Pattern Recognition and Machine Learning

Assignment 1
Report

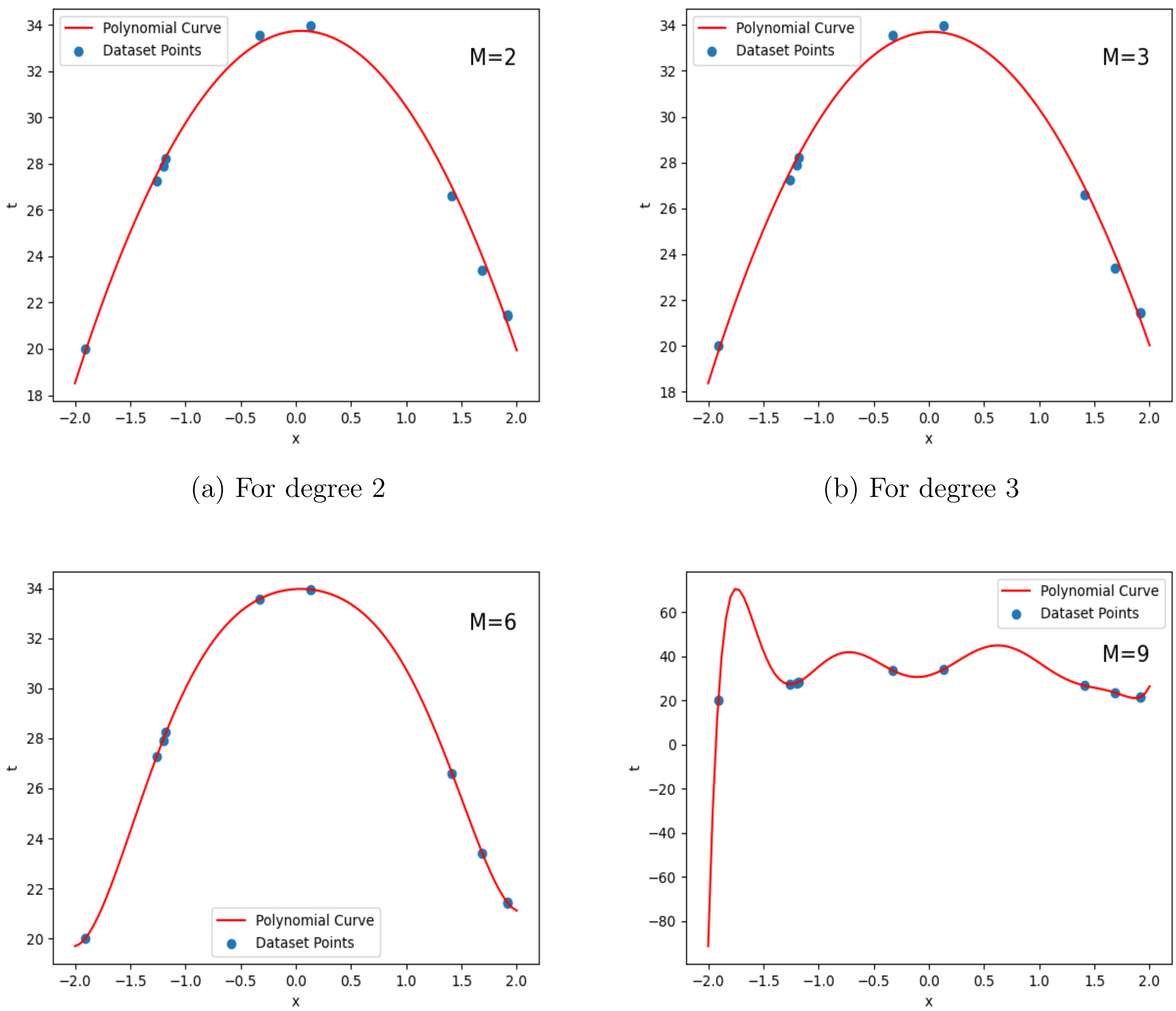
Group No. 19

Ranjith Tevnan	Sai Bandawar	Jay shah
EE18B146	EE18B150	EE18B158

17,March 2021

1 Task 1: Polynomial curve fitting for 1- Dimensional Dataset

1.1 Curve Fitting without regularisation



(a) For degree 2

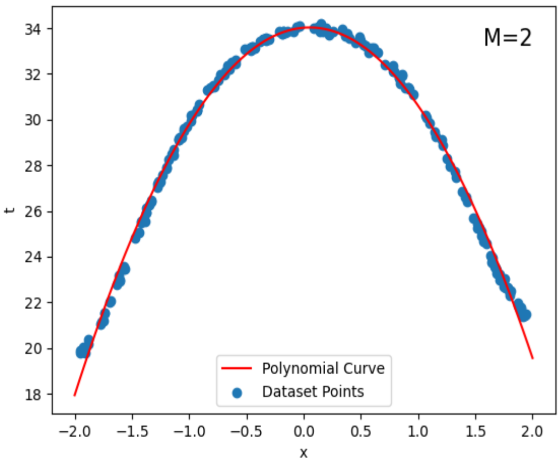
(b) For degree 3

(c) For degree 6

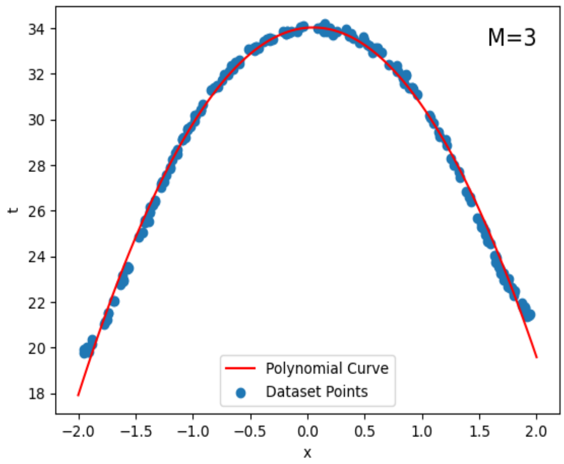
(d) For degree 9

Figure 1: Approximated functions obtained using training datasets size $N = 10$

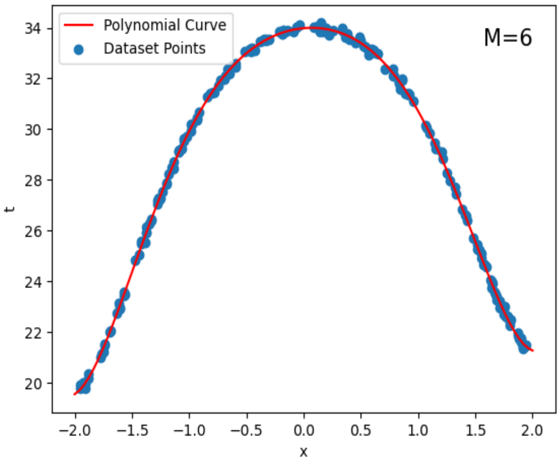
Here, We can see that for a small Dataset Training size (N=10), Degree 9 Polynomial clearly overfits the data and passes through all the points. With decreasing degree, the curve fitting of data reduces. For N=200, The polynomial curves of all the degrees fit the points in almost the same way. This is because for higher number of data points, the over-fitting of the curves on data reduces.



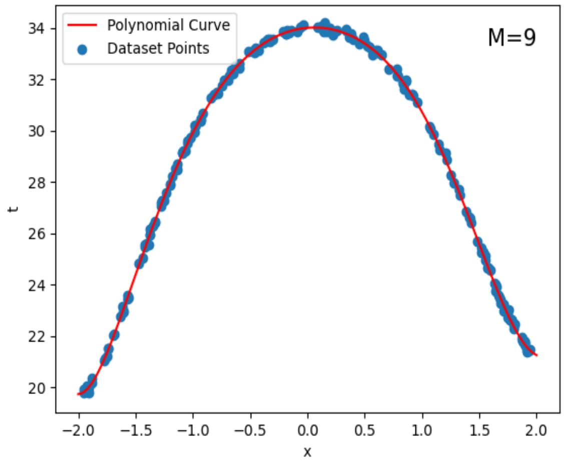
(a) For degree 2



(b) For degree 3



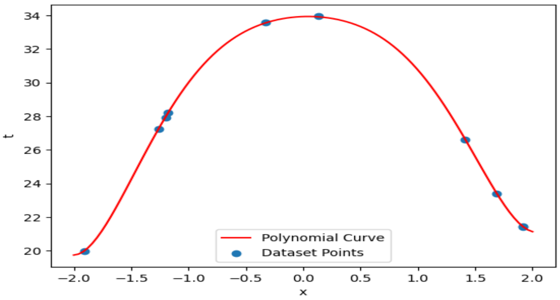
(c) For degree 6



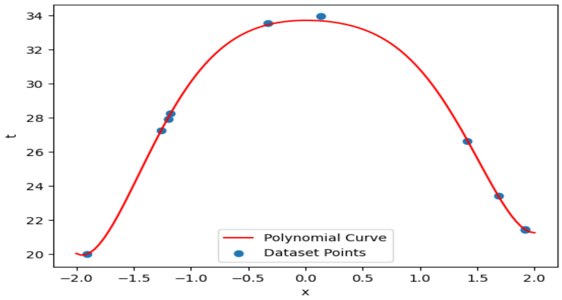
(d) For degree 9

Figure 2: Approximated functions obtained using training datasets size $\boldsymbol{N = 200}$

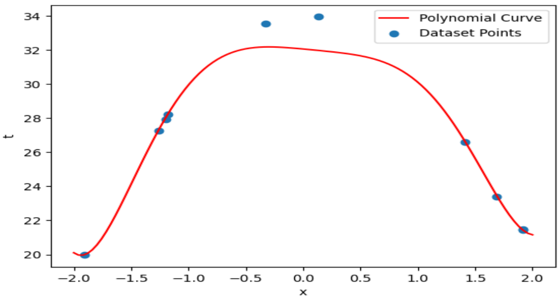
1.2 Curve Fitting with regularisation (dataset size 10)



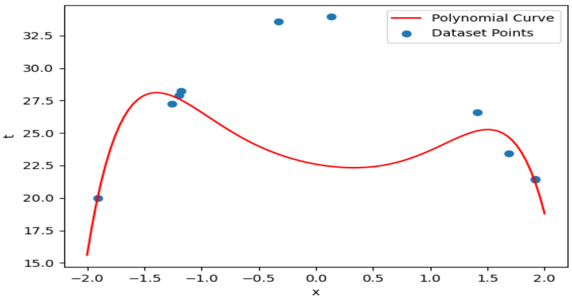
(a) $\lambda = 0.001$



(b) $\lambda = 0.01$

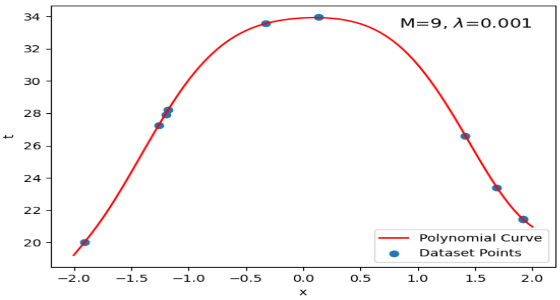


(c) $\lambda = 0.1$

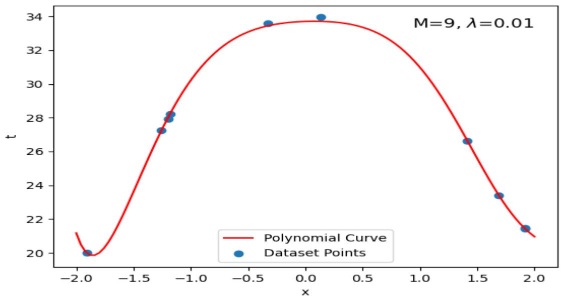


(d) $\lambda = 1$

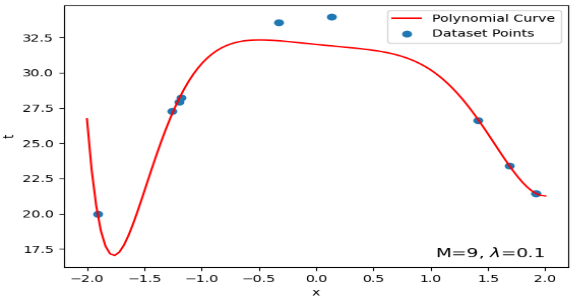
Figure 3: Training Dataset $\boldsymbol{size} = \mathbf{10}$ & $\boldsymbol{order} = \mathbf{6}$ with regularisation parameter $\boldsymbol{\lambda}$



