FOUNDATION OF DATA SCIENCE
CS F320

ASSIGNMENT 1

Members-

Aayush Hedaoo 2020A3PS0547H
Kaushik Kalakonda 2020A3PS0468H

1-A Prior and Posterior Distributions-

Initially, it is given that s is the probability of a customer liking a new update, and it is assumed that it follows a beta distribution with parameters a,b = (2,2). The posterior distribution of s can be obtained by multiplying prior beta distribution (a, b) and likelihood function (l,m) and normalizing it.

We can observe that the posterior distribution of s will be beta distribution with parameters (a +l,b +m).

Where l,m denote the number of people who liked and disliked the update that is our likelihood distribution is the Bernoulli distribution. In total, we did three surveys so

The beta parameters for the prior and posterior parameters after each survey are

After 1st survey: beta(s, a, b) and beta (s,a+40, b+10)

After 2nd survey: beta(s, a+40, b+10) and beta (s,a+40+13, b+10+17)

After 3rd survey: beta(s, a+40+13,b+10+17) and beta (s,a+40+13+70,b+10+17+30)

Proof of relationship between posterior, prior, and likelihood function-

Assignment-1 :-

posterior distribution — survey likelihood function — prior distribution

$$\overbrace{P\left(S/D\right)}^{\text{posterior distribution}} = \frac{\overbrace{P\left(D/s\right)}^{\text{survey likelihood function}}\ \overbrace{P(S)}^{\text{prior distribution}}}{P(D)}$$

$$P\left(S/D\right) \propto P\left(D/S\right)\, p(S)$$

$$P\left(\frac{D}{S}\right) = S^{f}(1-S)^{m} \quad , \; p(S) = \frac{\gamma(a+b)}{\gamma(a)\gamma(b)} S^{a-1}(1-S)^{b-1}$$

$$P\left(S/D\right) \propto S^{f}(1-S)^{m}\left(\frac{\gamma(a+b)}{\gamma(a)\gamma(b)} S^{a-1}(1-S)^{b-1}\right)$$

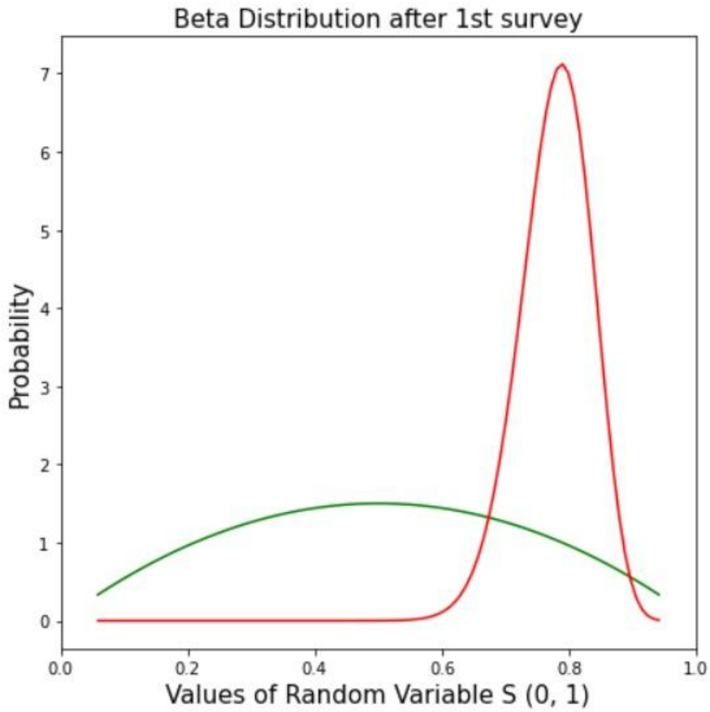
$$P\left(S/D\right) \propto S^{f+a-1}(1-S)^{m+b-1}$$

After normalizing.

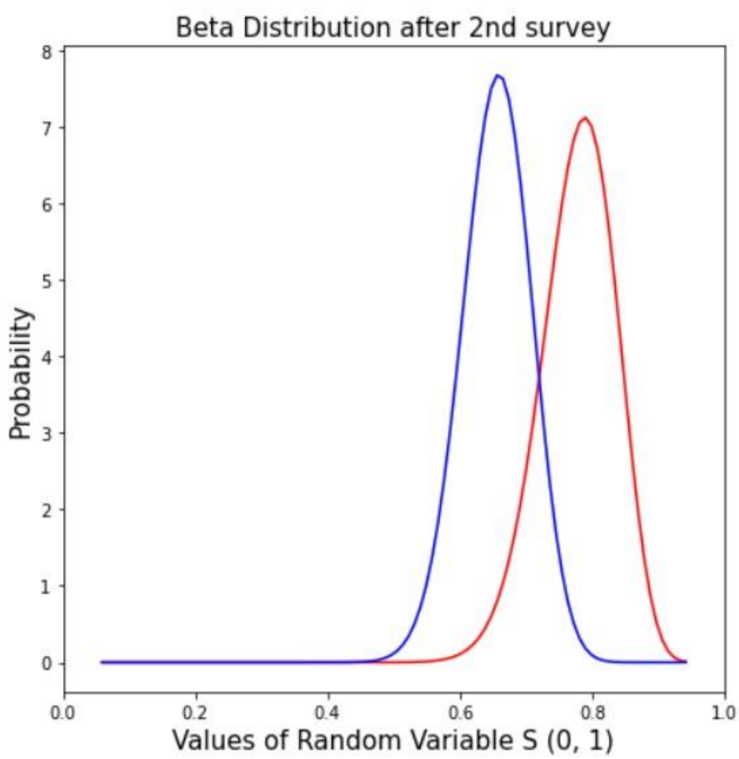
$$P\left(S/D\right) = S^{f+a-1}(1-S^{m+b-1})\ \frac{\gamma(a+f+b+m)}{\gamma(a+f)\gamma(b+m)} = Beta(S,a+f,b+m$$

$$P(S) = Beta(S,a,b)$$

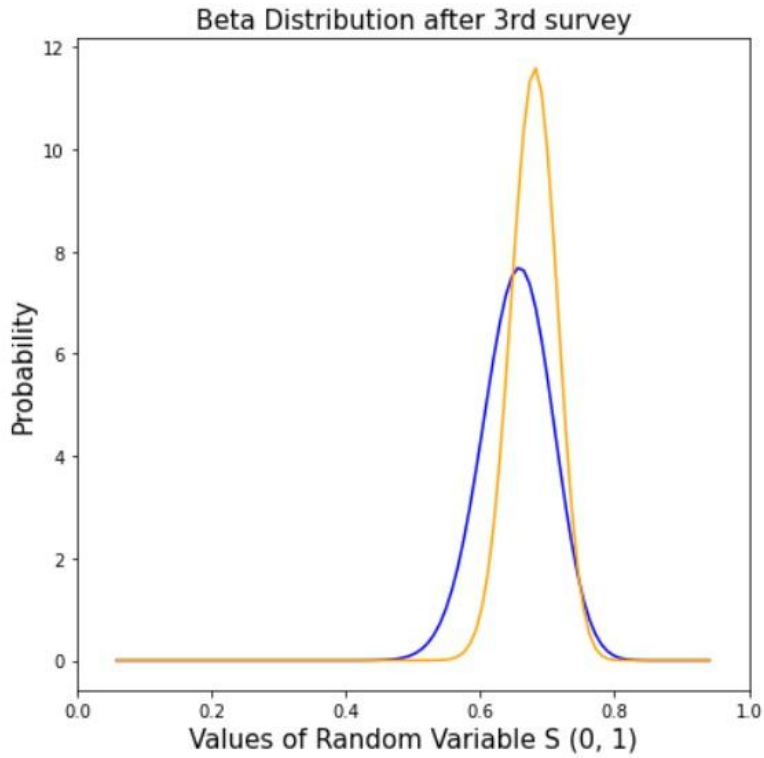
Plot for prior and posterior distribution of 1st survey:



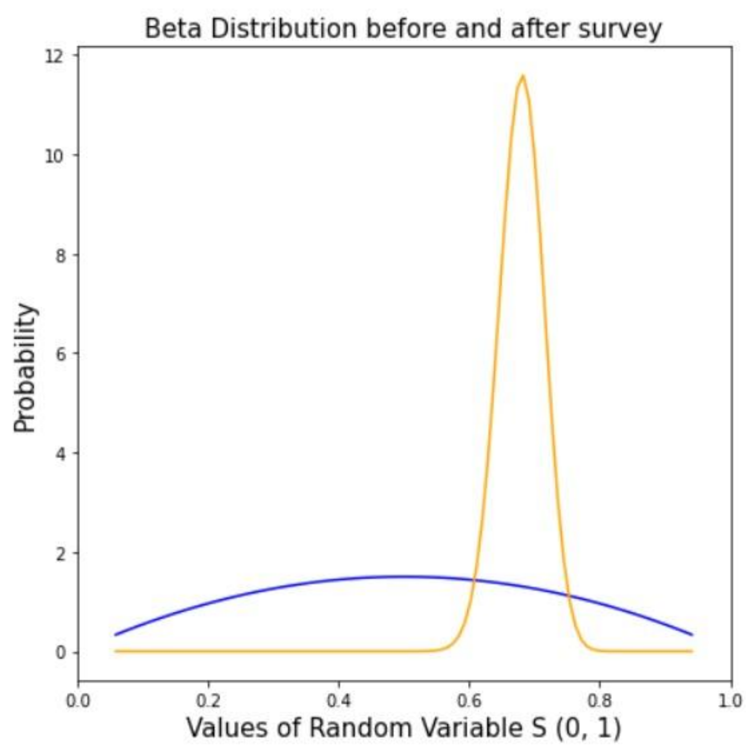
Plot for prior and posterior distribution of 2nd survey:



Plot for prior and posterior distribution of 3rd survey:



Plot for distribution of s before and after survey:

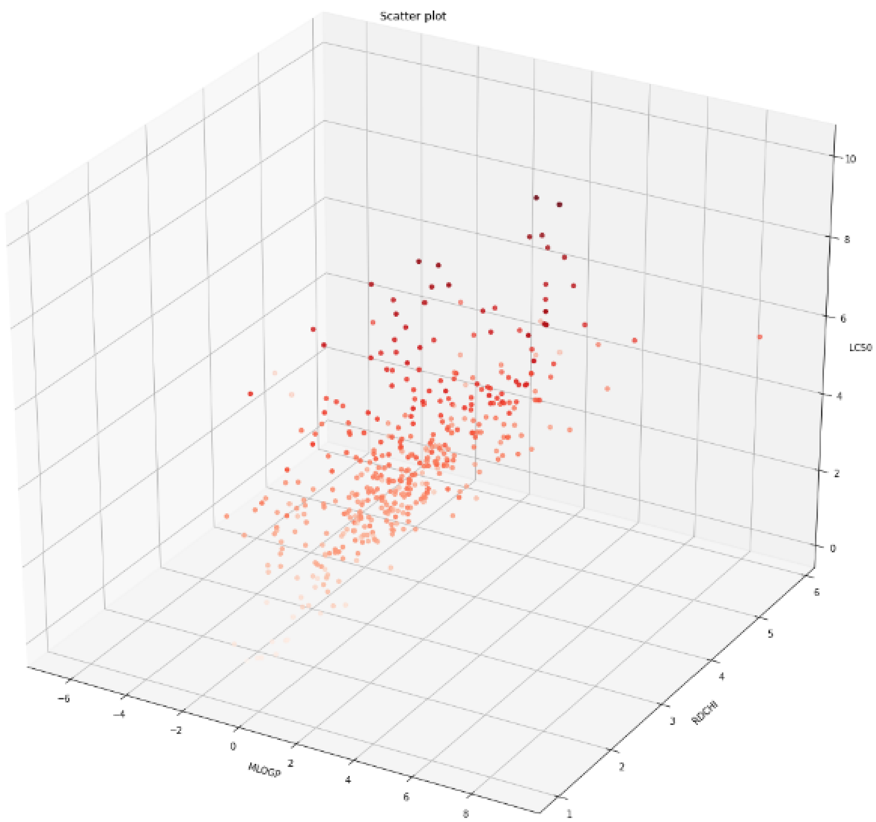


1-B Polynomial Regression and Regularization-

Splitting of the dataset in the ratio of 80:20 to get training and testing data, respectively-

```
p = 0.8                                 # 80-20 split for training-testing
p*X.shape[0]
n_train = math.floor(p * X.shape[0])    #train data
n_test = math.ceil((1-p) * X.shape[0])  #test data
X_train = X[:n_train]                   # MLOGP and RDCHI
y_train = y[:n_train]                   # LC50
```

Scatter plot of data-



Assignment - 1 :-

1 - B) Training data $\Rightarrow (x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

$x_i = (x_{i1}, x_{i2}) \; \forall \, i \in data$

$x_{i1} \rightarrow$ MLOGP, $x_{i2} \rightarrow$ GATS1i

$y_i = LC50$ (Qualitative response of LC50)

$t_n = \sum_{j=0}^{\gamma} w_j \, x_1^{\alpha} x_2^{\beta}$ $\quad \alpha + \beta = 0, 1, \ldots, k$ where k is the assumed degree of model.

$\gamma = \frac{(k+1)(k+2)}{2} = {}^{k+2}C_2$

Loss (or) Error function $= \frac{1}{2} \sum_{n=1}^{N} \left((t_n - y_n)^2 \right) = E$ ($N \leftarrow$ training data points.)

Gradient Descent Algorithm:-

Initialize $(w_0^{(1)}, w_1^{(1)}, \ldots) = (0.1, \ldots)$

$k = 1$, itr = no. of iterations.

Repeat until convergence (or) itr {

$$w_0^{(k+1)} = w_0^{(k)} - \eta \left(\nabla_{w_0} E \right) \Big|_{w = (w_0^{(k)}, w_1^{(k)}, \ldots)}$$

($\eta \leftarrow$ learning rate)

$\vdots$

$$w_j^{(k+1)} = w_j^{(k)} - \eta \left(\nabla_{w_j} E \right) \Big|_{w = (w_0^{(k)}, w_1^{(k)}, \ldots)}$$

till $j \leq {}^{k+2}C_2$

$k = k+1$

} return $(w_0^{(k)}, w_1^{(k)}, \ldots)$

$\rightarrow$ Convergence can be checked by seeing whether our error is increasing or decreasing after updating weights if it is increasing we will stop updating the weights.

Stocastic Gradient Descent:-

→ Shuffling the training data.

