Assignment 2

BY

Ravisanker E
2017A7PS0433H

S Ankit
2017A7PS0297H

CONTENTS

1. **PART A ………………………………………………………3**

   Degree 1 Polynomial …………………………………..3

   Degree 2 Polynomial …………………………………..4

   Degree 3 Polynomial …………………………………..5

   Degree 4 Polynomial …………………………………..6

   Degree 5 Polynomial …………………………………..7

   Degree 6 Polynomial …………………………………..8

2. **PART B……………………………………………………..9**

3. **PART C ……………………………………………………...11**

Part A

Degree 1 polynomial

$$y = w_0 + w_1 x_1 + w_2 x_2$$

The model obtained after gradient descent:

```
Degree 1


Model:

  [[22.17800962]
 [ 2.79440781]
 [-3.5682518 ]]

Half sum of squares:  58541036.335190825
Mean Sqaured Error:  168.2706894986198


On the Test Data
RMSE:  18.46987496271496
R2:  0.025401222950555913
```

Degree 2 polynomial

$$y = w_0 + w_1x_1 + w_2x_2 + \ w_3{x_1}^2 + w_4x_1x_2 + w_5{x_1}^2$$

The model obtained after gradient descent:

$$y = 24.02277375 + 0.63379332x_1 - 2.94218532x_2 - 2.86585605{x_1}^2 - 0.22390063x_1x_2 + 1.14640535{x_1}^2$$

```
Degree 2

100%

Model:

  [[24.02277375]
 [ 0.63379332]
 [-2.94218532]
 [-2.86585605]
 [-0.22390063]
 [ 1.14640535]]

Half sum of squares:  56462460.5916632
Mean Sqaured Error:  162.29601949900027


On the Test Data
RMSE:  18.135646279178207
R2:  0.06035452943351216
```

Degree 3 polynomial

$$y = w_0 + w_1x_1 + w_2x_2 + \ w_3{x_1}^2 + w_4x_1x_2 + w_5{x_2}^2 + w_6{x_1}^3 + w_7{x_1}^2x_2 + w_8x_1{x_2}^2 + w_9{x_2}^3$$

The model obtained after gradient descent:

```
Degree 3

100%

Model:

  [[22.20393839]
 [ 9.37133271]
 [-1.09507781]
 [-3.83076422]
 [ 0.49422711]
 [ 3.44714142]
 [-5.80879048]
 [10.55099555]
 [-5.66479562]
 [-2.53010432]]

Half sum of squares:  51069842.49983322
Mean Sqaured Error:  146.79544722830605


On the Test Data
RMSE:  17.25790103862147
R2:  0.1491090590340417
```

Degree 4 polynomial

$$y = w_0 + w_1x_1 + w_2x_2 + \; w_3{x_1}^2 + w_4x_1x_2 + w_5{x_2}^2 + w_6{x_1}^3 + w_7{x_1}^2x_2 + w_8x_1{x_2}^2 + w_9{x_2}^3 + w_{10}{x_1}^4 + w_{11}{x_1}^3{x_2}^1 + w_{12}{x_1}^2{x_2}^2 + w_{13}{x_1}^1{x_2}^3 + w_{14}{x_2}^4$$

For degree 4 the learning rate was set as 3×10^{-7}

No of iterations were 10000

The model obtained after gradient descent:

```
Degree 4

100%                                    10C


Model:

  [[18.65883539]
 [ 6.58190269]
 [-1.73307347]
 [-1.13720036]
 [ 3.35126208]
 [ 9.20912023]
 [-3.84891786]
 [ 4.29843853]
 [ 0.07982977]
 [-1.9564289 ]
 [-0.81286172]
 [ 2.27123293]
 [-4.68952504]
 [ 0.23089582]
 [-1.76501414]]

Half sum of squares:  48684115.259346455
Mean Sqaured Error:  139.9378992099594


On the Test Data
RMSE:  16.84879429026381
R2:  0.18897244976283656
```

