# Introduction to Machine Learning Project 1.2 Report

**UBID:** sppandey@buffalo.edu

**PERSON NO: 50288608**

**Overview :**

The goal of this project is to use machine learning to solve a problem that arises in Information Retrieval, one known as the Learning to Rank (LeToR) problem. We formulate this as a problem of linear regression where we map an input vector x to a real-valued scalar target y (x, w).

There are two tasks:
1. Train a linear regression model on LeToR dataset using a **closed-form solution**.
2. Train a linear regression model on the LeToR dataset using **stochastic gradient descent (SGD).**

**Project Content:**

The Report consists of statistics and different Root mean Squared values obtained by tuning the hyper- parameters such as regularizer, which in turn impacts BigSigma Output, variations in learning rate for Gradient Descent Model, learning rate. Concepts of Radial Basis Functions, K-means clustering and how Grid search is a time-expensive yet an optimal way to tune few hyperparameters have been touched upon on the report.

## Models:

### Linear Regression Model:

Linear Regression is the basic model used in this project. Conceptually linear regression is a method that allows to study and compare relationship between two variables expressed in a mathematical equation. The variables are aligned across the X-axis and Y-axis can be called as the independent variable and dependent variable respectively. For a positive linear regression, the value of dependent variable increases as the value of independent variable increases and vice and versa.

In project 1.2 the form of our linear regression will be;

$$y(x, w) = w^T \phi(x)$$

Where $w$ is the weight and is the vector of basis functions. The basis function used in this project will be the gaussian radial basis function which is explained as below.

### Gaussian Radial Basis Function:

Scientific Computing with Radial Basis Functions focuses on the reconstruction of unknown functions from known data. Gaussian rbf used in project results in the basis vectors which is then used to further identify and minimize the Error mean square values obtained from different hyper parameters

### Closed Form Solution:

For any given problem in terms if mathematical operations and function if can be solved by an equation using a given accepted set then such solutions are called closed form solution. Summing numbers till infinity is a very common example which can be mistaken for a closed form but is not. In the current project we are finding the closed-form solution with least-squared regularization which will then be compared to the root mean square output obtained by the Stochastic gradient Descent solution for performance evaluation.

**Observation of Different $E_{rms}$ values over different Hyper-Parameter values:**

**Scenario 1 :**

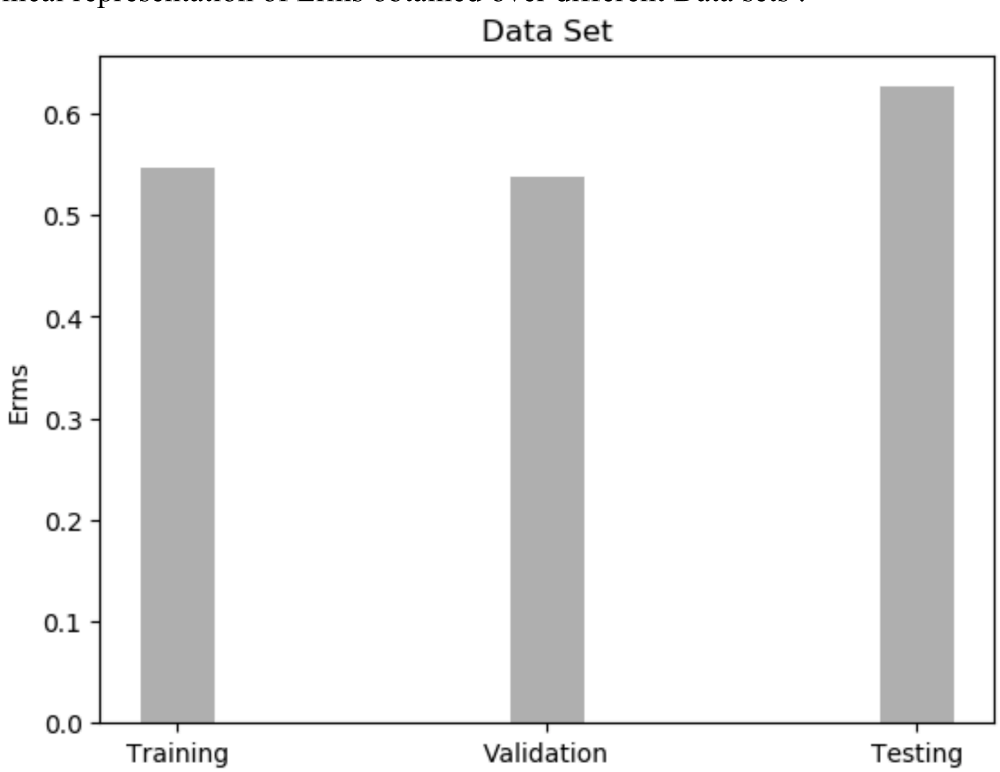
K-means clusters = 20
$\lambda$ = 0.03
Learning Rate = 0.001
Range – 400

