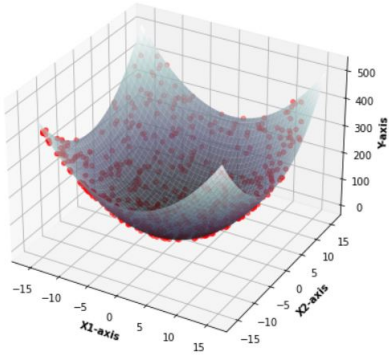
Figure 2.2.2: Approximated surface vs Training points for dataset2, training size = 200, $\lambda = e^{-18}$

2.2.3 Datapoints in training dataset - 500, $\lambda = e^{-18}$

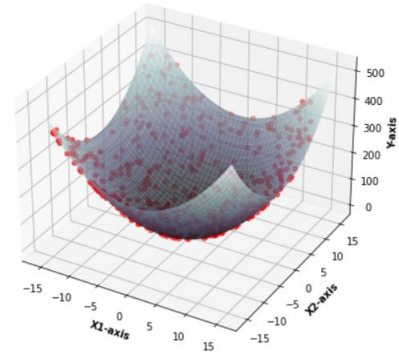
Hyperparameters $M = \{2, 3, 6\}$, $\lambda = e^{-18}$

M (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
2	$5.0186 * 10^{-13}$	$5.7271 * 10^{-13}$	$5.6067 * 10^{-13}$
3	$3.7982 * 10^{-13}$	$4.0030 * 10^{-13}$	$3.8668 * 10^{-13}$
6	$1.7098 * 10^{-10}$	$1.6708 * 10^{-10}$	$1.7894 * 10^{-10}$

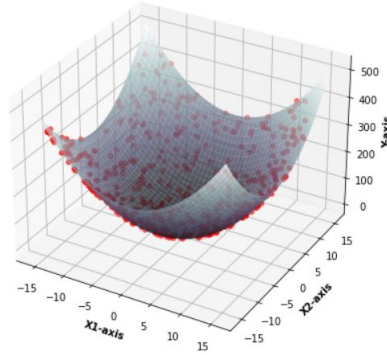
Table 2.2.3: ERMS for dataset2, training size = 500, $\lambda = e^{-18}$



(a) M=2



(b) M=3



(c) M=6

Figure 2.2.3: Approximated surface vs Training points for dataset2, training size = 500, $\lambda = e^{-18}$

2.2.4 Scatter plots, training size = 50, $\lambda = e^{-18}$

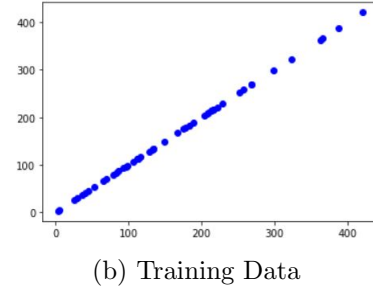
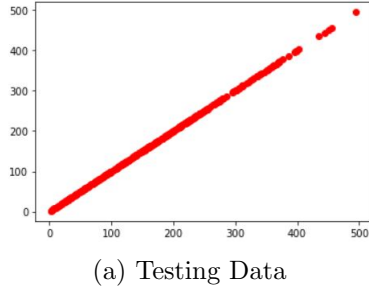


Figure 2.2.4: Model output vs Target output for dataset2, training size = 50, $\lambda = e^{-18}$, $M = 2$

2.2.5 Scatter plots, training size = 200, $\lambda = e^{-18}$

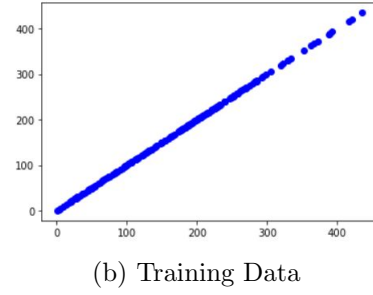
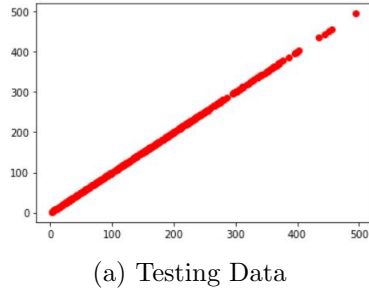


