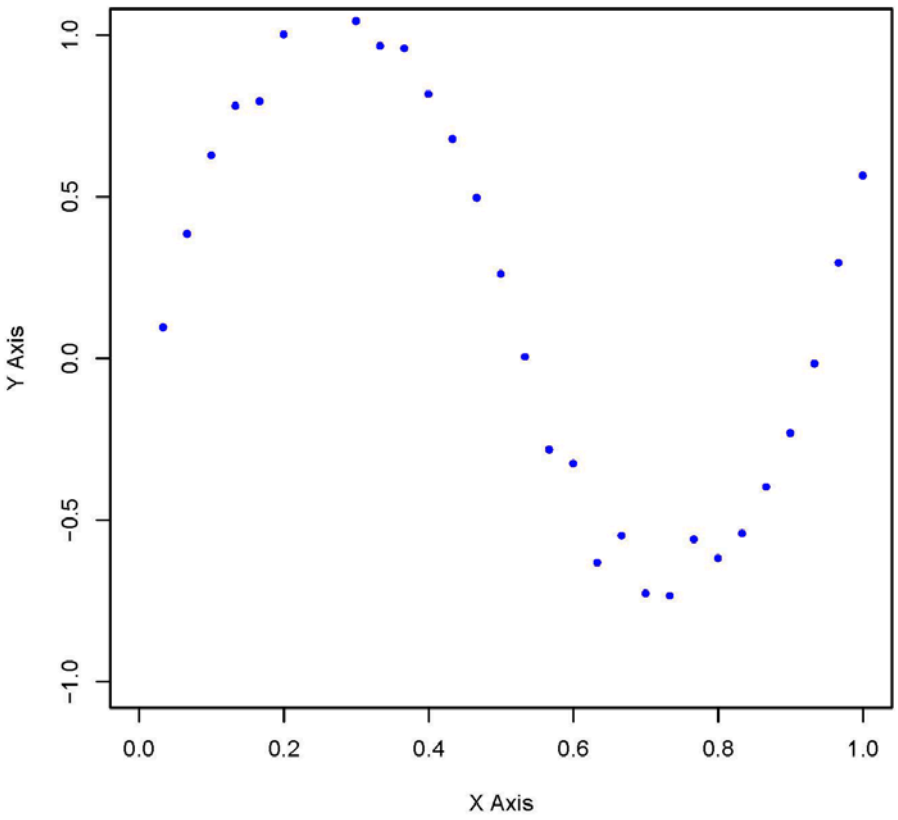
Machine Learning -- Assignment 1

Chinmaya R Naguri – cn262114

5.