Degree 5 polynomial

$$y = w_0 + w_1x_1 + w_2x_2 + \ w_3{x_1}^2 + w_4x_1x_2 + w_5{x_2}^2 + w_6{x_1}^3 + w_7{x_1}^2x_2 + w_8x_1{x_2}^2 \\ + w_9{x_2}^3 + w_{10}{x_1}^4 + w_{11}{x_1}^3{x_2}^1 + w_{12}{x_1}^2{x_2}^2 + w_{13}{x_1}^1{x_2}^3 + w_{14}{x_2}^4 \\ + w_{15}{x_1}^5 + w_{16}{x_1}^4{x_2}^1 + w_{17}{x_1}^3{x_2}^2 + w_{18}{x_1}^2{x_2}^2 + w_{19}{x_1}^1{x_2}^4 + w_{20}{x_2}^5$$

For degree 5 the learning rate was set as 3×10^{-9}

No of iterations were 10000

The model obtained after gradient descent:

```
Degree 5

100% 1

Model:

  [[18.47677972]
 [ 6.66098648]
 [-1.61347102]
 [ 0.23769576]
 [ 4.11609876]
 [10.39884517]
 [-3.8543762 ]
 [ 4.03530782]
 [ 1.54657758]
 [-4.81654756]
 [-0.97449485]
 [ 1.55643582]
 [-4.12719387]
 [-1.79769317]
 [-2.05390847]
 [-0.07241136]
 [ 1.80059876]
 [-4.93672726]
 [ 2.6042914 ]
 [-0.59796443]
 [ 1.2218109 ]]

Half sum of squares:  46825469.91046058
Mean Sqaured Error:  134.5953983939562

On the Test Data
RMSE:  16.52764649257954
R2:  0.2195951104164772
```

Degree 6 polynomial

$$
\begin{aligned}
y = w_0 + w_1 x_1 + w_2 x_2 + \ & w_3 {x_1}^2 + w_4 x_1 x_2 + w_5 {x_2}^2 + w_6 {x_1}^3 + w_7 {x_1}^2 x_2 + w_8 x_1 {x_2}^2 + w_9 {x_2}^3 \\
& + w_{10} {x_1}^4 + w_{11} {x_1}^3 {x_2}^1 + w_{12} {x_1}^2 {x_2}^2 + w_{13} {x_1}^1 {x_2}^3 + w_{14} {x_2}^4 + w_{15} {x_1}^5 + w_{16} {x_1}^4 {x_2}^1 \\
& + w_{17} {x_1}^3 {x_2}^2 + w_{18} {x_1}^2 {x_2}^2 + w_{19} {x_1}^1 {x_2}^4 + w_{20} {x_2}^5 + w_{21} {x_1}^6 + w_{22} {x_1}^5 {x_2}^1 + w_{23} {x_1}^4 {x_2}^2 \\
& + w_{24} {x_1}^3 {x_2}^3 + w_{25} {x_1}^2 {x_2}^4 + w_{26} {x_1}^1 {x_2}^5 + w_{27} {x_2}^6
\end{aligned}
$$

The model obtained after gradient descent:

```
Degree 6

100%

Model:

  [[18.79563462]
 [ 6.39508928]
 [-0.60631386]
 [ 0.7755895 ]
 [ 3.43241543]
 [11.05088848]
 [-3.88484325]
 [ 3.65387577]
 [ 2.57191541]
 [-4.06851492]
 [-1.17209202]
 [ 1.31955922]
 [-4.03210399]
 [-3.38551069]
 [-2.40337543]
 [-0.07525286]
 [ 1.84853262]
 [-4.44730419]
 [ 0.42030288]
 [ 0.56436537]
 [ 0.44152644]
 [-0.05112137]
 [ 0.81315445]
 [-1.90505858]
 [ 1.62907907]
 [-1.33310221]
 [ 1.6137336 ]
 [-0.16477425]]

Half sum of squares:  46279174.51703377
Mean Sqaured Error:  133.02512379212806


On the Test Data
RMSE:  16.424498687359254
R2:  0.2293056113691272
```

Part B

Comparison of the Models developed

Degree	RMSE on Validation Set	R2 on Validation Set
1	18.4698	0.02540
2	18.1356	0.06035
3	17.2579	0.14910
4	16.8487	0.18897
5	16.5276	0.21959
6	16.4244	0.2293

Inference from RMSE

We observe that the RMSE metric decreases with increase in degree with RMSE lowest at degree = 6. So among the models created in Part A, the model of degree 6 best fits the data. Since the training error is lowest at degree 6, we infer that none of the models created in Part A overfits the data.

