Indian Institute Of Technology Madras

CS5691 - Pattern Recognition and Machine Learning

Assignment 1 Report

Team 5

Jeswant Krishna K - AE17B030

Advait Walsangikar - AE17B101

Ananthu Nair S. - AE17B109

19th March 2021

Contents

1 Task 1

1.1 Objective

To fit a polynomial function for the univariate case for different sized subsets of a given dataset. The model complexities, and the regularization constants are variables which are to be changed and cross-validation to be performed so as to obtain the best performing model. The predictions of each models are to be plotted and inferences drawn.

1.2 Method

Dataset 1 (function 1) is a Univariate input data which is used here. In this task we are going to perform Polynomial regression on the given 1D input data matrix. The degree is also taken as input and based on it the feature matrix is constructed i.e. the new features are power of the input column vector from zero until the given degree which is appended together to create the feature matrix. We say that the model is fitted/learnt if the parameters of the model are found and here it is done using Normal equation.

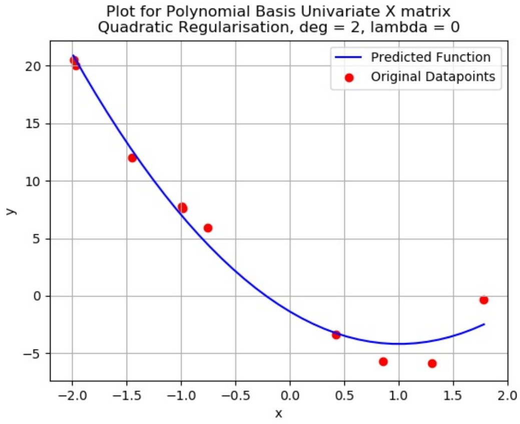
But along with regression we are also performing regularisation of the parameters which are being learnt i.e. in regularization the normal equation is modified to give penalty for higher values of the parameter and it is added to the before no regularisation normal equation after multiplying it by some λ which is called regularisation hyperparameter.

Since there are multiple ways to give the penalty to the higher values of the parameters there are different types of regularisation methods which were also implemented like Quadratic regularisation or Tikhonov regularisation. But mainly in this task it is either no regularisation($\lambda = 0$) or Quadratic regularisation which was analysed.

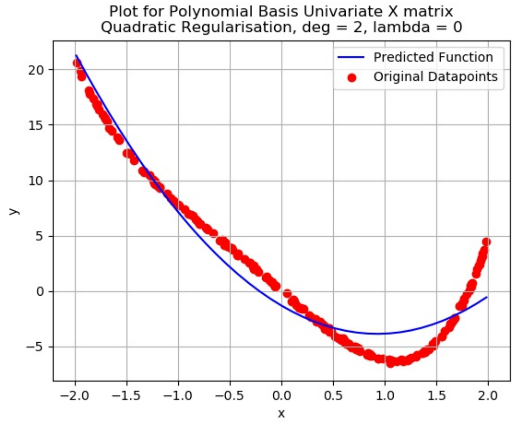
Training of Model

- For training the dataset was split as (80-20%) Train-Test splitting %.
- The Train part of the split used to train using KFold cross validation where K=3.
- The best model from the KFold cross validation is chosen based on the lowest validation error in cross validation.
- The chosen model from above is used to predict the values for test data inputted.
- Finally the Erms Error is computed for chosen model on Test dataset using predicted output values and true output values.

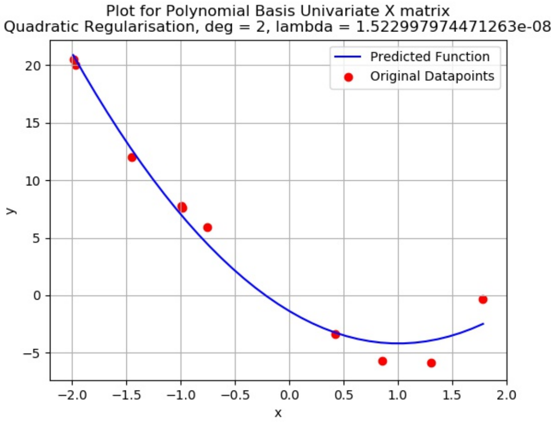
1.3 Plots



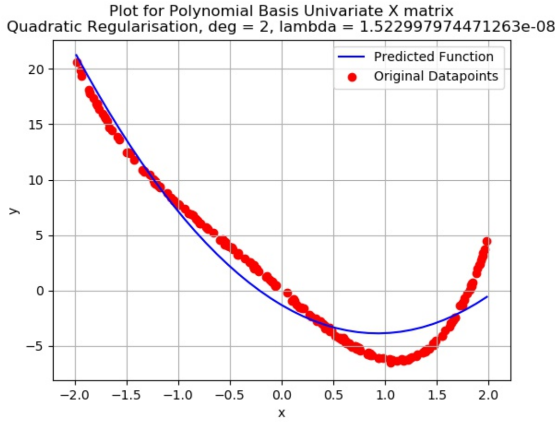
(a) No regularisation. Dataset of size 10.



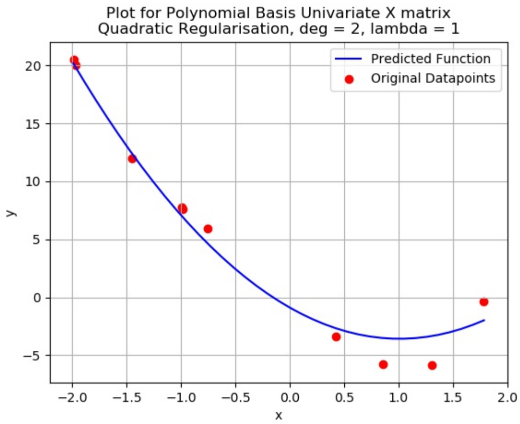
(b) No regularisation. Dataset of size 200.



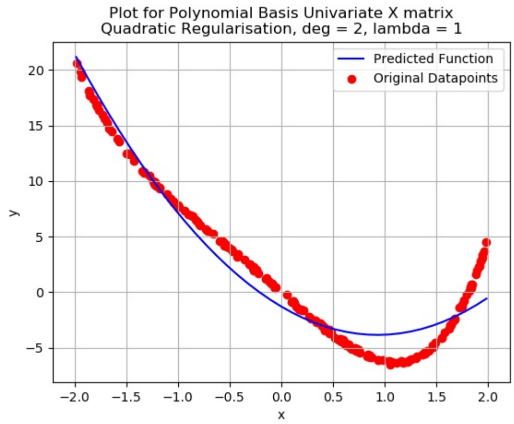
(c) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 10.



(d) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 200.

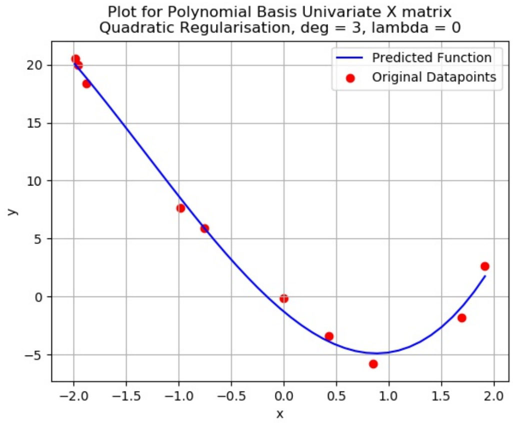


(e) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 10.

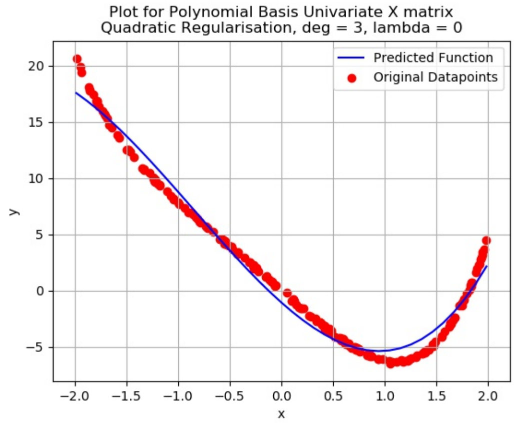


(f) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 200.

