CSE474/574: Introduction to Machine Learning (Fall 2018)
Project 1.2 :LeToR

UBIT: sassikes
Person no: 50291558

Introduction:

The primary objective of this project is to solve the Learning to Rank problem using Linear regression. The two task that required to be done are

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).

Linear Regression:

Regression is a method of modelling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

Machine learning, more specifically the field of predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. In applied machine learning we will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends.

As such, linear regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm.

Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x).

When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.

Different techniques can be used to prepare or train the linear regression equation from data, the most common of which is called Ordinary Least Squares. It is common to therefore refer to a model prepared this way as Ordinary Least Squares Linear Regression or just Least Squares Regression.

Linear Regression is made with an assumption that there’s a linear relationship between X and Y.

Form of Linear Regression

Mathematically, we can write a linear relationship as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Model Coefficients/Parameters

When training a linear regression model, it’s way to say we are trying to find out a coefficients for the linear function that best describe the input variables.

Sassikes
50291558

Cost Function (Loss Function)

When building a linear model, it’s said that we are trying to minimize the error an algorithm does making predictions, and we got that by choosing a function to help us measure the error also called cost function.

Closed Form Solution:

An equation is said to be a closed-form solution if it solves a given problem in terms of functions and mathematical operations from a given generally accepted set. For example, an infinite sum would generally not be considered closed-form. However, the choice of what to call closed-form and what not is rather arbitrary since a new "closed-form" function could simply be defined in terms of the infinite sum.

In mathematics, an expression is said to be a closed-form expression if it can be expressed analytically in terms of a finite number of certain "well-known" functions. Typically, these well-known functions are defined to be elementary functions—constants, one variable x, elementary operations of arithmetic (+ − × ÷), nth roots, exponent and logarithm (which thus also include trigonometric functions and inverse trigonometric functions). Often problems are said to be tractable if they can be solved in terms of a closed-form expression.

Gradient Decent:

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.

List of hyperparameters:

The most important hyper parameters are,

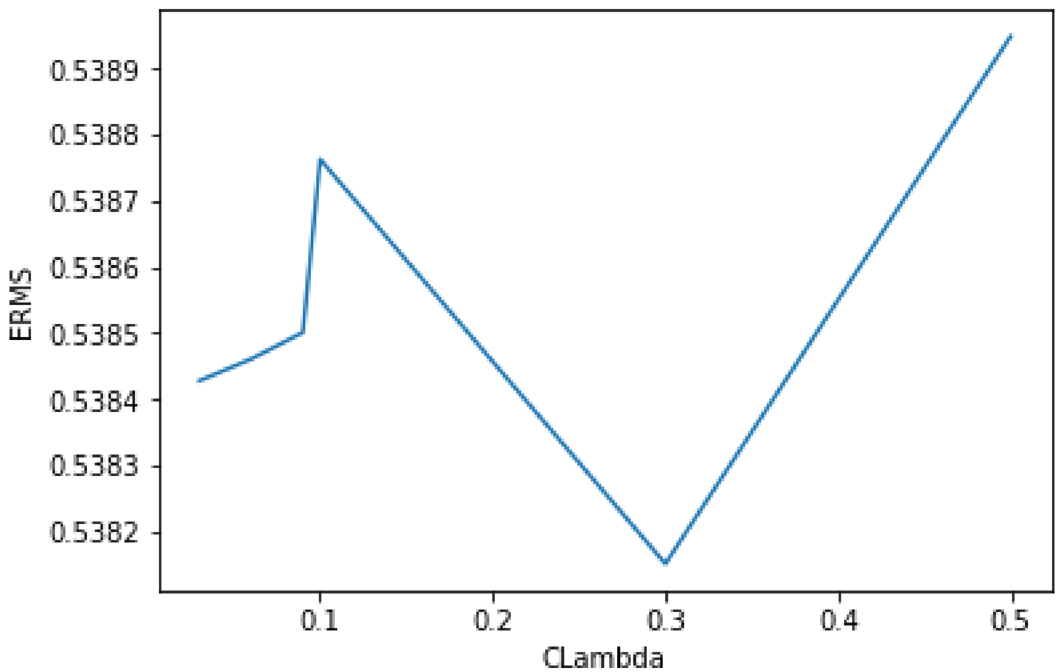
1. C_Lambda
2. W_Now
3. La
4. learningRate

Now let us see how the ERMS changes as we modify each parameter.

