## 

## **ASSIGNMENT 2**

Vrutik Halani

ID No. 2017A7PS1732H

Smit Sheth

ID No. 2017A7PS1666H

Yug Ajmera

ID No. 2017A4PS0644H

## 

## 

## 

## 

## 

## **Assignment 2**

**Developed polynomial regression models to predict the altitude values from the latitude and longitude values given in the dataset.**

**The dataset has 434874 coordinate points. We have normalized the entire dataset using the formula:**

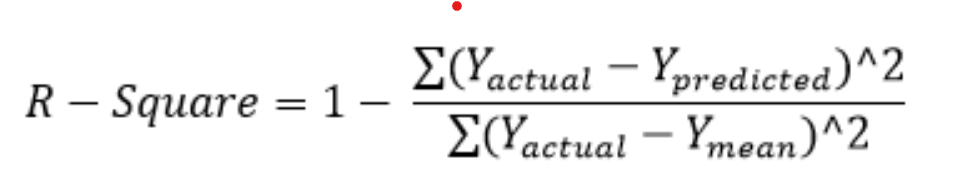
****

**We have used test\_train\_split to randomly split this dataset into 70% training data and 30% test data for degree 1 , 2 , 3 , 4 , 5 and 6.**

**We have used the following linear regression models to predict the altitude in this assignment:**

1. **Gradient Descent**
2. **Gradient Descent with regularization**
3. **L1 Regularization (Lasso)**
4. **L2 Regularization (Ridge)**

**To check error values in the prediction model we have used R-square score and RMSE.**

****

**Max Iteration : 5000**

**Stopping Criteria : (E - E` )<= 1e-5**

**Following is the Comparison of different Regression Models :**

1. **Degree : 1**

Features : ['1', 'x0', 'x1']

Coefficients : [ 0.201 , 0.09471633 , -0.10008946]

RMSE Error : 0.12844211399884353

R-square Score : 0.026609495535749983

Training Error : 2519.4929487682152

Validation Error : 1080.2545627608704

1. **Degree : 2**

Features : ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2']

Coefficients : [ 0.1424 , 0.52852481 , -0.17674444 , -0.47858433 ,

-0.02334835 , 0.1238497 ]

RMSE Error: 0.12609122550893637

R-square Score : 0.06191552836581482

Training Error : 2421.219239212934

Validation Error : 1037.1874469279

1. **Degree : 3**

Features : ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2', 'x0^3', 'x0^2 x1', 'x0 x1^2',

'x1^3']

Coefficients : [ 0.2295 , -0.28066551 , -0.54607023 , 3.66296098

-4.5869653 3.54356053 , -4.72180876 , 7.13853521 ,

-3.22729369 , -1.14330758]

RMSE Error : 0.11994890951827324

R-square Score : 0.1510837955682789

Training Error : 2239.5317944931057

Validation Error : 959.7601599918426

1. **Degree : 4**

Features : ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2', 'x0^3', 'x0^2 x1', 'x0 x1^2',

'x1^3', 'x0^4', 'x0^3 x1', 'x0^2 x1^2', 'x0 x1^3', 'x1^4']

Coefficients : [ 1.00000000e-08, -1.88803365e-01 , 2.09872184e+00 ,

3.48574325e+00 , -6.45769177e+00 -5.76634087e+00 ,

-9.60190852e-01 -3.98447685e+00 , 1.41512872e+01 ,

6.18805065e+00 , -4.95300502e+00 , 1.49074054e+01,

-1.38872391e+01 , -1.76688271e+00 , -3.09775989e+00]

RMSE Error : 0.11606668346644405

R-square Score : 0.20514599768586905

Training Error : 2254.0290596195805

Validation Error : 960.8730510217588

1. **Degree : 5**

Features : ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2', 'x0^3', 'x0^2 x1', 'x0 x1^2',

'x1^3', 'x0^4', 'x0^3 x1', 'x0^2 x1^2', 'x0 x1^3', 'x1^4', 'x0^5',

'x0^4 x1', 'x0^3 x1^2', 'x0^2 x1^3', 'x0 x1^4', 'x1^5']

Coefficients: [ 1.00000000e-08 , 2.39516171e+00 , 1.17953663e+00 ,

-2.75775404e+01 , 2.50697637e+01 , -2.06643384e+01 ,

1.03153602e+02 , -7.67970632e+01 , -6.05519957e+01 ,

8.40153898e+01 , -1.49427174e+02 , 1.36837423e+02 ,

-1.21070183e+01 , 1.26531217e+02 , -1.36413819e+02 ,

7.38686256e+01 , -1.13002148e+02 , 1.58375845e+02 ,

-2.37538421e+02 8.57282431e+01 , 3.44676775e+01]

RMSE Error : 0.10693071787777099

R-square Score : 0.3253520882003068

Training Error : 2200.5831010026695

Validation Error : 935.315805045403

1. **Degree : 6**

Features: ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2', 'x0^3', 'x0^2 x1', 'x0 x1^2',

'x1^3', 'x0^4', 'x0^3 x1', 'x0^2 x1^2', 'x0 x1^3', 'x1^4', 'x0^5',

'x0^4 x1', 'x0^3 x1^2', 'x0^2 x1^3', 'x0 x1^4', 'x1^5', 'x0^6', 'x0^5

x1', 'x0^4 x1^2', 'x0^3 x1^3', 'x0^2 x1^4', 'x0 x1^5', 'x1^6']