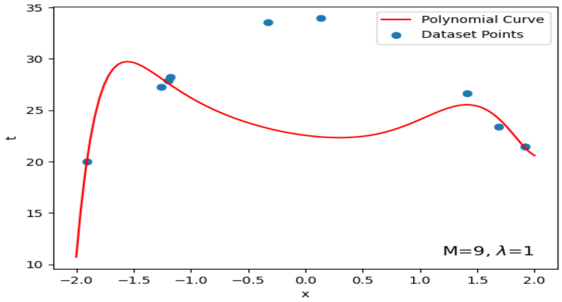
(a) $\lambda = 0.001$



(b) $\lambda = 0.01$



(c) $\lambda = 0.1$



(d) $\lambda = 1$

Figure 4: Training Dataset $\boldsymbol{size} = \mathbf{10}$ & $\boldsymbol{order} = \mathbf{9}$ with regularisation parameter $\boldsymbol{\lambda}$

It can be seen that in both cases as λ increases the curve fit worsens because of more weight to the regularisation term and less to the sum of errors term. Hence a tradeoff ensues with λ around 0.001 giving the best fit

1.3 E_{rms} using different model complexities (w/o regularisation)

Polynomial Order	Training Data	Validation Data	Testing Data
2.0	0.3137	0.5088	0.336
3.0	0.3076	0.5854	0.3988
6.0	0.0392	0.1518	0.1003
9.0	0.0	6.6656	0.8775

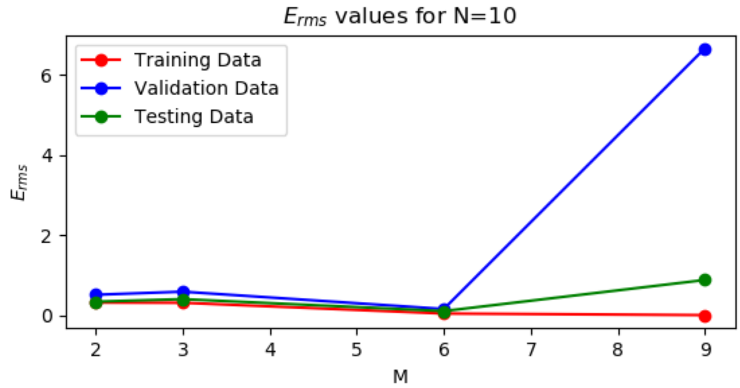


Figure 5: $Erms$ for N=10 without regularisation

It can be observed that:

- The training error decreases with model complexity with degree 9 giving 0 error.
- However, since the sample size is small (10) the **degree 9 model over-fits the data as it has comparatively high validation and test error**.Hence the need for regularisation.
- Degrees 2 and 3 underfit the data while degree fares appears better and generalises more.

Polynomial Order	Training Data	Validation Data	Testing Data
2.0	0.3407	0.3481	0.4472
3.0	0.3406	0.3481	0.4466
6.0	0.1047	0.1061	0.0989
9.0	0.1036	0.1072	0.1007

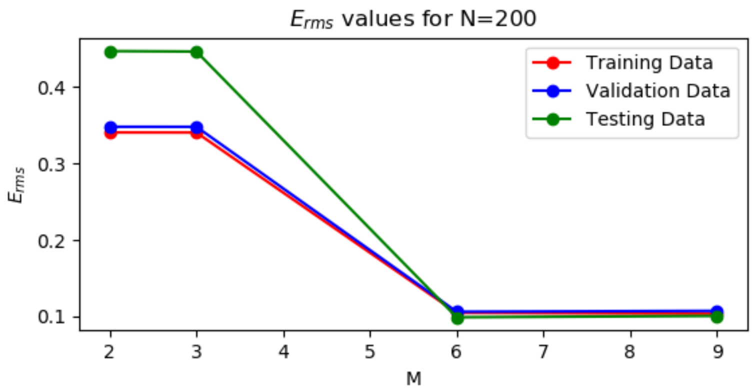
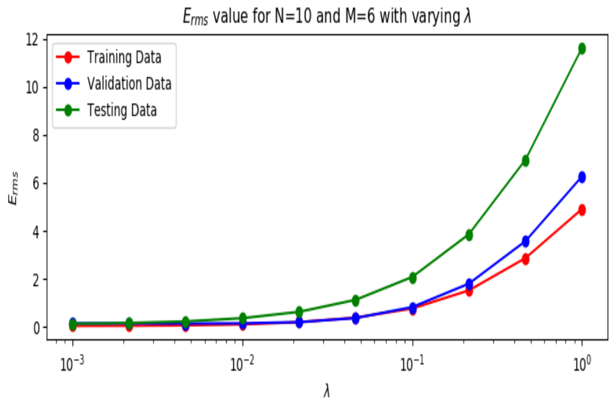


Figure 6: E_{rms} for N=200 without regularisation

In the case of a relatively large sample size (200) all given models underfit and so degree 9 fares the best having small errors as compared to other models

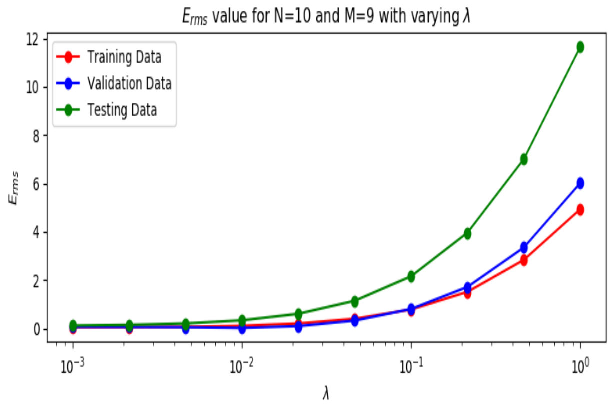
1.4 E_{rms} with regularisation

λ	Training Data	Validation Data	Testing Data
0.001	0.0406	0.1465	0.1286
0.0022	0.045	0.1422	0.1605
0.0046	0.0607	0.1384	0.2268
0.01	0.1024	0.1461	0.3608
0.0215	0.1937	0.1952	0.6229
0.0464	0.3811	0.3649	1.1237
0.1	0.7638	0.8208	2.0752
0.2154	1.5171	1.797	3.8546
0.4642	2.8639	3.5702	6.9599
1.0	4.9113	6.2703	11.6279



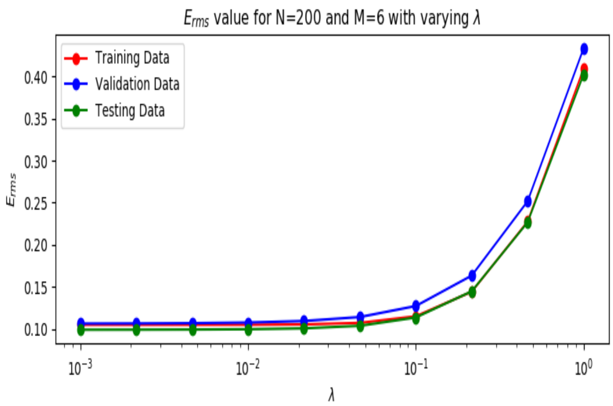
(a) $Erms$ for N=10 & M=6

λ	Training Data	Validation Data	Testing Data
0.001	0.0402	0.0569	0.1117
0.0022	0.0442	0.0553	0.14
0.0046	0.0591	0.0434	0.2005
0.01	0.1013	0.0124	0.3293
0.0215	0.1975	0.0895	0.5987
0.0464	0.3938	0.3176	1.1377
0.1	0.777	0.7985	2.1516
0.2154	1.5085	1.7118	3.9569
0.4642	2.8322	3.3447	7.0074
1.0	4.9293	6.0133	11.6517



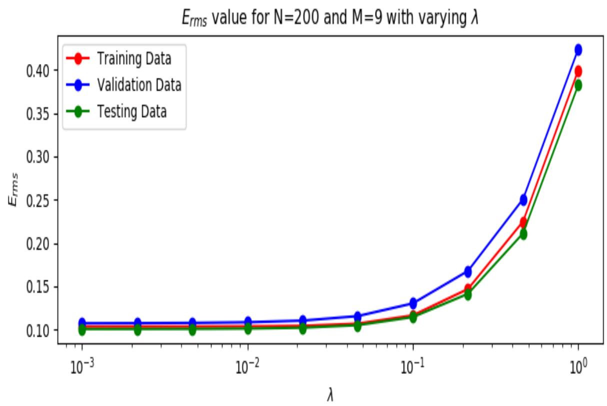
(b) $Erms$ for N=10 & M=9

λ	Training Data	Validation Data	Testing Data
0.001	0.1047	0.1063	0.0989
0.0022	0.1047	0.1064	0.099
0.0046	0.1048	0.1067	0.0992
0.01	0.1049	0.1075	0.0995
0.0215	0.1053	0.1093	0.1006
0.0464	0.1071	0.114	0.1036
0.1	0.1149	0.1271	0.1135
0.2154	0.1441	0.1633	0.1446
0.4642	0.2274	0.2516	0.2266
1.0	0.409	0.4332	0.4022



(c) $Erms$ for N=200 & M=6

λ	Training Data	Validation Data	Testing Data
0.001	0.1036	0.1074	0.1007
0.0022	0.1036	0.1075	0.1008
0.0046	0.1037	0.1078	0.1009
0.01	0.1038	0.1085	0.1012
0.0215	0.1044	0.1104	0.1022
0.0464	0.1069	0.1156	0.1052
0.1	0.1164	0.1303	0.1145
0.2154	0.1469	0.1677	0.1413
0.4642	0.2248	0.2507	0.2112
1.0	0.3993	0.4238	0.3821



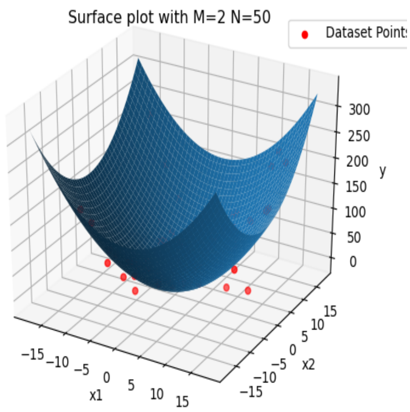
(d) $Erms$ for N=200 & M=9

Figure 7: $Erms$ on Training Dataset with regularisation parameter $\boldsymbol{\lambda}$

