Assignment 1

Machine Learning

- Mayank Chauhan, MT18008

1. Linear Regression

1.1. Linear Regression on Abalone dataset

a) Root mean squared error (RMSE) vs. gradient descent iterations

This dataset has 8 features and 1 output label. The output label plus 1.5 indicates the age of abalone. The model used is linear regression, which is trained using gradient descent algorithm. First, the dataset is divided into five folds. Then, rmse is calculated for different no. of iterations of gradient descent. Mean rmse at each iteration, across all the five folds.

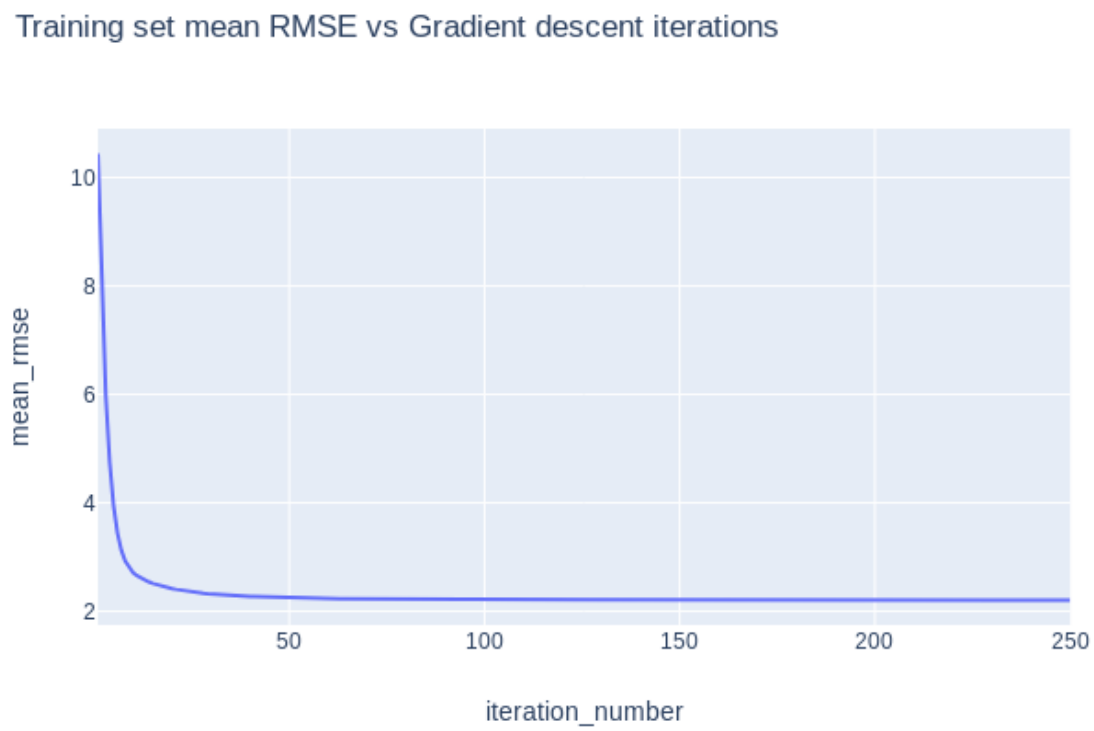


Figure 1: Trainin set mean rmse vs gradient descent iterations

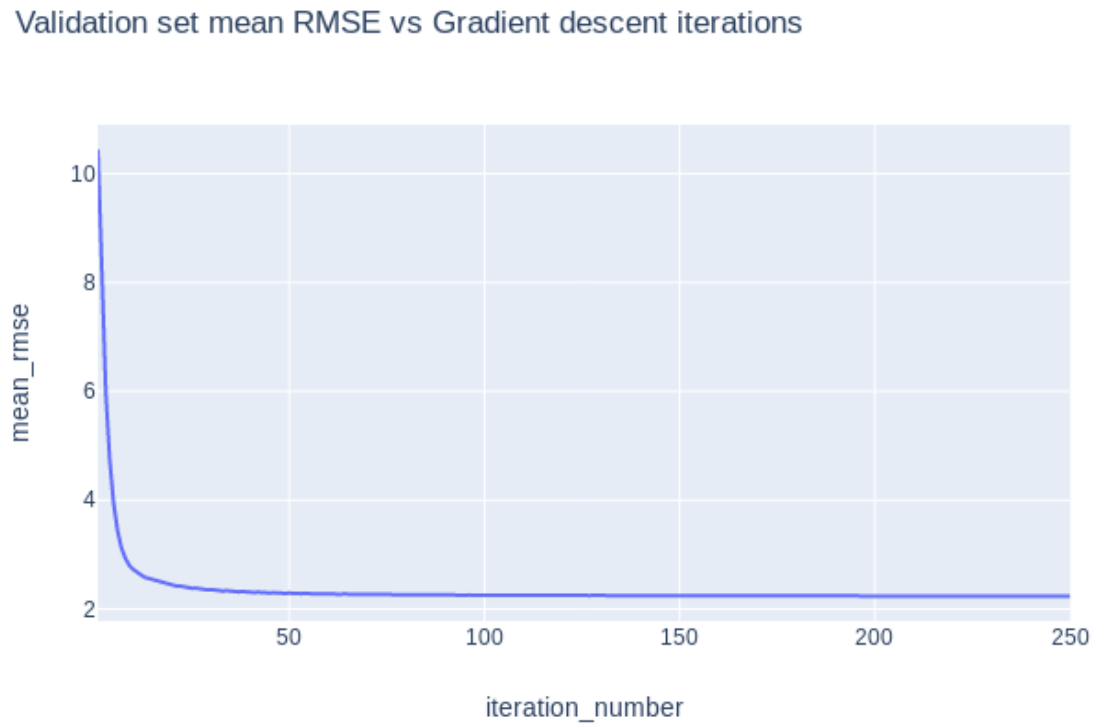


Figure 2: Validation set mean rmse vs gradient descent iterations

b) Implementing using normal equations

Table 1: RMSE for all the folds using Normal Equations

K-Fold Number	Training RMSE	Validation RMSE
0	2.195168	2.185749
1	2.173975	2.340820
2	2.188144	2.210625
3	2.171389	2.279727
4	2.219315	2.079537

b) Comparing RMSE of gradient descent and normal equations

Table 2: Comparing RMSE of both the methods

	Training	Validation
Average Training (normal eq.)	2.189598	2.219291
Gradient Descent (after convergence)	2.205675	2.236787

The rmse obtained using normal is lower than what was obtained using gradient descent. Because gradient descent is an optimization method, and the result depends on the no. of iterations used.

1.2. Regularization

Optimal regularization parameter value for Ridge and Lasso is 2.25 and 0.002 respectively.

