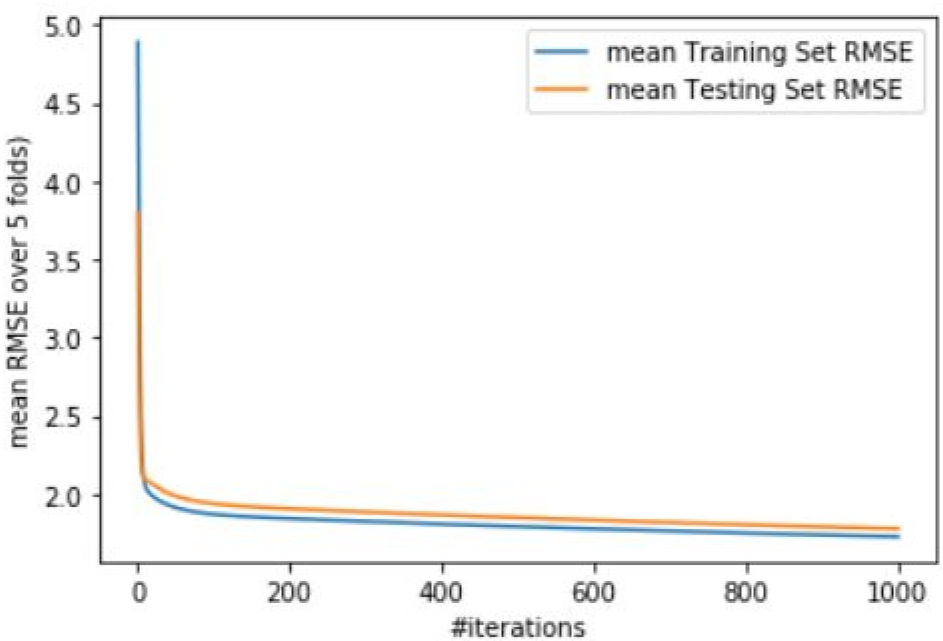
ML Assignment 1 Report: { 2017218, Anand }

1) Linear Regression

1.1a) RMSE vs Gradient Descent for 100 iterations on 5 folds:

```
Fold [ 1 ]
Training Error:  1.5185236353010159
Testing Error :  2.630491901006289
Fold [ 2 ]
Training Error:  1.8289740034115824
Testing Error :  1.365511062502657
Fold [ 3 ]
Training Error:  1.7115200228623348
Testing Error :  1.8718501996807793
Fold [ 4 ]
Training Error:  1.7956903756241243
Testing Error :  1.4821202469779067
Fold [ 5 ]
Training Error:  1.7783364260757404
Testing Error :  1.5376802391846953
Mean Training Error:  1.7266088926549596
Mean Testing Error :  1.7775307298704655
```



1.1b) Normal Equation for Linear Regression RMSEs:

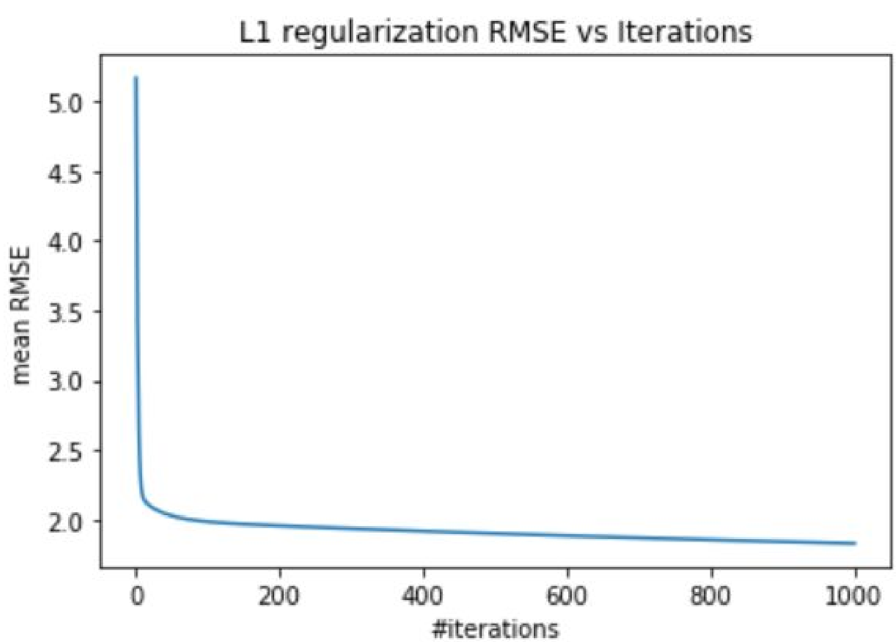
Fold #	RMSE Training	RMSE Testing
1	1.3853232925460048	2.245990457100311
2	1.6441264724086302	1.255424346323494
3	1.5347771731706987	1.7266986106603013
4	1.6110657101936057	1.3862608396230545
5	1.5996648854815512	1.4377242653609477