→ Initialize $(w_0^{(1)}, w_1^{(1)}, \ldots)$

$k = 1$

Repeat until convergence or no. of iterations (K)

for $(n = 1, 2, \ldots, N)$ {

$$w_0^{(k+1)} = w_0^{(k)} - \eta \left[\nabla_{w_0} \left(\frac{1}{2} (y_n - t_n)^2 \right) \right] \Big|_{w = (w_0^{(k)}, w_1^{(k)}, \ldots)}$$

$$\vdots$$

$$w_j^{(k+1)} = w_j^{(k)} - \eta \left[\nabla_{w_j} \left(\frac{1}{2} (y_n - t_n)^2 \right) \Big|_{w = (w_0^{(k)}, w_1^{(k)}, \ldots)} \right]$$

$k = k + 1$

}

Convergence can be checked by seeing our error function value is increasing or decreasing after updating weights, if error function value is ~~de~~ increasing then we will stop updating weights.

After getting our model for each degree,we will calculate the error or loss function for testing data and the model with the least error value will be our best fit model for the given dataset.

Gradient Descent-

Degree of polynomial	Training error	Testing error
0	0.499	0.495
1	0.347	0.394
2	0.345	0.375
3	0.343	0.373
4	453.653	5151.224
5	2.839e+06	3.735e+07
6	8.9702e+08	1.7056e+10
7	4.694e+12	1.275e+14
8	2.5268e+16	9.729e+17
9	1.397e+20	7.557e+21

Degree 2 polynomial regression gives the least testing error according to Occam's Razor.

Polynomial -
Y = (w0) + (w1)(x1) + (w2)(x2) + (w3)(x1)^2 + (w4)(x2)^2 + (w5)(x1)(x2)

Stochastic Gradient Descent-

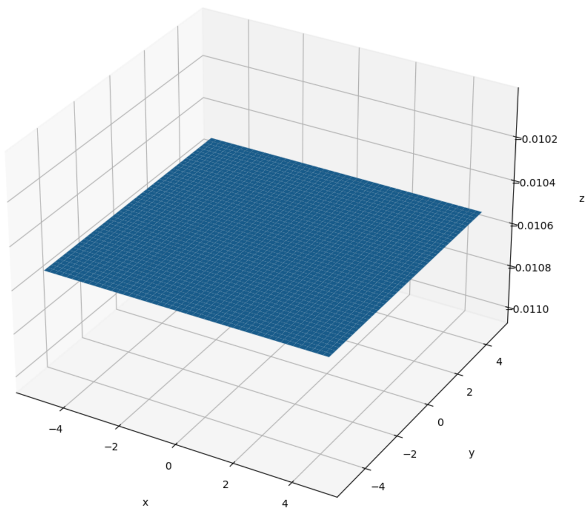
Degree of polynomial	Training error	Testing error
0	13.307	13.570
1	17.765	18.769
2	15.167	9.891
3	119.212	18.764
4	8746.556	80275.5964
5	185468.397	2.577e+06
6	4.1025e+11	7.773e+12
7	8.421e+09	2.558e+11
8	1.752e+12	6.763e+13
9	1.4496e+18	8.200e+19

Training and testing errors are small till degree 4, and larger for remaining degrees. Hence SGD does not give proper answer.