1) **Closed Form Solution :**

$E_{rms}$ values obtained for the Training, Validation and Testing data.

```
----------------------------------------------------
------------------LeToR Data------------------------
----------------------------------------------------
-------Closed Form with Radial Basis Function-------
----------------------------------------------------
M = 10
Lambda = 0.9
E_rms Training   = 0.5461233210909291
E_rms Validation = 0.5383600398443413
E_rms Testing    = 0.6261902077808426
```

Graphical representation of Erms obtained over different Data sets :

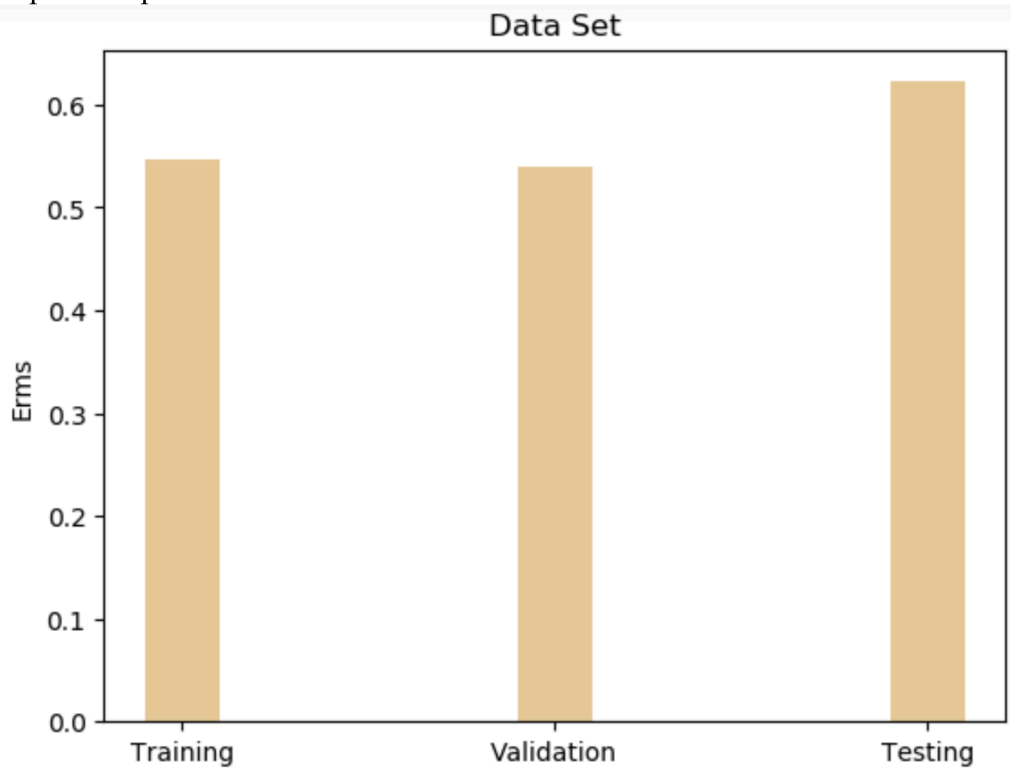


## 2) Gradient Descent Solution for linear Regression:

Lambda for Delta E calculation = 2

```
E_rms Training   = 0.54621
E_rms Validation = 0.53869
E_rms Testing    = 0.6221
```

Graphical representation of Erms obtained over different Data sets:




Observation :- For the standard given input the output Erms values are not the best possible solutions and can be further improved by either tweaking the learning rate or the regularizer value.