1.1c) RMSEs obtained over 5 folds using Gradient Descent methods are higher as the Gradient Descent has not converged fully after 100 iterations, it is decreasing very slowly and requires more iterations to become equal to the RMSEs obtained using normal equation.

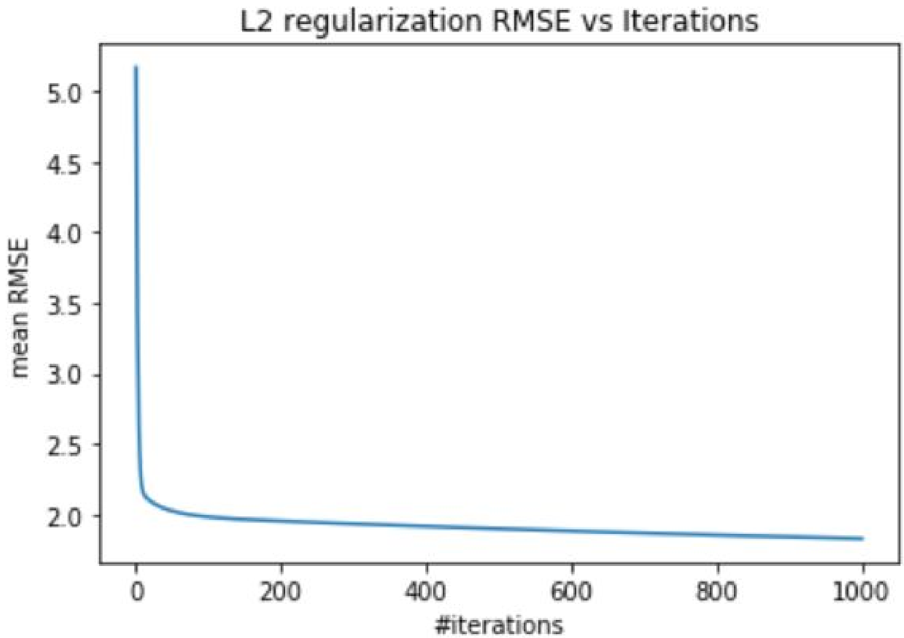
1.2) Regularization

1.2a) L1 optimal regularization parameter = 0.001
b) L2 optimal regularization parameter = 1