Coefficients: [0.16209113947472165, 0.09996167390438612,

0.03441374578372098, -0.01005499892486365,

-0.0369545091166129, -0.05164872754372274,

-0.058589837380639116, -0.01913041729478206,

-0.005124651325202706, -0.011626248210226587,

-0.021551212667985854, -0.02947630848322382,

-0.03444525574802474, -0.0015420225483188405,

-0.0011342059661773665, -0.008353983932195084,

-0.01563306146481123, -0.02081409570600366,

0.0028563924977215134, -0.003155996325001775,

-0.00990798963849666, -0.014885688508900348,

-0.004425986496932112, -0.010054508143572057,

-0.014214221294307485, -0.014469257367811659,

-0.01730281842461692, -0.02329860562683845]

RMSE Error : 0.10055696875031997

R-square Score : 0.4033817113791195

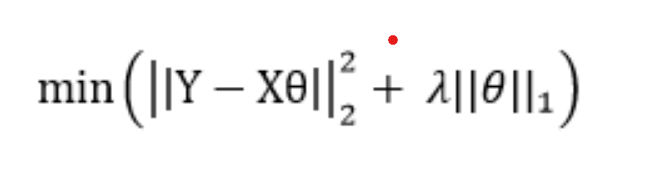
Training Error : 2040.1371044663774

Validation Error : 1724.3058572258158

**Degree - 6 Regularization L1 (Lasso):**

Minimization objective = LS Obj + α \* (sum of absolute value of

coefficients)



Features: ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2', 'x0^3', 'x0^2 x1', 'x0 x1^2',

'x1^3', 'x0^4', 'x0^3 x1', 'x0^2 x1^2', 'x0 x1^3', 'x1^4', 'x0^5',

'x0^4 x1', 'x0^3 x1^2', 'x0^2 x1^3', 'x0 x1^4', 'x1^5', 'x0^6', 'x0^5

x1', 'x0^4 x1^2', 'x0^3 x1^3', 'x0^2 x1^4', 'x0 x1^5', 'x1^6']

Coefficients: [0.08824125813688263, 0.023060247900390907,

-0.022343662196507873, -0.03161254073635382,

-0.03408566966991815, -0.02300816816592562,

-0.016205939887315432, -0.05676642787208582,

-0.03208517949054775, -0.03373146855219865,

-0.028280508417212174, -0.02162200364943641,

-0.012286417829582019, -0.030215167743847297,

-0.02275144198938446, -0.026204645250295564,

-0.018324002461147418, -0.00893320282537293,

-0.03343445597802352, -0.02180411546938499,

-0.012676598010308167, -0.015394667191482506,

-0.026020566438000083, -0.014742134667179007,

-0.015906731475413667, -0.022643522423598413,

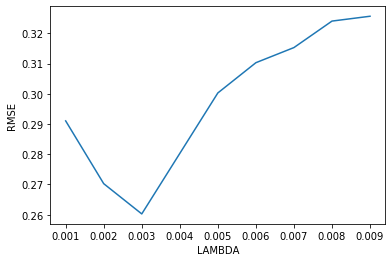
-0.021476573437008782, -0.018362666940422295]

Lambda = 0.003

RMSE Error : 0.2602696073231428

R-square Score : 0.323482001335434

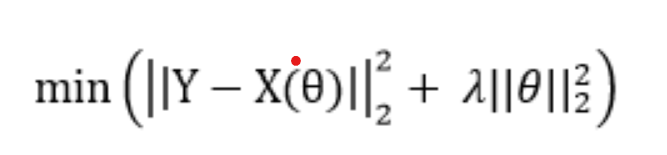
Lambda vs RMSE graph :



**Degree - 6 Regularization L2 (Ridge):**

Minimization objective = LS Obj + α \* (sum of square of

coefficients)



Features: ['1', 'x0', 'x1', 'x0^2', 'x0 x1', 'x1^2', 'x0^3', 'x0^2 x1', 'x0 x1^2',

'x1^3', 'x0^4', 'x0^3 x1', 'x0^2 x1^2', 'x0 x1^3', 'x1^4', 'x0^5',

'x0^4 x1', 'x0^3 x1^2', 'x0^2 x1^3', 'x0 x1^4', 'x1^5', 'x0^6', 'x0^5

x1', 'x0^4 x1^2', 'x0^3 x1^3', 'x0^2 x1^4', 'x0 x1^5', 'x1^6']

Coefficients: [0.16209113947472165, 0.09996167390438612,

0.03441374578372098, -0.01005499892486365,

-0.0369545091166129, -0.05164872754372274,

-0.058589837380639116, -0.01913041729478206,

-0.005124651325202706, -0.011626248210226587,

-0.021551212667985854, -0.02947630848322382,

-0.03444525574802474, -0.0015420225483188405,

-0.0011342059661773665, -0.008353983932195084,

-0.01563306146481123, -0.02081409570600366,

0.0028563924977215134, -0.003155996325001775,

-0.00990798963849666, -0.014885688508900348,

-0.004425986496932112, -0.010054508143572057,

-0.014214221294307485, -0.014469257367811659,

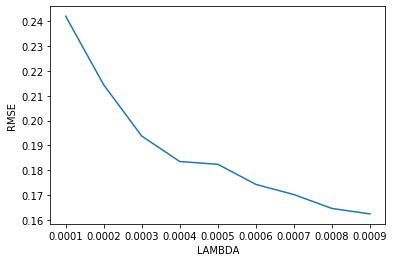
-0.01730281842461692, -0.02329860562683845]

Lambda = 0.0009

RMSE Error : 0.16234263290

R-square Score : 0.4283792838927

LAMBDA vs RMSE :



Following are the conclusions from data :

As higher degree polynomials are fit to the data, the training error decreases while the R2 error increases, indicating some overfitting.

Over fit: degree 6 polynomial Best fit: degree 4 polynomial