**Scenario 2 :**

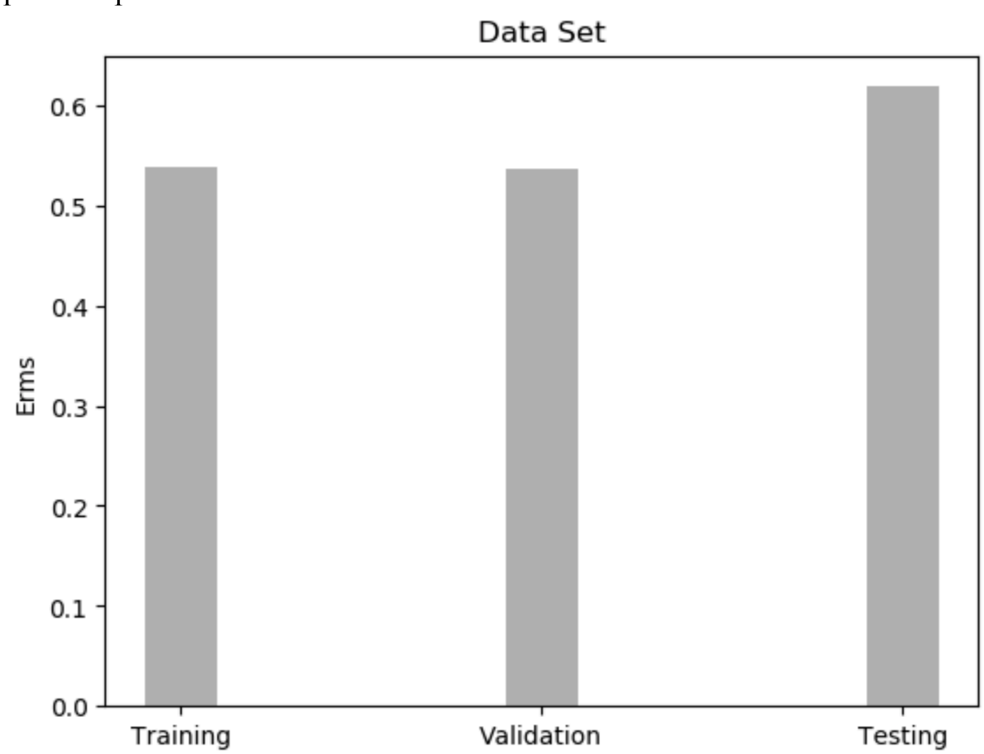
K-means clusters = 56
$\lambda$ = 0.003
Learning Rate = 0.0001
Range – 700

1) **Closed Form Solution :**

$E_{rms}$ values obtained for the Training, Validation and Testing data.

```
----------------------------------------------------
------------------LeToR Data------------------------
----------------------------------------------------
-------Closed Form with Radial Basis Function-------
----------------------------------------------------
M = 10
Lambda = 0.9
E_rms Training   = 0.538383223545505
E_rms Validation = 0.5356687739794361
E_rms Testing    = 0.6183637175240994
```

Graphical representation of Erms obtained over different Data sets :



## 2) Gradient Descent Solution for linear Regression:

Lambda for Delta E calculation = 3

```
----------Gradient Descent Solution--------------------
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training   = 32.83068
E_rms Validation = 32.57208
E_rms Testing    = 33.77846
```

Graphical representation of Erms obtained over different Data sets :


Observation:- A significant change in the cluster size , learning rate and the lambda value leads to a rather decent Erms value. This is not the optimal solution though and thus with the exhaustive of available options the optimal solution can be achieved.

**Scenario 3 :** This scenario was tested over a constant set of hyper parameters but for different lambda values in the gradient descent model.