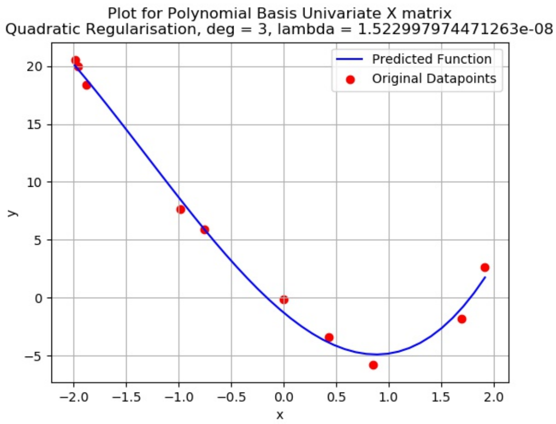
Figure 1: Plot for polynomial fitting for univariate data using polynomial basis functions of degree 2. (With and without regularisation for different sized data sets)



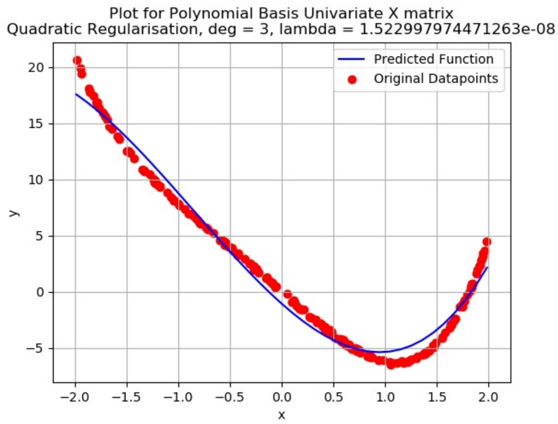
(a) No regularisation. Dataset of size 10.



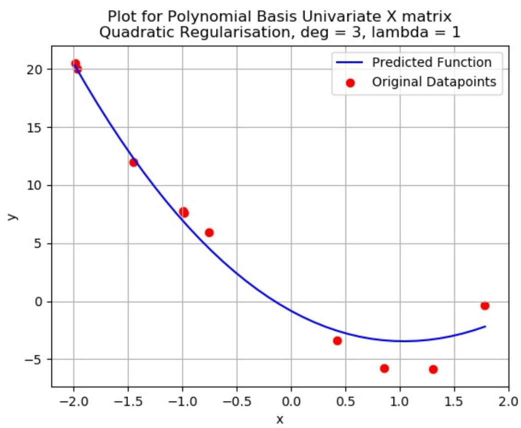
(b) No regularisation. Dataset of size 200.



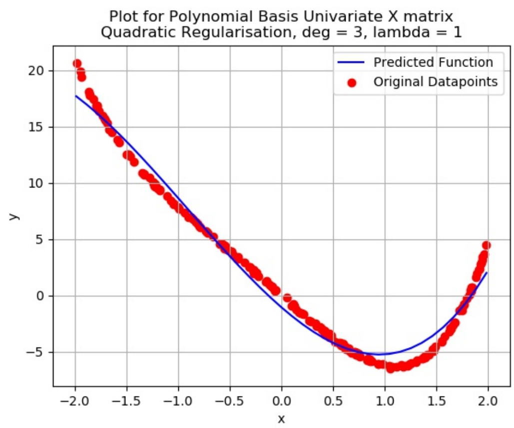
(c) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 10.



(d) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 200.

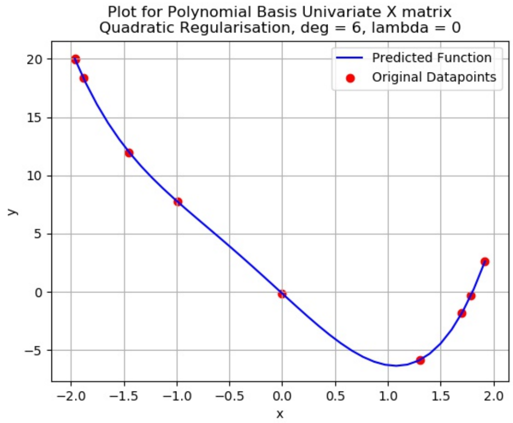


(e) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 10.

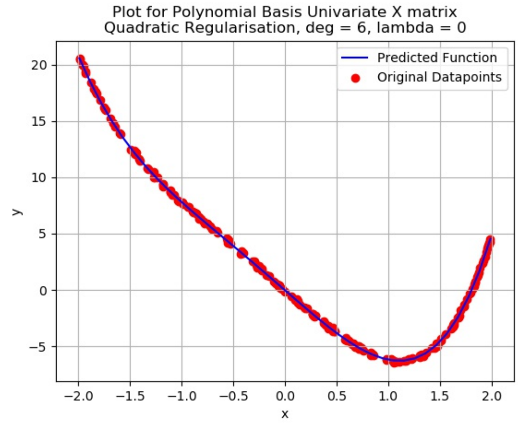


(f) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 200.

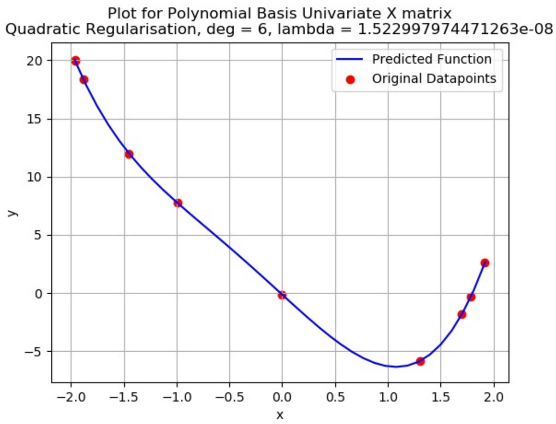
Figure 2: Plot for polynomial fitting for univariate data using polynomial basis functions of degree 3. (With and without regularisation for different sized data sets)



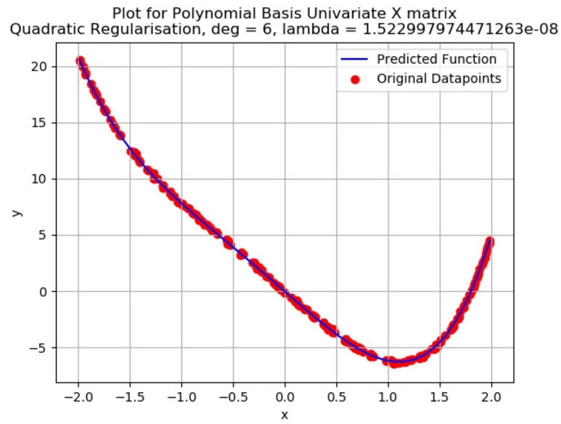
(a) No regularisation. Dataset of size 10.



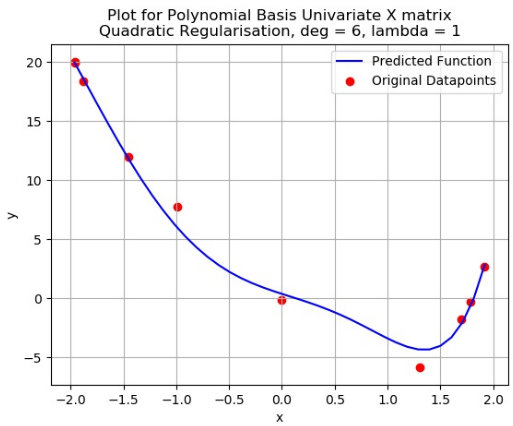
(b) No regularisation. Dataset of size 200.



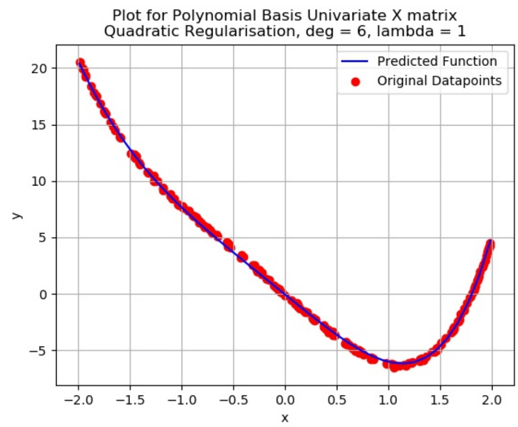
(c) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 10.



(d) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 200.