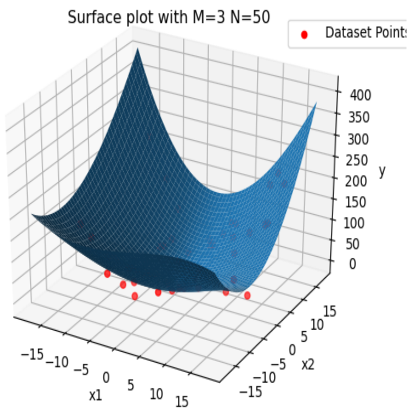
It can be seen that for both dataset sizes E_{rms} increases with increasing λ with an optimum value of around 0.001.
However, in Fig7(c) for degree 9 this regularised model slightly outperforms the unregularised one as seen in the validation error.
In the case of dataset size 200, regularisation does not perform better than the original model as all models underfit the data

2 Task 2: Linear Regression using polynomial basis functions for 2-Dimensional DataSet

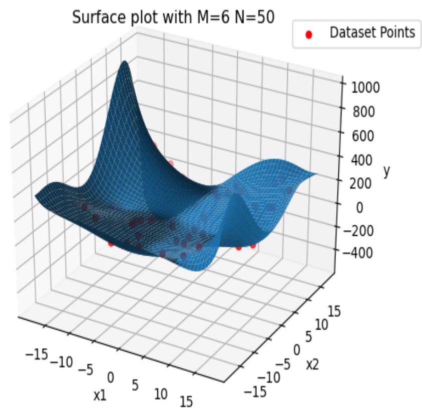
2.1 Surface Plots with different model complexities



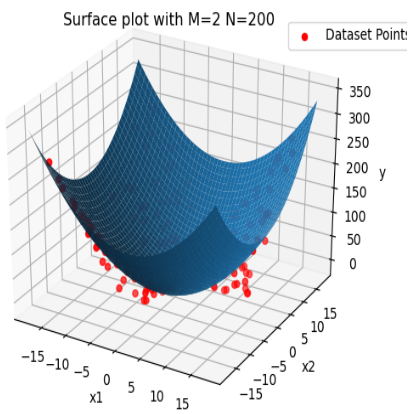
(a) M=2 & N=50



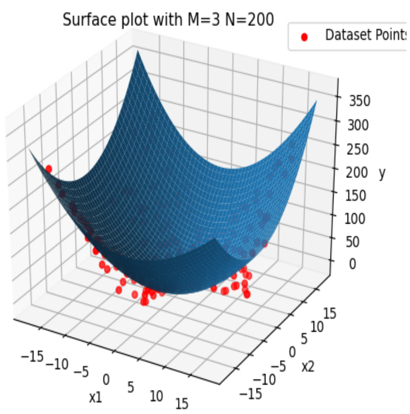
(b) M=3 & N=50



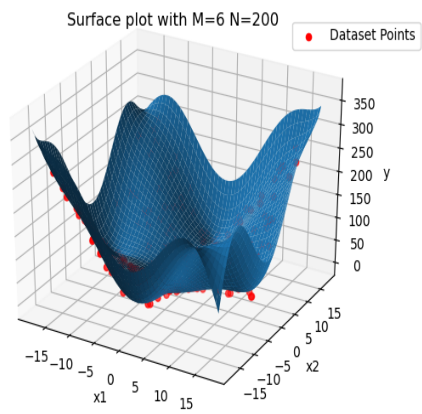
(c) M=6 & N=50



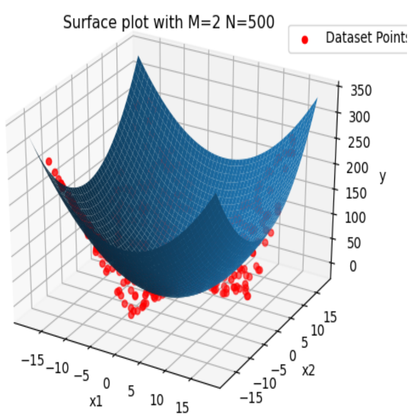
(d) M=2 & N=200



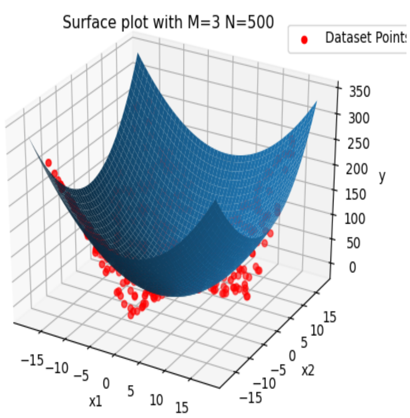
(e) M=3 & N=200



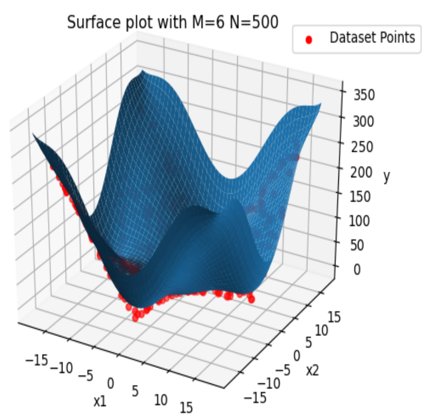
(f) M=6 & N=200



(g) M=2 & N=500



(h) M=3 & N=500

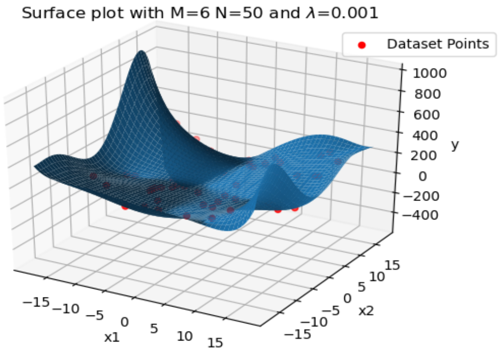


(i) M=9 & N=500

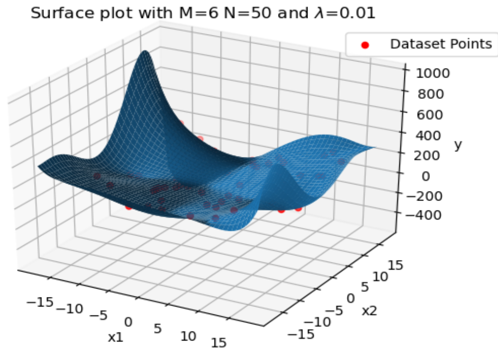
Figure 8: Surface plots with different dataset size (N) and degrees (M)

It can be observed that M=6, with N=(50,200) overfits slightly as seen the the undulations on the surface plots

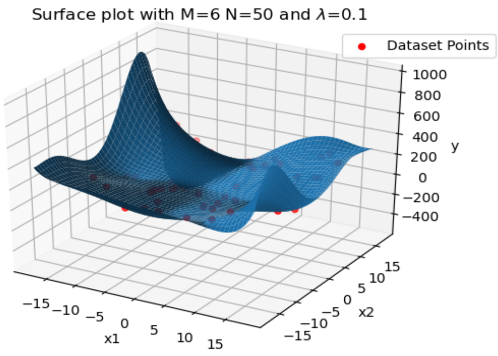
2.2 Surface plots with regularisation with degree 6



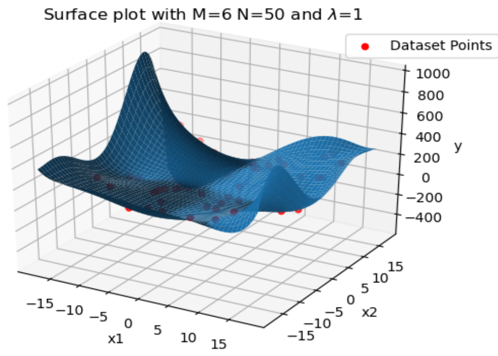
(a) λ=0.001 & N=50



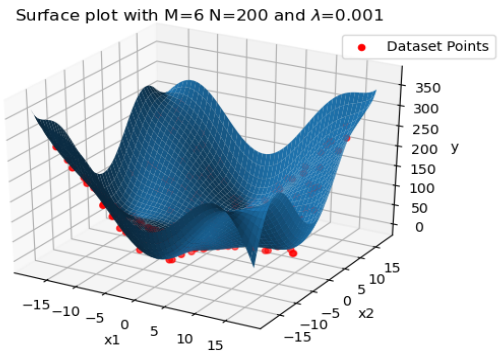
(b) λ=0.01 & N=50



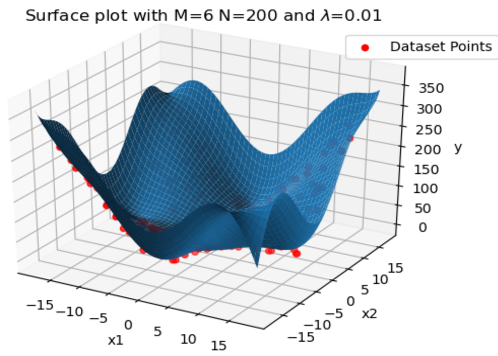
(c) λ=0.1 & N=50



(d) λ=1 & N=50

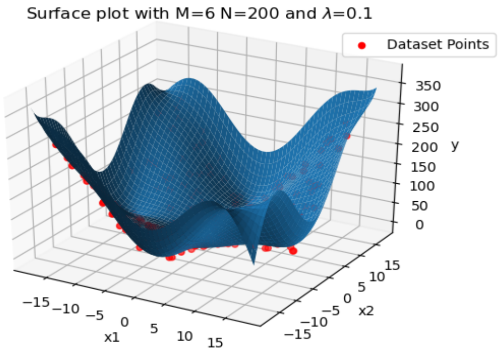


(e) λ=0.001 & N=200

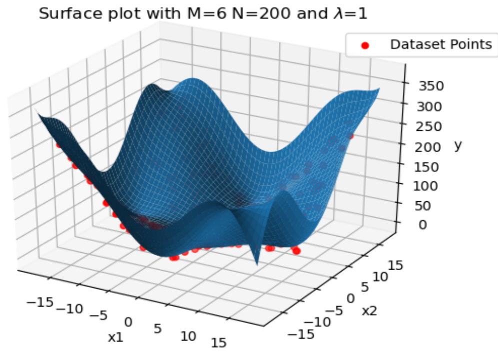


(f) λ=0.01 & N=200

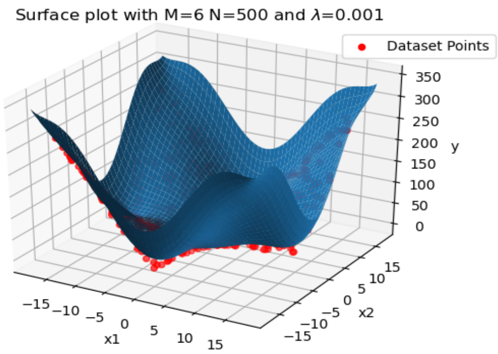
Figure 9: Surface plots with degree 6 and varying regularisation parameter λ & sample size(N)