RMSE on test set using Ridge: 2.103649

RMSE on test set using Lasso: 2.102273

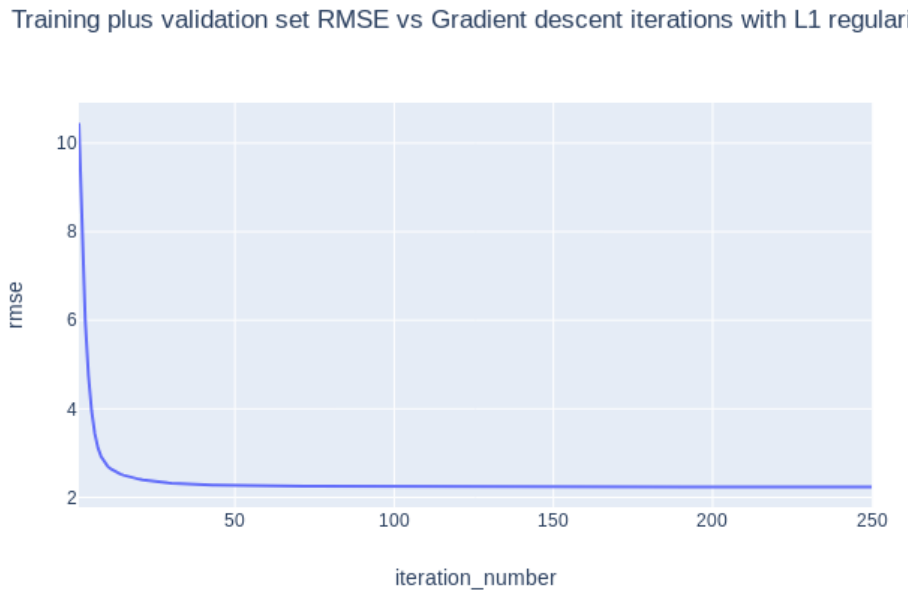


Figure 3 RMSE vs iteraitons

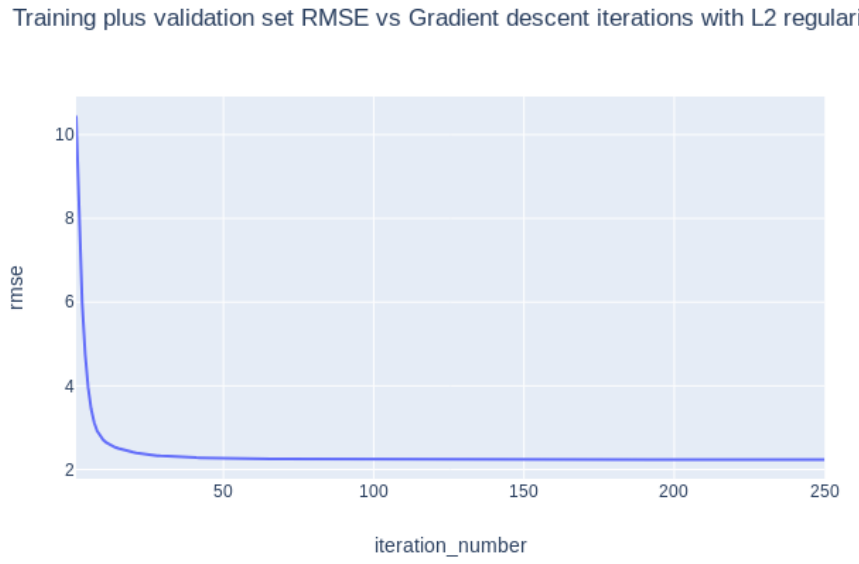


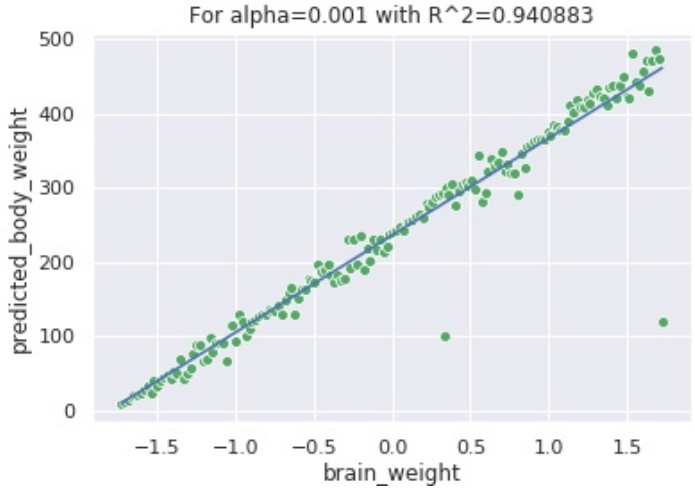
Figure 4 RMSE vs iterations

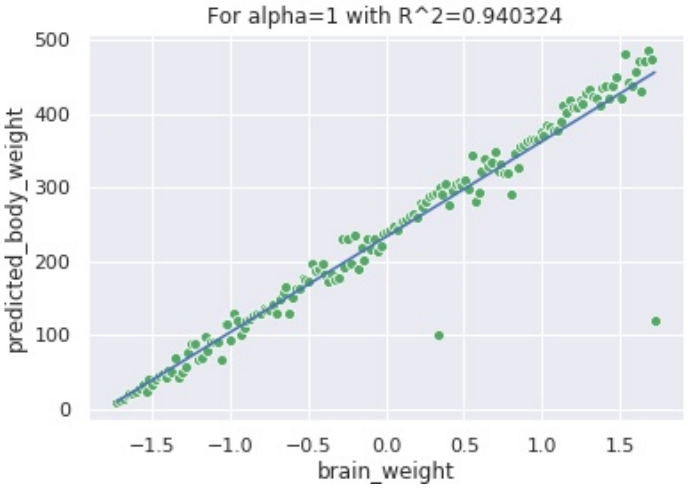
1.3 Best Fit Line

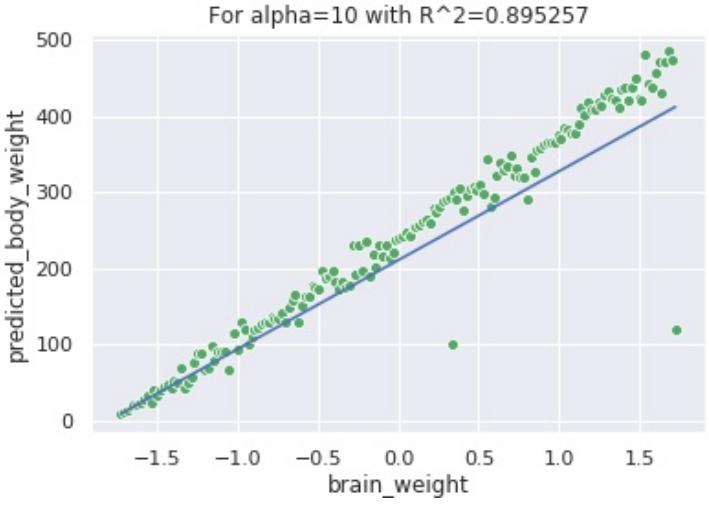
a) Without Regularization

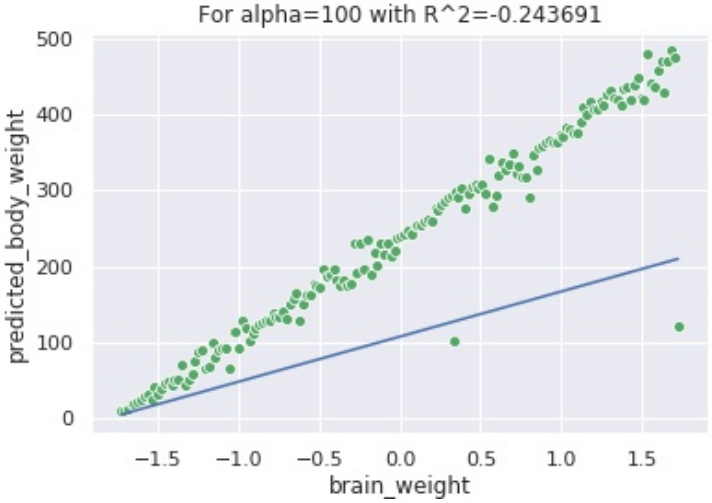