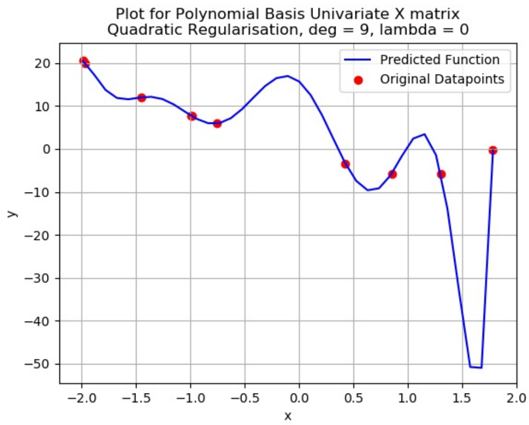


(e) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 10.

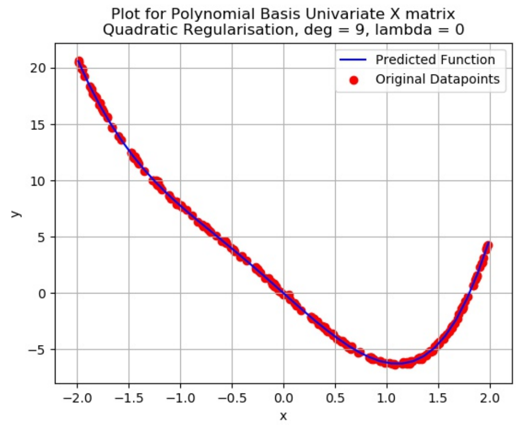


(f) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 200.

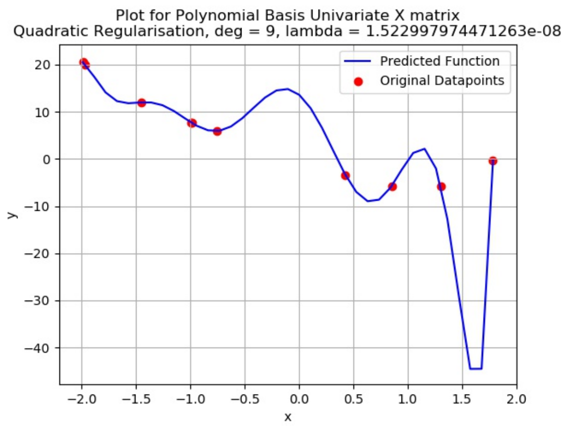
Figure 3: Plot for polynomial fitting for univariate data using polynomial basis functions of degree 6. (With and without regularisation for different sized data sets)



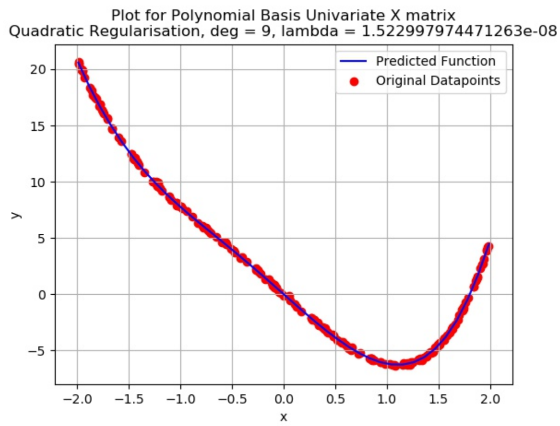
(a) No regularisation. Dataset of size 10.



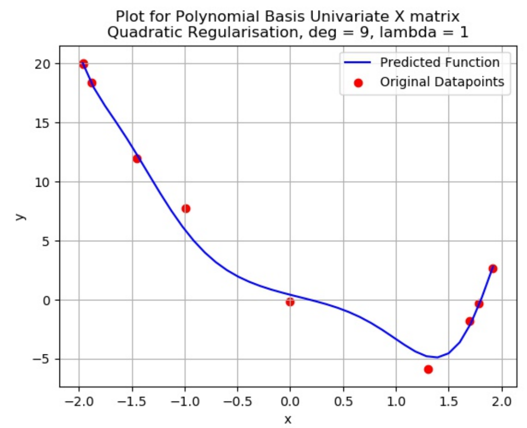
(b) No regularisation. Dataset of size 200.



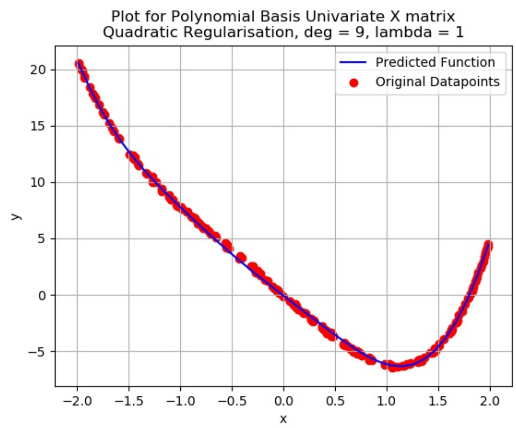
(c) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 10.



(d) Polynomial regularisation with regularisation constant, $ln\lambda = -18$. Dataset of size 200.



(e) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 10.



(f) Polynomial regularisation with regularisation constant, $\lambda = 1$. Dataset of size 200.

Figure 4: Plot for polynomial fitting for univariate data using polynomial basis functions of degree 9. (With and without regularisation for different sized data sets)

1.4 Inferences

Size	Degree	Lambda	Training Error	Validation Error	Test Error
10	2	0	1.223	1.734	1.513
10	2	e^{-18}	1.223	1.734	1.513
10	2	1	1.317	1.764	2.061
10	3	0	0.718	1.096	1.561
10	3	e^{-18}	0.718	1.096	1.561
10	3	1	1.398	1.922	2.13
10	6	0	0.017	0.094	0.23
10	6	e^{-18}	0.017	0.094	0.23
10	6	1	0.77	1.204	2.293
10	9	0	0.0	35.47	4.658
10	9	e^{-18}	0.002	30.762	4.028
10	9	1	0.719	1.229	2.407
200	2	0	1.624	1.678	1.845
200	2	e^{-18}	1.624	1.678	1.845
200	2	1	1.624	1.678	1.847
200	3	0	0.94	1.039	0.946
200	3	e^{-18}	0.94	1.039	0.946
200	3	1	0.944	1.042	0.972
200	6	0	0.093	0.103	0.091
200	6	e^{-18}	0.093	0.103	0.091
200	6	1	0.174	0.183	0.181
200	9	0	0.091	0.106	0.098
200	9	e^{-18}	0.091	0.106	0.098
200	9	1	0.176	0.187	0.185

Table 1: Table showing the E_{RMS} on the training data, validation data and the test data for different models

As can be observed from the plots and the tabular data, without any regularisation, when the model gets too complex for the given data, it tends to overfit the data, i.e, the training error keeps lowering while the validation error keeps increasing. This problem is seen to go away with a higher value of the regularisation constant using a quadratic regularisation (ridge regression). We choose the best performing model by looking at the lowest value of the validation rms error.

2 Task 2

2.1 Objective

To fit a polynomial function for a bivariate case for different sized subsets of a given dataset. The model complexities, and the regularization constants are variables which are to be changed and cross-validaion to be performed so as to obtain the best performing model. The predictions of each models are to be plotted and inferences drawn.

2.2 Method

Dataset 2 (function1 2d) which is a Bivariate input data is used. In this task we are going to perform regression using Polynomial basis functions on the given 2D input data matrix. The degree is also taken as input and based on it the feature matrix is constructed in here the basis functions are constructed as combinations of powers of the 2 column vectors in the input where the sum of their powers is equal to the given input degree. And the model is fitted on this feature constructed matrix.

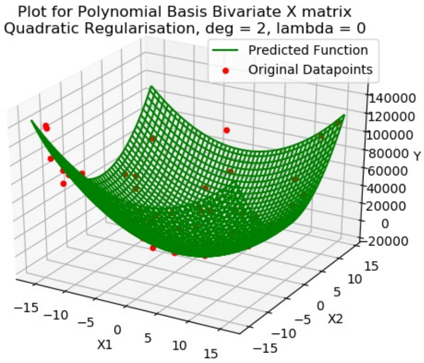
Again here also we are performing regularisation of the parameters which are being learnt and of the regularisation methods we specified before here Quadratic regularisation was applied by modifying the normal equation.