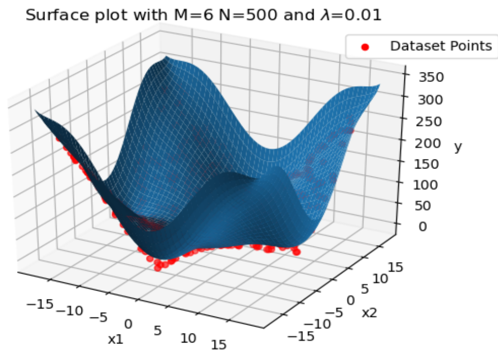
(a) λ=0.1 & N=200



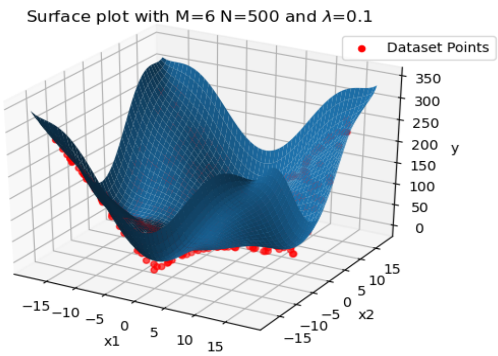
(b) λ=1 & N=200



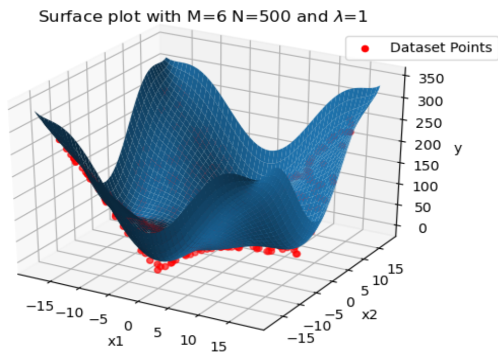
(c) λ=0.001 & N=500



(d) λ=0.01 & N=500



(e) λ=0.1 & N=500



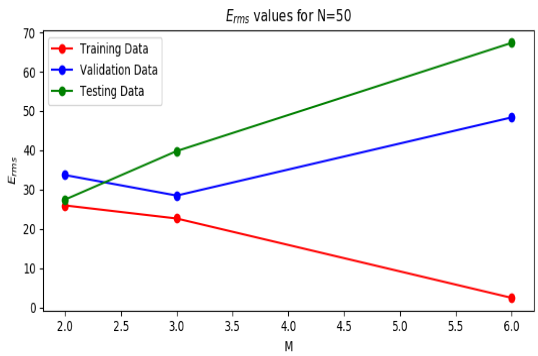
(f) λ=1 & N=500

Figure 10: Surface plots with degree 6 and varying regularisation parameter λ & sample size(N)

We can see in the above plots that the curves do not change when we use a higher lambda value. This is because the lambda value used is not big enough to cause any major curve fitting change.

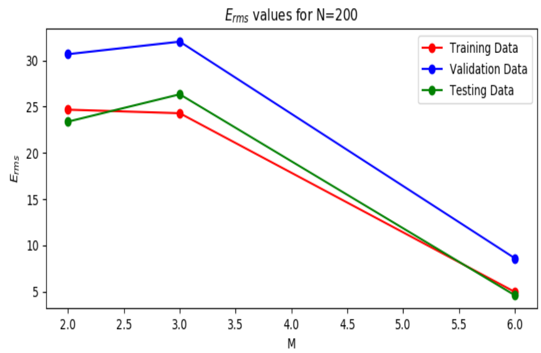
2.3 E_{rms} using different model complexities

Polynomial Order	Training Data	Validation Data	Testing Data
2.0	25.9528	33.6829	27.4625
3.0	22.6333	28.47	39.805
6.0	2.444	48.3576	67.333



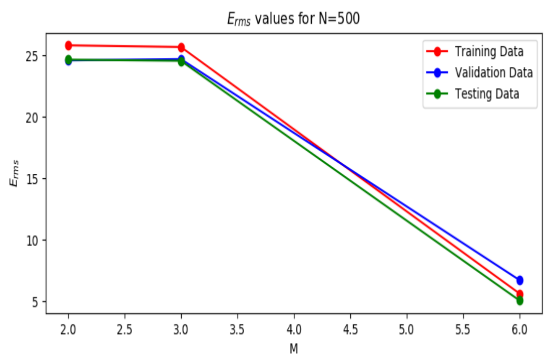
(a) *Erms* for N=50

Polynomial Order	Training Data	Validation Data	Testing Data
2.0	24.6806	30.684	23.3886
3.0	24.2959	32.0384	26.3432
6.0	4.9715	8.6149	4.6109



(b) *Erms* for N=200

Polynomial Order	Training Data	Validation Data	Testing Data
2.0	25.8138	24.5896	24.6487
3.0	25.6786	24.6954	24.5568
6.0	5.6139	6.7087	5.048

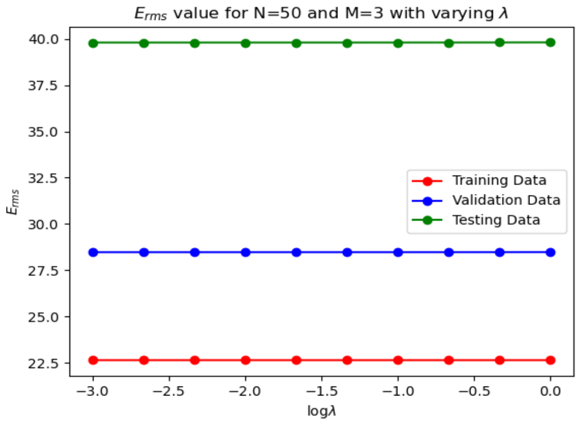


(c) *Erms* for N=500

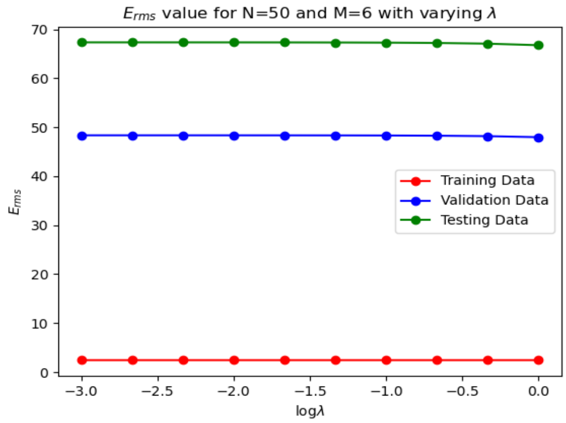
Figure 11: *Erms* on Training Dataset size N=(50,200,500) without regularisation

Here we can see that for N=50, the best performing model is for M=3 as the validation data rms is the lowest. Also for N=200 and N=500, M=6 is the best performing model as the validation data rms is least among all the other degrees

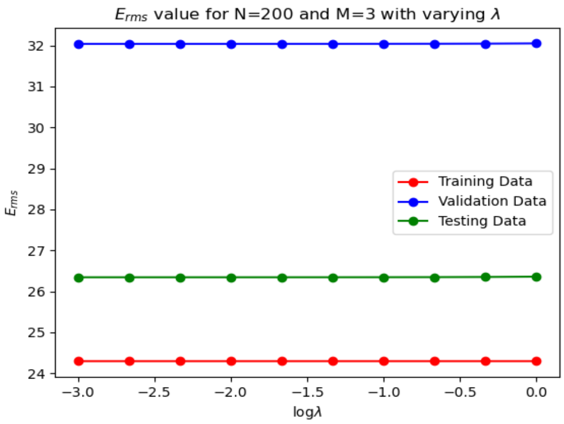
2.3.1 E_{rms} with small regularisation parameter λ



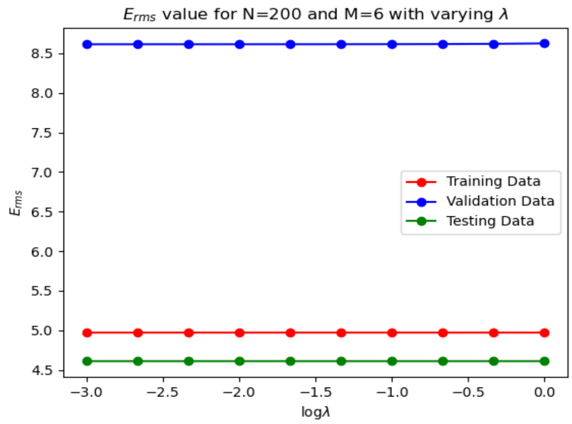
(a) Caption1



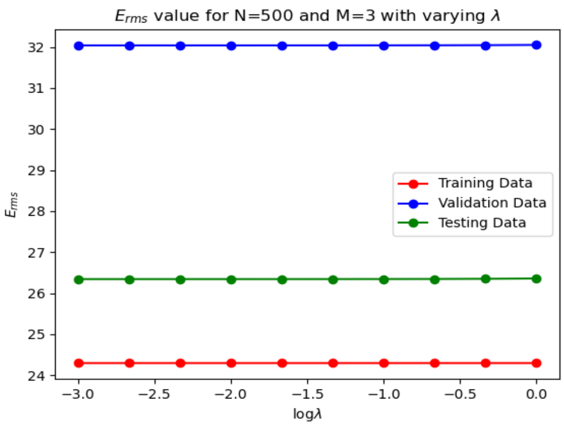
(b) Caption 2



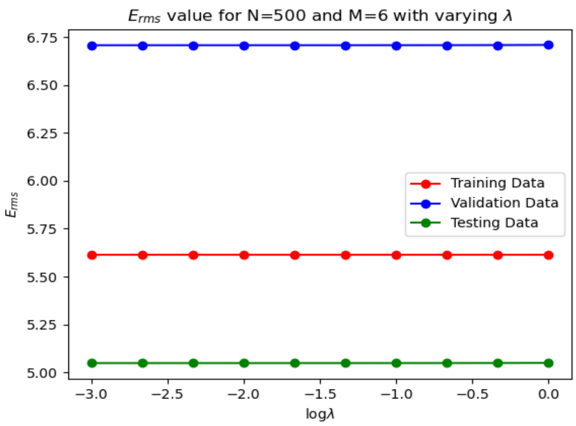
(c) Caption3



(d) Caption 4



(e) Caption5



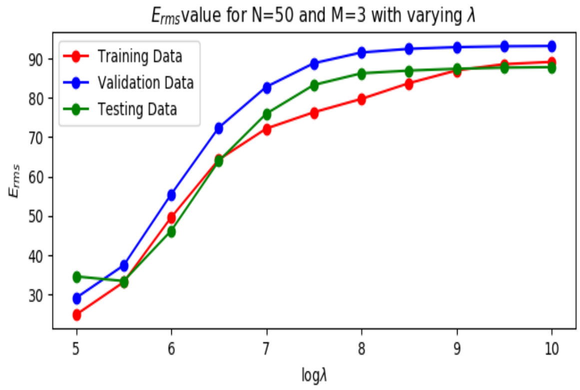
(f) Caption 6

Figure 12: $Erms$ on Training Dataset size N=(50,200,500) & M = (3,6) with small regularisation parameter λ

Here we can clearly see that change in λ when it is small makes no difference in the E_{rms} values of data.