Figure 2.2.5: Model output vs Target output for dataset2, training size = 200, $\lambda = e^{-18}$, $M = 3$

2.2.6 Scatter plots, training size = 500, $\lambda = e^{-18}$

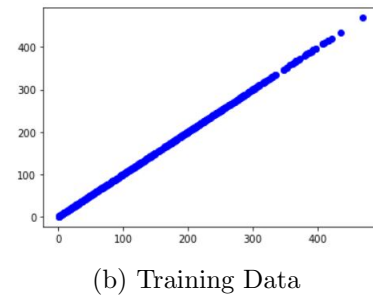
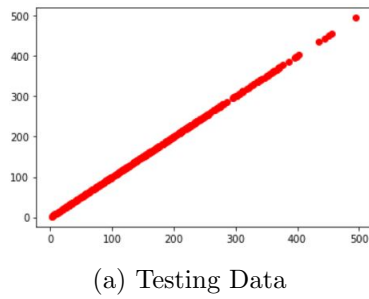


Figure 2.2.6: Model output vs Target output for dataset2, training size = 500, $\lambda = e^{-18}$, $M = 3$

3 Linear model for regression using Gaussian Basis function

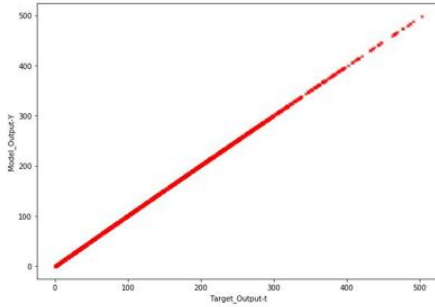
3.1 Dataset 2 : function2-2d.csv

3.1.1 Without regularization

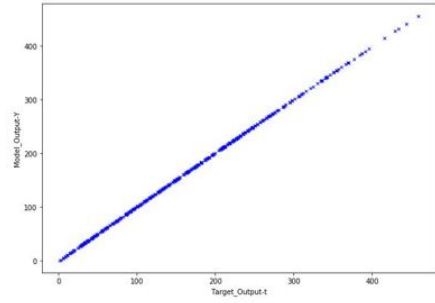
Hyperparameters : $\sigma = 20$, K = Number of clusters = $\{24, 26, 28\}$

K (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
24	0.7490	0.6947	0.6464
26	0.4903	0.5323	0.5270
28	3.1736	3.1996	3.1926

Table 3.1.1: ERMS for dataset2, $\lambda = 0$, $\sigma = 20$



(a) Testing Data



(b) Training Data

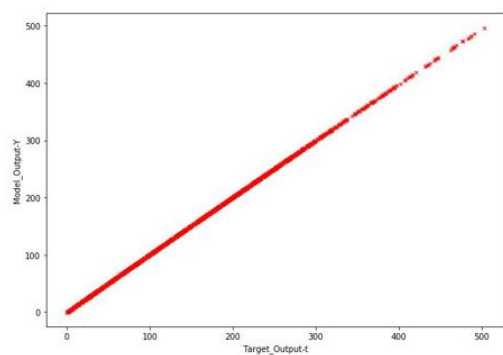
Figure 3.1.1: Model output vs Target output for dataset2, $\lambda = 0$, $\sigma = 20$, $K = 26$