Here the training procedure of the model is same as that of Task 1 case.

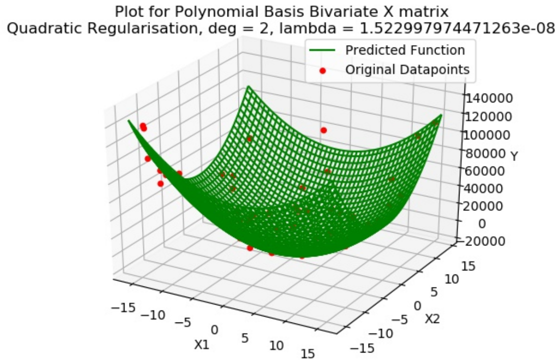
Training of Model

- For training the dataset was split as (80-20%) Train-Test splitting %.
- The Train part of the split used to train using KFold cross validation where K=3.
- The best model from the KFold cross validation is chosen based on the lowest validation error in cross validation.
- The chosen model from above is used to predict the values for test data inputted.
- Finally the Erms Error is computed for chosen model on Test dataset using predicted output values and true output values.

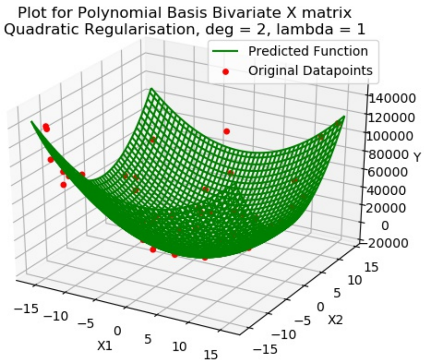
2.3 Plots



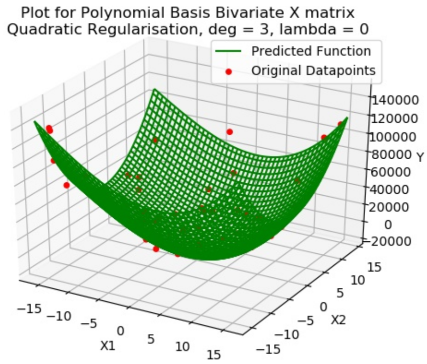
(a) Degree of the polynomial basis = 2. No regularisation.



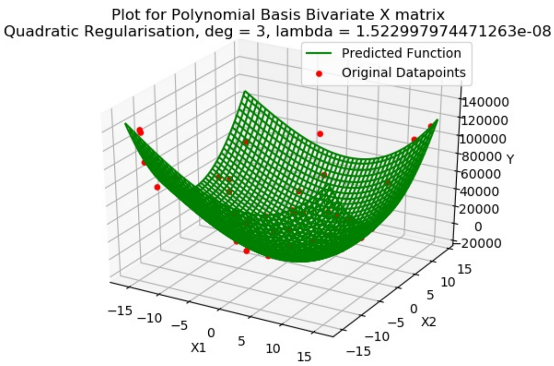
(b) Degree of the polynomial basis = 2. Regularisation constant, $ln\lambda = -18$



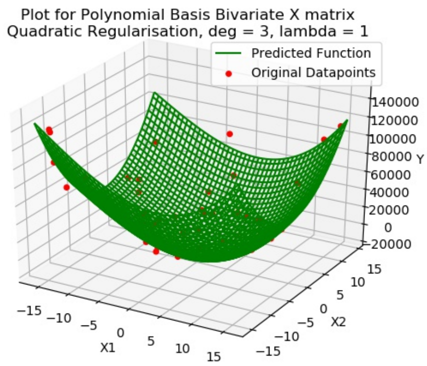
(c) Degree of the polynomial basis = 2. Regularisation constant, $\lambda = 1$



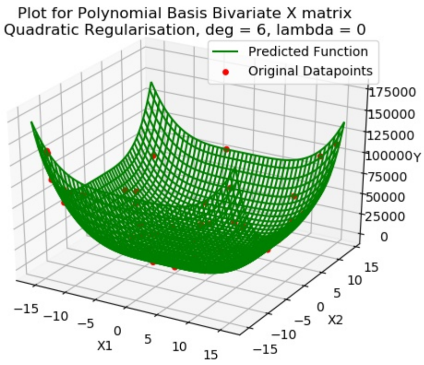
(d) Degree of the polynomial basis = 3. No regularisation.



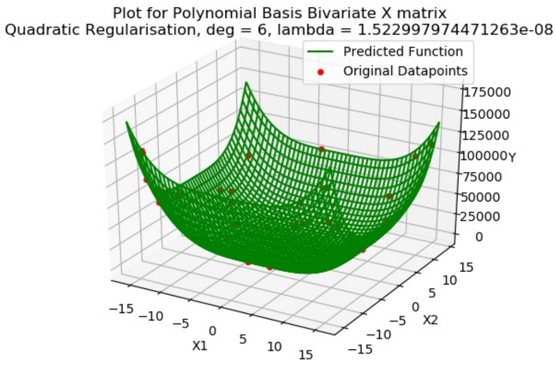
(e) Degree of the polynomial basis = 3. Regularisation constant, $ln\lambda = -18$



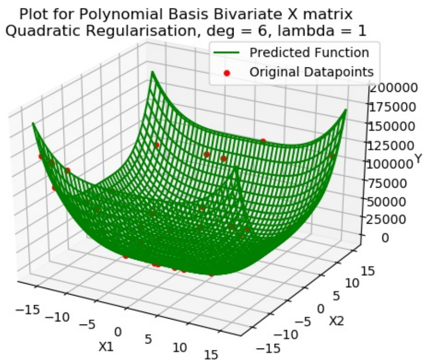
(f) Degree of the polynomial basis = 3. Regularisation constant, $\lambda = 1$



(g) Degree of the polynomial basis = 6. No regularisation.

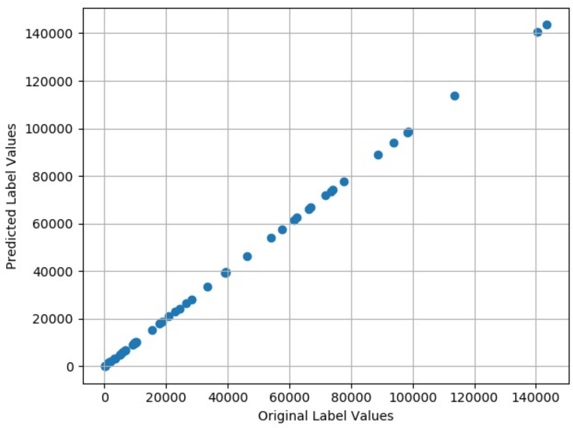


(h) Degree of the polynomial basis = 6. Regularisation constant, $ln\lambda = -18$

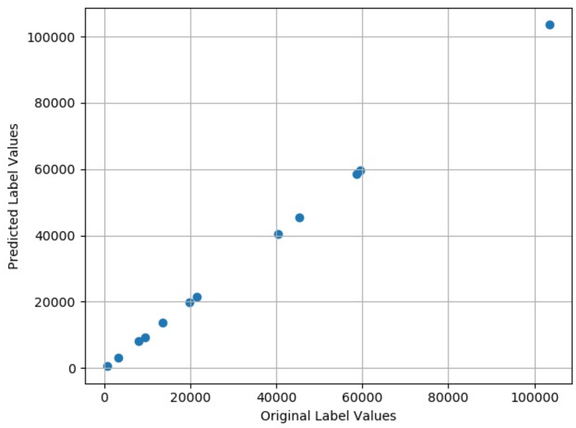


(i) Degree of the polynomial basis = 6. Regularisation constant, $\lambda = 1$

Figure 5: Plot for a bivariate data fit and predicted using polynomial basis functions, with a data set of size 50. (For different regularisation and degree)

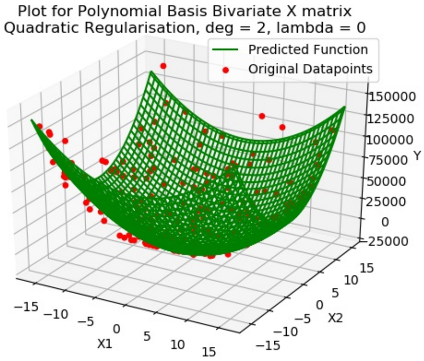


(a) Target output is obtained using the training data.