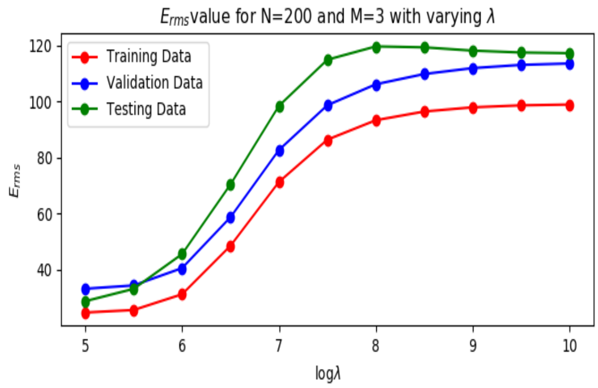
2.3.2 E_{rms} with regularisation using degree 3

$log\lambda$	Training Data	Validation Data	Testing Data
5.0	24.8828	29.1363	34.5946
5.5	33.1405	37.41	33.4295
6.0	49.6294	55.5024	46.2306
6.5	64.3128	72.4341	63.9334
7.0	72.19	82.8003	75.9812
7.5	76.3406	88.8083	83.288
8.0	79.7293	91.5385	86.243
8.5	83.7292	92.4625	86.936
9.0	86.9754	92.8969	87.3925
9.5	88.5601	93.1041	87.6863
10.0	89.1481	93.1827	87.8085



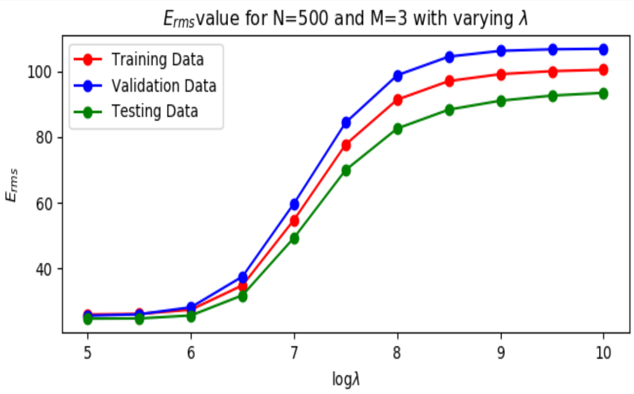
(a) $Erms$ for N=50

$log\lambda$	Training Data	Validation Data	Testing Data
5.0	24.7574	33.2446	28.7985
5.5	25.6616	34.4375	33.2106
6.0	31.2815	40.6172	45.6225
6.5	48.5916	58.7909	70.4963
7.0	71.3692	82.7751	98.5333
7.5	86.4569	98.7454	114.9763
8.0	93.3846	106.2138	119.7287
8.5	96.4956	109.88	119.3903
9.0	97.9906	111.9824	118.217
9.5	98.6944	113.1327	117.551
10.0	98.9809	113.6321	117.3067



(b) $Erms$ for N=200

$log\lambda$	Training Data	Validation Data	Testing Data
5.0	26.0458	25.6684	24.7669
5.5	26.1994	26.095	24.8291
6.0	27.4829	28.2271	25.7455
6.5	34.9008	37.5742	31.8831
7.0	54.8376	59.8239	49.3846
7.5	77.7046	84.3705	69.9455
8.0	91.3426	98.746	82.5612
8.5	96.9763	104.4042	88.2957
9.0	99.0826	106.1511	91.0086
9.5	99.9864	106.6265	92.5573
10.0	100.4194	106.7918	93.3984



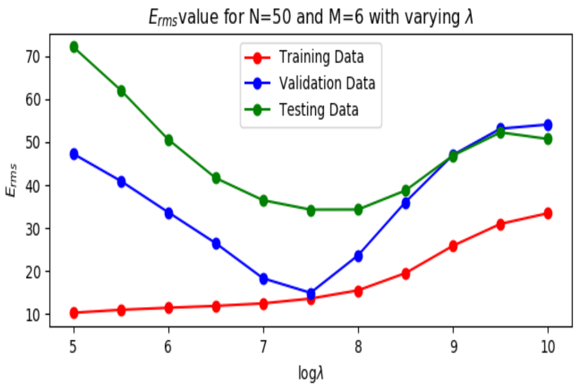
(c) $Erms$ for N=500

Figure 13: $Erms$ on Training Dataset size N=(50,200,500) & M = 3 with regularisation

Here with increase in higher values of λ, The E_{rms} values increase suggest that the model underfits the data as regularisation makes a marginal difference

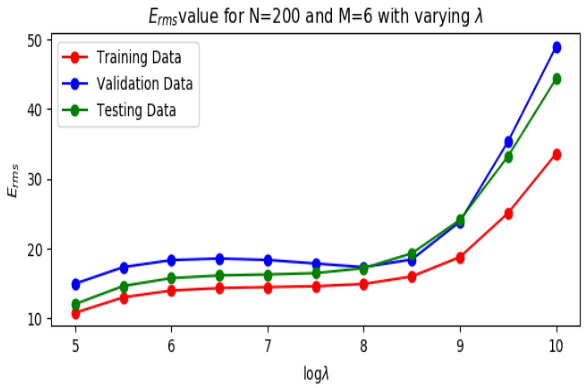
2.3.3 E_{rms} with regularisation using degree 6

logλ	Training Data	Validation Data	Testing Data
5.0	10.3393	47.2956	72.1379
5.5	11.0247	40.9867	62.1055
6.0	11.5024	33.7152	50.6394
6.5	11.9293	26.5333	41.6684
7.0	12.5153	18.354	36.4923
7.5	13.6366	14.9699	34.3111
8.0	15.5493	23.6781	34.3308
8.5	19.5575	36.068	38.7588
9.0	25.9074	47.0173	46.8804
9.5	30.9931	53.1288	52.2611
10.0	33.4736	54.0895	50.7283



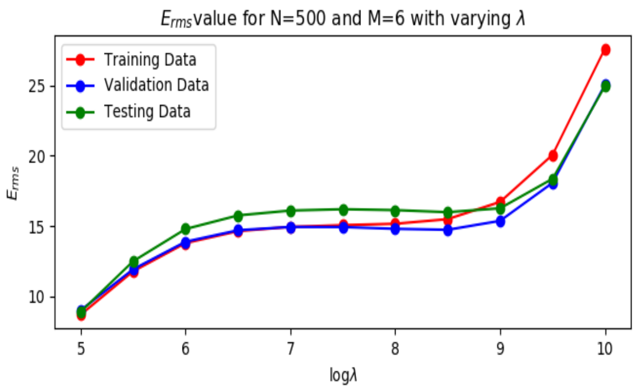
(a) *Erms* for N=50

logλ	Training Data	Validation Data	Testing Data
5.0	10.8198	14.9939	12.0588
5.5	13.0252	17.3591	14.6525
6.0	14.002	18.3647	15.7794
6.5	14.3567	18.5899	16.1645
7.0	14.4991	18.385	16.2978
7.5	14.6212	17.8736	16.4931
8.0	14.9399	17.3511	17.2069
8.5	16.0238	18.4577	19.3385
9.0	18.8013	23.8725	24.1199
9.5	25.1065	35.3627	33.2474
10.0	33.6358	48.9845	44.3674



(b) *Erms* for N=200

logλ	Training Data	Validation Data	Testing Data
5.0	8.7107	9.0273	8.9189
5.5	11.7929	11.9379	12.507
6.0	13.7973	13.8942	14.7964
6.5	14.646	14.7087	15.7639
7.0	14.9508	14.9468	16.11
7.5	15.0718	14.9386	16.2053
8.0	15.1714	14.811	16.1397
8.5	15.4978	14.7418	15.9941
9.0	16.7245	15.3793	16.2714
9.5	20.0653	18.0763	18.3666
10.0	27.6137	25.0619	24.9425

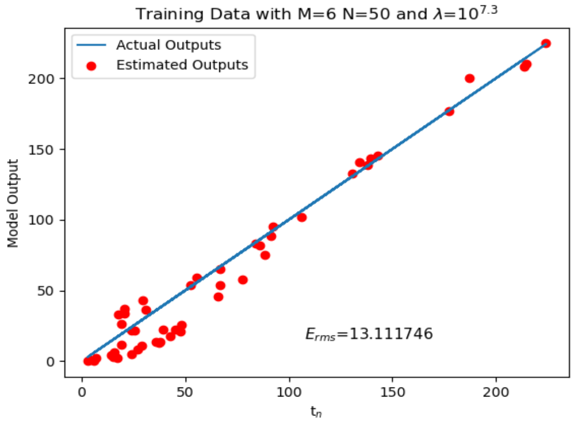


(c) *Erms* for N=500

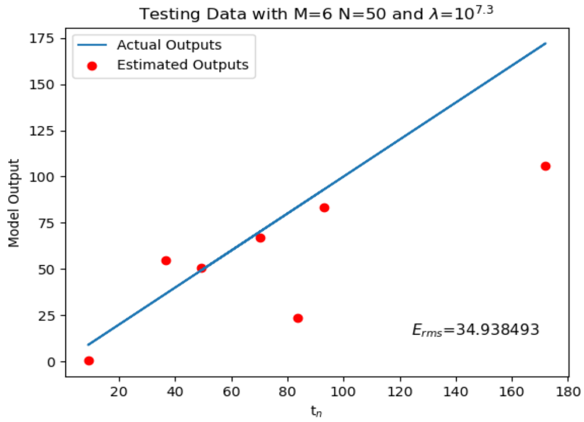
Figure 14: *Erms* on Training Dataset size N=(50,200,500) & M = 6 with regularisation

The N=50 Plot suggests that that the E_{rms} is lowest for logλ=7.4 and hence it gives the best model. Both for N=200 and N=500, the E_{rms} increases with λ and hence no regularisation is the best model in those cases.

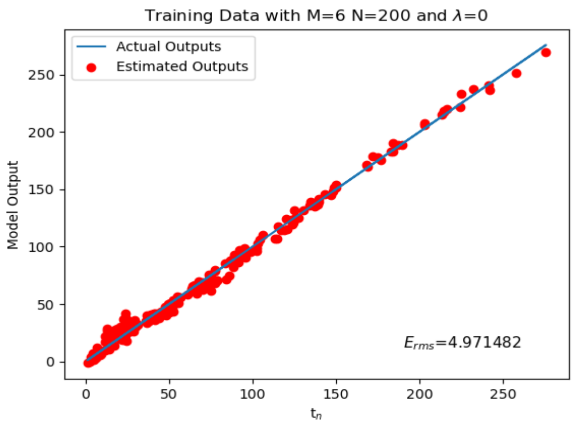
2.4 Scatter plots for the best models for each dataset size