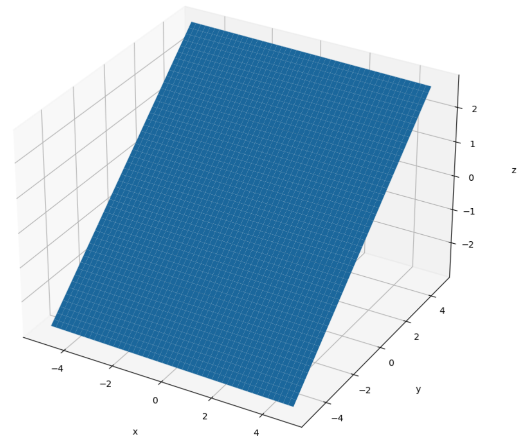
Gradient Descent Plots-

Degree 0

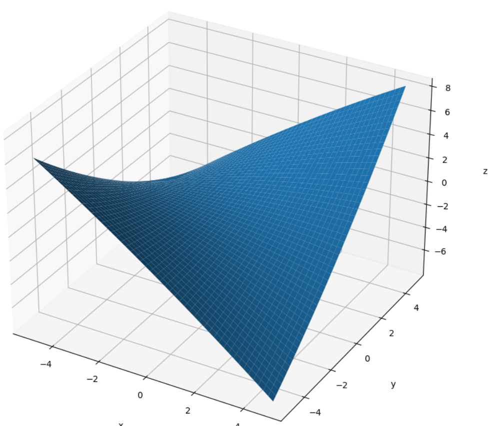
Degree 1



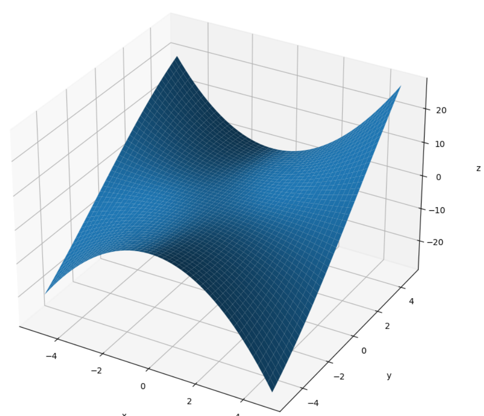
Degree 2



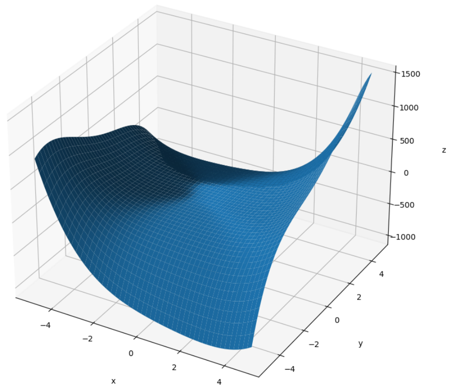
Degree 3

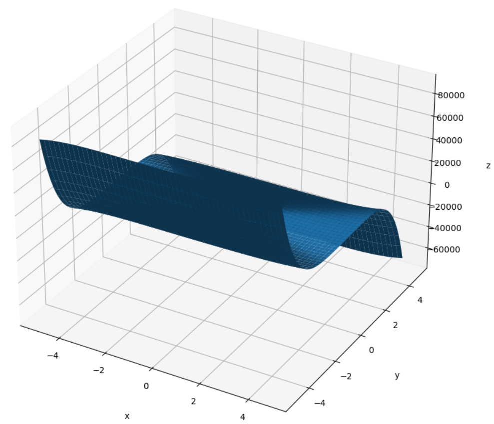


Degree 4



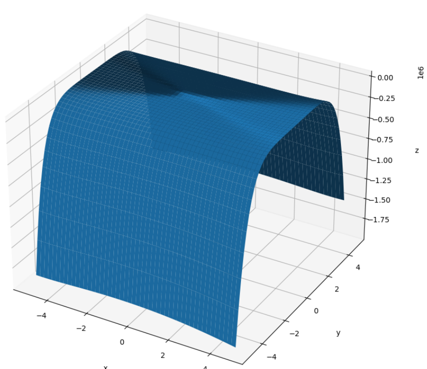
Degree 5

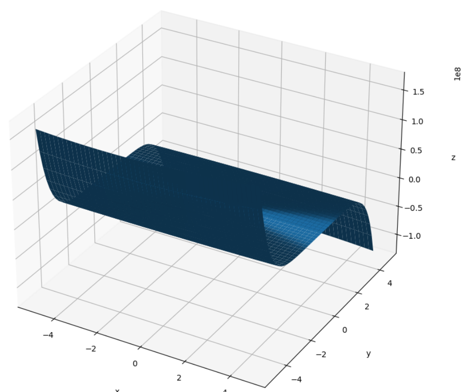




Degree 6

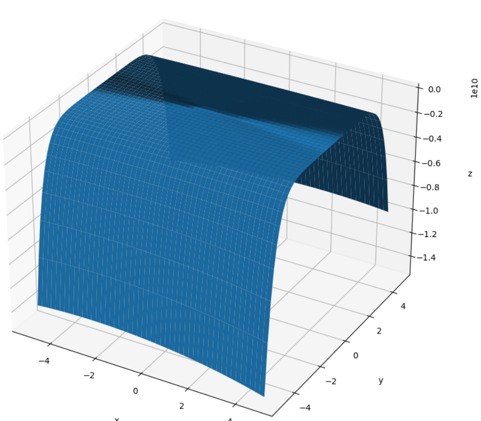
Degree 7

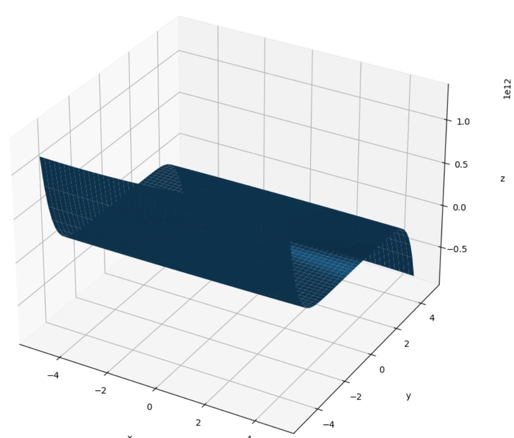




Degree 8

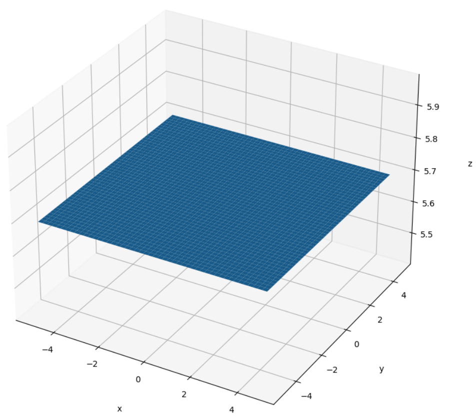
Degree 9



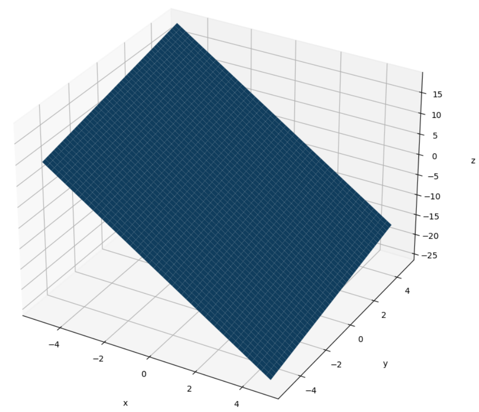


Stochastic Gradient Descent Plots-

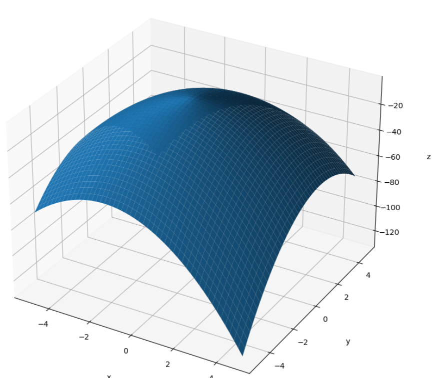
Degree 0



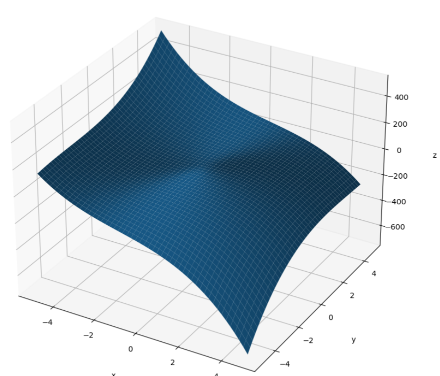
Degree 1



Degree 2

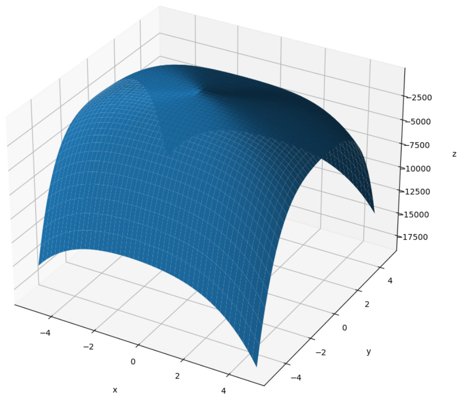


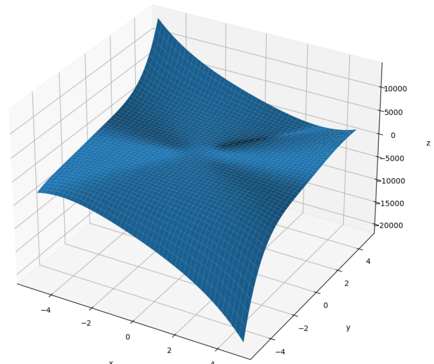
Degree 3



