Abhijay Gupta
Student ID: 260708548

Comp 551 Assignment 1

Question 1: Sampling

1.2

On average the comparison would turn out to be something like this –
Out of 100 days: Movies – 20 days, Comp 551 – 40 days, Playing – 10 days, Studying – 30 days
Out of 1000 days: Movies – 200 days, Comp 551 – 400 days, Playing – 100 days, Studying – 300 days

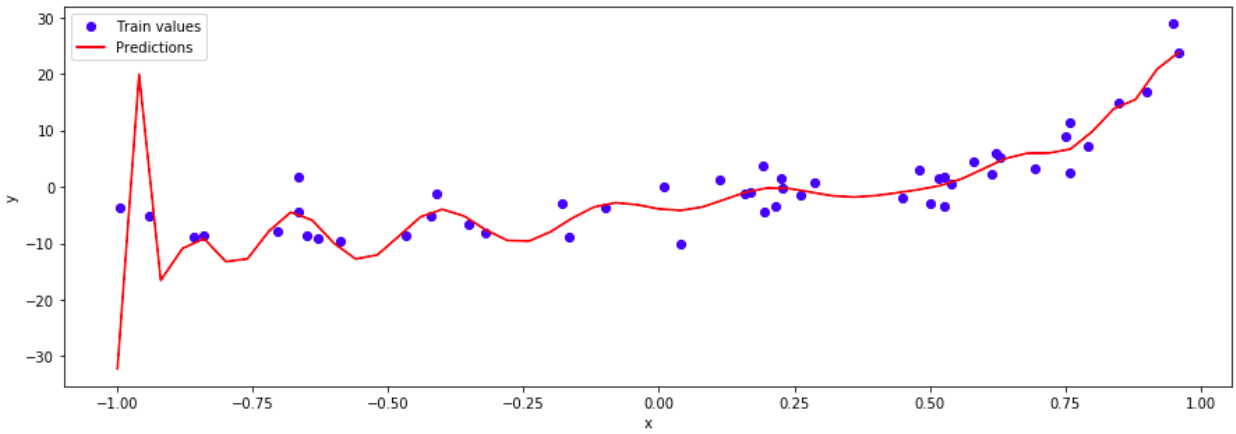
Question 2: Model Selection

2.1

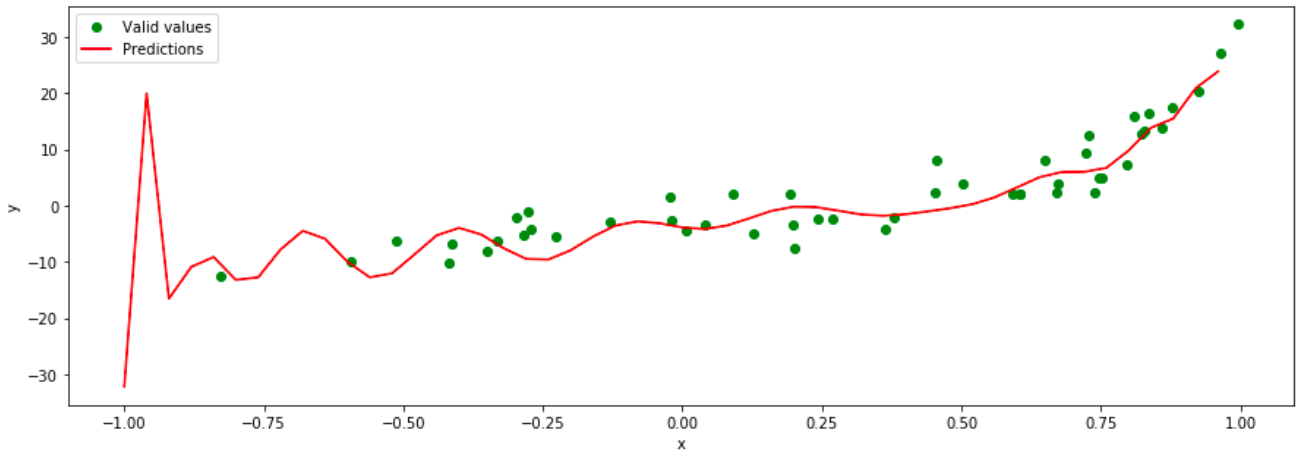
(a)
Training MSE = 6.474766080931443
Validation MSE = 1419.5725161143287b)