3.1.2 With Quadratic regularization

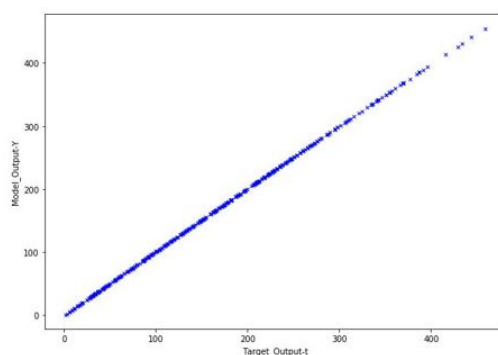
Hyperparameters : $\sigma = 20$, $K = \text{Number of clusters} = \{24, 26, 28, 44\}$, $\lambda = e^{-18}$

K (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
24	1.5278	1.2489	1.4309
26	1.3595	1.1126	1.2563
28	1.2026	0.9737	1.1920
44	0.6999	0.5889	0.6377

Table 3.1.2: ERMS for dataset2, $\lambda = e^{-18}$, $\sigma = 20$



(a) Testing Data



(b) Training Data

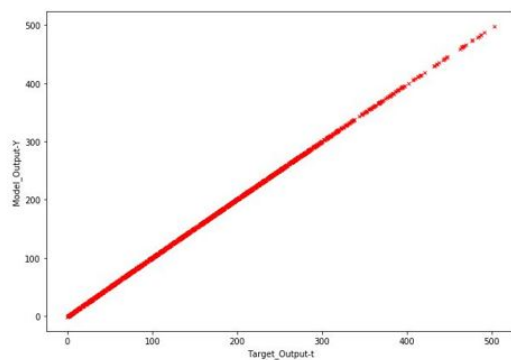
Figure 3.1.2: Model output vs Target output for dataset2, $\lambda = e^{-18}$, $\sigma = 20$, $K = 44$

3.1.3 With Tikhonov regularization

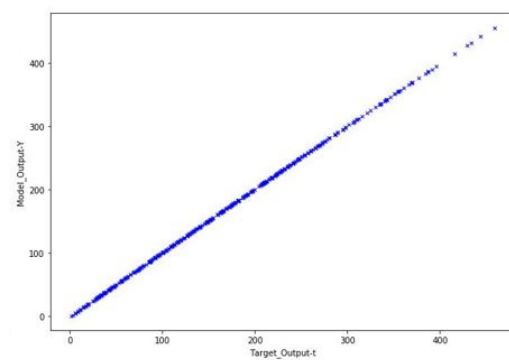
Hyperparameters : $\sigma = 20$, $K = \text{Number of clusters} = \{24, 26, 28\}$, $\lambda = e^{-18}$

K (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
24	0.7488	0.6955	0.6482
26	0.4957	0.5384	0.5369
28	4.70524	4.7112	4.7020

Table 3.1.3: ERMS for dataset2, $\lambda = e^{-18}$, $\sigma = 20$



(a) Testing Data



(b) Training Data

Figure 3.1.3: Model output vs Target output for dataset2, $\lambda = e^{-18}$, $\sigma = 20$, $K = 26$

3.2 Dataset 3 : 2-music.txt

3.2.1 Without regularization

Hyperparameters : $\sigma = \{100, 250\}$; K = Number of clusters

σ	K (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
100	3	0.0534366	0.0616985	0.0440546
100	5	0.050401	0.0550517	0.0429060
100	8	0.019156	0.019807	0.015414
100	10	0.020862	0.022438	0.017509
250	3	0.054272	0.062679	0.043885
250	5	0.047456	0.050827	0.036754
250	8	0.021510	0.023942	0.019601
250	10(<i>Best</i>)	0.016433	0.015821	0.013432

