



# CSE 574: INTRODUCTION TO MACHINE LEARNING

Programming Assignment 1

Sanjay Bandlamudi Shailendra - 50290962

Aswathy Kumar - 50293990

[sbandlam@buffalo.edu](mailto:sbandlam@buffalo.edu), [aswathyk@buffalo.edu](mailto:aswathyk@buffalo.edu)

# Programming Assignment – 1

## 1. Linear Regression

### Training Data

RMSE without intercept 159.47902331066288

RMSE with intercept 53.77549791572339

### Test Data

RMSE without intercept 168.16910425224083

RMSE with intercept 53.931667955092294

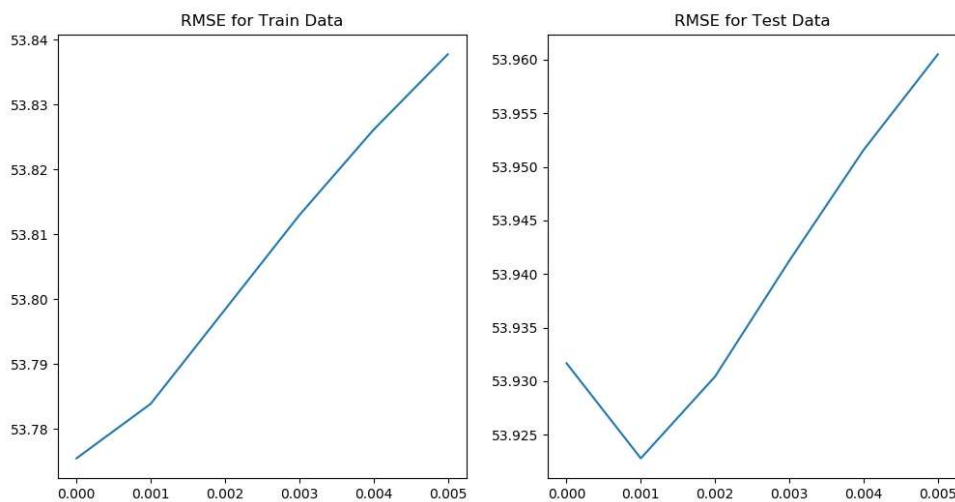
We observe that in case of Training as well as Test data the RMSE is lower when an intercept or bias is used. Thus, using an Intercept gives better performance.

## 2. Ridge Regression

Varying  $\lambda$  from 0 to 0.005 in steps of 0.001 we get the following results

lambda	RMSE for Training Data	RMSE for Training Data
0	53.77549792	53.93166796
0.001	53.78390648	53.92279448
0.002	53.79842442	53.93043156
0.003	53.81295917	53.94124377
0.004	53.82612652	53.95158206
0.005	53.83775681	53.96050654

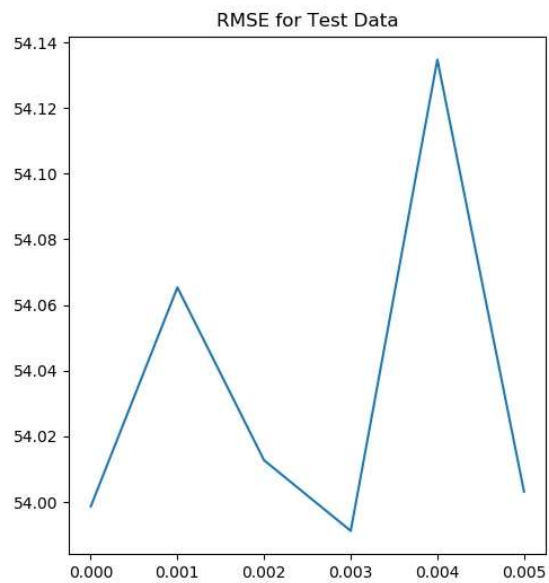
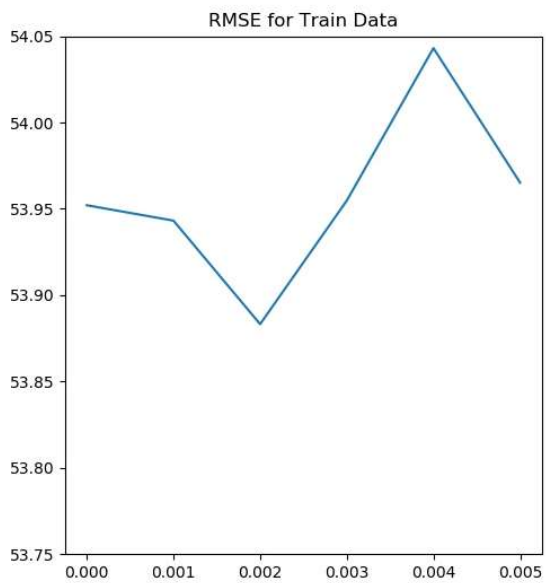
RMSE value is lowest in the case of  $\lambda = 0.001$  and hence can be used. But to get the optimal value cross validation using validation dataset must be done.