Figure 5: Best Fit Line

b) L2 Regularization

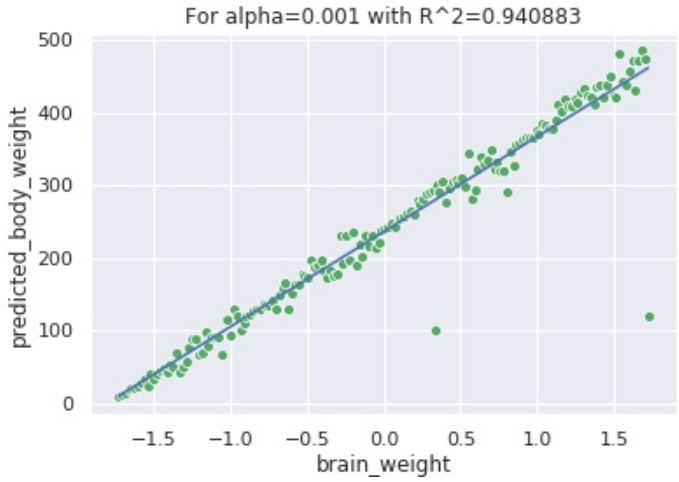


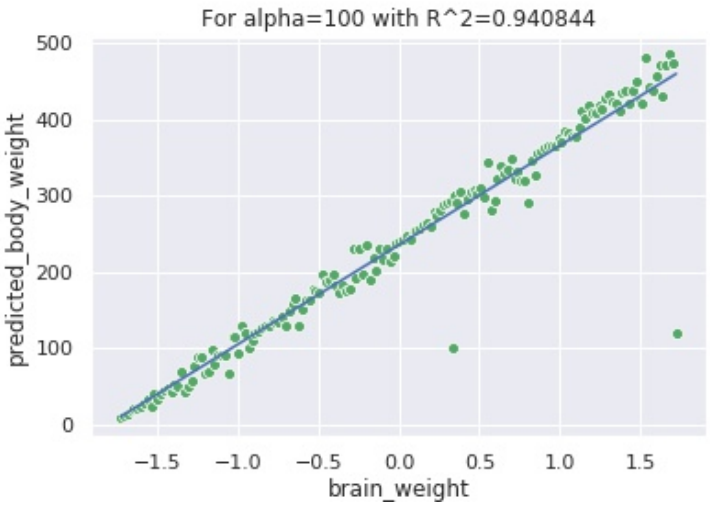


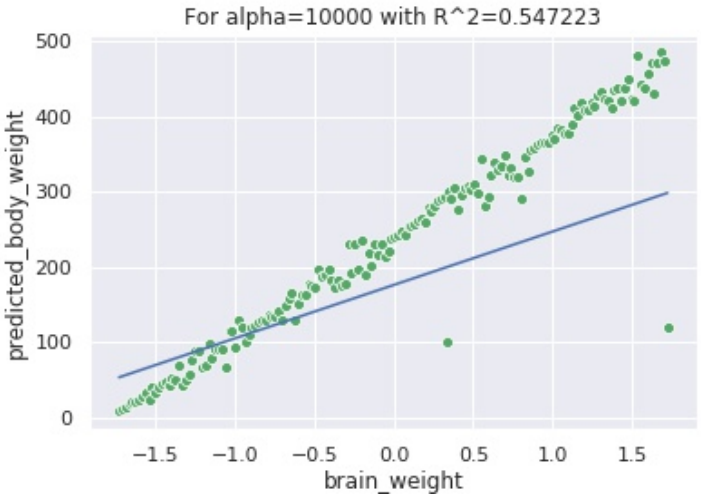




c) L1 Regularization







L1 regularisation performs when there are outliers. Because the L1 regularisation restricts the absolute value of coefficients, but in the L2 case it is the sum of squares of those coeff.

2. Logistic Regression

There are 104 features after converting the categorical features. Out of them 6 are of continuous types and the rest features contain just contain values 0 and 1. We can remove some categorical features whose variance is very low. It was found that there are 48 features have variance lower than 0.1. After removing those features, the accuracy achieved was similar to that using all of the features.