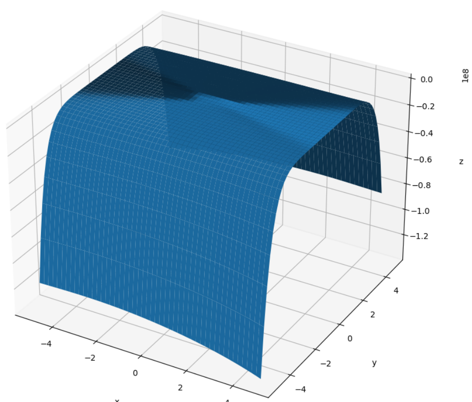
Degree 4

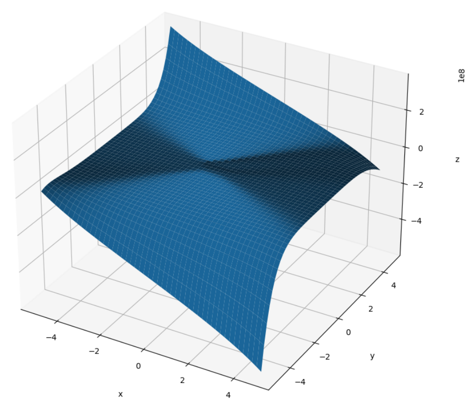
Degree 5



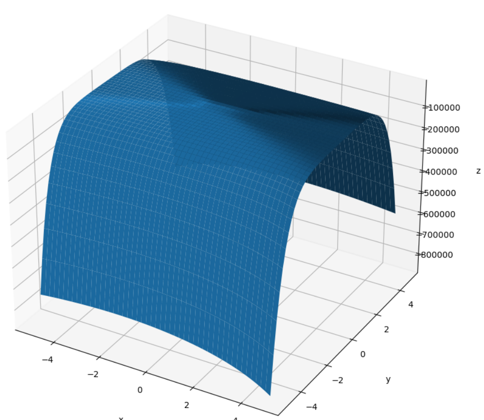


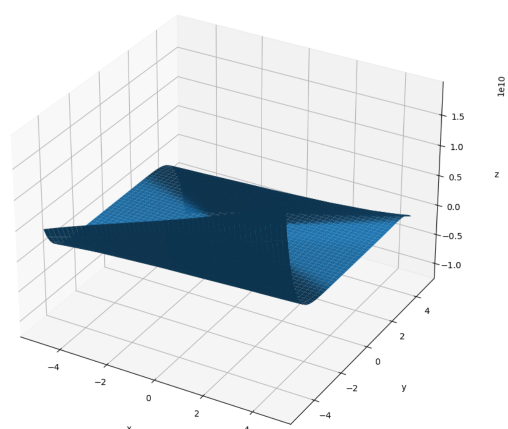
Degree 6 Degree 7





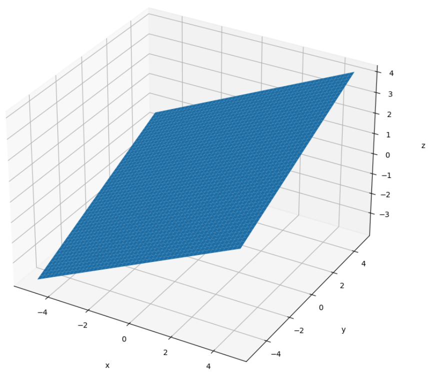
Degree 8 Degree 9

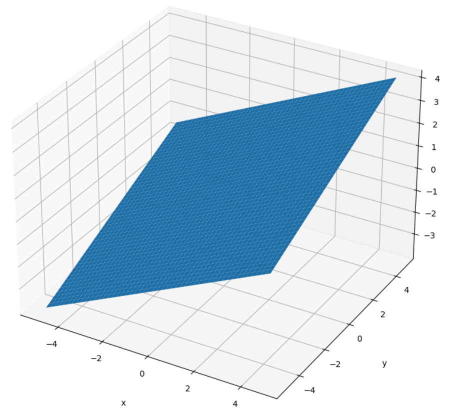




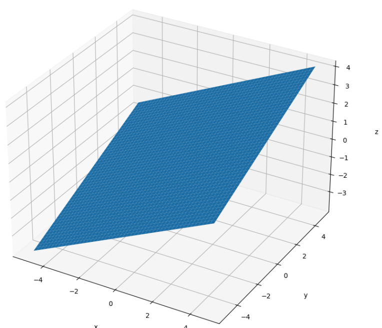
<u>Regularization-</u>

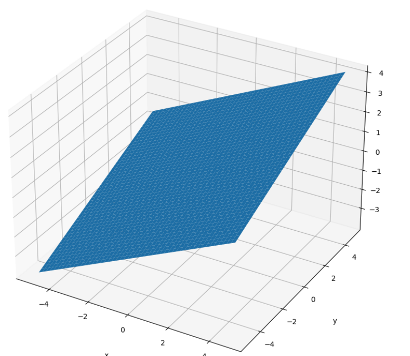
q=0.5,λ=1 q=0.5,λ=0



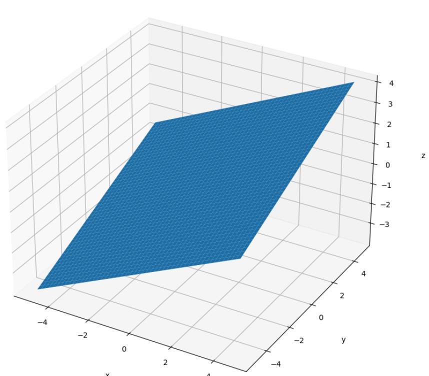


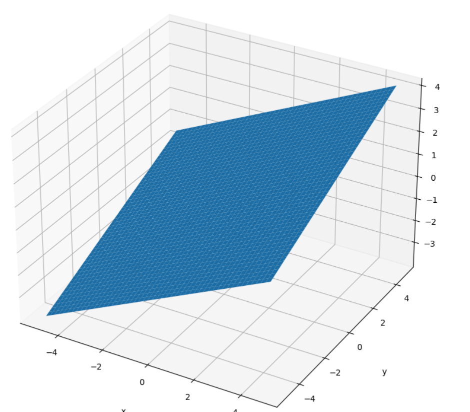
q=0.5,λ=0.01 q=0.5,λ=0.0001



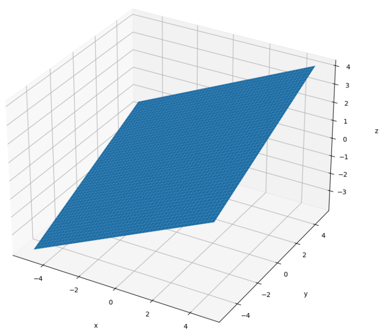
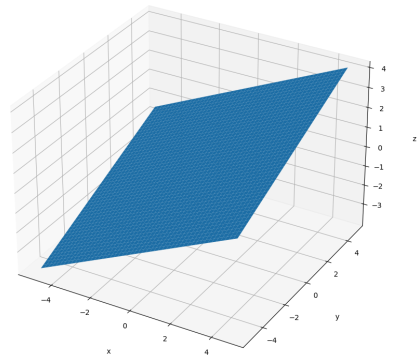


q=1,λ=1 q=1,λ=0

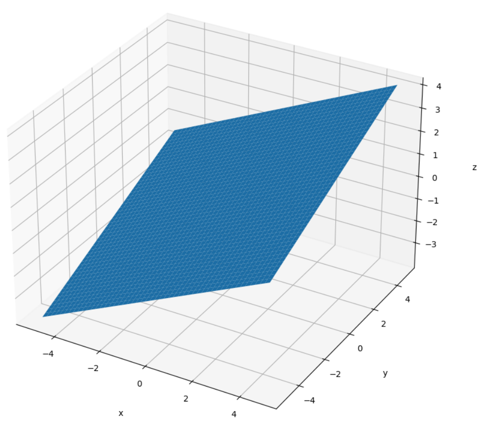
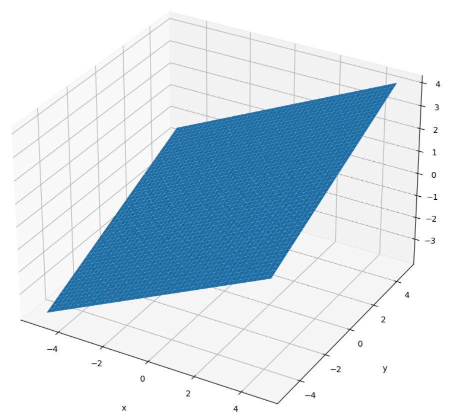




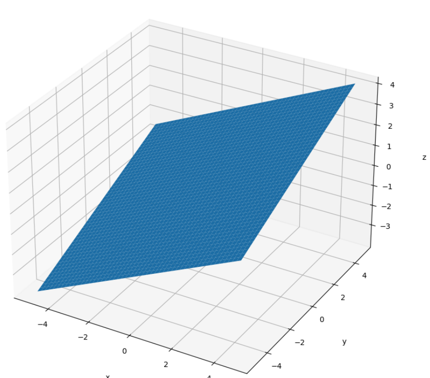
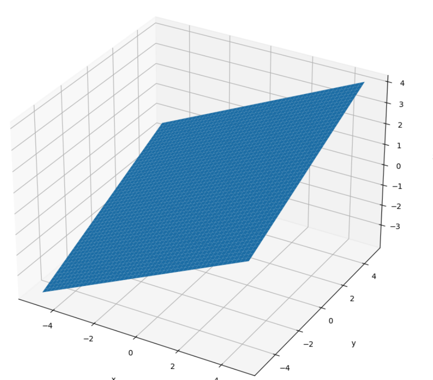
q=1,λ=0.01 q=1,λ=0.0001