RMSE on Train+Val Set using L1 regularization: 1.8289746563442435

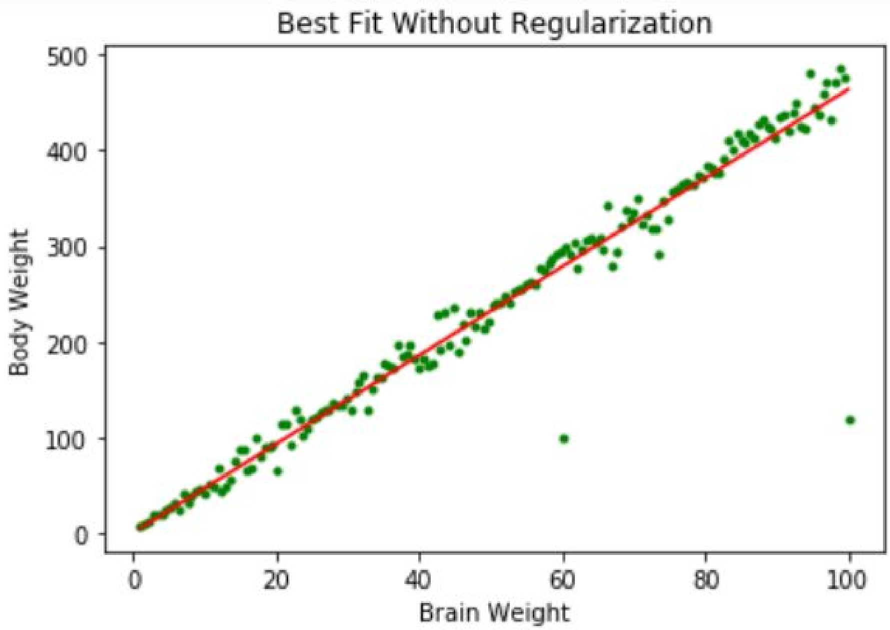


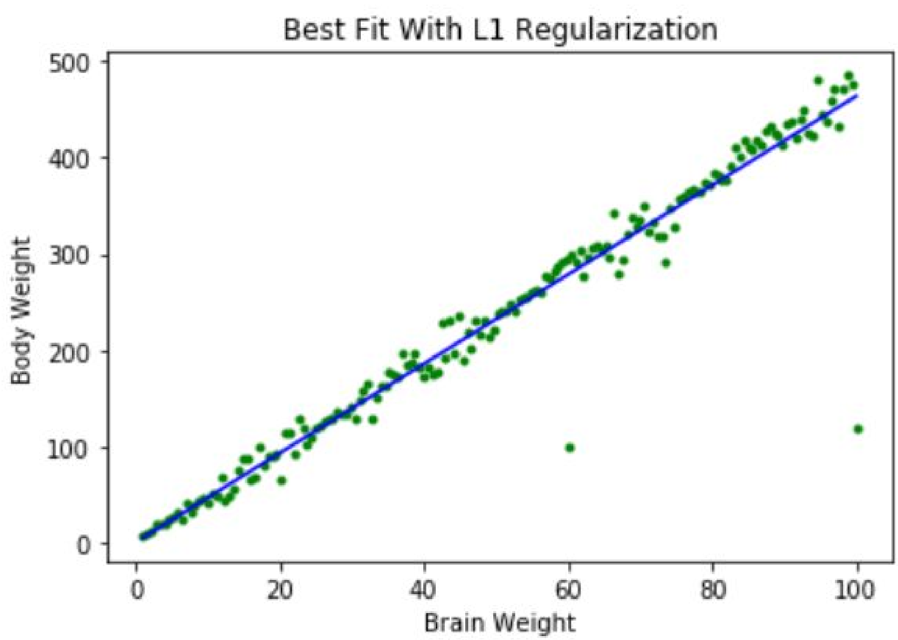
RMSE on Train+Val Set using L2 regularization: 1.8310346970256142