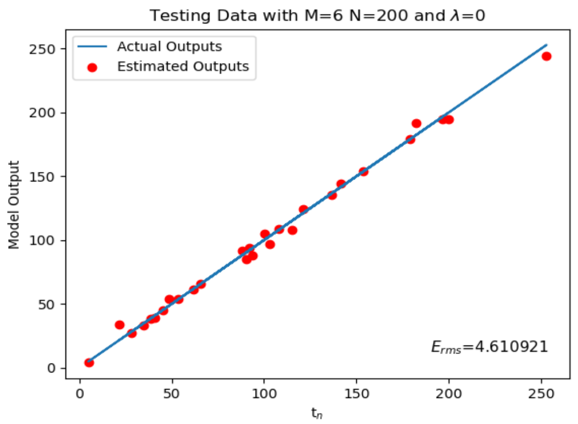
(a) Caption1



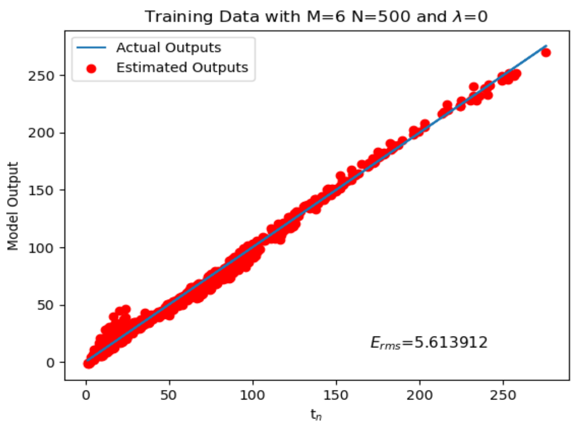
(b) Caption 2



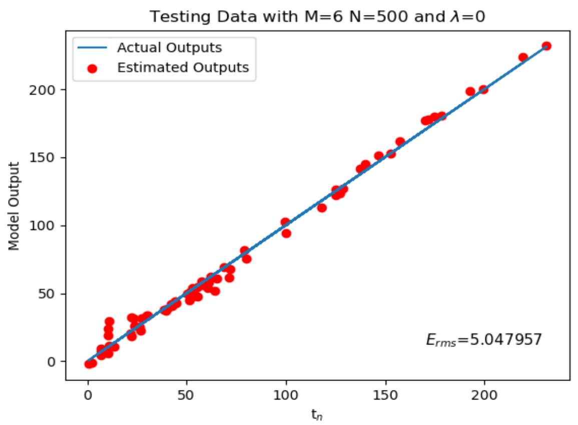
(c) Caption3



(d) Caption 4



(e) Caption5

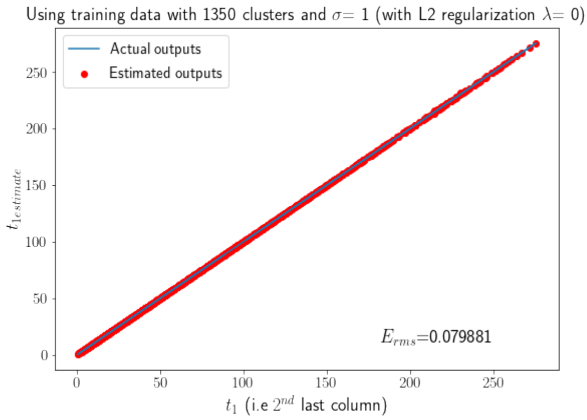


(f) Caption 6

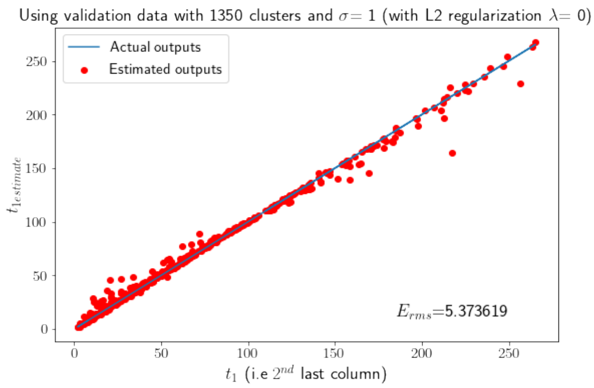
Figure 15: ERMS with regularisation

From the above plots, we can see that models fit the data better when N is higher.

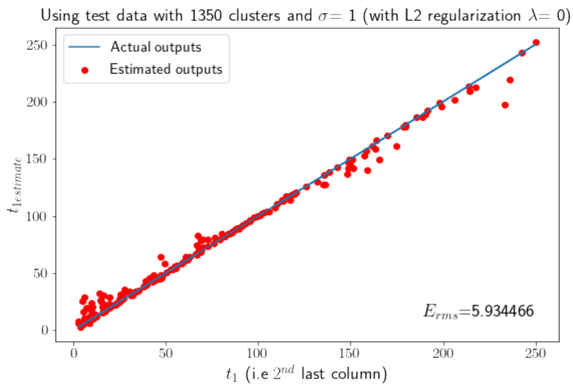
2.5 Scatter plots guassian basis function



(a) Using Training data for t_1



(b) Using Validation data for t_1

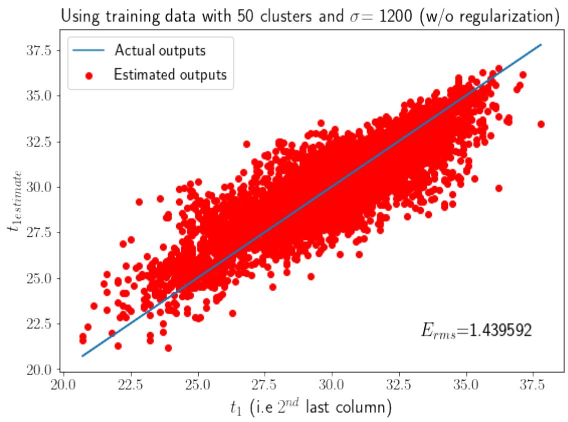


(c) Using Test data for t_1

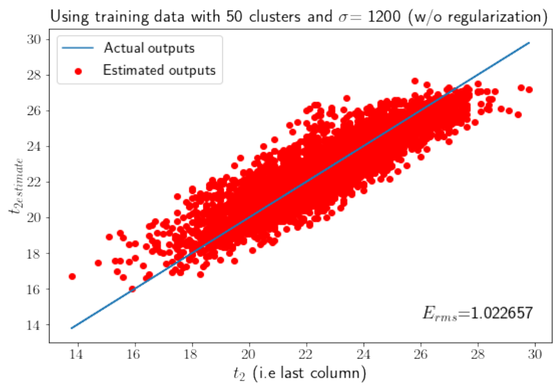
Figure 16: P

3 Task 3: Linear model for regression using Gaussian basis functions for Real World Dataset

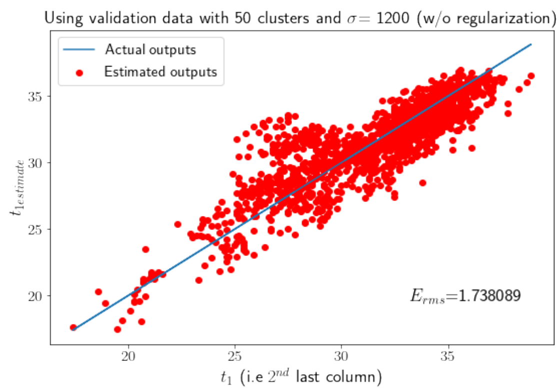
3.1 Scatter plots for a few model complexities



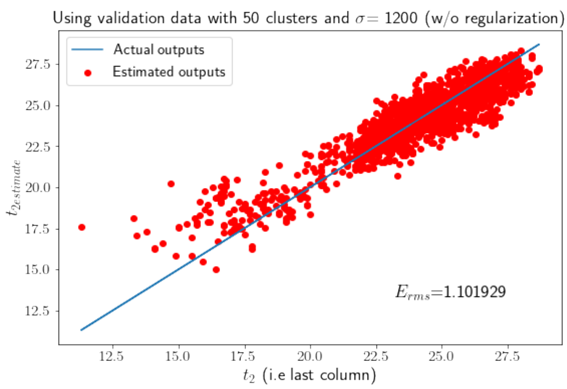
(a) Using Training data for t_1



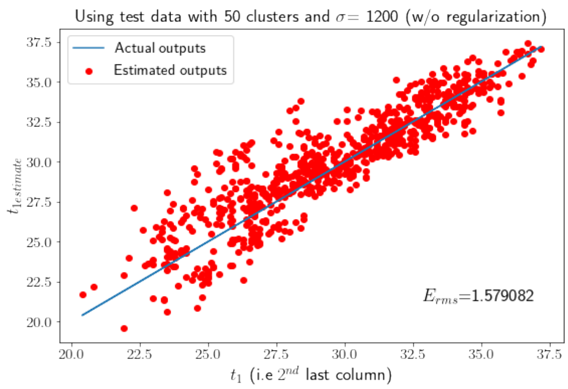
(b) Using Training data for t_2



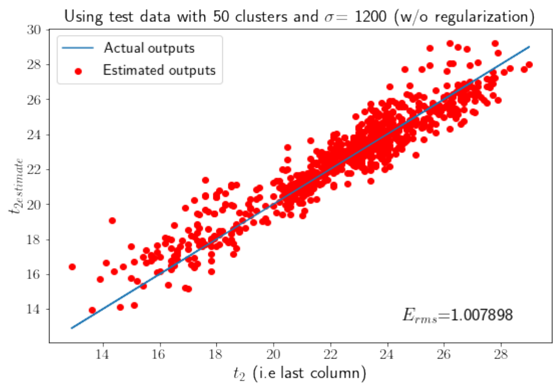
(c) Using Validation data for t_1



(d) Using Validation data for t_2



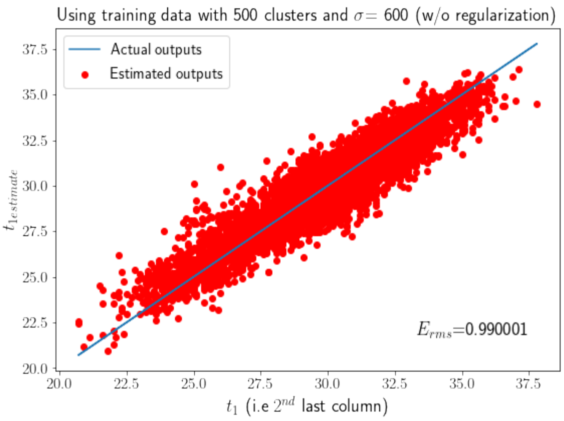
(e) Using Test data for t_1



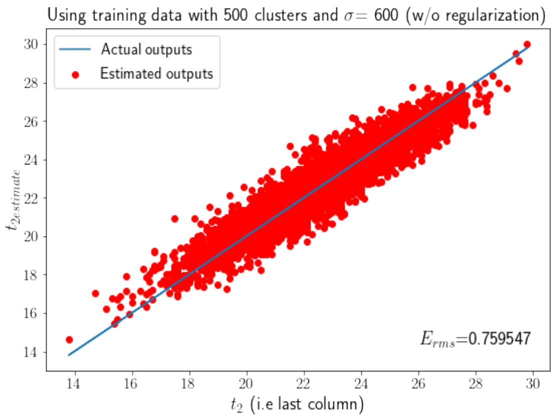
(f) Using Test data for t_2

Figure 17: Plots for 50 clusters and sigma=1200

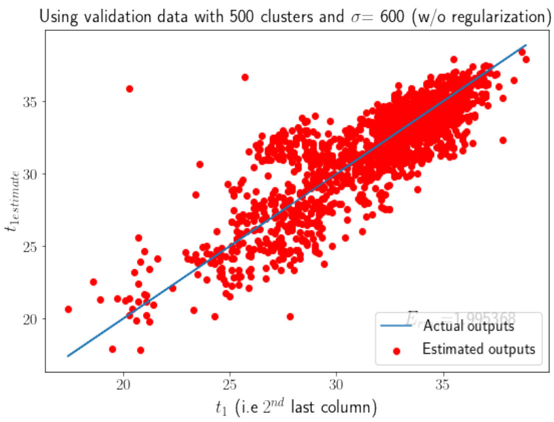
This model seems to be the best model(though being simple) as it generalizes more to the validation and test data.



