ASSIGNMENT 1

CS5691 Pattern Recognition and Machine Learning

CS5691 Assignment 1

Team No. 26

Team Members:

CS24M046 Shreyas Rathod

CS24M020 Grishma Karekar

Indian institute of Technology, Madras

1. Task 1

1.1. Mathematical Formulation

The data for univariate polynomial regression is obtained by raising it to the required degree. In case of univariate polynomial regression of degree d, the dependent variable, of size (d, 1) is assumed to have the form

$$\vec{y}_{n\times1} = \phi_{n\times d}W_{d\times1}$$

The weights corresponding to a given degree is then calculated by using the closed form solution for univariate polynomial regression:

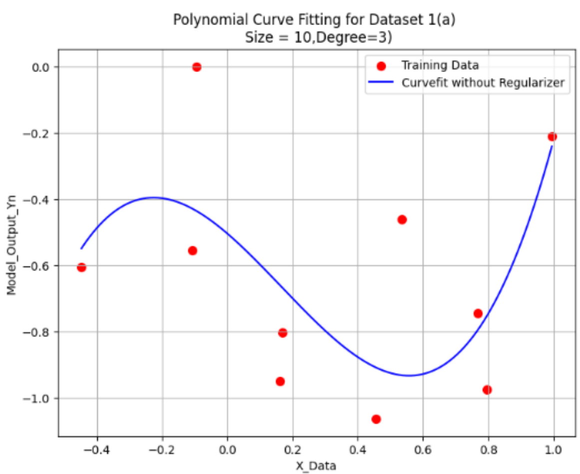
$$W = (\phi^T\phi + \lambda I)^{-1}\phi^T\vec{y}$$

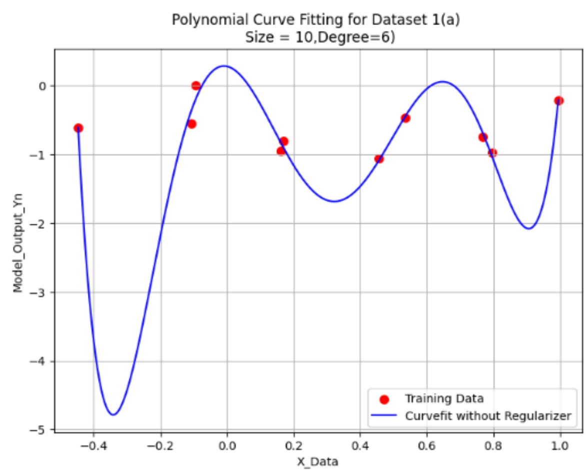
Where, λI is the regularization term.

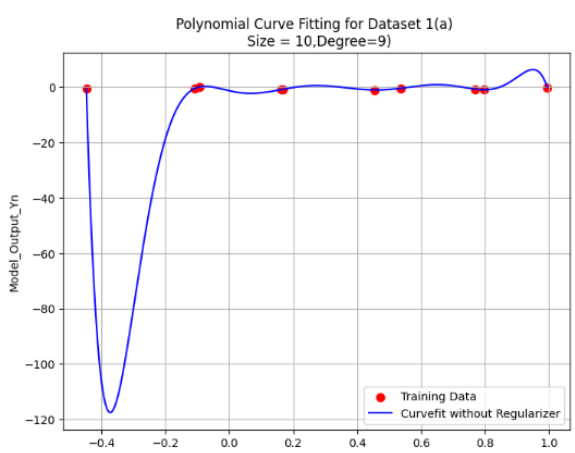
1.2. Model fits with no regularization

Plots of the approximated functions (curves) obtained using training datasets of different sizes (10 and 50), for different model complexities with no regularization is as follows:

1.2.1. Dataset 1a (Sample size: 10)





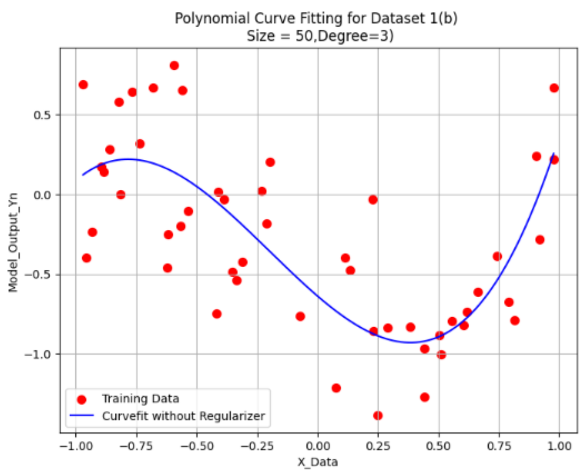


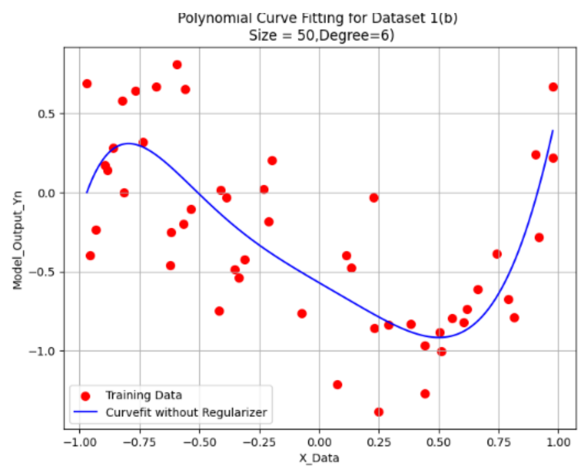
Inference

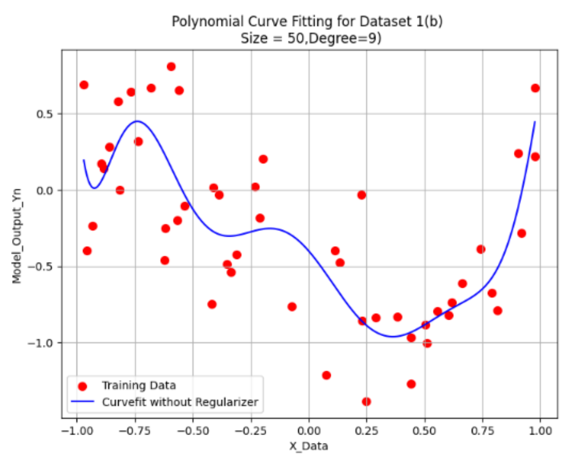
From the above plots, we can observe that:

- Lower degree polynomial curves aren't able to model the dataset well. The curves don't pass through all the data points.
- Higher degree polynomials are able to fit the dataset well. The curves pass through all the data points.
- However, the polynomial degree 9 curves seem to have a lot more variance along the y-axis than the remaining polynomial degrees.

1.2.2. Dataset 1b (Sample size: 50)







Inference

From the above plots, we can observe that:

- A clear difference between the degree 9 fit when the dataset size was 10 to that when the dataset size is 50. The increase in dataset helped decrease the variance and potential overfitting.

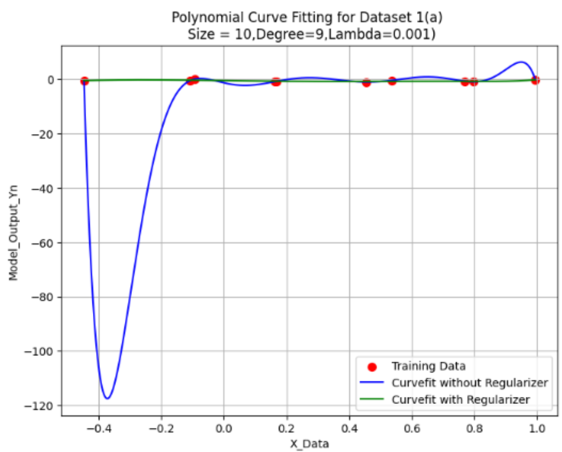
Train Size	λ	Degree	E_{rms} Train	E_{rms} Validation	E_{rms} Test

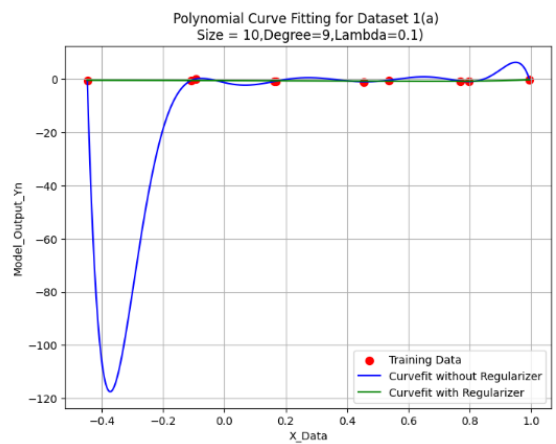
Database 1a (10)	0	3	0.24728789357316405	1.4506414020423941	0.6631540393157739
	0	6	0.1003339665455664	288.4457259928818	55.99530348711504
	0	9	7.8324578496179e-11	55170.44193064778	7257.502947135737
Database 1b (50)	0	3	0.3555989569346142	0.5153733501771435	0.3480371443180693
	0	6	0.3475324577826156	0.5139206369405441	0.35230810125931933
	0	9	0.33432715382422545	0.48932118293868615	0.36987954963546105

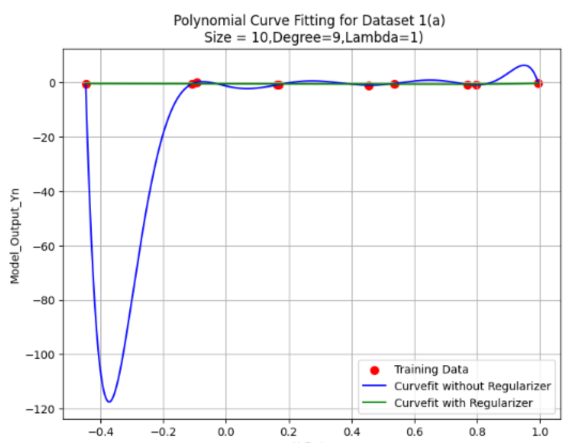
1.3. Effects of Regularization

Plots of the approximated functions (curves) obtained using training datasets of different sizes (10 and 50), for different model complexities with regularization is as follows:

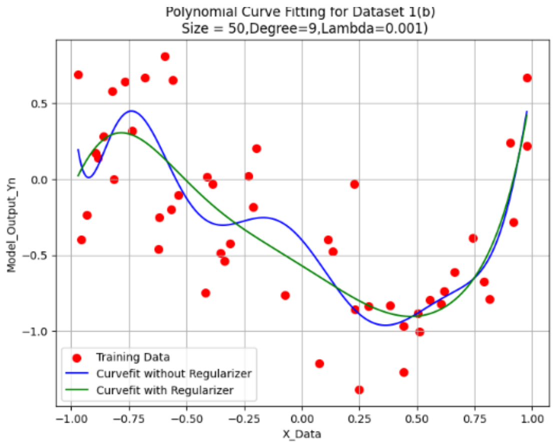
1.3.1. Dataset 1a (Sample size: 10 & Degree: 9)

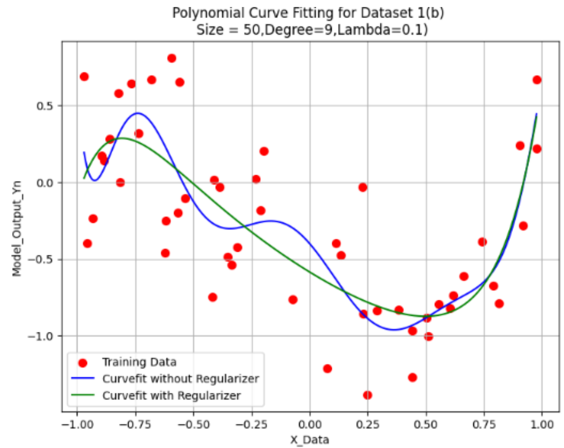


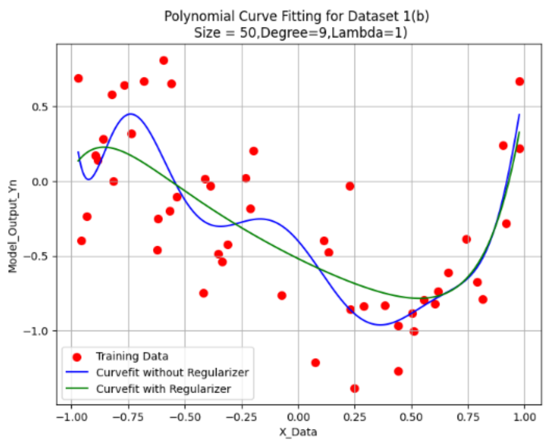




1.3.2. Dataset 1b (Sample size: 50 & Degree: 9)







Inference

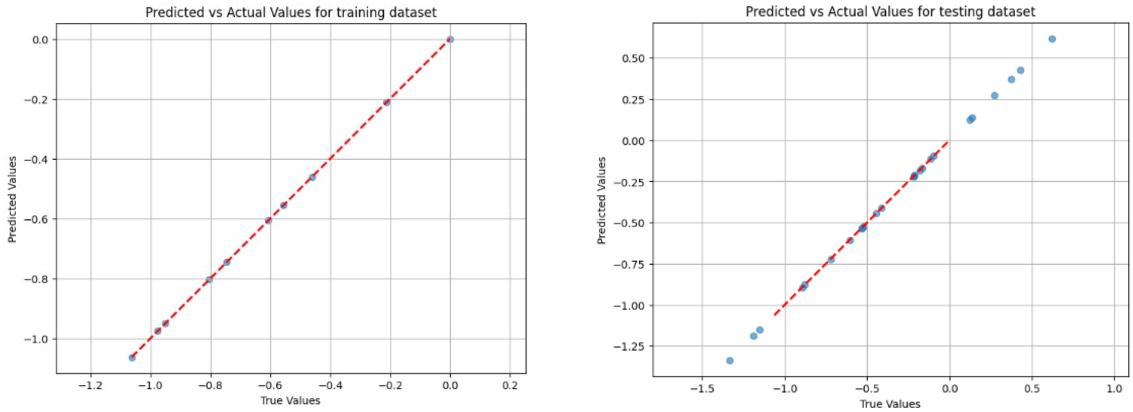
From the above plots, we can see that:

- Regularization was only applied to the degree 9 polynomial, with 10 data points as it had the same number of data points and parameters.
- We can see that; the curve starts becoming more flatter with increasing value of the regularization parameter λ.