a. Plot for the complete Dataset

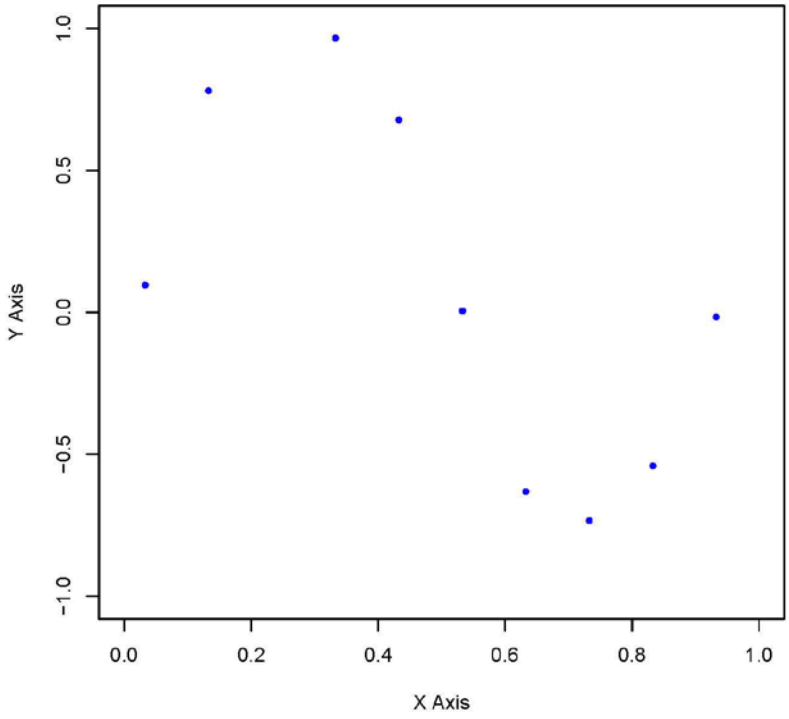
Rplots_Dataset.pdf



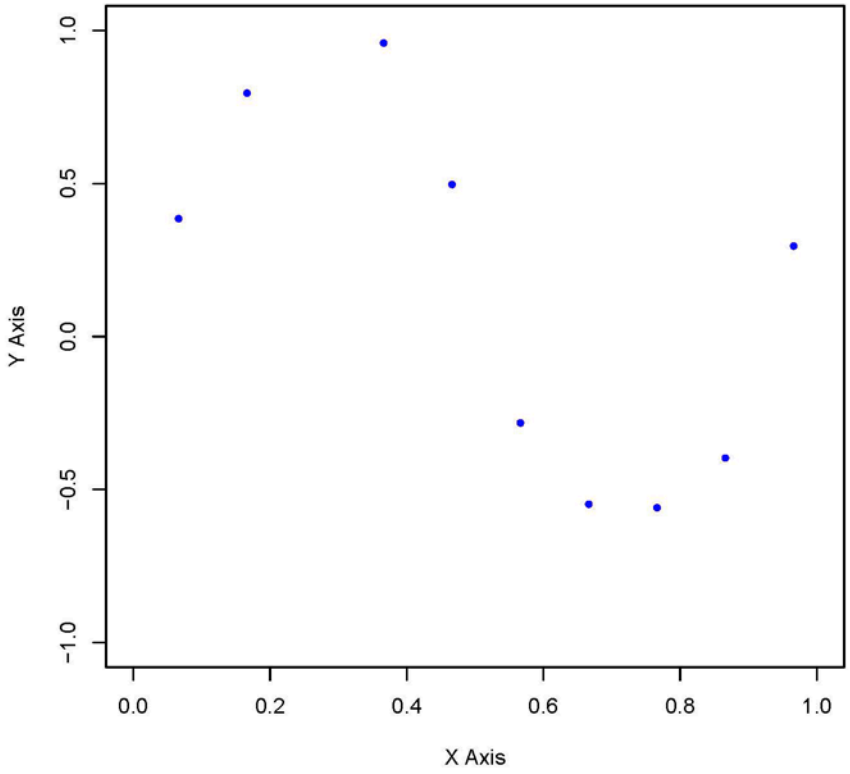
b. Plot for Train, Test and Validation data