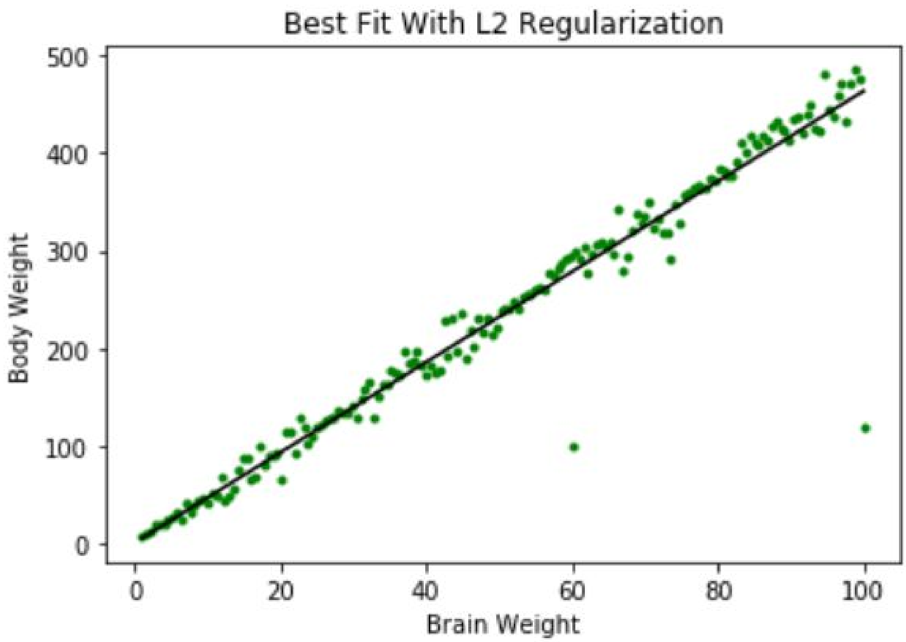


- **RMSE on Test Data using L1:** 0.8000915671369132
- **RMSE on Test Data using L2:** 0.8013507825994648

1.3) Best Fit Line





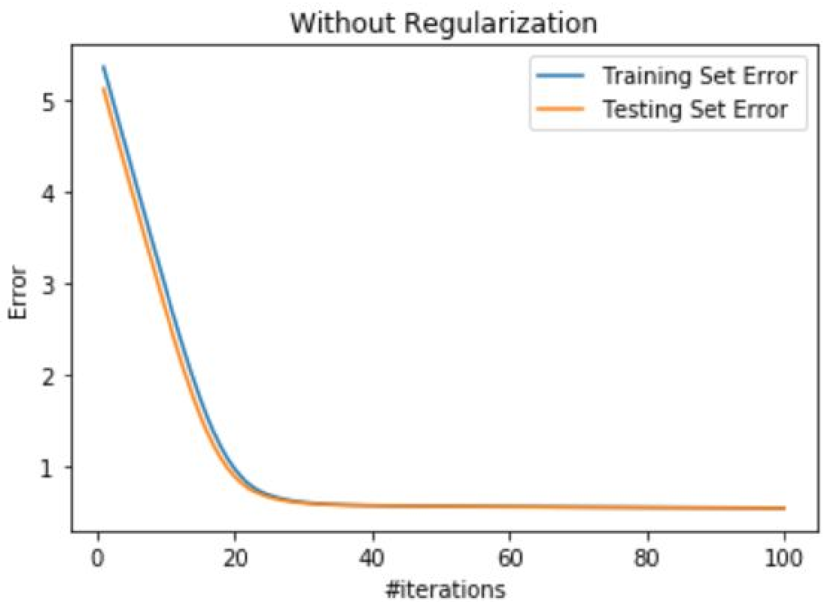


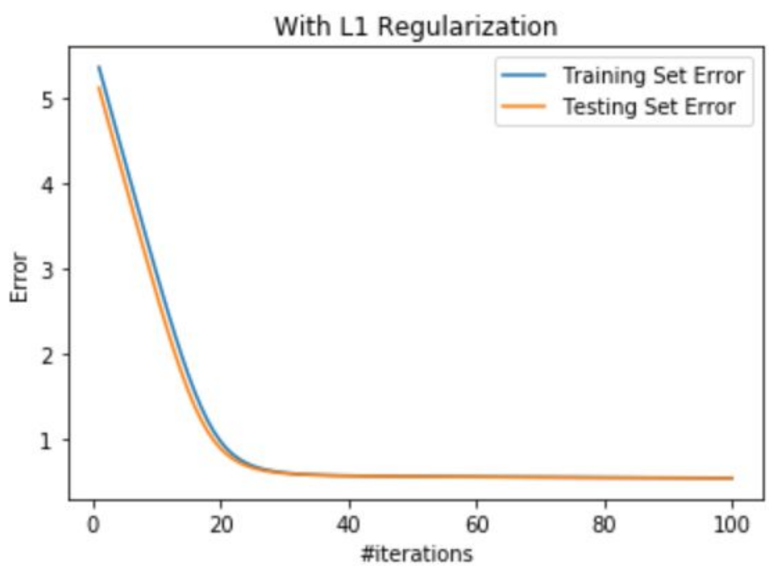
Without Regularization RMSE is: 23.241507111052897
With L1 Regularization RMSE is: 23.24162374784743
With L2 Regularization RMSE is: 23.241647293550106
Without Regularization Parameters: [1.06331913 4.62164566]
L1 Regularization Parameters   : [1.05742645 4.62155399]
L2 Regularization Parameters   : [1.05740731 4.62091247]