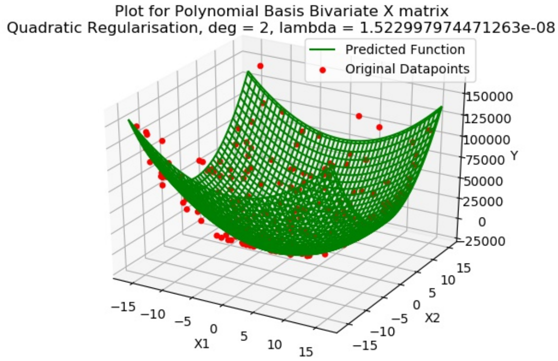


(b) Target output is obtained using the test data.

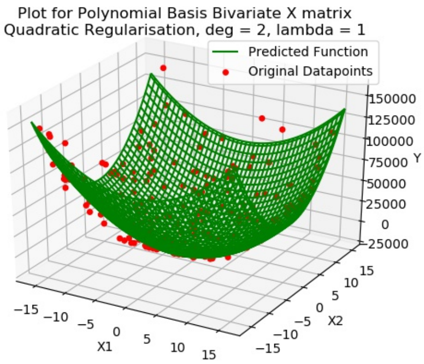
Figure 6: A scatter plot showing the trends of the target output vs the model output for the best performing model trained on a dataset size of 50.



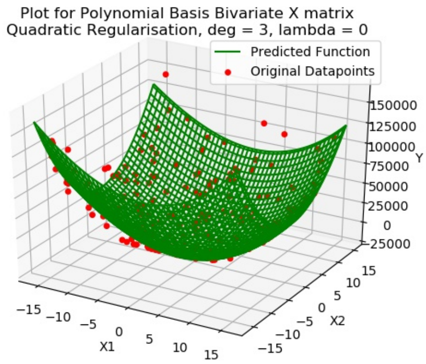
(a) Degree of the polynomial basis = 2. No regularisation.



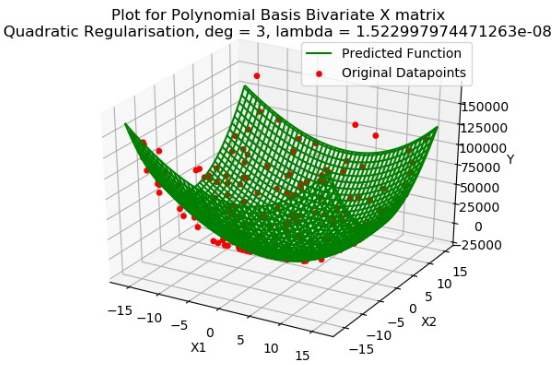
(b) Degree of the polynomial basis = 2. Regularisation constant, $ln\lambda = -18$



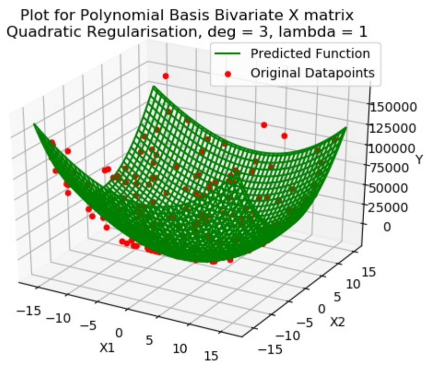
(c) Degree of the polynomial basis = 2. Regularisation constant, $\lambda = 1$



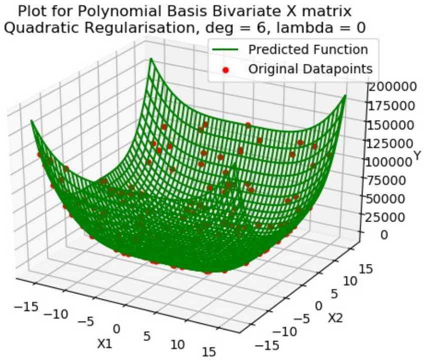
(d) Degree of the polynomial basis = 3. No regularisation.



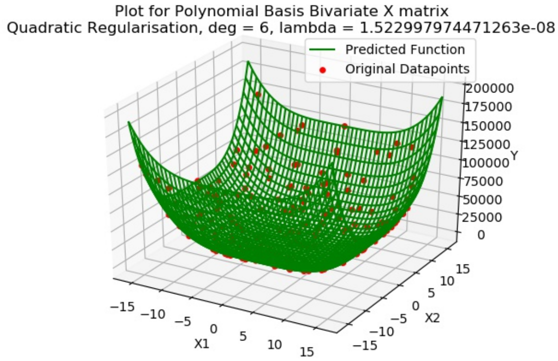
(e) Degree of the polynomial basis = 3. Regularisation constant, $ln\lambda = -18$



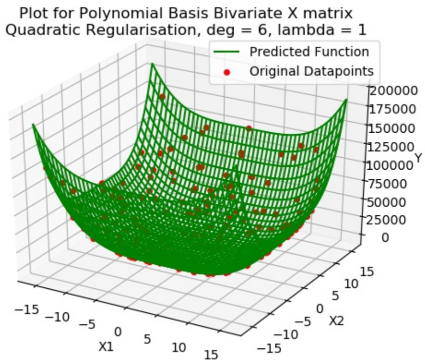
(f) Degree of the polynomial basis = 3. Regularisation constant, $\lambda = 1$



(g) Degree of the polynomial basis = 6. No regularisation.

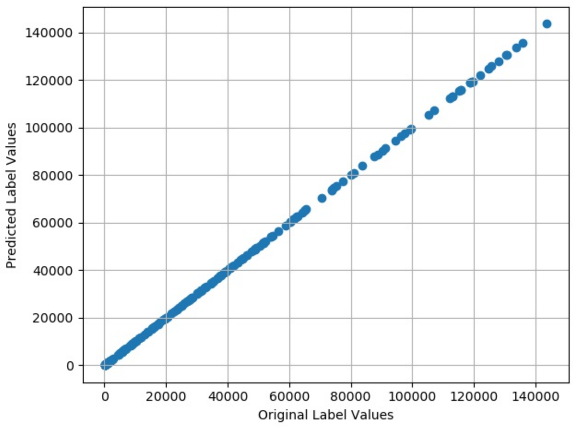


(h) Degree of the polynomial basis = 6. Regularisation constant, $ln\lambda = -18$

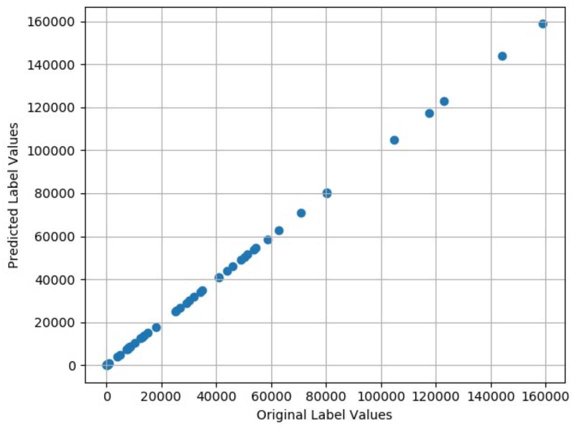


(i) Degree of the polynomial basis = 6. Regularisation constant, $\lambda = 1$

Figure 7: Plot for a bivariate data fit and predicted using polynomial basis functions, with a data set of size 200. (For different regularisation and degree)

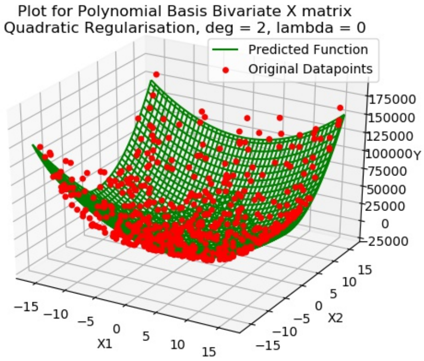


(a) Target output is obtained using the training data.