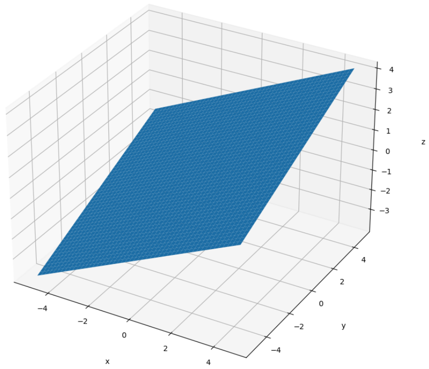
q=2,λ=1 q=2,λ=0

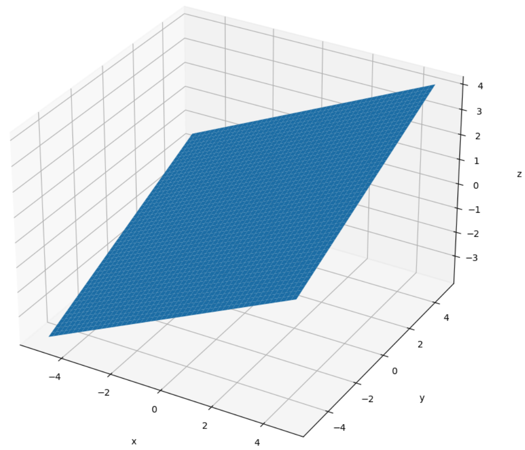



q=2,λ=0.01 q=2,λ=0.0001

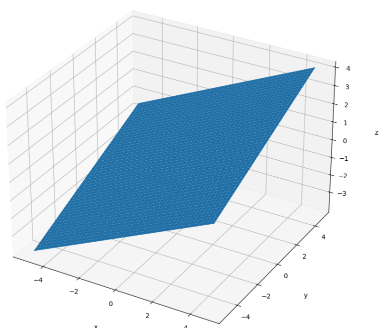
q=4,λ=1 q=4,λ=0

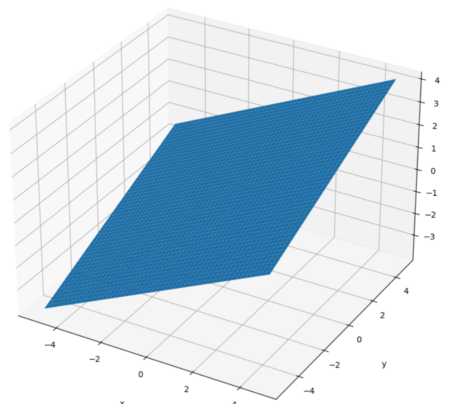




q=4,λ=0.01

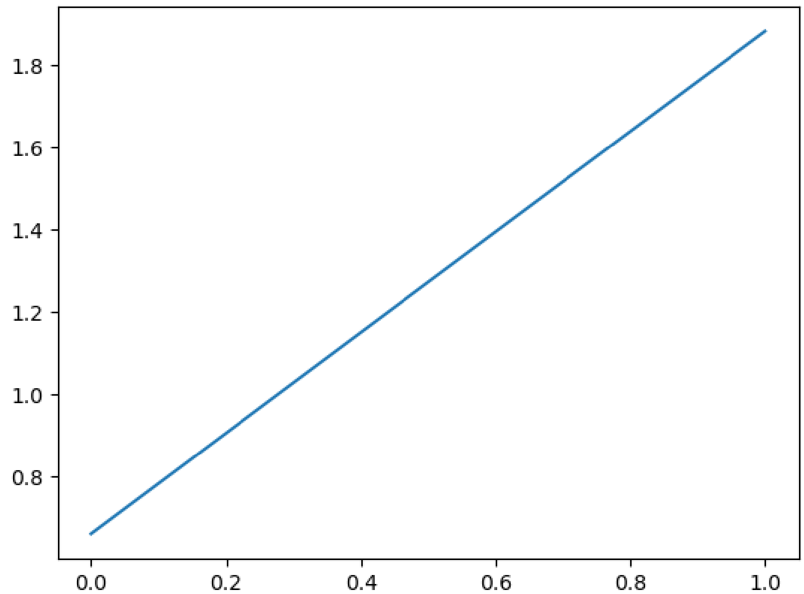
q=4,λ=0.0001



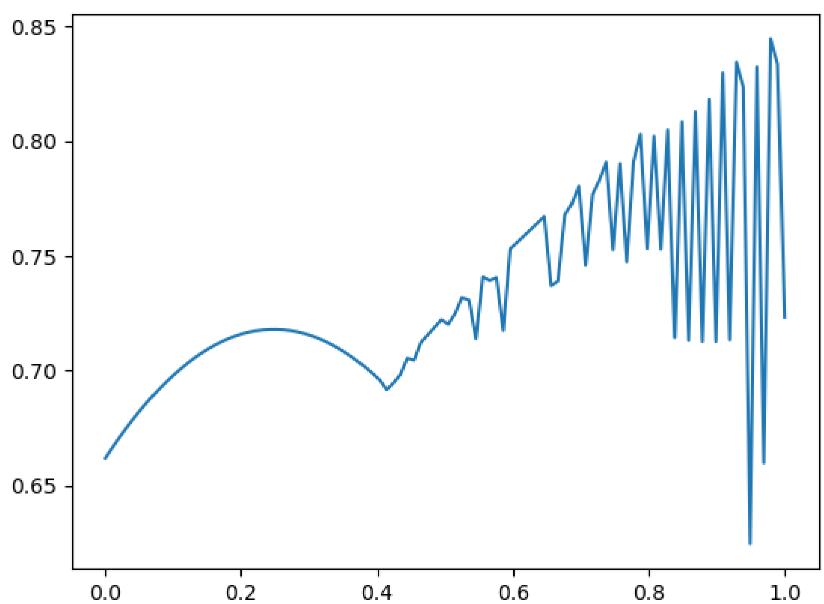


Cost vs λ plots-

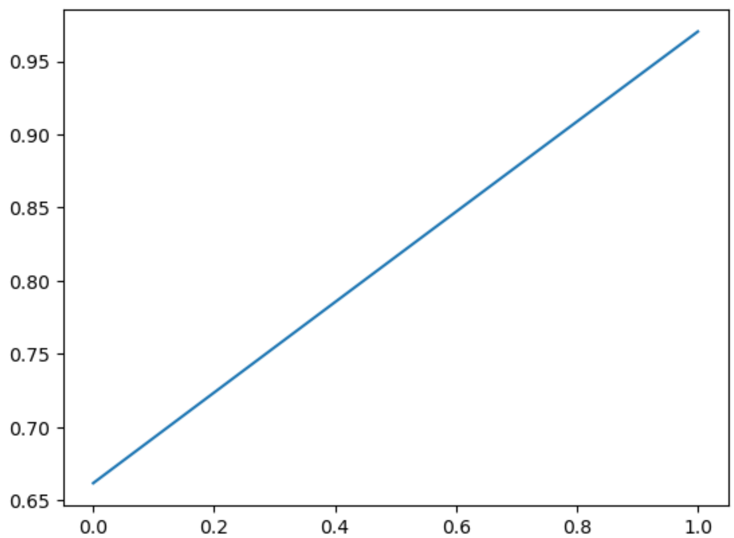
q=0.5



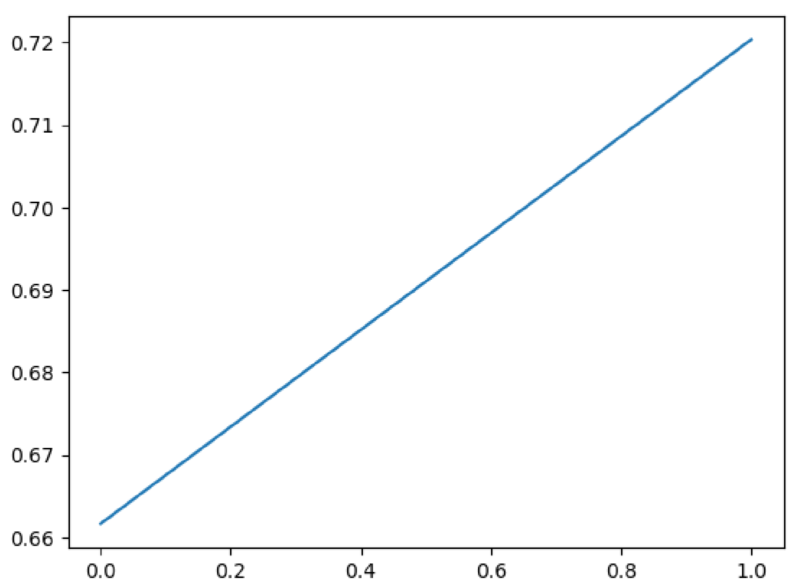
q=1



q=2