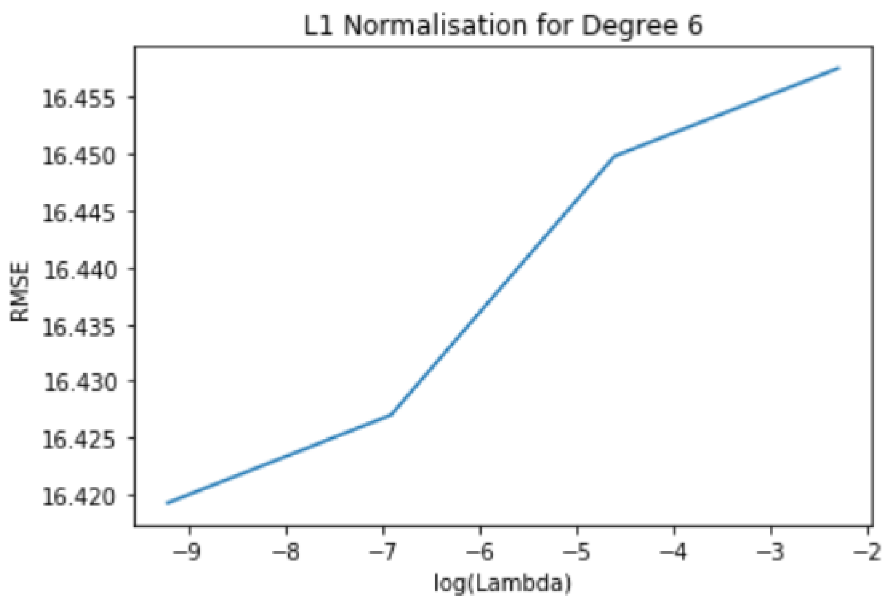
Inference from R^2

We observe that the R^2 metric steadily increases with increase in degree with R^2 highest (closest to 1) at degree = 6. So among the models created in Part A, the model of degree 6 best fits the data. Since the training error is lowest at degree 6, we infer that none of the models created in Part A overfits the data.

In fact the model of degree 6 could be underfitting the data.

Part C

L1 regularisation



Inference:

We plotted the RMSE to log(λ) and we observed that RMSE decreases with decreasing λ. This is in accordance with our results from Part B that model of degree 6 does not overfit the data and hence regularization has no significant effect on the model. The lowest RMSE we recorded was corresponding to log(λ) = -9 (λ = 1e-9). So we have obtained the L1 regularization for the regression model of degree 6 with λ = 1e-9.

The model obtained from L1 regularization for the regression model of degree 6 with λ = 1e-9 is

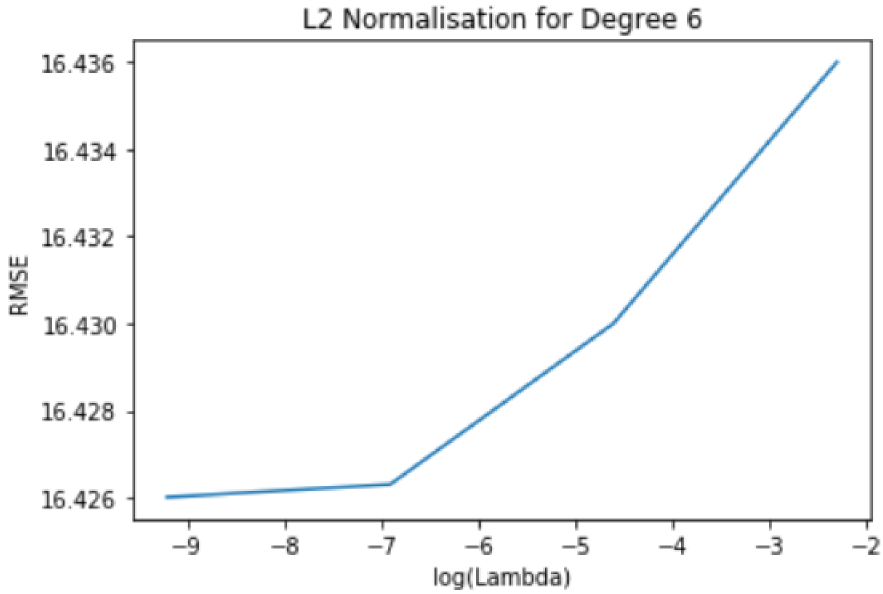
```
Model:

 [[19.02231368]
 [ 6.46576268]
 [-0.6671071 ]
 [ 0.55784914]
 [ 3.64960071]
 [10.24318711]
 [-3.83586674]
 [ 3.58008618]
 [ 2.25429736]
 [-4.08393878]
 [-1.12957665]
 [ 1.45243143]
 [-4.18470913]
 [-3.51768806]
 [-1.81011914]
 [-0.08037901]
 [ 1.78546555]
 [-4.24515547]
 [ 0.49732235]
 [ 0.52867699]
 [ 0.4930985 ]
 [-0.04711567]
 [ 0.76631041]
 [-1.90080624]
 [ 1.72587122]
 [-1.22584183]
 [ 1.52549056]
 [-0.26111988]]

Half sum of squares:  [[46441056.12472574]]
Mean Sqaured Error:  [[133.49043721]]


On the Test Data
RMSE:  [[16.45392981]]
R2:  0.22654111622864148
```