Table 3.2.1: ERMS for dataset3, $\lambda = 0$

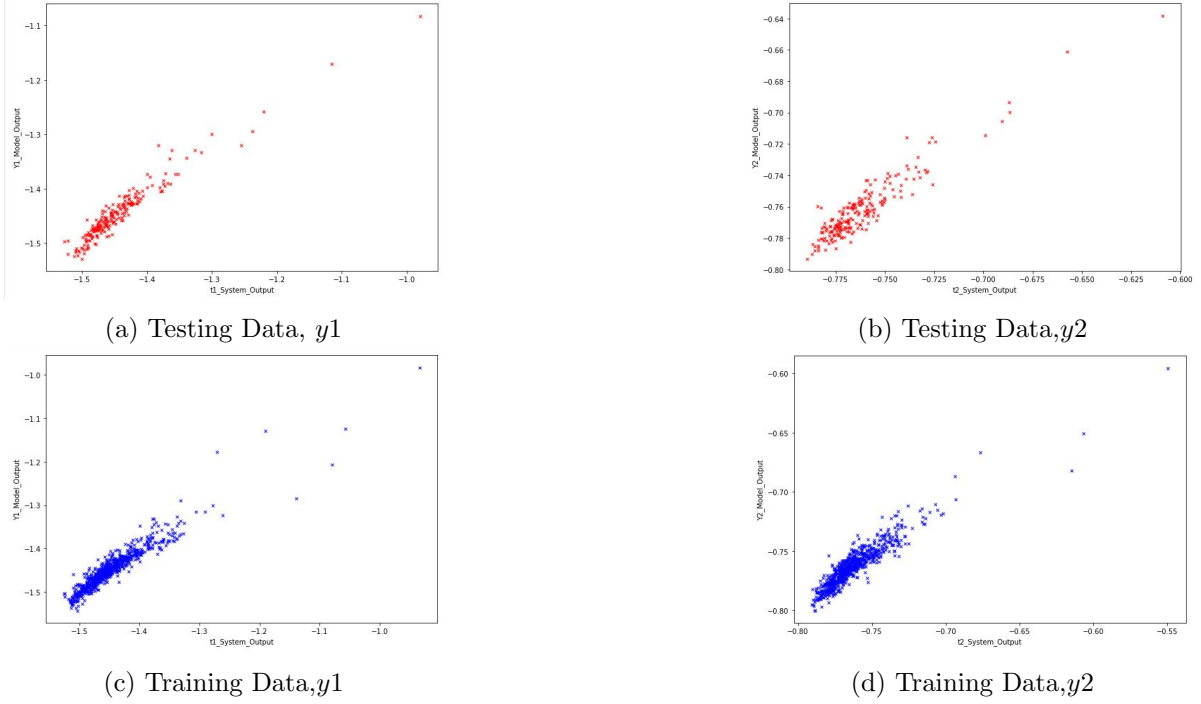


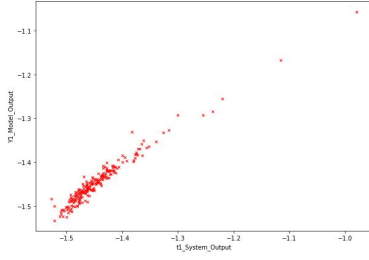
Figure 3.2.1: Model output vs Target output for dataset3, $\lambda = 0$

3.2.2 With Quadratic regularization

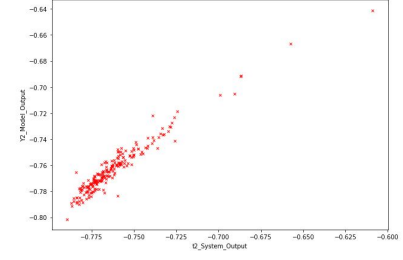
Hyperparameters : $\sigma = \{100, 250\}$; K = Number of clusters, $\lambda = e^{-18}$

σ	K (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
100	5	0.034362	0.040501	0.030973
100	8	0.018606	0.019096	0.015125
100	10	0.018795	0.019301	0.015040
100	14	0.017408	0.017247	0.015577
100	18(<i>Best</i>)	0.0137221	0.012451	0.013255
250	5	0.049459	0.052720	0.041285
250	8	0.018979	0.019356	0.015727
250	10	0.171402	0.0175077	0.015188
250	14	0.01633	0.016545	0.013944
250	18	0.015304	0.014845	0.013648