(b)

Training set vs. Prediction function



Validation set vs. Prediction function



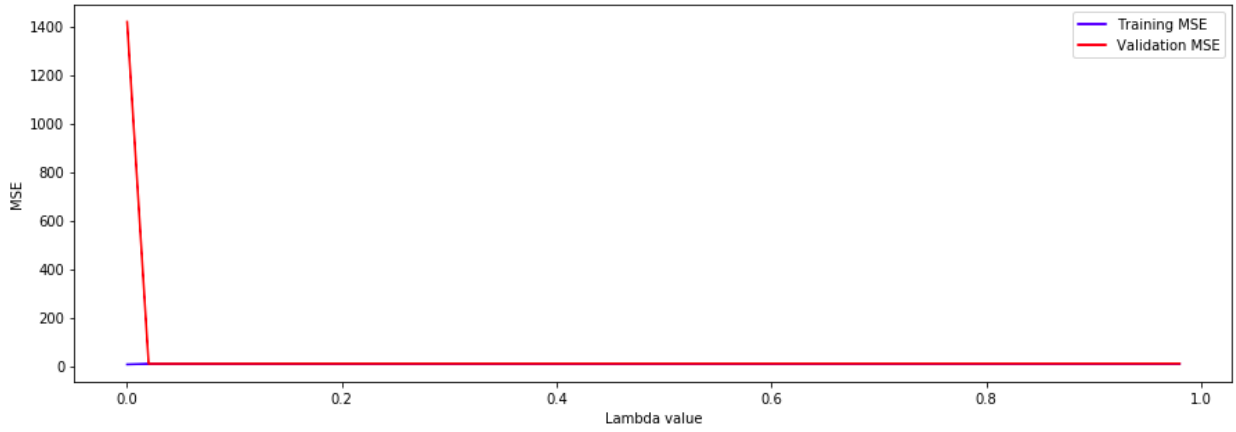
(c)

The model is overfitting since the validation error (1419.5725161143287) is very large compared to the training error (6.474766080931443). This is a clear sign that the model is highly tuned to the training data set since when we introduced the model to new data (from the validation set), the error increased dramatically.

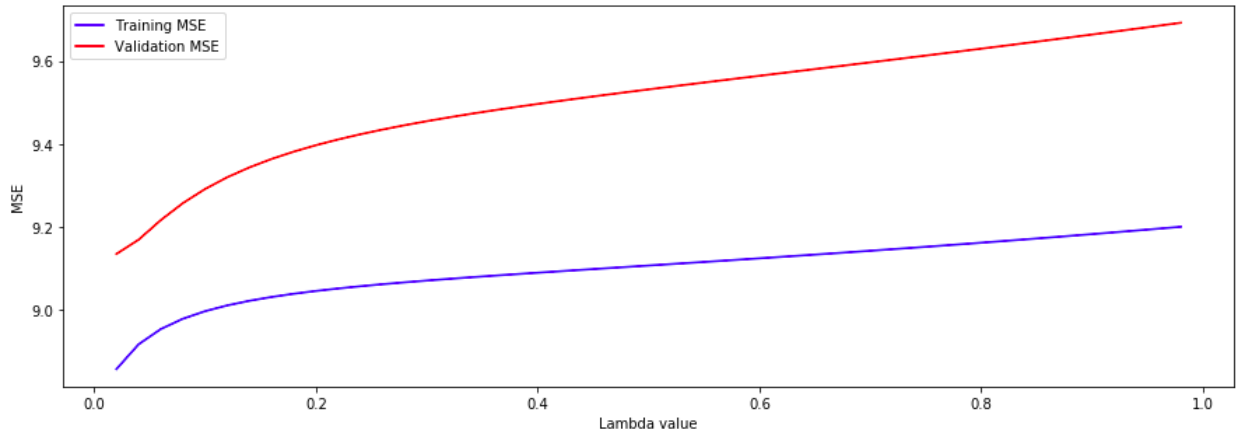
2.2

(a)