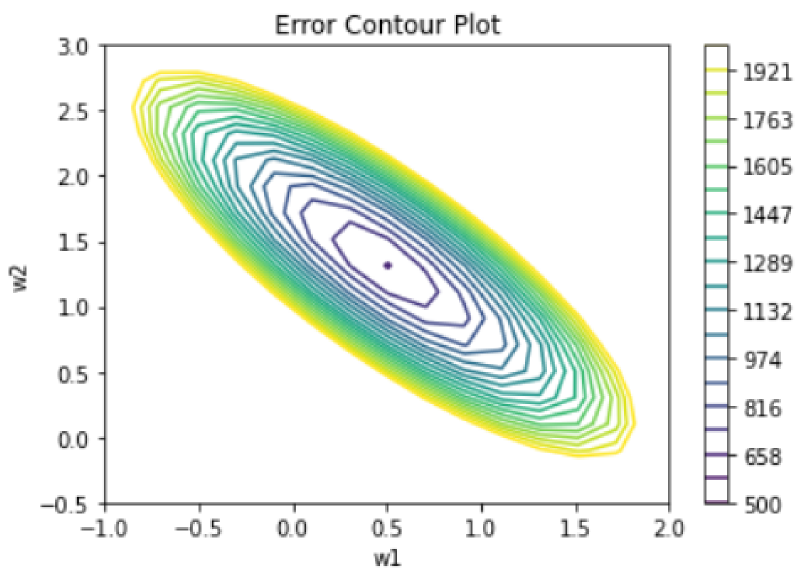

0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.0
0.2
0.4
0.6
0.8
1.0

q=4

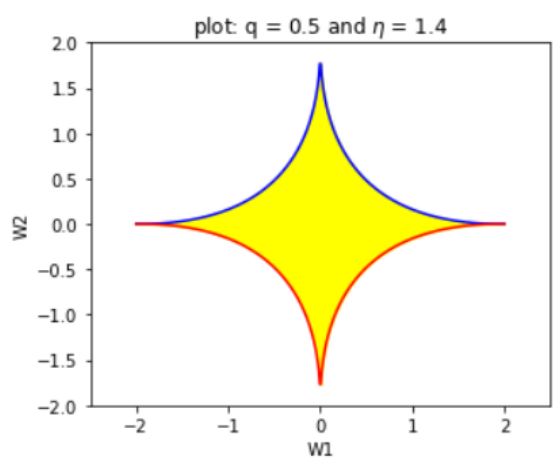

0.72
0.71
0.70
0.69
0.68
0.67
0.66
0.0
0.2
0.4
0.6
0.8
1.0

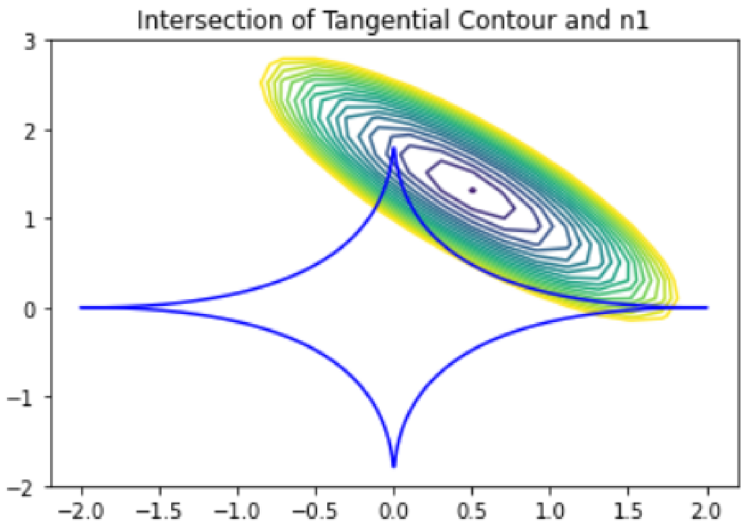
1-C Visualizing Regularization-

1) Error contour plot-

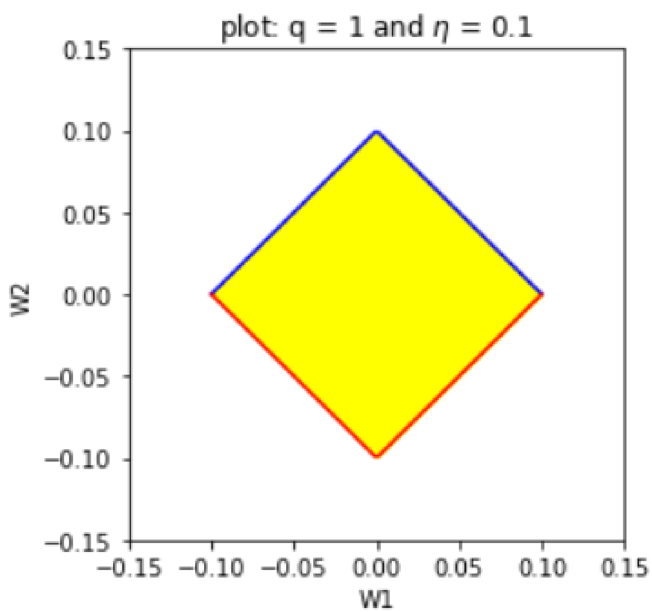


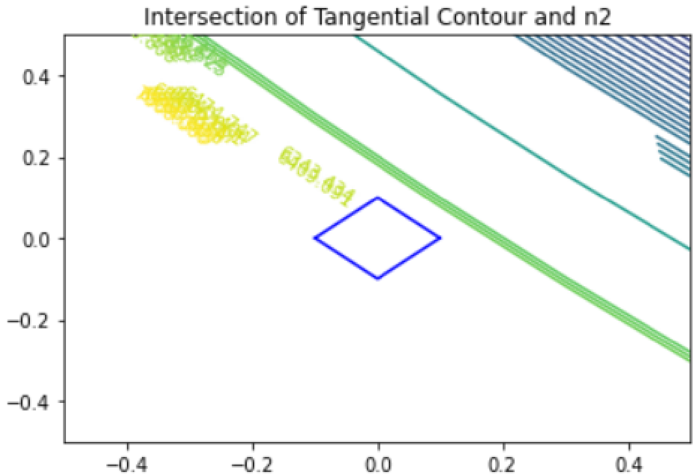
2) Constraint regions and Intersection-
q=0.5 ,n=1.4