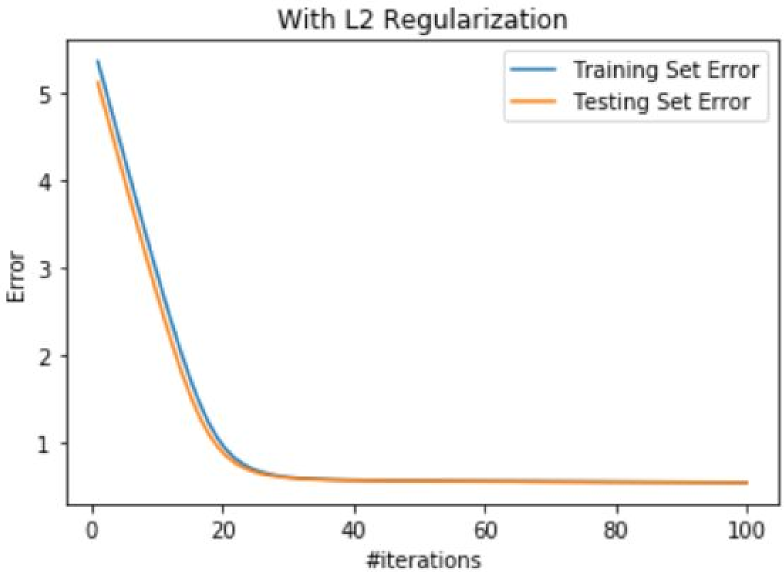
Conclusion for best fit line: It doesn't change much visually, but the error increases after using L1, L2 regularization, being the most in L2. Hence, L2 gave the worst fit, and Unregularized fit was the best.

2. Logistic Regression

L1 Regularisation performs better here because in L2 regression, the coefficients are just reduced factor of the coefficients in case of no regularization. But in L1, coefficients within some range can become zero, this would essentially remove unnecessary features. But in our case the difference is too small and therefore both work almost the same.







With L2 Regularization
Training Set Error
Testing Set Error
Error
#iterations
0
20
40
60
80
100
1
2
3
4
5

ANAND
2017218

ML Assignment 1

Ans: 1 (i) $P(y=1 \mid x, w) = g(w_0 + w_1 x)$

where $g(z) = \frac{1}{1+e^{-(z)}} = \frac{1}{1+e^{-(w_0+w_1 x)}}$

where $x \in (-\infty, \infty)$

$\therefore \; g(z) \in (0,1)$

range of P = range of $g(z)$

$\therefore \; \underline{0 < P < 1}$

(ii) $l(x) = \log(\text{odds}) = \log\left(\frac{P}{1-P}\right)$ where $P \in (0,1)$

$\therefore$ odds $\in (0, \infty)$

$\therefore \; l(x) = \text{logit}(x)$.

$\therefore$ range of $l(x)$ is $(-\infty, \infty)$ $\forall x = [0,1]$.

Ans: 4 (i) $MAE = \frac{1}{m} \sum_{i=0}^{m} \left| h_\theta(x^{(i)}) - y^{(i)} \right|$

$MSE = \frac{1}{2m} \sum_{i=0}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

$\frac{\partial}{\partial \theta_j}(MAE) = \frac{1}{m} \sum_{i=0}^{m} \left(\left(\text{sign}\left(h_\theta(x^{(i)}) - y^{(i)} \right) \right) x_j^{(i)} \right)$

→ MAE is not differentiable at $h_\theta(x^{(i)}) = y^{(i)}$

→ If we use MAE during optimization & , the gradient magnitude only depends on the sign of $(h_\theta(x^{(i)}) - y^{(i)})$, which can increase gradient magnitude, even if the error is small, therefore, the algorithm may find it hard to converge.

(ii) We use MAE when we don't want the outliers influence our error/algoritm significantly.
(MSE will increase the error significantly if outliers are present).
Also, if the data is noisy, MAE performs better.

(iii) Quantile loss performs better than MAE if there are some outliers in our data. It is also useful when we want to predict intervals instead of a single value.

☆ Quantile loss helps to discover better relationship b/w variables.