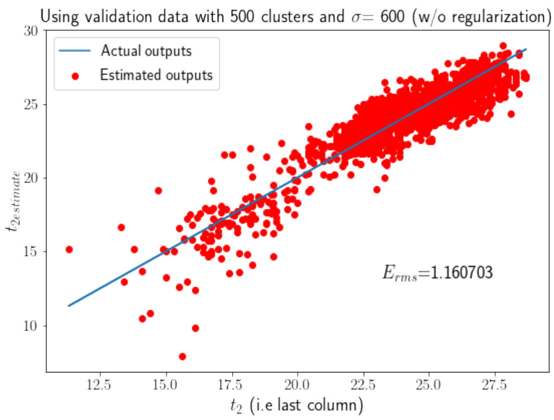
(a) Using Training data for t_1



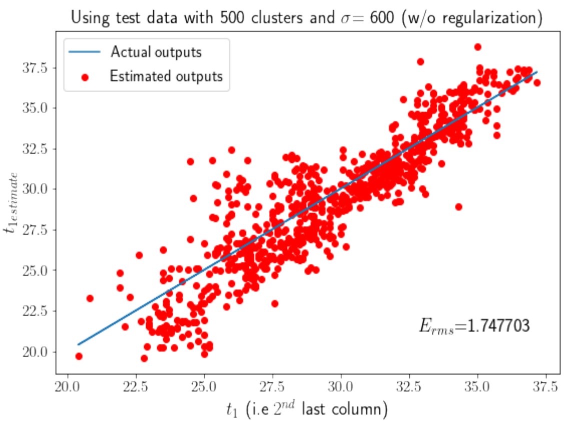
(b) Using Training data for t_2



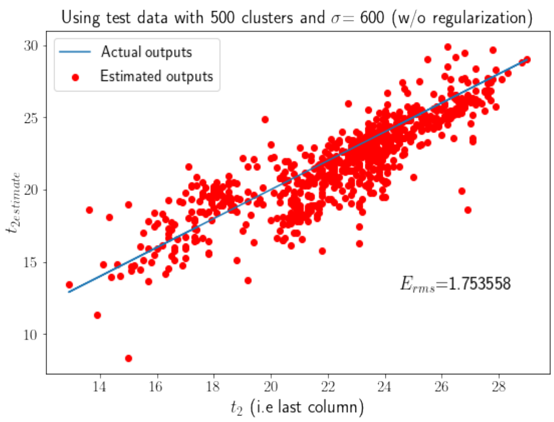
(c) Using Validation data for t_1



(d) Using Validation data for t_2

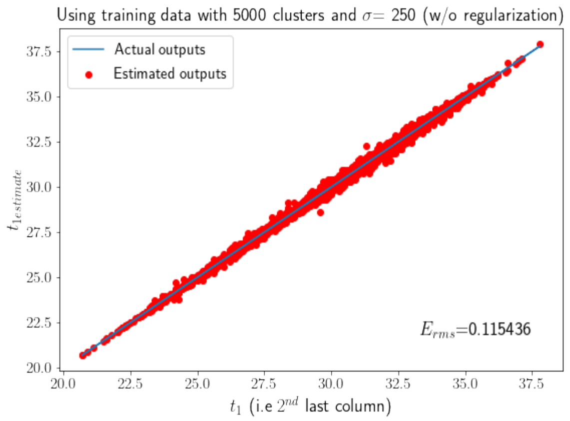


(e) Using Test data for t_1

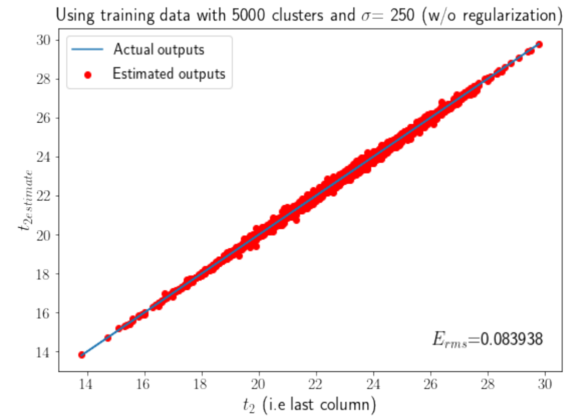


(f) Using Test data for t_2

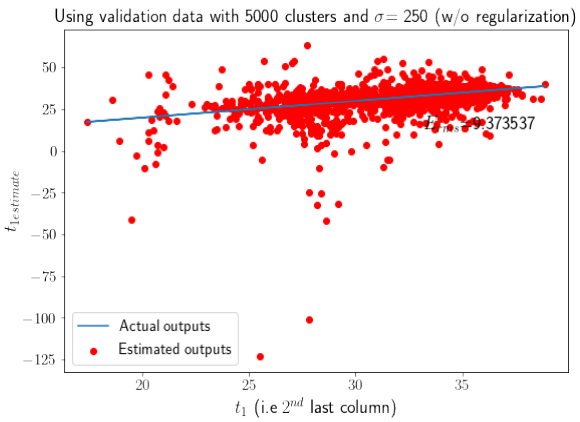
Figure 18: Plots for 500 clusters and sigma=600



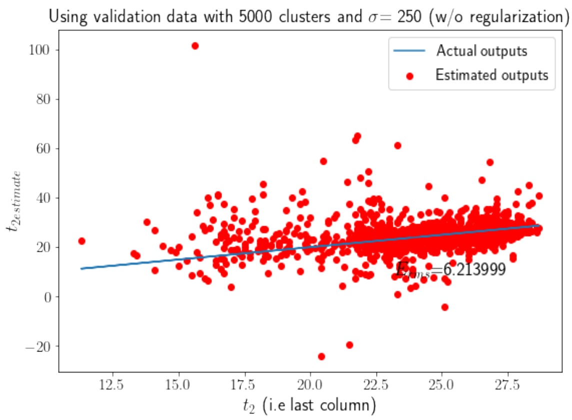
(a) Using Training data for t_1



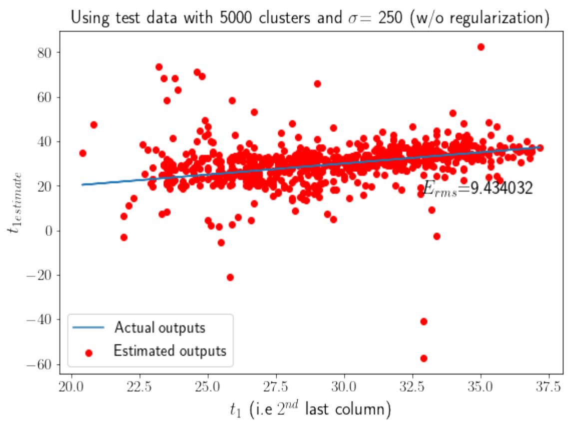
(b) Using Training data for t_2



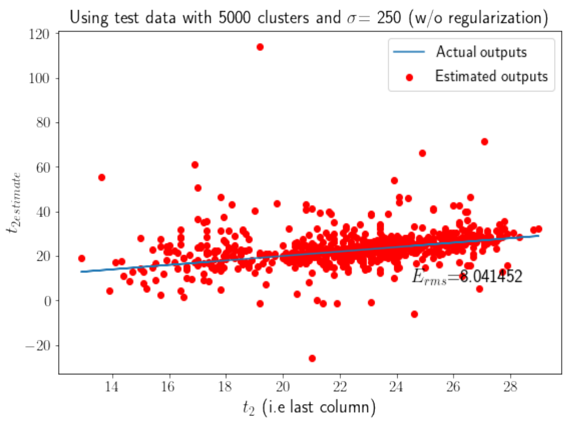
(c) Using Validation data for t_1



(d) Using Validation data for t_2



(e) Using Test data for t_1

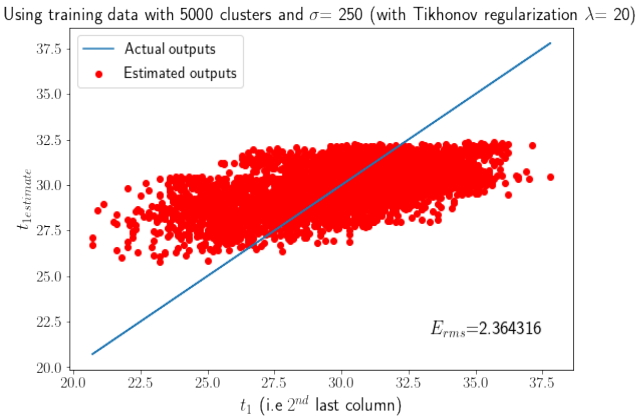


(f) Using Test data for t_2

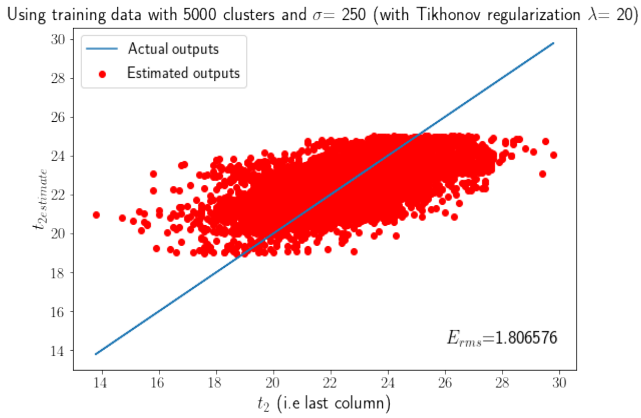
Figure 19: Plots for 5000 clusters and sigma=250

This model gives the minimum error in term of training data but has very high validation and test data error. It overfits and hence regularisation is required

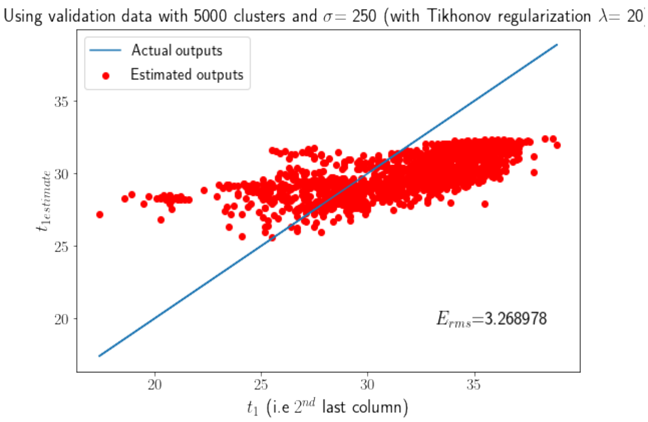
3.2 Scatter plots for 5000 clusters with Tikhonov Regularisation