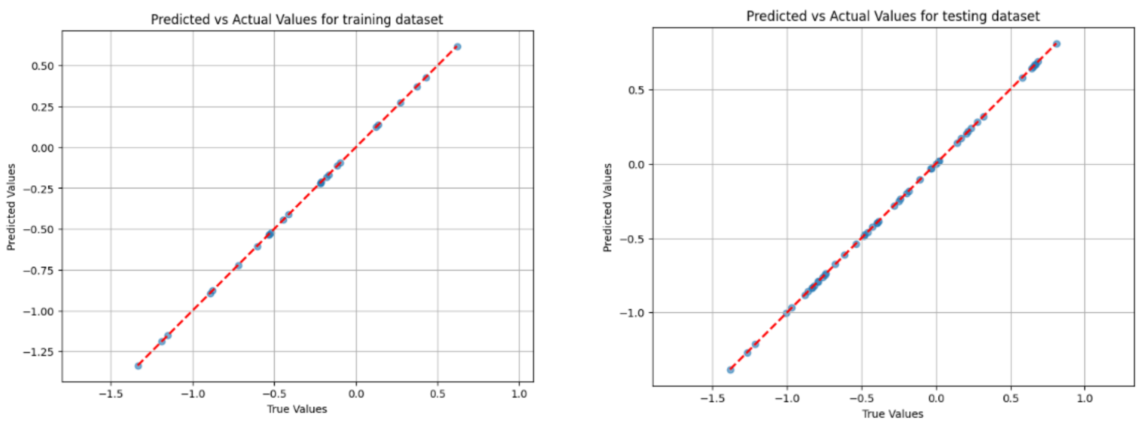
Train Size	λ	Degree	E_{rms}Train	E_{rms} Validation	E_{rms}Test
Database 1a (10)	.001	9	0.2055993266205511	4.154812951101543	1.1896530832182064
	0.1	9	0.25949774582295937	0.6550517756327734	0.4872806639047138
	1	9	0.2825282043391326	0.6104744272253448	0.4874032053129575
Database 1b (50)	.001	9	0.34546175568014037	0.5067500378201477	0.3551492963619565
	0.1	9	0.34817589373446933	0.5064134300288917	0.3559760253240707
	1	9	0.35750003504432565	0.49788108604482334	0.35783376516831056

1.4. Best Model

The best fit for 1a, d: 9 and λ: 0.1 is visualized as follows:



The best fit for 1b, d: 9 and λ: 1 is visualized as follows:



Train Size	$\boldsymbol{\lambda}$	**Degree**	$\boldsymbol{E_{rms}}$ **Train**	$\boldsymbol{E_{rms}}$ **Validation**	$\boldsymbol{E_{rms}}$ **Test**
10	0.1	9	0.25949774582295937	0.6550517756327734	0.4872806639047138
50	1	9	0.35750003504432565	0.49788108604482334	0.35783376516831056

2. Task 2

2.1. Mathematical formulation

We assume that the target variable is of the form:

$$y = \sum_{i=0}^{M} w_i \phi_i(x_1, x_2) + \epsilon$$

The function for regularization is then used to obtain optimum values of $\vec{w}$

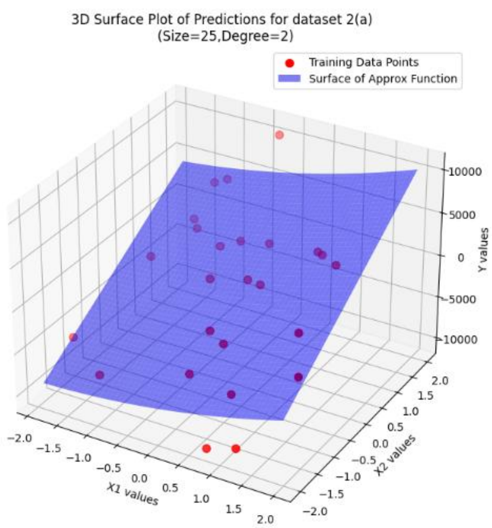
$$\vec{w} = [(X^T X + \lambda I)^{-1} X^T].y$$

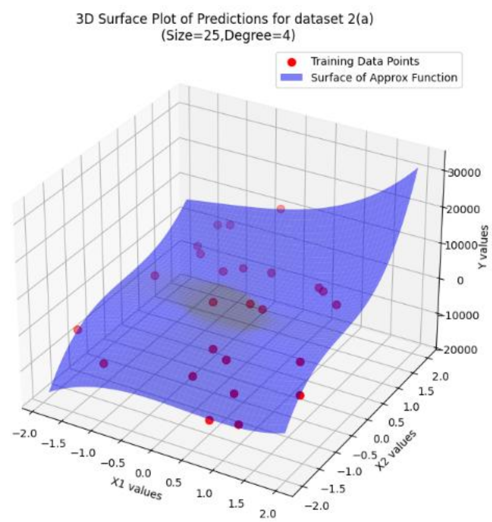
$$y_{prediction} = X\vec{w}$$

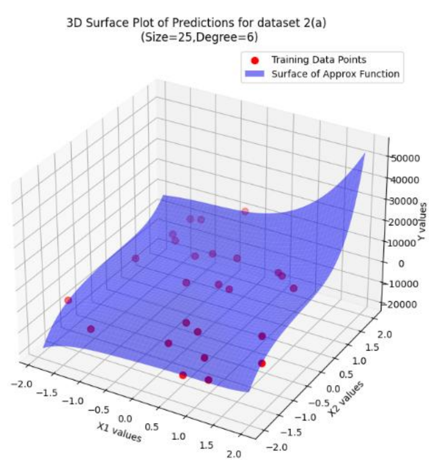
2.2. Model fits with no regularization

Plots of the approximated functions (curves) obtained using training datasets of different sizes (25 and 100), for different model complexities with no regularization is as follows:

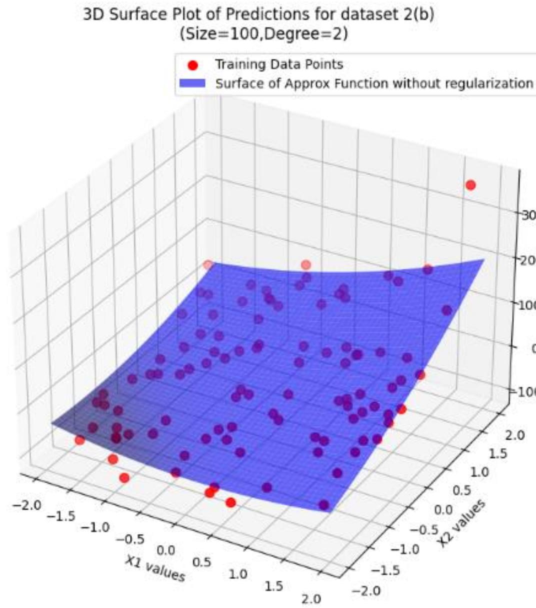
2.2.1. Dataset 2a

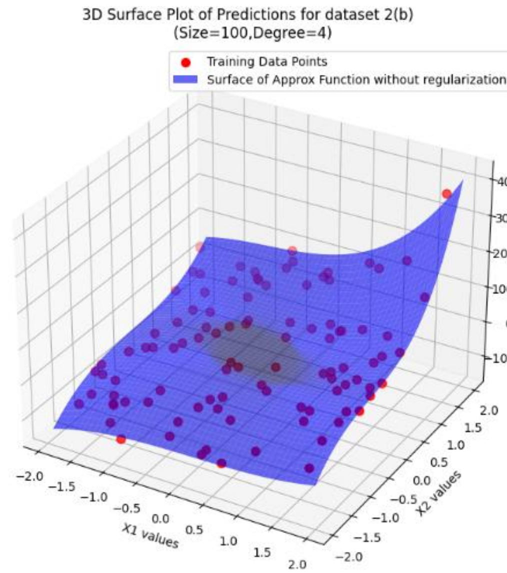


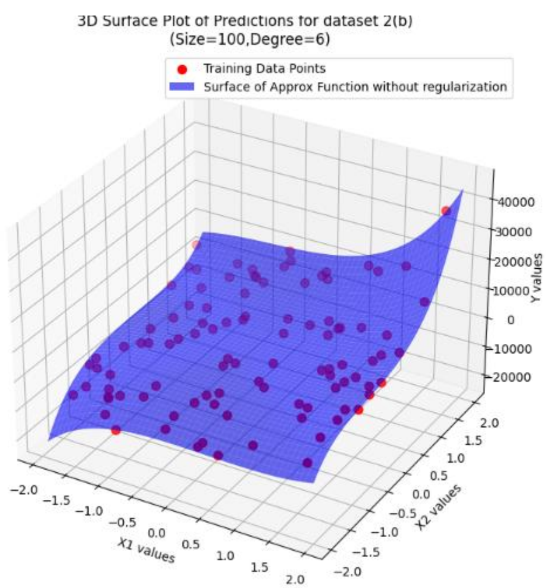




2.2.2. Dataset 2b



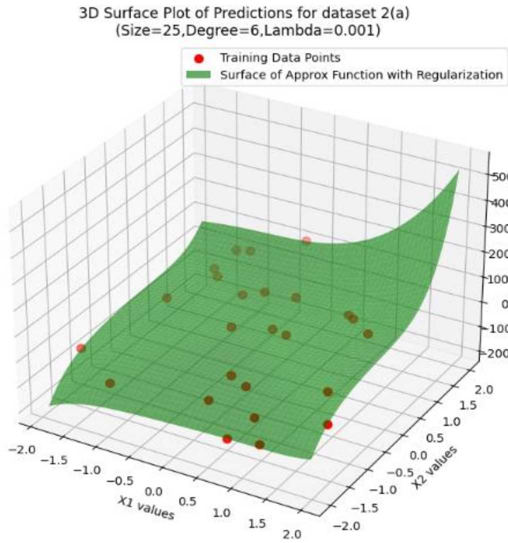


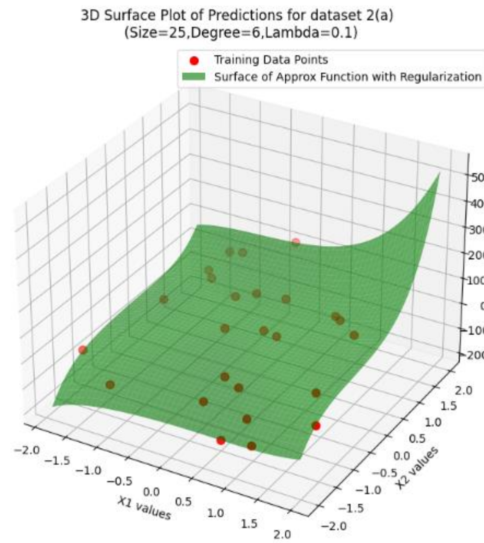


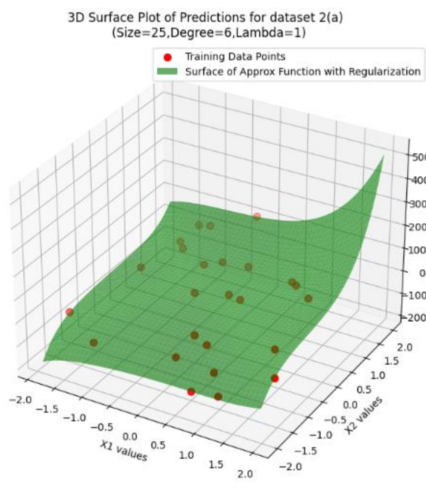
Train Size	λ	Degree	E_{rms}Train	E_{rms} Validation	E_{rms}Test
Database 2a (25)	0	2	1940.1698253536626	3341.7569439537133	1921.6102417115383
	0	4	194.58742196964099	800.4667178168404	437.00322280988985
	0	6	2.9842190347114e-11	751.7609143905504	368.32817310666576
Database 2b (100)	0	2	2684.628719074299	2456.211136642522	2188.0021662310387
	0	4	436.26023407252245	452.33301413953825	412.83918990881824
	0	6	12.13508043409006	9.361333182892624	5.469279239785253