Rplots_Train.pdf Rplots_Test.pdf Rplots_Valid.pdf

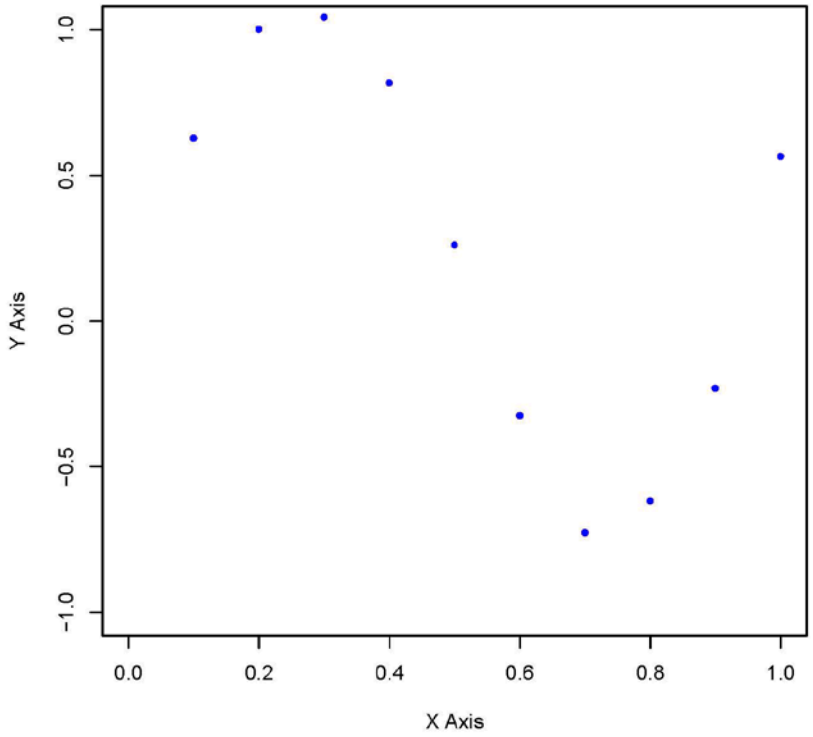
Training Data



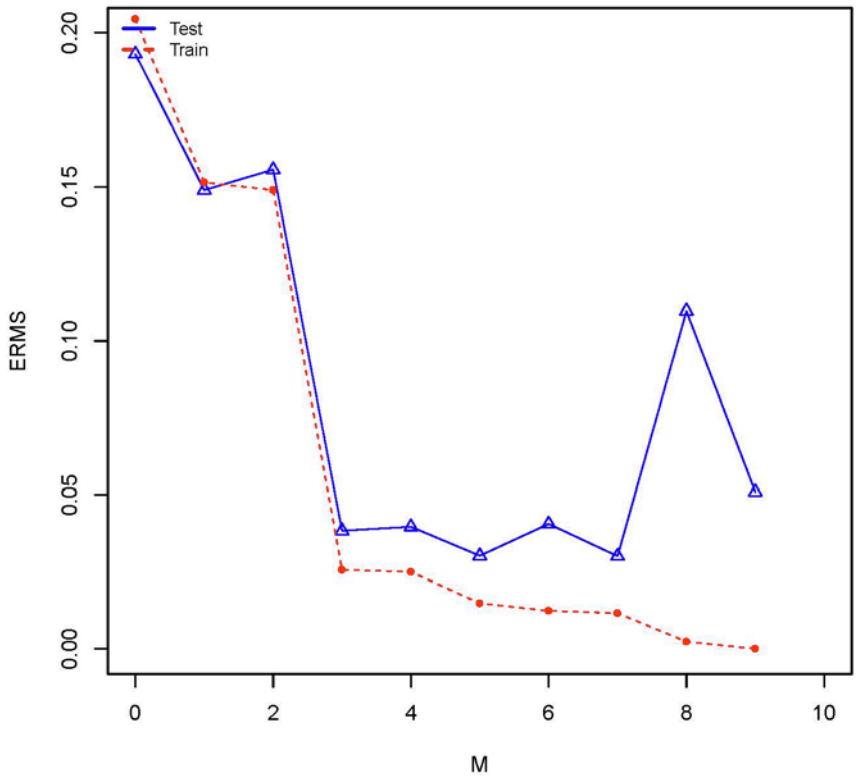
Test Data



Validation Data

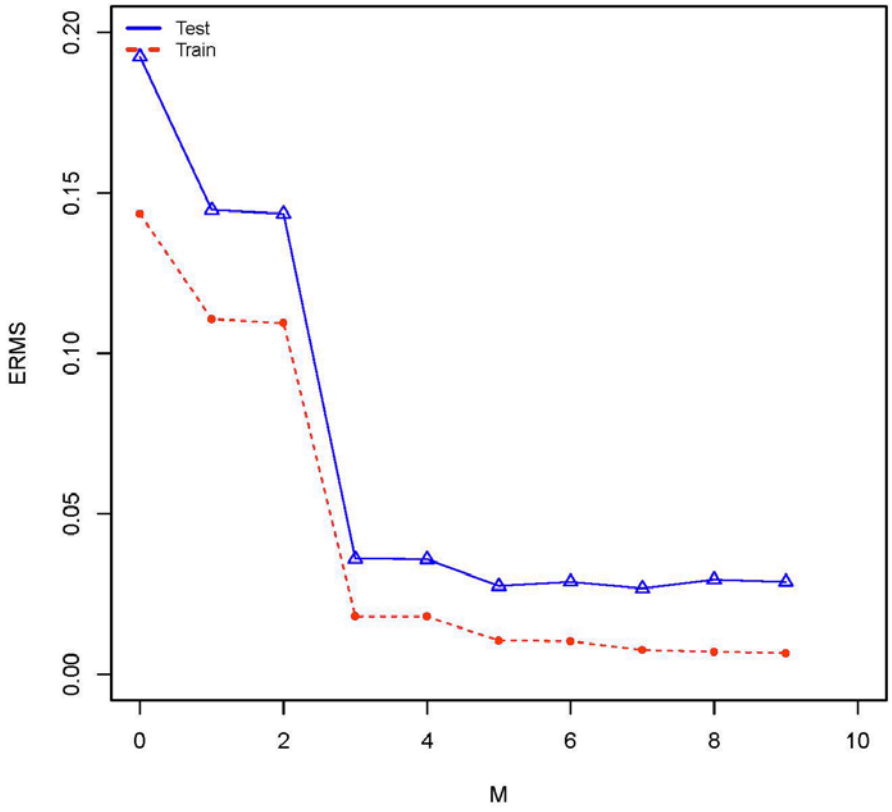


c. 1. **Without Regularization – M vs ERMS values for Training and Test Data**



Rplots_MvsERMS_
WithoutReg_Trainin

c. 2. **<u>Without Regularization – M vs ERMS values for Training + Validation Data and Test Data</u>**