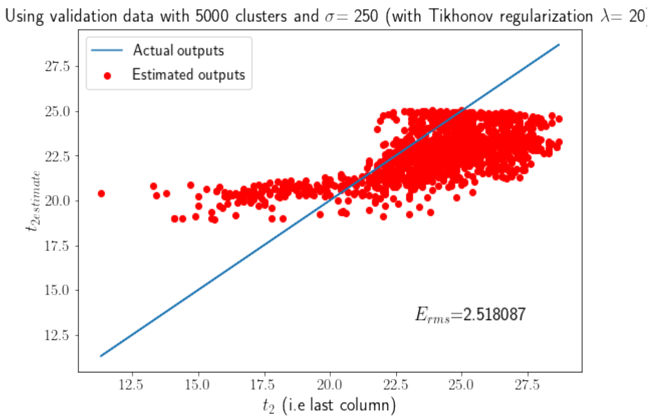
(a) Using Training data for t_1



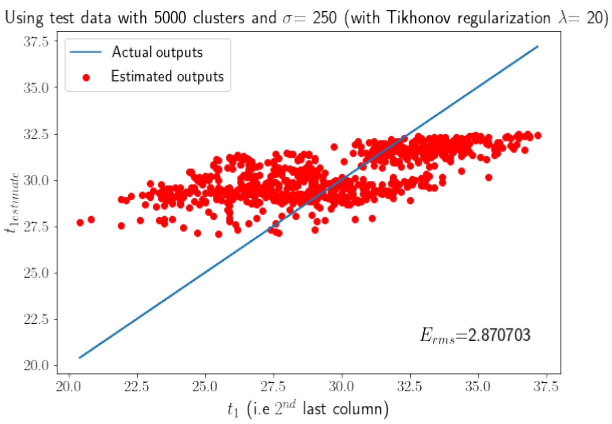
(b) Using Training data for t_2



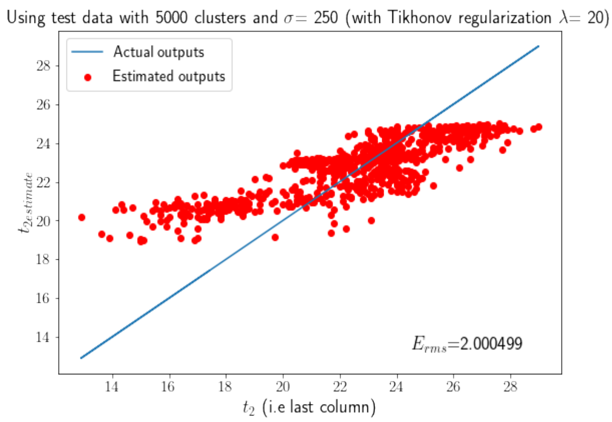
(c) Using Validation data for t_1



(d) Using Validation data for t_2



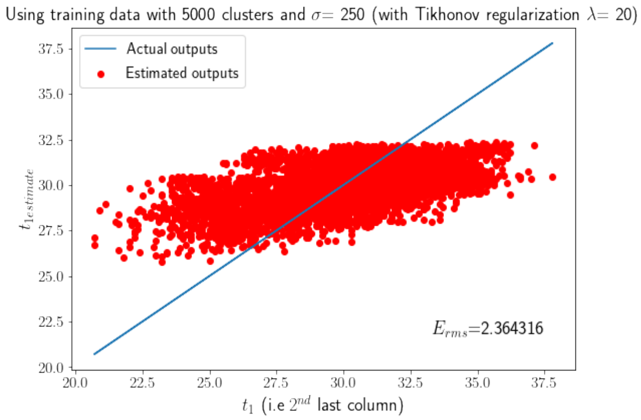
(e) Using Test data for t_1



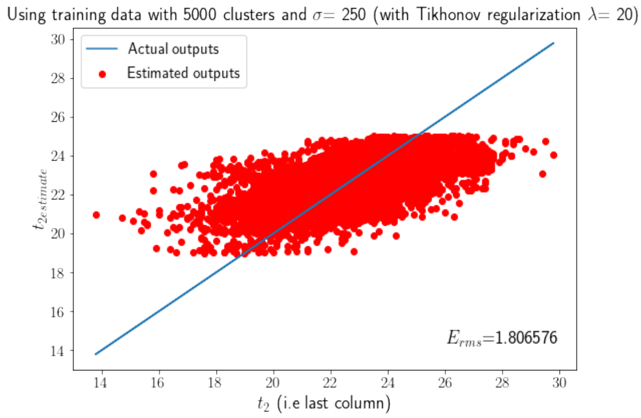
(f) Using Test data for t_2

Figure 20: Plots for 5000 clusters and sigma=250

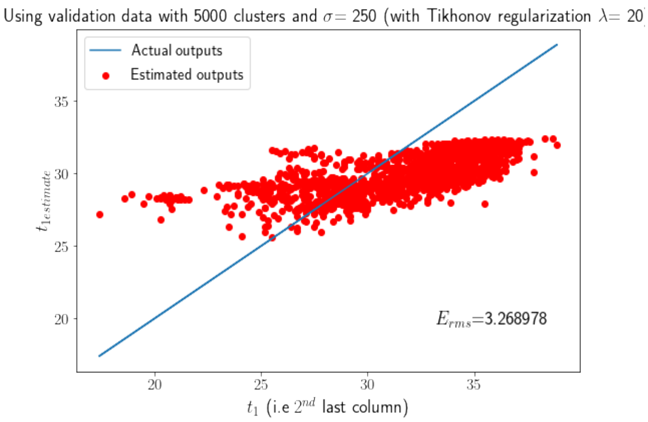
3.3 Scatter plots for 5000 clusters with L2 Regularisation



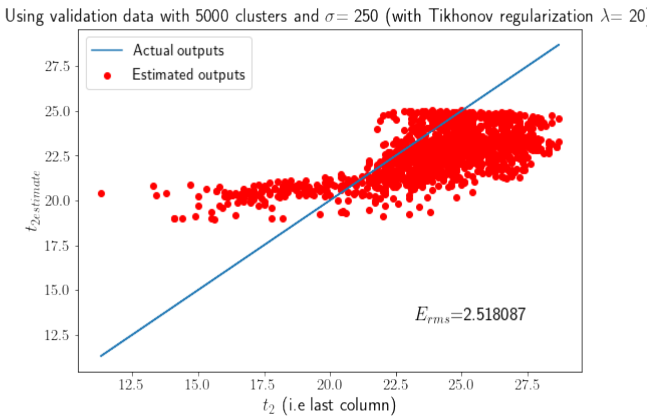
(a) Using Training data for t_1



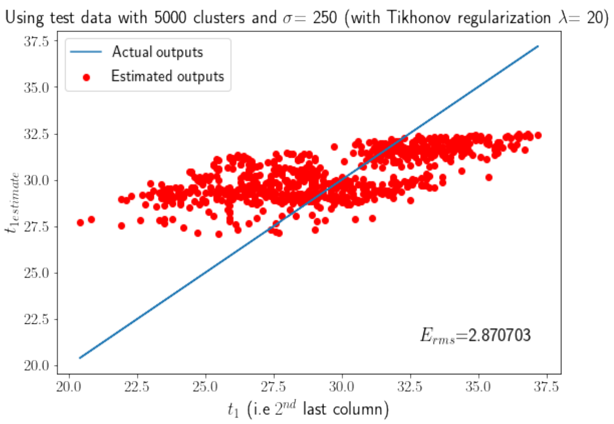
(b) Using Training data for t_2



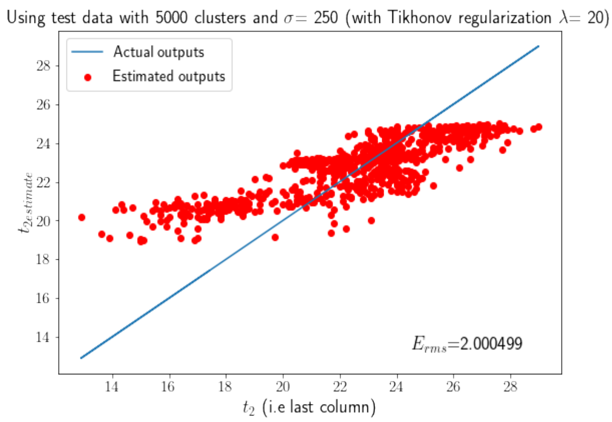
(c) Using Validation data for t_1



(d) Using Validation data for t_2



(e) Using Test data for t_1



(f) Using Test data for t_2

Figure 21: Plots for 5000 clusters and sigma=250

This shows the expected effect of regularisation. It curtails the validation and test error (i.e generalises the model) but at the same time reduces the training error Regularisation for other smaller clusters have not been plotted as they do not overfit the data

3.4 Table showing the errors of different models (clusters) without regularisation

Sr.No.	Clusters	Sigma	Training Data		Validation Data		Test Data	
			Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)
1	20	2000	1.537	1.175	1.861	1.161	1.619	1.070
2	20	3000	1.495	1.151	1.773	1.153	1.525	1.057
3	20	4000	1.489	1.132	1.750	1.140	1.522	1.070
4	50	500	1.570	1.211	2.217	1.615	2.325	1.530
5	50	1000	1.465	1.043	1.720	1.205	1.600	1.070
6	50	1200	1.439	1.022	1.738	1.101	1.579	1.007
7	50	2000	1.393	1.006	1.750	1.030	1.500	0.964
8	200	1000	1.177	0.902	2.225	1.155	1.607	1.013
9	200	2000	1.173	0.896	2.251	1.261	1.566	1.129
10	200	3000	1.173	0.891	2.362	1.214	1.491	1.155
11	500	100	1.320	0.995	2.414	1.794	2.507	1.6458
12	500	300	1.029	0.777	2.325	1.294	1.691	1.631
13	500	600	0.990	0.759	1.995	1.160	1.747	1.753
14	500	1000	0.966	0.744	2.268	1.261	2.129	1.555
15	1000	150	0.866	0.670	2.320	1.327	2.049	1.454
16	1000	200	0.825	0.653	2.491	1.377	2.120	1.487
17	1000	300	0.801	0.635	2.690	1.533	2.751	1.821
18	1000	600	0.784	0.602	2.548	1.927	2.667	2.565

Sr.No.	Clusters	Sigma	Training Data		Validation Data		Test Data	
			Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)
19	1000	1000	0.759	0.583	5.048	2.418	2.998	2.195
20	1000	1200	0.750	0.583	4.685	2.728	2.874	2.316
21	1000	1500	0.796	0.612	2.990	2.991	3.093	2.425
22	3000	25	0.855	0.704	3.033	2.356	3.044	2.345
23	3000	50	0.613	0.457	2.725	1.75	2.598	1.769
24	3000	75	0.517	0.391	2.767	1.740	2.252	1.663
25	3000	100	0.474	0.361	3.071	1.862	2.748	1.843
26	3000	150	0.435	0.333	3.539	2.152	3.186	2.421
27	3000	250	0.412	0.316	3.935	2.707	4.532	2.975
28	3000	450	0.408	0.309	10.878	5.727	6.199	4.209
29	3000	500	0.411	0.309	13.078	7.438	6.198	4.267
30	3000	600	0.440	0.324	11.164	4.783	6.680	4.685
31	5000	50	0.176	0.147	2.892	2.371	2.465	2.213
32	5000	200	0.115	0.084	7.623	6.903	6.467	6.786
33	5000	250	0.115	0.083	9.373	6.213	9.434	8.041
34	5000	500	0.326	0.195	14.864	10.434	13.768	7.363

It can be observed that:

- With increasing number of clusters the training error reduces.However, the validation and test error increases.
- Models having smaller clusters 20, 50, 200, 500 generalise more than those with higher clusters. 3000+ clusters tend to overfit and get more tuned to the errors in the data
- It can be noted that models having clusters around 50 seem to give the best validation and test errors

3.5 Table showing the errors for 5000 clusters with Tikhonov regularisation

Sr.No.	Clusters	Sigma	λ	Training Data		Validation Data		Test Data	
				Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)
1	5000	250	0.01	2394.24	1814.25	2420.63	1835.029	2626.485	1996.74
2	5000	250	2		2.9781	4.056	3.095	2.822	2.144
3	5000	250	10	2.388	1.872	2.888	2.151	3.311	2.340
4	5000	250	30	2.424	1.822	3.400	2.589	2.881	2.038
5	5000	250	50	2.511	1.846	3.490	2.647	2.924	2.074
6	5000	250	100	2.6029	1.871	3.549	2.673	2.950	2.106
7	5000	250	200	2.808	1.998	3.935	2.981	2.988	2.244
8	5000	100	20	2.179	1.711	3.309	2.537	2.874	2.123
9	5000	150	20	2.248	1.754	3.299	2.523	2.862	2.074
10	5000	200	20	2.300	1.779	3.316	2.530	2.875	2.071
11	5000	250	20	2.364	1.806	3.268	2.518	2.870	2.000
12	5000	300	20	2.933	2.226	3.784	3.018	2.952	1.929
13	5000	500	20	273.79	208.28	279.20	212.38	328.61	248.98

3.6 Table showing the errors for 5000 clusters with L2 regularisation

Sr.No.	Clusters	Sigma	λ	Training Data		Validation Data		Test Data	
				Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)	Next-t_{max} (t_1)	Next-t_{min} (t_2)
1	5000	250	0.01	1.499	1.059	2.121	1.566	2.333	1.337
2	5000	250	0.1	1.694	1.250	2.448	1.740	2.598	1.599
3	5000	250	1	1.962	1.584	2.902	2.204	2.688	1.905
4	5000	250	10	2.127	1.774	3.143	2.466	2.745	2.097
5	5000	250	20	2.200	1.822	3.220	2.526	2.828	2.148
6	5000	250	50	2.333	1.898	3.336	2.610	3.015	2.232

4 Conclusion

We have shown that regularisation does reduce the validation and test error of the overfitted model having 5000 clusters but at the cost of increasing the training error.
And for this real-world dataset both regularisation perform similarly

But in the final analysis simple model of 50 clusters gives a better result overall and genralises better