2.2.3. Effect of Regularization

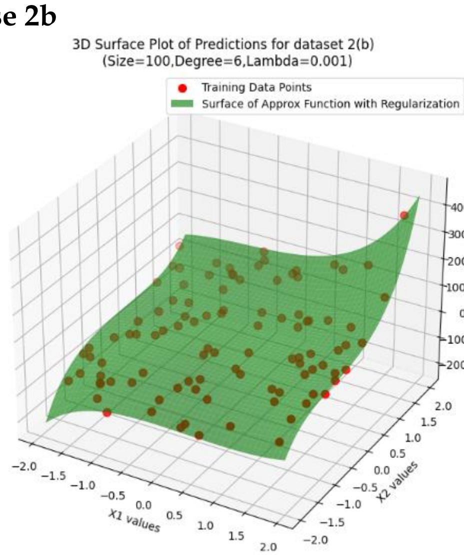
Database 2a

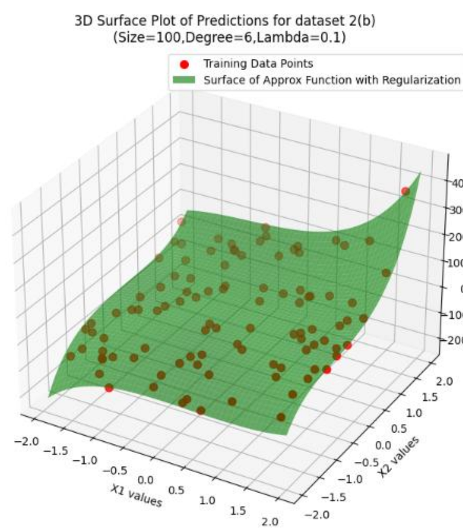


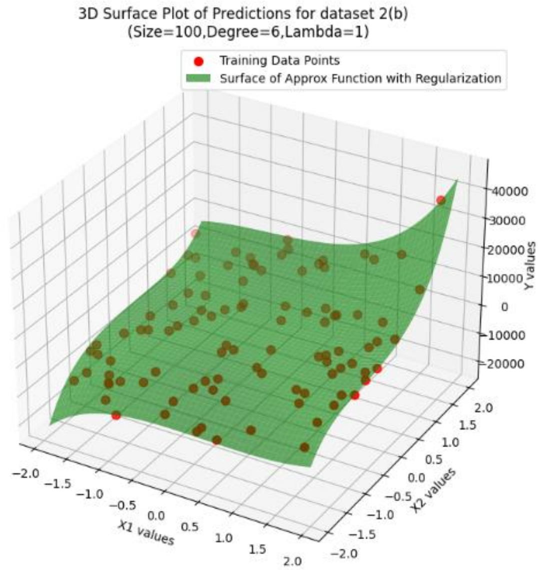




Database 2b



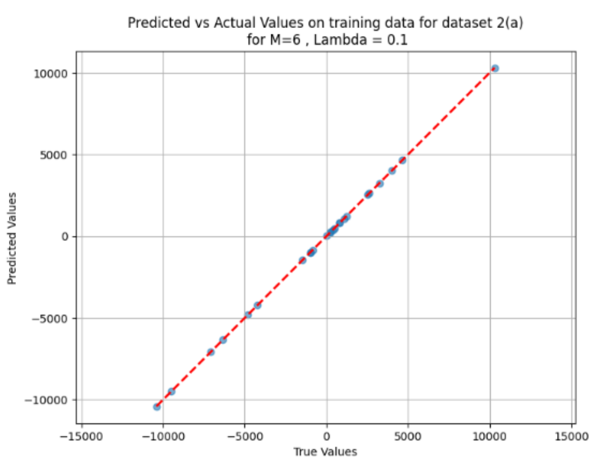


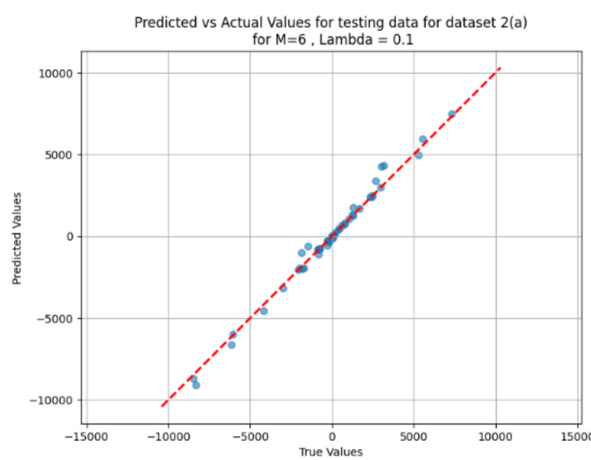


Train Size	λ	Degree	E_{rms}Train	E_{rms} Validation	E_{rms}Test
Database 2a (25)	0.001	6	0.6693777819202148	767.0996265069608	327.9566215084053
	0.1	6	8.766316729446649	413.09356498799633	290.24267449163057
	1	6	35.50170501716778	348.3630311072226	345.9031544907109
Database 2b (100)	0.001	6	12.135087525901127	9.37059786993278	5.470621175872881
	0.1	6	12.203155925920438	10.346465149552957	5.834791082506698
	1	6	16.64157970690196	19.155804240860327	15.864842387858012

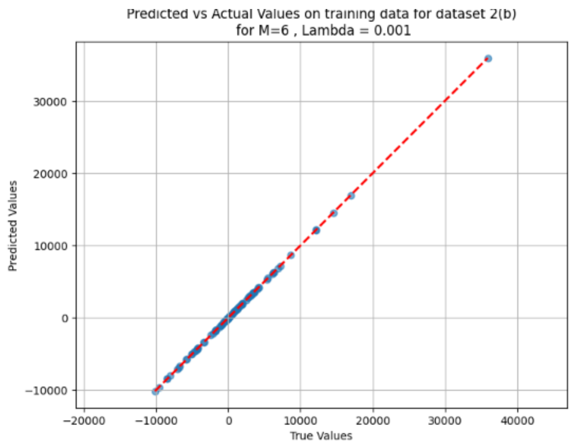
2.3. Best Model

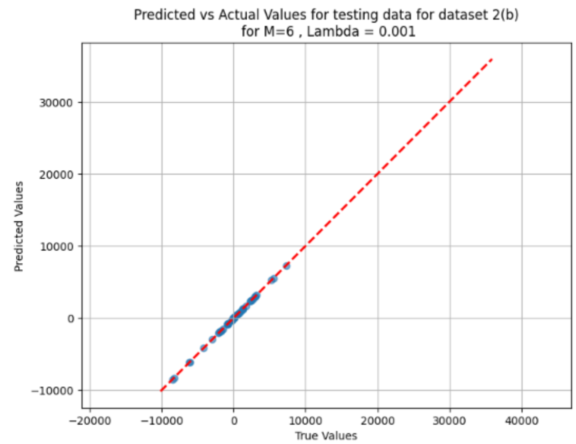