Rplots_MvsERMS_
WithoutReg_Train&

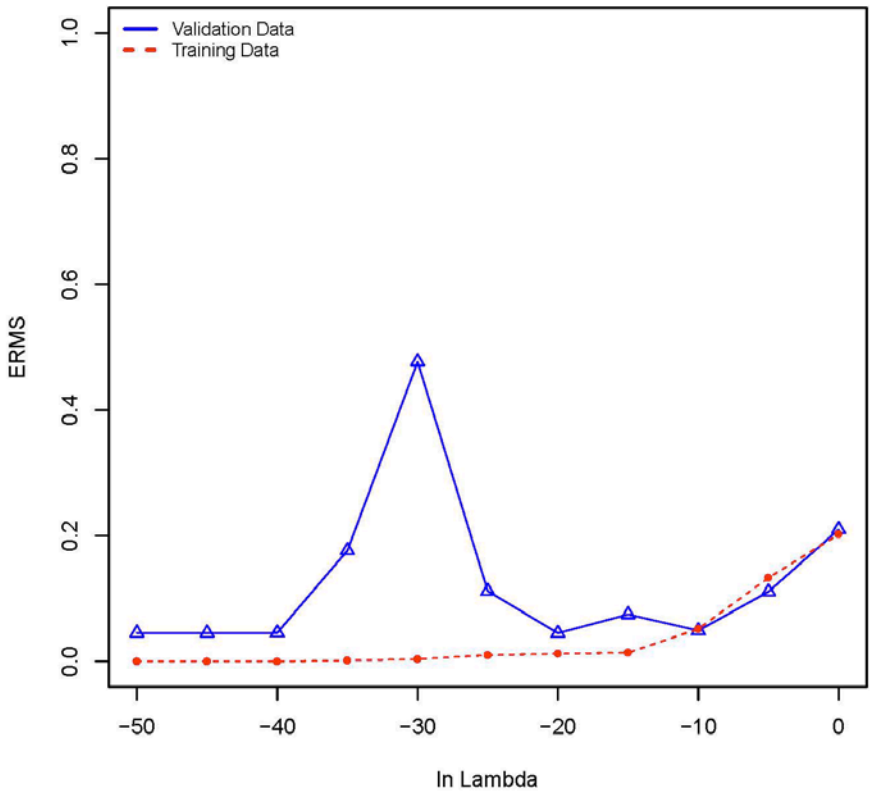
In Without Regularization, with just use *Train.dat* (i.e. training data). We can see that M-ERMS Graph for Test data is consistent in M = [3 to 7] with some distortion. But when the training data is increased to Train + Validation data these distortions are gone & is smooth.

When M = 8/9 with just *Train.dat* data, ERMS value is high but after introducing validation data into training we can observe that it is smoothed out and is consistent with the rest of graph.

One more observation would be the ERMS values of the training data prediction is considerably reduced after introducing validation data into training data set.

So, we can see that introducing more data reduced the ERMS values i.e. deviation of the predicted value from the original is getting reduced. It even stabilized the error rate as well from M = 3 to M = 9 for this data set.

c. 3.



Rplots_ERMSvsLnLa
mbda_Train_Validatic

With Regularization – ln λ = -50 & ERMS value for Training Data & Test Data (M=9)

```
Given Training File: C:\Users\cn262114\Documents\eclipse-
workspace\ML_A1\data\train.dat
Given Test File: C:\Users\cn262114\Documents\eclipse-
workspace\ML_A1\data\test.dat
ln Lambda: -50
ERMS Value for the Testing Predictions: 0.050945142543143876
```

Even with less data i.e. nearly 10 data points, ERMS value is 0.051 on test data which is nearly negligible. This clearly shows our usage of λ really paid off with better prediction even with less data.

With Regularization – ln λ = -50 & ERMS value for Training + Validation Data & Test Data (M=9)

```
Given Training File: C:\Users\cn262114\Documents\eclipse-
workspace\ML_A1\data\train_valid.dat
Given Test File: C:\Users\cn262114\Documents\eclipse-
workspace\ML_A1\data\test.dat
ln Lambda: -50
ERMS Value for the Testing Predictions: 0.028811278123970326
```

When the training data is increased (here doubled by including validation data), ERMS value is 0.029 on test data which is even better improvement from the previous case. So, having more examples seems to help in most times.

We usually expect that we can reduce the ERMS value after using Regularization in any model. But, we can observe that for the given data, ERMS value at M=9 does not change much in with & without regularization models. This is a behavior we see with this data.