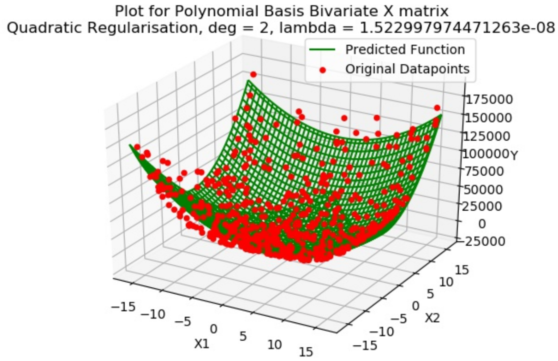


(b) Target output is obtained using the test data.

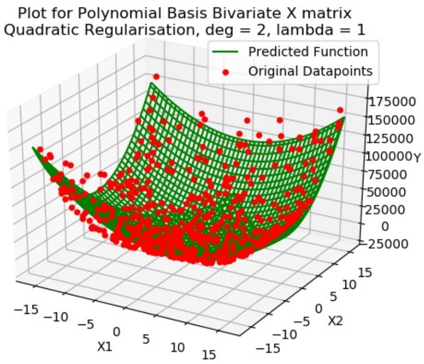
Figure 8: A scatter plot showing the trends of the target output vs the model output for the best performing model trained on a dataset size of 200.



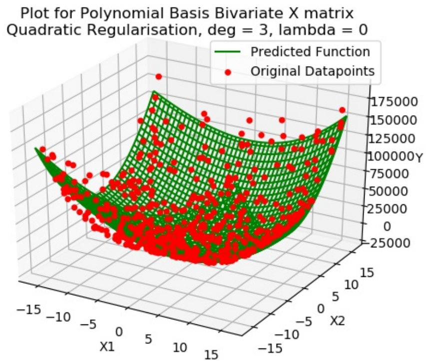
(a) Degree of the polynomial basis = 2. No regularisation.



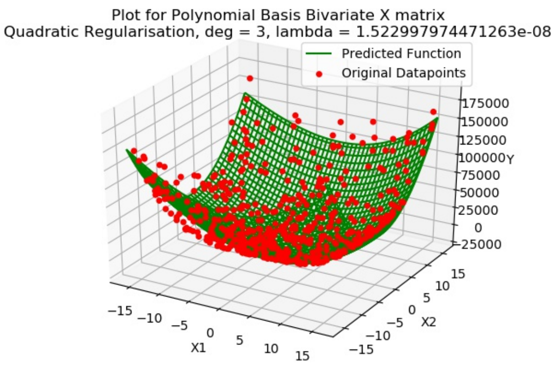
(b) Degree of the polynomial basis = 2. Regularisation constant, $ln\lambda = -18$



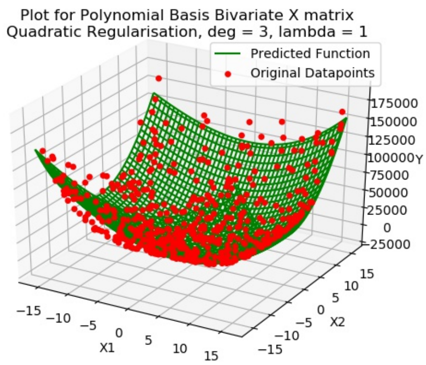
(c) Degree of the polynomial basis = 2. Regularisation constant, $\lambda = 1$



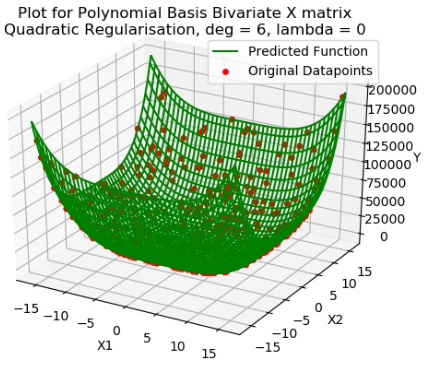
(d) Degree of the polynomial basis = 3. No regularisation.



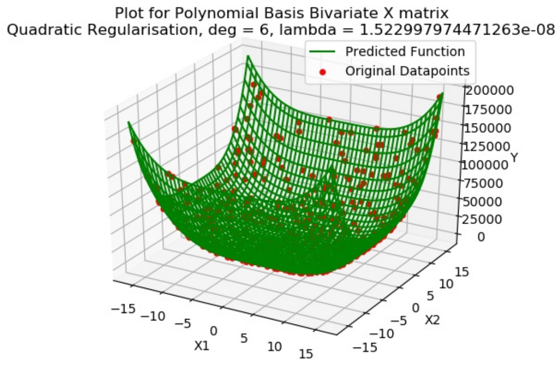
(e) Degree of the polynomial basis = 3. Regularisation constant, $ln\lambda = -18$



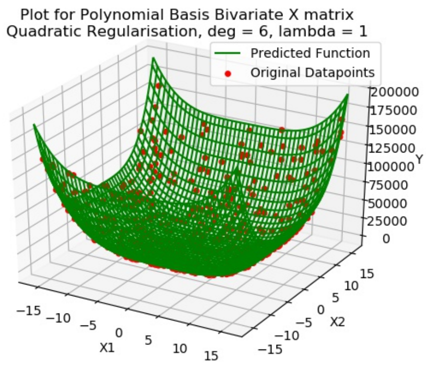
(f) Degree of the polynomial basis = 3. Regularisation constant, $\lambda = 1$



(g) Degree of the polynomial basis = 6. No regularisation.

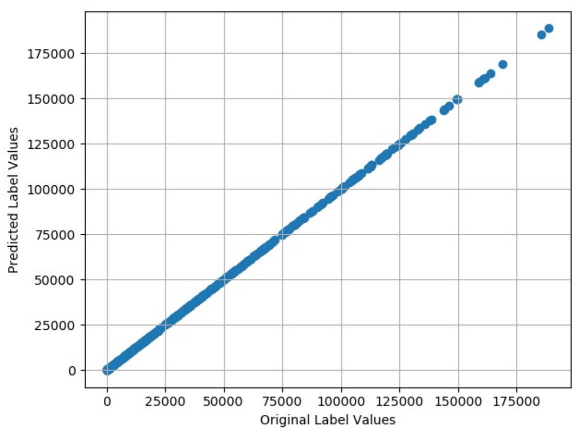


(h) Degree of the polynomial basis = 6. Regularisation constant, $ln\lambda = -18$

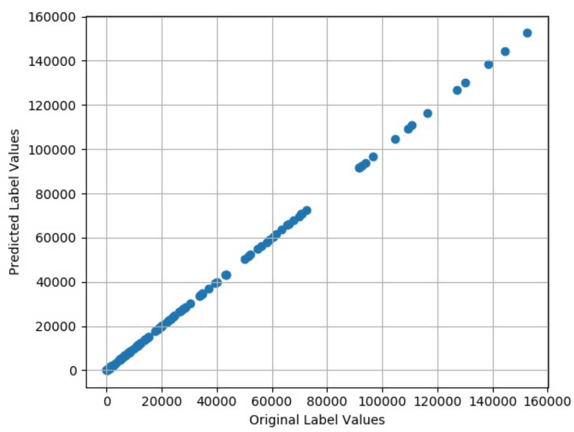


(i) Degree of the polynomial basis = 6. Regularisation constant, $\lambda = 1$

Figure 9: Plot for a bivariate data fit and predicted using polynomial basis functions, with a data set of size 500. (For different regularisation and degree)



(a) Target output is obtained using the training data.



(b) Target output is obtained using the test data.

Figure 10: A scatter plot showing the trends of the target output vs the model output for the best performing model trained on a dataset size of 500.

2.4 Inferences

Size	Degree	Lambda	Training Error	Validation Error	Test Error
50	2	0	7974.865	10345.677	9599.846
50	2	e^{-18}	7974.865	10345.677	9599.846
50	2	1	7990.337	10376.375	9810.157
50	3	0	7129.033	10390.529	10599.068
50	3	e^{-18}	7129.033	10390.529	10599.068
50	3	1	7146.318	10310.328	10794.562
50	6	0	0.0	0.0	0.0
50	6	e^{-18}	0.0	0.0	0.0
50	6	1	0.0	0.001	0.0
200	2	0	10215.104	11235.906	13104.006
200	2	e^{-18}	10215.104	11235.906	13104.006
200	2	1	10216.331	11241.444	13095.701
200	3	0	9837.442	11249.388	12819.288
200	3	e^{-18}	9837.442	11249.388	12819.288
200	3	1	9838.644	11256.76	12813.379
200	6	0	0.0	0.0	0.0
200	6	e^{-18}	0.0	0.0	0.0
200	6	1	0.0	0.0	0.0
500	2	0	10954.365	11442.56	10122.637
500	2	e^{-18}	10954.365	11442.56	10122.637
500	2	1	10954.521	11443.483	10120.853
500	3	0	10734.615	11310.072	10252.917
500	3	e^{-18}	10734.615	11310.072	10252.917
500	3	1	10734.774	11311.525	10250.64
500	6	0	0.0	0.0	0.0
500	6	e^{-18}	0.0	0.0	0.0
500	6	1	0.0	0.0	0.0

Table 2: Table showing the E_{RMS} on the training data, validation data and the test data for different models

From the plots and the tables above, we see that a degree 6 polynomial basis function with a little regularization performs the best (lowest validation error) irrespective of the training dataset sizes (50, 200, 500). While the degree 2 and degree 3 basis functions are underfitting the data.