2.1. L2 Regularization

Table 3 Finding optimal alpha

Regularization parameter value	Train	Validation	Test
0.0001	0.847741	0.850629	0.845551
0.1	0.847782	0.850629	0.845551
0.9	0.847658	0.850795	0.845551
1	0.847658	0.850795	0.845551
12	0.847948	0.850961	0.845551
50	0.847161	0.850464	0.845351

2.2. L1 Regularization

Table 4 Finding optimal alpha

Regularization parameter value	Train	Validation	Test
0.0001	0.847741	0.850629	0.845551
0.001	0.847741	0.850629	0.845551
0.1	0.847741	0.850629	0.845551
0.8	0.847824	0.851127	0.845551
1	0.847824	0.850961	0.845683
2	0.847948	0.850795	0.846082
10	0.847865	0.850464	0.846812

L1 regularizations is working better on the dataset, as it deletes some of the features from the model which are not important.

2.3. Without Regularization

	Train	Validation	Test
Without Regularization	0.847741	0.850629	0.845551
L2(alpha = 12)	0.847948	0.850961	0.845551
L1(alpha = 0.8)	0.847824	0.851127	0.845551

2.4. Accuracy Vs no. of Iterations for L1 and L2

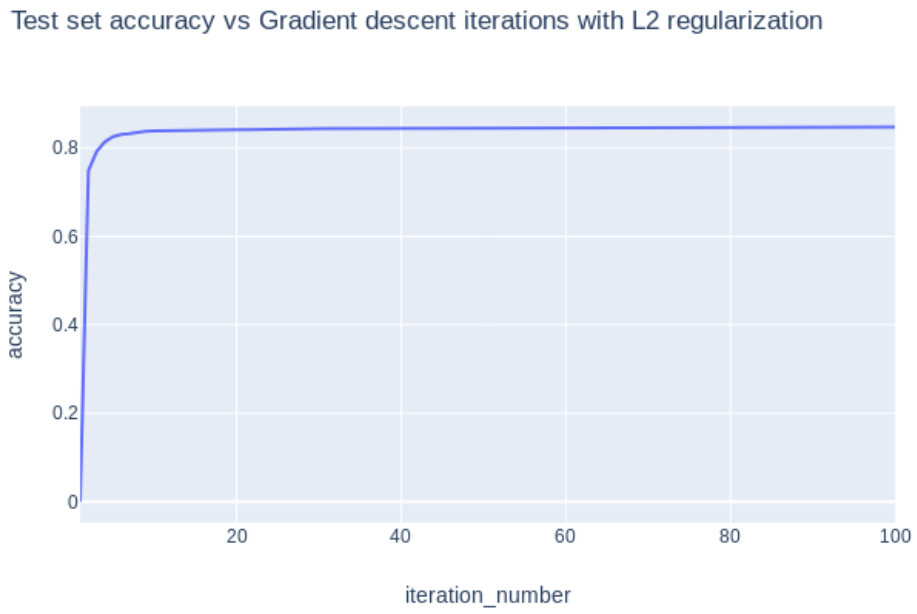


Figure 6 Accuracy vs iterations

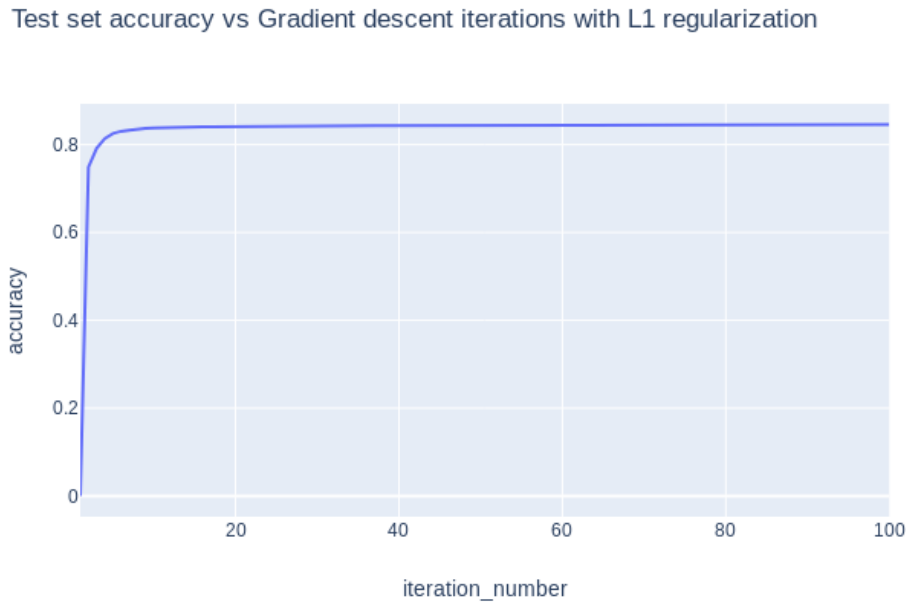
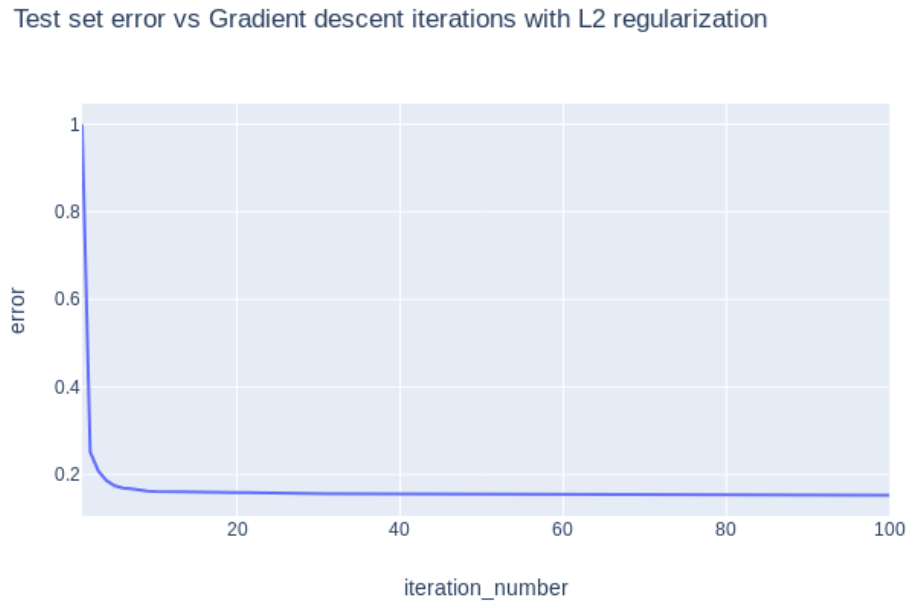
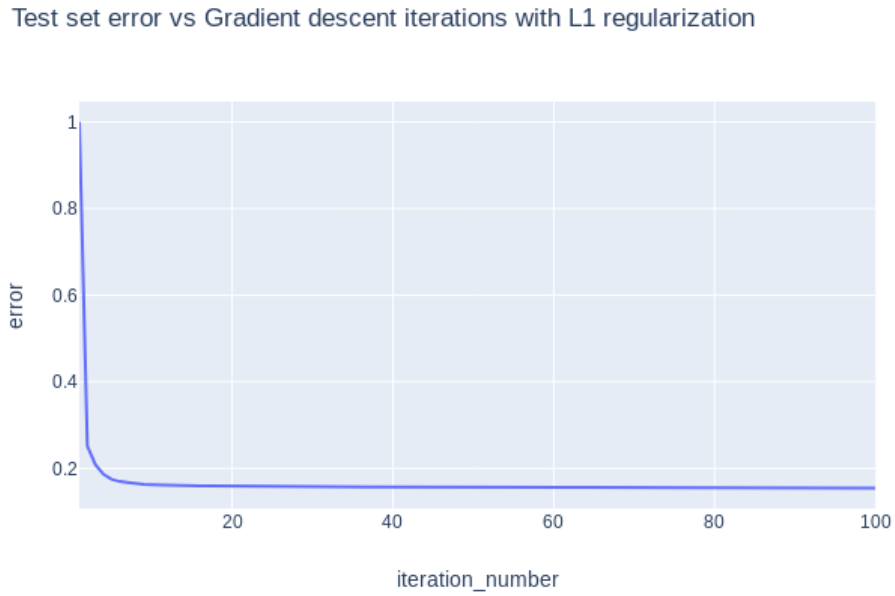