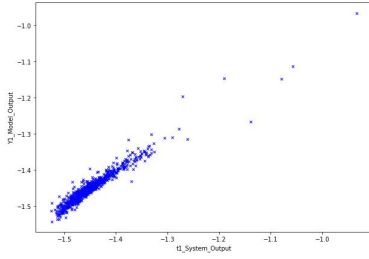
Table 3.2.2: ERMS for dataset3, $\lambda = e^{-18}$



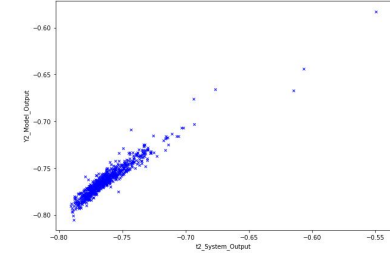
(a) Testing Data, y_1



(b) Testing Data, y_2



(c) Training Data, y_1



(d) Training Data, y_2

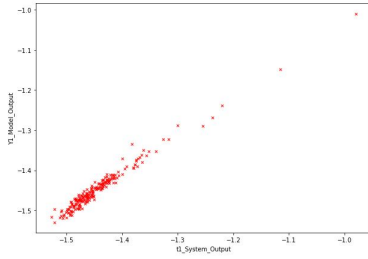
Figure 3.2.2: Model output vs Target output for dataset3, $\lambda = e^{-18}$

3.2.3 With Tikhonov regularization

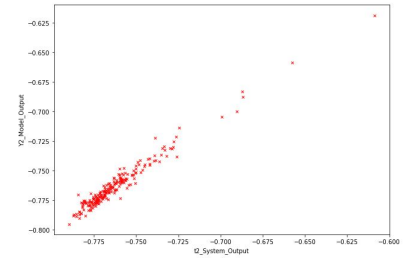
Hyperparameters : $\sigma = \{100, 250\}$; K = Number of clusters, $\lambda = e^{-18}$

σ	K (Hyperparameter)	Training Dataset	Test Dataset	Validation Dataset
100	5	0.047063	0.049364	0.040235
100	8	0.019724	0.021607	0.015492
100	10	0.019228	0.019893	0.016137
100	14	0.0172312	0.017366	0.015561
100	16	0.013487	0.012603	0.0132219
250	5	0.053574	0.060502	0.043764
250	10	0.020576	0.021579	0.016228
250	14	0.0178025	0.01837	0.0015628
250	18(<i>Best</i>)	0.012148	0.011746	0.010630

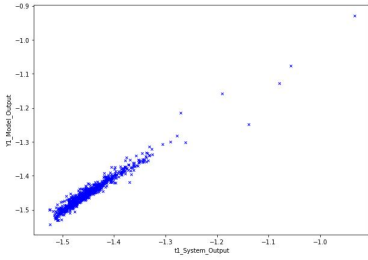
Table 3.2.3: ERMS for dataset3, $\lambda = e^{-18}$ using Tikhonov regularization



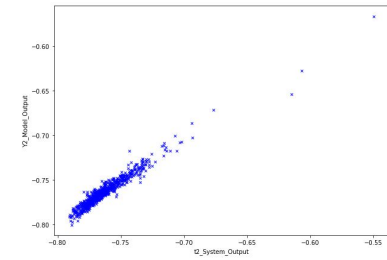
(a) Testing Data, y_1



(b) Testing Data, y_2



(c) Training Data, y_1



(d) Training Data, y_2

Figure 3.2.3: Model output vs Target output for dataset3, $\lambda = e^{-18}$ using Tikhonov regularization

Observation :

Here since we were using 70% of the data we had the liberty to go for a complex model i.e. for higher value of K with the performance of model still increasing at a good rate.