3 Task 3 Dataset 2

3.1 Objective

To approximate a dependent variable (depending on bivariate input) using Gaussian basis functions and Tikhonov regularization.

3.2 Method

Dataset 2 is used again. In this task we are going to perform regression on Gaussian basis functions on the given 2D input data matrix. Here the no of basis functions is taken as input and again based on it the feature matrix is constructed. The i^{th} basis is constructed from the Gaussian function based on μ_i and σ. But the μ_i is found using the KMeans clustering and σ was also derived from it which is the RMS of the average distance of each data point from its respective cluster centre. And then model is fitted on this feature constructed matrix.

Again here also we are performing regularisation of the parameters which are being learnt and of the regularisation methods we specified before here Quadratic regularisation was applied by modifying the normal equation. Also the model was also trained using the Tikhonov regularisation which was also implemented.

Training of Model

- For training the dataset was split as (70-20-10%) Train-Validation-Test splitting %.
- The Train part of the split was used to train without using the cross-validation.
- Then the trained model was used for validation and this validation error was used for tuning of the hyperparameters i.e. no. of basis function and λ in regulariser.
- Then the tuned model from above is used to predict the values for test data inputted.
- Finally the Erms Error is computed for chosen model on Test dataset using predicted output values and true output values.

3.3 Plots - Best performing model

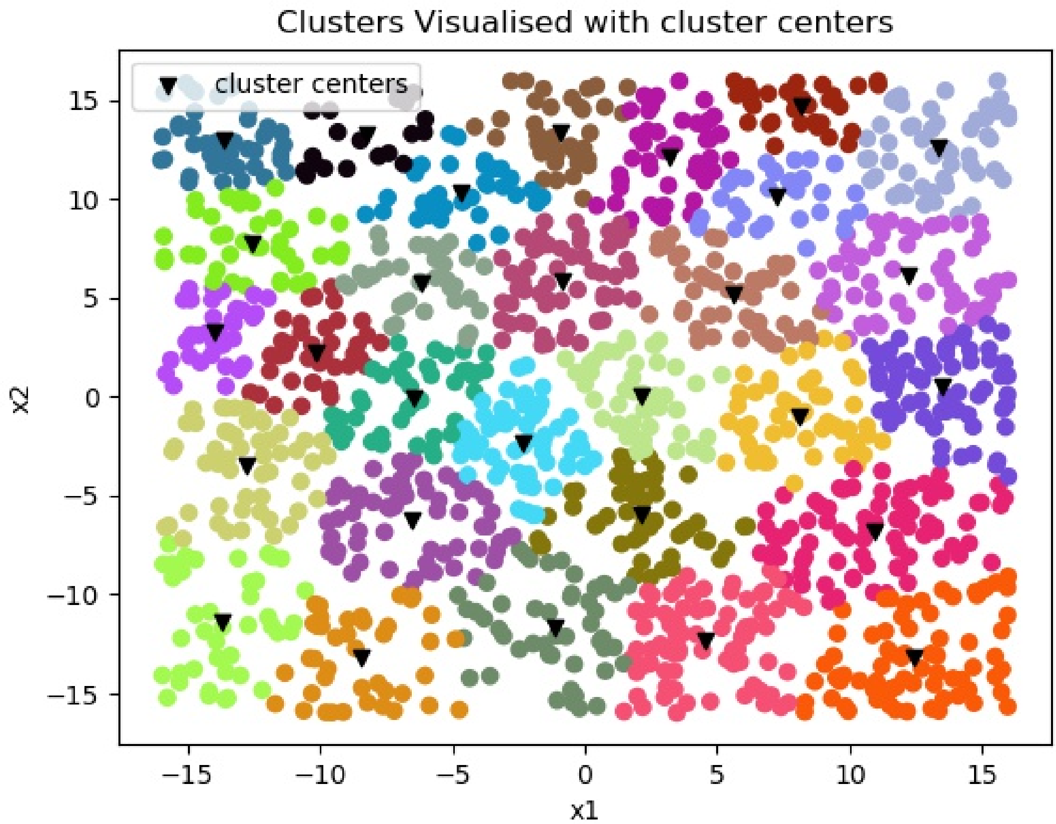


Figure 11: Visualizing cluster quality for 30 clusters - Dataset 2

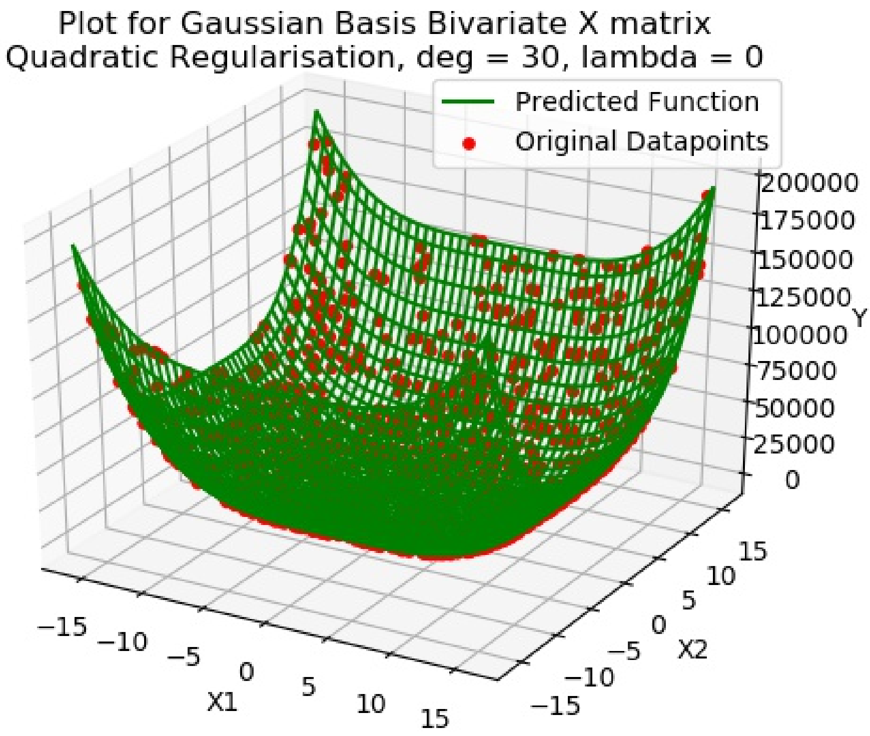
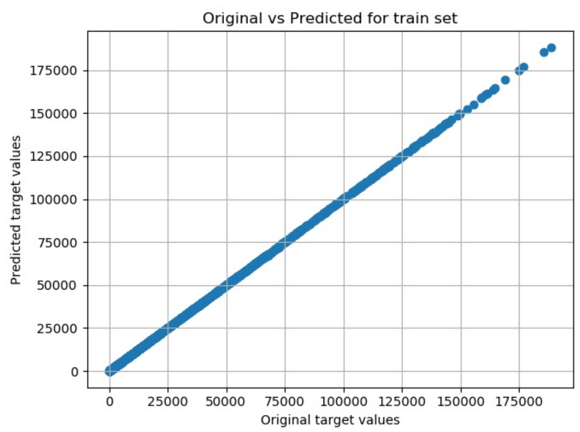
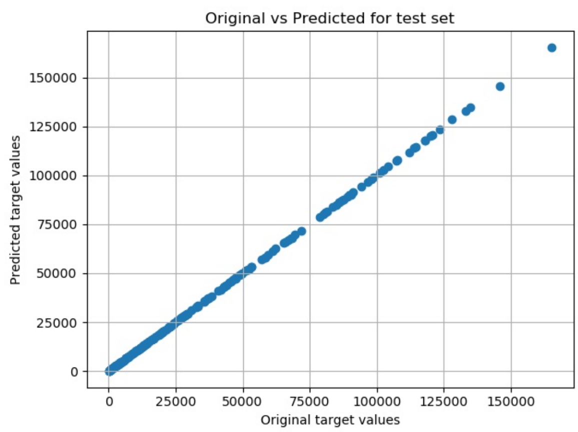


Figure 12: Bivariate Input data. Target variable predicted using 30 Gaussian basis functions. No regularization.