L2 regularisation



Inference:

We plotted the RMSE to log(λ) and we observed that RMSE decreases with decreasing λ. This is in accordance with our results from Part B that model of degree 6 does not overfit the data and hence regularization has no significant effect on the model. The lowest RMSE we recorded was corresponding to log(λ) = -9 (λ = 1e-9). So we have obtained the L2 regularization for the regression model of degree 6 with λ = 1e-9.

The model obtained from L2 regularization for the regression model of degree 6 with λ = 1e-9 is

```
Model:

  [[18.94446578]
 [ 6.73317104]
 [-0.66605041]
 [ 0.60715593]
 [ 3.51833642]
 [10.53453898]
 [-4.11672982]
 [ 3.70063708]
 [ 2.33523913]
 [-4.1380352 ]
 [-1.21748604]
 [ 1.45532766]
 [-3.98451189]
 [-3.35984927]
 [-2.13892089]
 [-0.05504146]
 [ 1.82207652]
 [-4.18404229]
 [ 0.33436061]
 [ 0.54233484]
 [ 0.49920549]
 [-0.03645738]
 [ 0.75421348]
 [-1.81083965]
 [ 1.64986308]
 [-1.36412006]
 [ 1.55501684]
 [-0.18726724]]

Half sum of squares:  [[46324176.81229524]]
Mean Sqaured Error:  [[133.15447865]]


On the Test Data
RMSE:  [[16.43356092]]
R2:  0.2284549139340426
```

Model	RMSE on Validation Set	R2 on Validation Set
L1 Regularisation λ=1e-9	16.4539	0.22654
L2 Regularisation λ=1e-9	16.433	0.2284
Degree 6 (without regularisation)	16.4244	0.2293

Inference

We observe that the model obtained from L2 regularisation performs slightly better than the model obtained from L1 regularisation. However both models have greater RMSE metrics and lower R^2 metric than the degree 6 model without regularisation. This is in accordance with our results from Part B, that model of degree 6 does not overfit the data. Hence regularisation doesn't improve the model. In fact the model of degree 6 could be underfitting the data.

Comment on the effect of regularization on the loss over the test set.

We observe that regularisation in our case does not reduce the loss over the test set since the model of degree 6 does not overfit the data. However the model obtained from L2 regularisation performs slightly better than the model obtained from L1 regularisation.

Comment on the effect of regularization on overfitting.

Since the model of degree 6 has no overfitting. There is no effect of regularisation in our case. However in case of an overfitting, regularisation could serve us better fit the model.

Compare the loss of a model with a degree 6 polynomial with regularization applied to that of a lower degree polynomial without regularization.

Degree	**RMSE on Validation Set**	**R2 on Validation Set**
1	18.4698	0.02540
2	18.1356	0.06035
3	17.2579	0.14910
4	16.8487	0.18897
5	16.5276	0.21959
6	16.4244	0.2293
6 **L1 Regularisation** λ=1e-9	16.4539	0.22654
L2 Regularisation λ=1e-9	16.433	0.2284

Inference

We observed that the models obtained from regularisation perform better than a lower degree polynomial without regularisation. Among the models prepared the model of degree 6 without regularisation provides the best result.