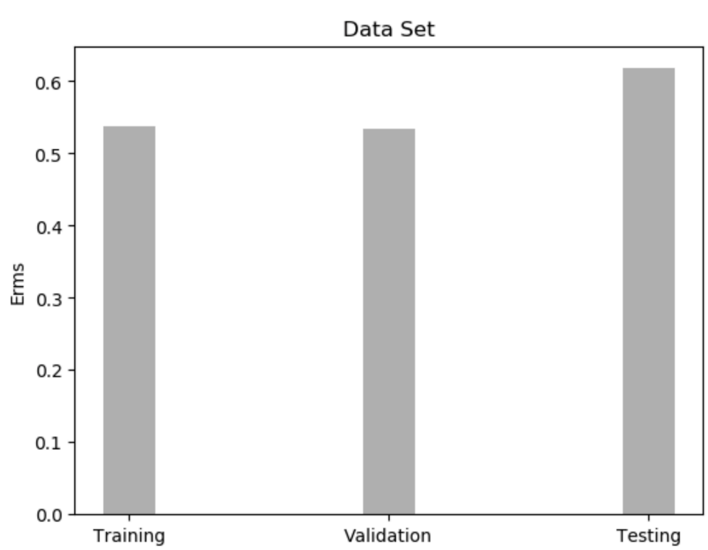
K-means clusters = 70
$\lambda$ = 0.0003
Learning Rate = 0.0001
Range – 700

1) **Closed Form Solution:**

$E_{rms}$ values obtained for the Training, Validation and Testing data.

```
------------------LeToR Data------------------------
----------------------------------------------------
-------Closed Form with Radial Basis Function-------
----------------------------------------------------
M = 10
Lambda = 0.9
E_rms Training   = 0.5370279988457431
E_rms Validation = 0.5341118475175701
E_rms Testing    = 0.6178740071546542
```

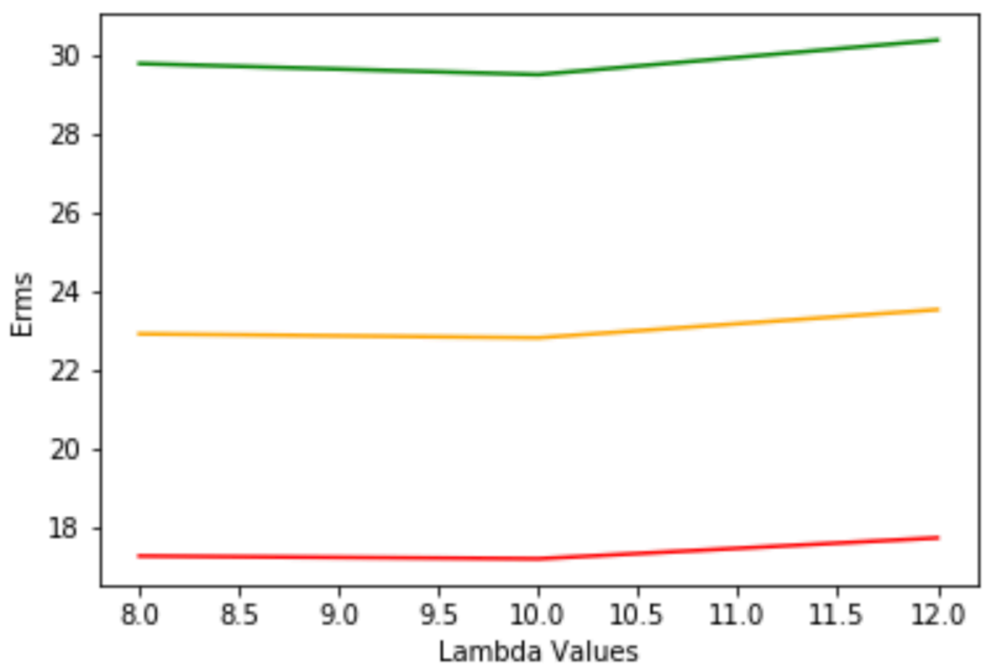
Graphical representation of Erms obtained over different Data sets :



## 2) Gradient Descent Solution for linear Regression:

| Lambda Value | Erms for Training Data | Erms for Valistaion Data | Erms for Testing Data |
|---|---|---|---|
| 8 | 29.76668 | 29.48682 | 30.36005 |
| 10 | 22.91538 | 22.81539 | 23.52838 |
| 12 | 17.283 | 17.2137 | 17.74285 |

The comparison Graph is shown Below :




Observation:- In this specific scenario we tried changing lambda recursively to obtain a lower Erms value. Keeping in mind the inverse relation Erms and accuracy has we have observed that the stochastic gradient method provides better accuracy and thus can be considered as an optimal model for the Erms minimisaton.