The database 2a best fit, d: 6 and λ: 0.1 is visualized as follows:





The database 2b best fit, d: 6 and λ: 1 is visualized as follows:





Train Size	λ	Degree	E_{rms} Train	E_{rms} Validation	E_{rms} Test
25	0.1	6	8.766316729446649	413.09356498799633	290.24267449163057
100	0	6	12.13508043409006	9.361333182892624	5.469279239785253

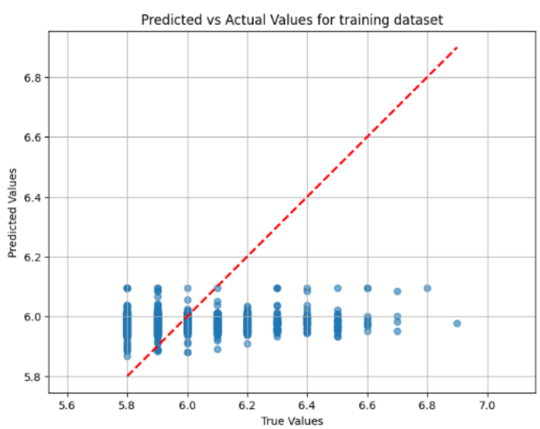
3. Task 3

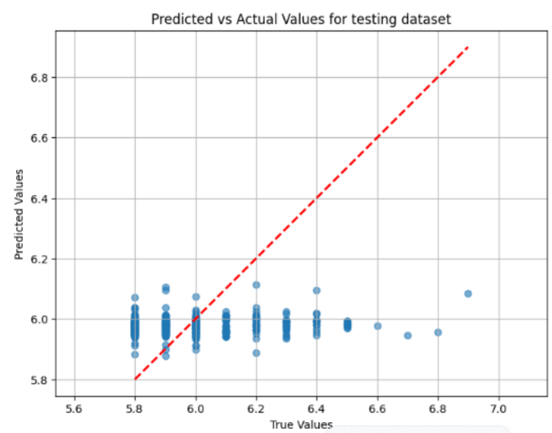
3.1. Dataset 3

Train Size	λ	Degree	E_{rms} Train	E_{rms} Validation	E_{rms} Test
Database 3	0	2	0.1878409268173666	0.1863690282128376	0.1928108750093735
	0	3	0.18735067965253804	0.18603518496195157	0.19019310349671306
	0.000001	3	0.18735067965253857	0.18603518343631667	0.19019310308297227
	0.0001	3	0.18735067965775407	0.1860350346104525	0.19019306601400646
	0.1	3	0.1873558869008548	0.1858905505637254	0.19016131385993024