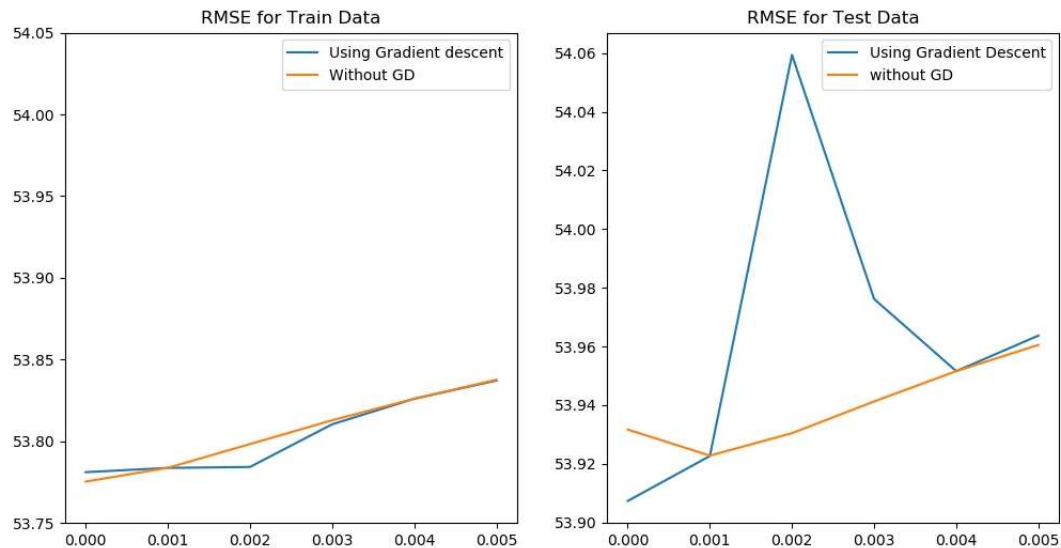


### 3. Using Gradient Descent for Logistic Regression

Varying  $\lambda$  from 0 to 0.005 in steps of 0.001 we get the following RMSE values for Training and Test data.

lambda	RMSE for Training Data	RMSE for Training Data
0	53.95209877	53.99863906
0.001	53.94319937	54.06535173
0.002	53.88328102	54.01271733
0.003	53.9548026	53.99121671
0.004	54.04303366	54.13471801
0.005	53.96518847	54.00318231





Comparing Ridge Regression using explicit formula and Ridge Regression using Gradient descent

On comparison we can observe that the RMSE values using explicit formula is lesser than using Gradient Descent.

#### 4. Using Gradient Descent for Logistic Regression

The following are the accuracy for Training and Test Data respectively

Training set Accuracy : 97.07941588317664%

Testing set Accuracy : 95.00441652762784%

#### 5. Support Vector Machine

Accuracy of predictions obtained after varying different parameters in SVM model.

Parameters	Training Accuracy	Testing Accuracy
Linear Kernel	92.41648330%	92.11349931%
RBF Gamma =1	98.08961792%	98.39914505%
RBF Gamma = default	92.82056411%	92.78960971%
C-Value=1	92.82056411%	92.78960971%
C-Value=10	95.80916183%	95.84301153%
C-Value=20	96.35927185%	96.82882411%
C-Value=30	96.72934587%	97.18868933%
C-Value=40	96.88537708%	97.37843644%
C-Value=50	97.02740548%	97.49512001%
C-Value=60	97.14942989%	97.60526058%
C-Value=70	97.23544709%	97.68050512%
C-Value=80	97.29545909%	97.69359113%
C-Value=90	97.34146829%	97.72957765%
C-Value=100	97.36947389%	97.78083118%