Training MSE and Validation MSE vs. Lambda value (Here lambda = 0 is part of the plot)

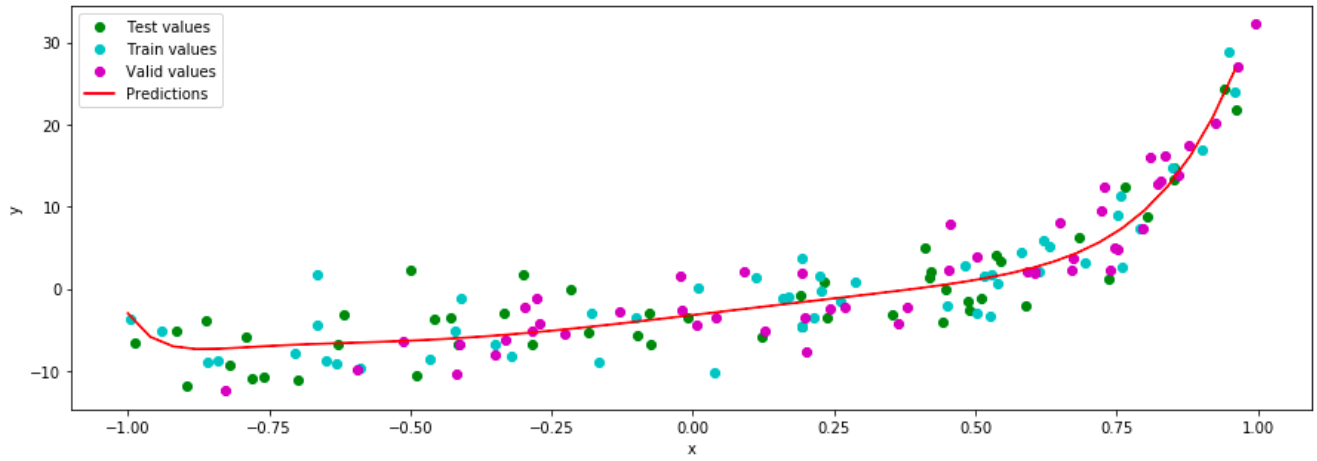


Training MSE and Validation MSE vs. Lambda value (Here lambda = 0 is **NOT** part of the plot)



(b)
Best lambda: 0.02
Test MSE (for corresponding model): 10.73021840092746

(c)



(d)
Best:-
Train MSE: 8.85765667763102
Valid MSE: 9.135098784694307
Test MSE: 10.73021840092746

Model is not overfitting/underfitting since the errors obtained for the 3 datasets are all in the same ball park. None of them is too high or too low compared to the others.

2.3

The degree of the source polynomial from the visualization produced in the previous question SEEMS to be 2. The function seems to have a minimum in the middle region and is increasing on either sides.

The parameters found in the end are thus:

0 : -3.1895175369382422
1 : 8.175883958132497
2 : 0.6112151148010956
3 : -5.4698920901688535
4 : 5.797558883882184
5 : 7.122666099360175
6 : 3.360904669983391
7 : 7.5188097698382546
8 : 1.8298297252901023
9 : 4.9222903402159
10 : 1.7299674708988768
11 : 2.2024896678153985
12 : 1.9859033354714088
13 : 0.06219120025704106
14 : 1.9550070267278925
15 : -1.3879381523052363
16 : 1.456386519831753
17 : -2.218418916326774
18 : 0.5389623554156411
19 : -2.5525946232371037
20 : -0.6736889899083081