(a) Target output is obtained using the training data.



(b) Target output is obtained using the test data.

Figure 13: Original vs Predicted value for target variable for the best performing model on the whole dataset

3.4 Inferences

Regularisation	Degree	Lambda	Training Error	Validation Error	Test Error
None	30	0	95.640	108.395	94.046
Tikhonov	30	e^{-18}	106.209	122.841	105.366

Table 3: Table showing the E_{RMS} on the training data, validation data and the test data for different models

We find that the calculated approximation to the function surface (with 30 basis functions and no regularization) is quite good as can be seen from Figure 13 which is an approximately 45 degree line.

4 Task 3 Dataset 3

4.1 Objective

To predict two dependent variables which depend on 68 independent variables taken from a real world dataset using Gaussian basis functions and Tikhonov and Quadratic regularization methods.

4.2 Method

Dataset 3 (music.txt) is used which is Multivariate real world data with 68 independent variables and 2 dependent/target variables. In this task again we are going to perform regression on Gaussian basis functions on the given N-Dimensional input data matrix. Similar to before here the no of basis functions is taken as input and again based on it the feature matrix is constructed. The i^{th} basis is constructed from the Gaussian function based on μ_i and σ. But the μ_i is found using the KMeans clustering and σ was also derived from it which is the RMS of the average distance of each data point from its respective cluster centre. And then model is fitted on this feature constructed matrix.

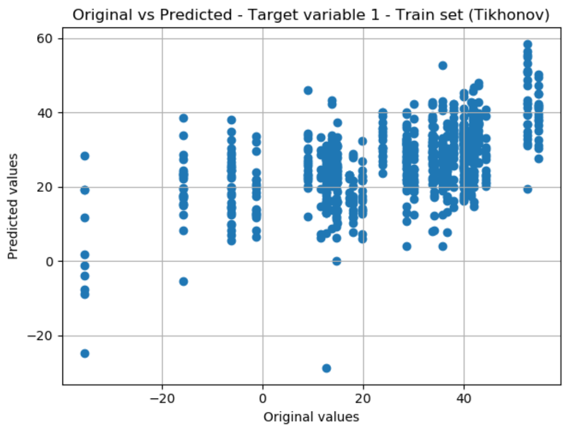
But here the output data is 2D vector unlike the other datasets which had only 1D output, but this doesn't change the normal equation. We also noticed that this data can be approached with a classification approach due to the fact that only 33 unique values for the first target variable and 31 unique for the second target variable out of the 1059 training examples present.(This is evident from Plots present in Figure 14)

Again here also we are performing regularisation of the parameters which are being learnt and of the regularisation methods we specified before here Quadratic regularisation was applied by modifying the normal equation. Also the model was also trained using the Tikhonov regularisation which was also implemented.

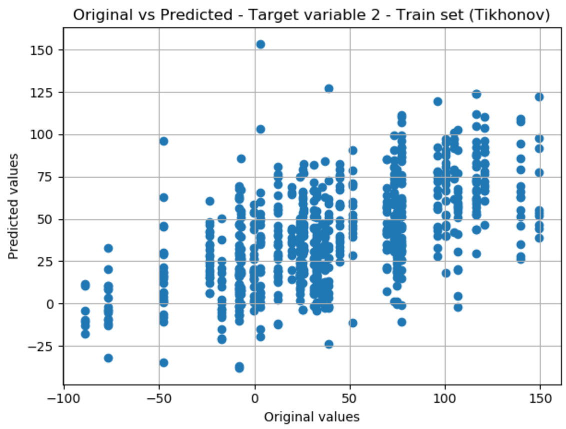
Training of Model

- For training the dataset was split as (70-20-10%) Train-Validation-Test splitting %.
- The Train part of the split was used to train without using the cross-validation.
- Then the trained model was used for validation and this validation error was used for tuning of the hyperparameters i.e. no. of basis function and λ in regulariser.
- Then the tuned model from above is used to predict the values for test data inputted.
- Finally the Erms Error is computed for chosen model on Test dataset using predicted output values and true output values.

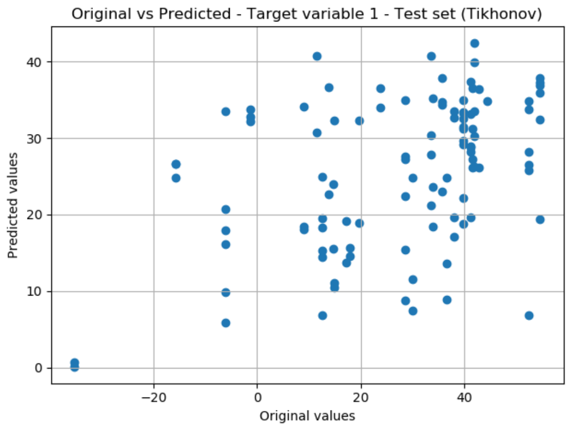
4.3 Plots - best performing model



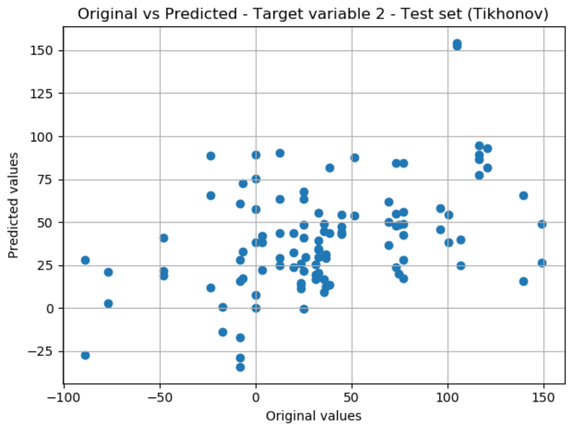
(a) Target output is obtained using the training data.



(b) Target output is obtained using the training data.



(c) Target output is obtained using the test data.



(d) Target output is obtained using the test data.

Figure 14: Original vs Predicted value for target variables for the best performing model(56 Gaussian basis functions, $\ln(\lambda)$ = -18, Tikhonov regularization) on the whole dataset

4.4 Inferences

Regularisation	Degree	Lambda	Training Error	Validation Error	Test Error
Quadratic	114	1	51.164	49.825	51.781
Tikhonov	56	e^{-18}	43.85	47.886	49.61

Table 4: Table showing the E_{RMS} on the training data, validation data and the test data for different models

We find the lowest rms error for 56 gaussian basis with a little Tikhonov regularization. Even though it's the smallest value, looking at figure 14, it looks like it is not as good a fit as expected. The reason for this could be that the target variables have only a few unique values which repeat over the training set. (33 for variable 1 and 31 for variable 2). A better result maybe found by treating this as a classification task.

A Appendix

A.1 Equations

A.1.1 No Regularisation

$$\bar{w}^* = (\phi^T \phi)^{-1} \phi^T \bar{t} \tag{1}$$

Above Equation-1 is called as Normal Equation.

A.1.2 Quadratic Regularisation

$$\bar{w}^* = (\phi^T \phi + \lambda I)^{-1} \phi^T \bar{t} \tag{2}$$

In Equation-2, λ is called as Quadratic regulariser factor.

A.1.3 Tikhonov Regularisation

$$\bar{w}^* = (\phi^T \phi + \lambda \tilde{\phi})^{-1} \phi^T \bar{t} \tag{3}$$

Where,

$$\tilde{\phi}_{ij} = e^{-\frac{||\mu_i - \mu_j||}{2\sigma^2}} \tag{4}$$

Similarly, In Equation-3, λ is called as Tikhonov regulariser factor.