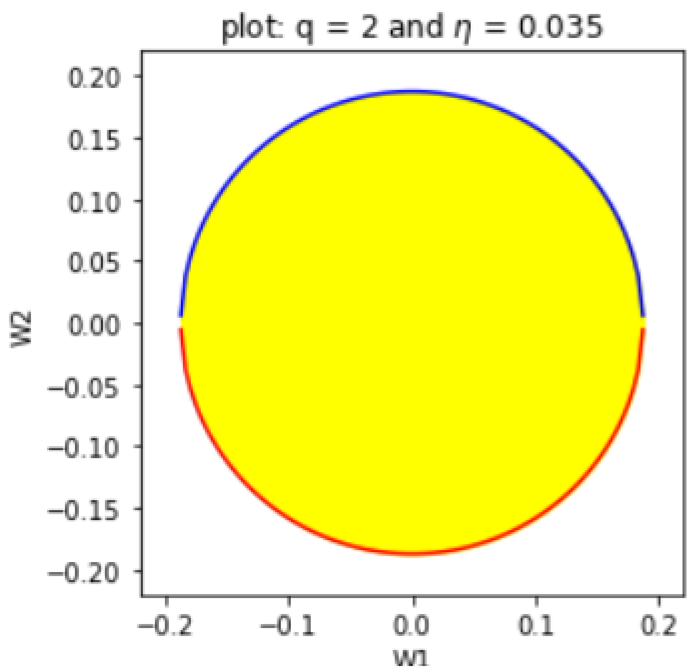

Intersection of Tangential Contour and n1
3
2
1
0
−1
−2
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

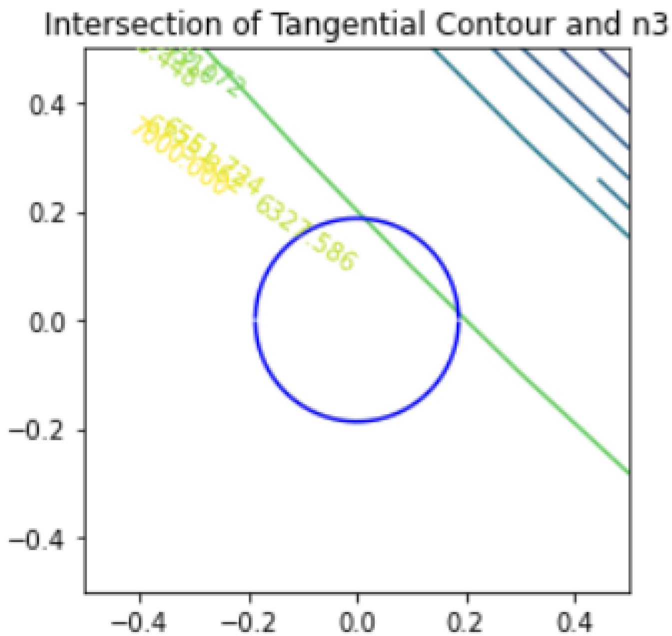
q=1, n=0.1


plot: q = 1 and η = 0.1
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
W2
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
W1

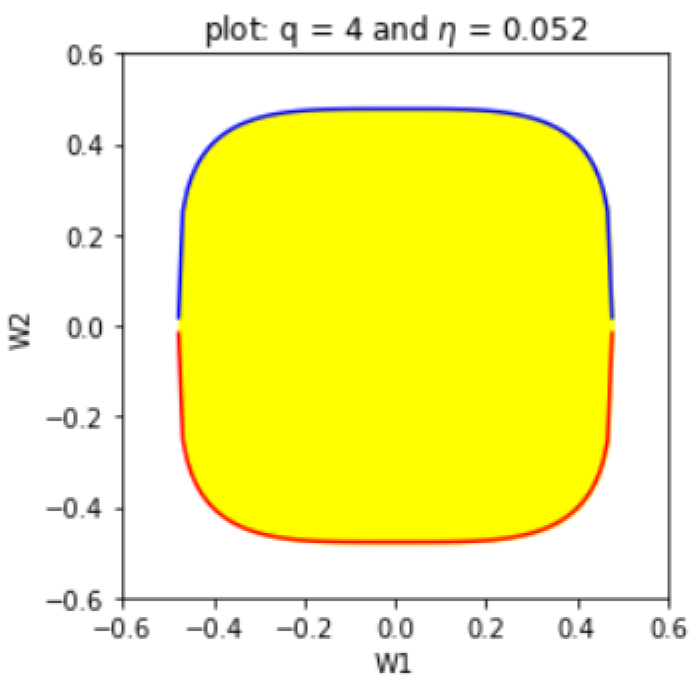

Intersection of Tangential Contour and n2

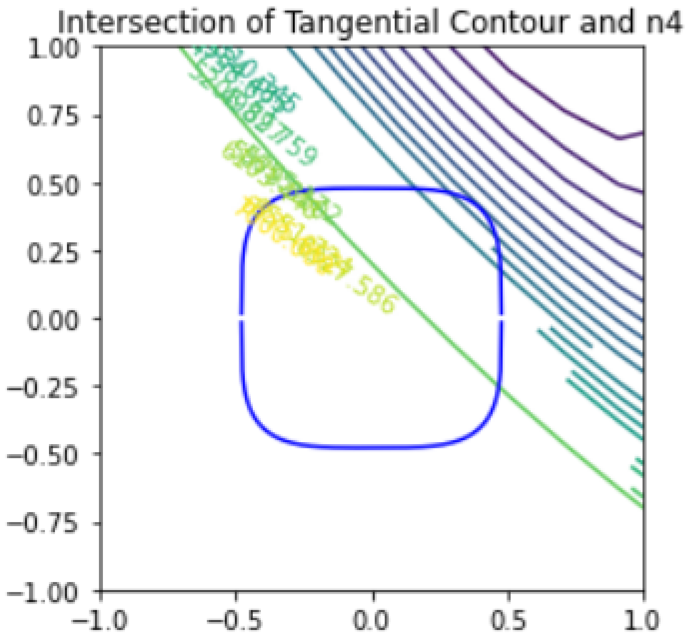
q=2, n=0.035


plot: q = 2 and η = 0.035
W2
W1


Intersection of Tangential Contour and n3
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4

q=4, n=0.052


plot: q = 4 and η = 0.052
W2
W1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6



3) Mean Squared Errors-

```
Mean Squared Error in case 1 is: 1.5034563324428771
Mean Squared Error in case 2 is: 1.516846676955036
Mean Squared Error in case 3 is: 1.5168466769550364
Mean Squared Error in case 4 is: 1.5168466769550444
```