Figure 7 Accuracy vs iteration

2.5. Error Vs no. Of Iterations for L1 and L2





3. Logistic Regression on MNIST

3.1. L1 Regularization

Class	Train	Test
0	0.9932	0.9917
1	0.9926	0.9935
2	0.9806	0.9792
3	0.9763	0.9763
4	0.9838	0.9819
5	0.9754	0.9749
6	0.9842	0.9842
7	0.9854	0.9837
8	0.9601	0.9600
9	0.9657	0.9636

3.2. L2 Regularization

Class	Train	Test
0	0.9934	0.9911
1	0.9930	0.9930
2	0.9810	0.9790
3	0.9764	0.9760
4	0.9844	0.9819
5	0.9750	0.9747
6	0.9886	0.9835
7	0.9856	0.9836
8	0.9606	0.9603
9	0.9656	0.9631

The model is not overfitting the data as the training and test are error are very low and close together. Also, it is not underfitting as the accuracy on the training set is high.

3.3. ROC

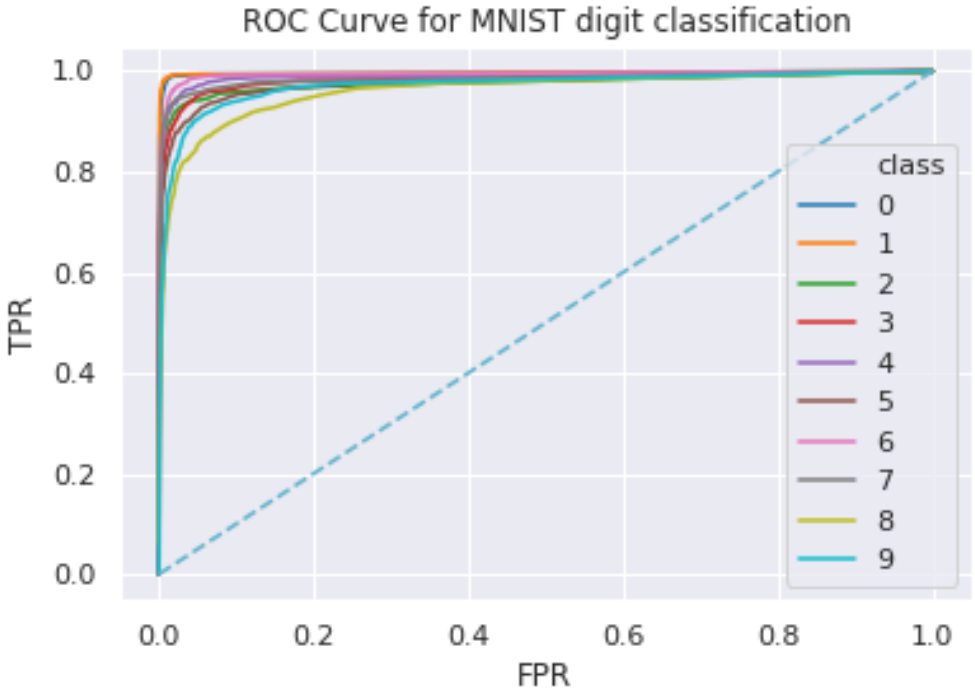


Figure 8ROC for all the classes 0 to 9