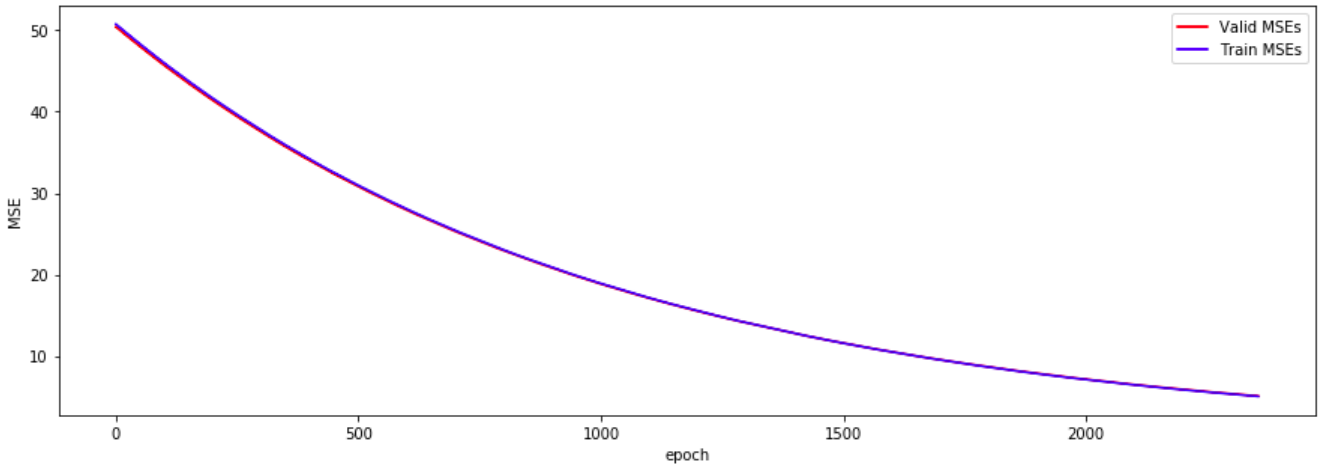
Question 3: Gradient Descent for Regression

3.1

(a)
The values below are in the format “Epoch number: Validation MSE” –

```
0 : 50.40127833857479
1 : 50.34956222298456
2 : 50.300488649391134
3 : 50.252083798481735
4 : 50.205382414140075
5 : 50.15714166032182
6 : 50.10899888977088
7 : 50.05952090412588
8 : 50.01110565054203
9 : 49.964157887698605
10 : 49.91336048597385
.
.
.
.
.
2346 : 5.1312503080151055
2347 : 5.126469334294431
2348 : 5.121596685454074
2349 : 5.116933041016027
2350 : 5.11211699652181
2351 : 5.107220440195193
2352 : 5.102486664471439
2353 : 5.097458354412147
2354 : 5.092746441775124
2355 : 5.087967757007468
```

(b)
Training and Validation MSE for every epoch. The two MSEs differ at the beginning but get closer and closer after every epoch.



3.2

(a)
Validation MSEs for different step sizes (Parameters also included)

	params	step_size	valid_mse
0	[4.297806614755417, 6.186818283084968]	1.000000	5.150194
1	[3.6820307541505404, 4.264015731429388]	0.500000	0.079552
2	[3.6799010650013018, 4.4031969644009585]	0.100000	0.096576
3	[3.585696118518043, 4.339331390811895]	0.050000	0.073554
4	[3.551652288704744, 4.324573494364959]	0.010000	0.074814
5	[3.7001009331773194, 4.192404420193073]	0.005000	0.082322
6	[3.9695614512871518, 3.7711909480284795]	0.001000	0.157460
7	[3.757061087493041, 3.401558371226435]	0.000500	0.546891
8	[3.39551710243344, 3.0692261500338445]	0.000100	1.705296
9	[3.392879474713817, 3.032554963386048]	0.000050	1.794170
10	[3.0653769108044706, 2.7269095296541175]	0.000010	3.576049
11	[2.851271020135174, 2.530437227511242]	0.000001	5.082183

(b)
Test MSE: 0.06981675209679272

Question 4: Real life dataset

4.1

(a)
Filling in the missing values by the mean of that attribute is not the best option. This would lead to repeated value and there would be no variation in the data set.

(b)
abs(mean (+ or - or 0)*SD)
[if this new value exceeds 1, make it equal to the mean]
Absolute value of (mean of attribute -> do nothing, add standard deviation or sub tract standard deviation)

(c)
This can be used to fill in the missing values. It would provide some degree of randomness to our data instead of filling in the same constant value. Since we have a lot of empty data points, taking SD will give us some range. It would better simulate the absent data points and our data set would be much closer to a data set wherein all values were filled in.

(d)
<<Submitting the completed data set>>

4.2

<<Submitted the test MSE and parameters learnt in a txt file>>