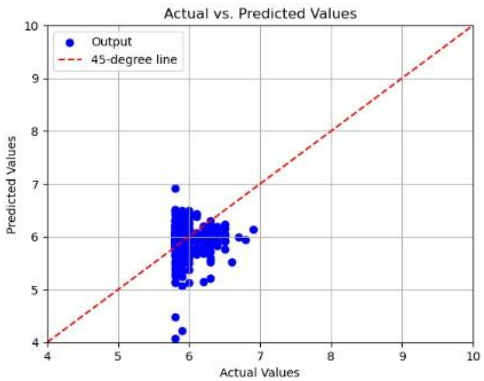
3.2. Best Models:

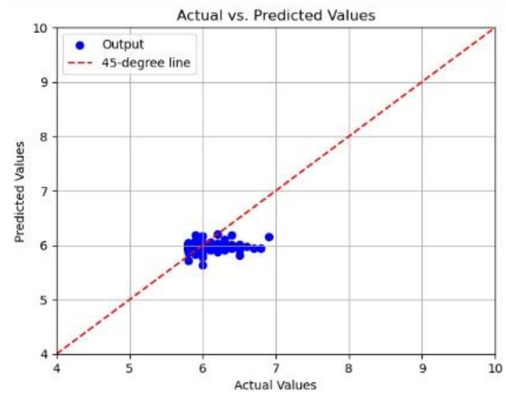
For best fit training and testing database

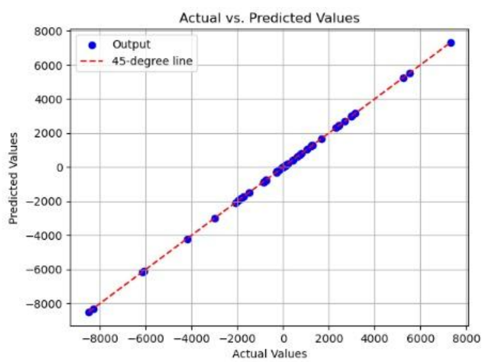


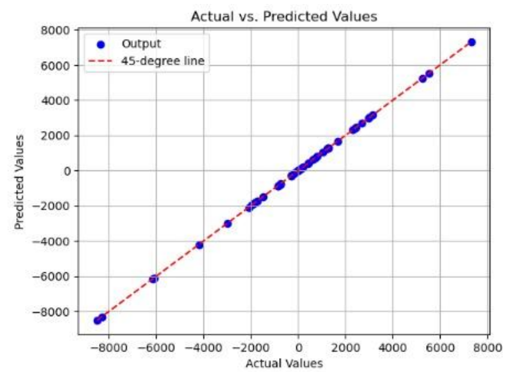


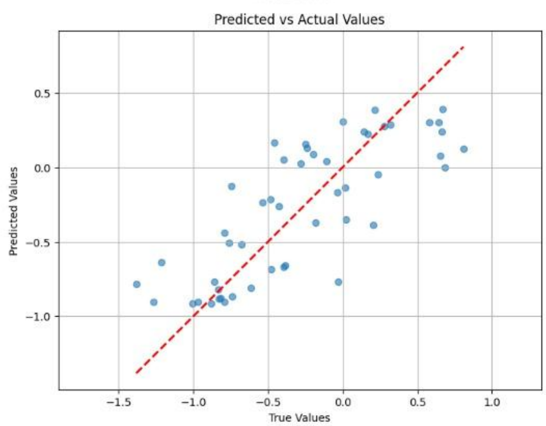
Train Size	λ	Degree	E_{rms} Train	E_{rms} Validation	E_{rms} Test
1524	0.1	3	0.1873558869008548	0.1858905505637254	0.19016131385993024

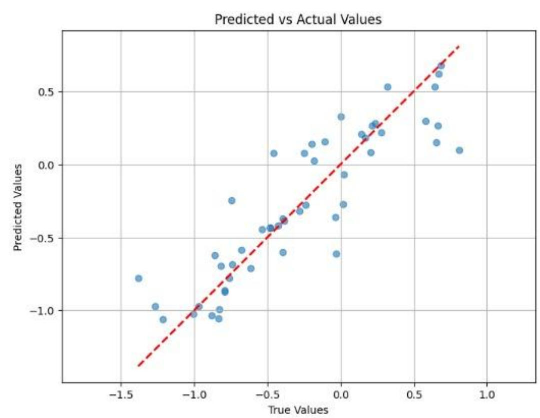

Actual vs. Predicted Values
Output
45-degree line
Predicted Values
Actual Values


Actual vs. Predicted Values
Output
45-degree line
Predicted Values
Actual Values


Actual vs. Predicted Values
Output
45-degree line
Predicted Values
Actual Values


Actual vs. Predicted Values
Output
45-degree line
Predicted Values
Actual Values


Predicted vs Actual Values
Predicted Values
True Values


Predicted vs Actual Values
Predicted Values
True Values