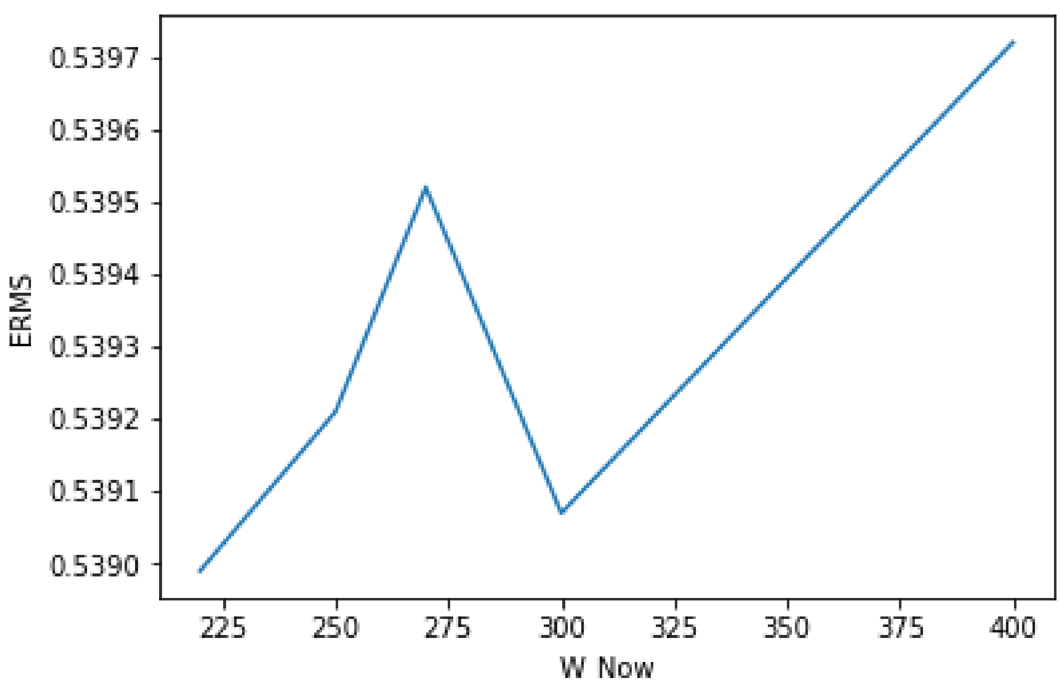
Sassikes
50291558

C_LAMDA vs ERMS

```
Closed form solution
Graph for Clamda and ERMS
```




```
Gradient Decent solution
Graph for W_Now and ERMS
```



Sassikes
50291558

Grid Search

With Grid search methodology we train multiple sets of models with finite no. of parameters and cross-validate it and choose the weights of best performing model in the set.

Learning Rate

After the result of the loss function, we use a small value to multiply it with the derivative of the loss function before subtracting it from the weights. Goal is to minimize the loss. If a higher learning rate is chosen results will easily diverge resulting in more less.it is important to choose an optimal learning rate. As shown below if we choose a learning rate of 0.3 we can see a drastic increase in the error rate.by default many start with the learning late of 0.01 and increase it gradually

S.No	Learning rate	RMS		
		Test set	Train set	Val set
1	0.02	0.63076	0.55869	0.54741
2	0.1	0.64879	0.58417	0.57105
3	0.2	0.63588	0.56518	0.55413
4	0.3	12.25598	12.36029	11.99335

1.5 Reference:

1. https://towardsdatascience.com
2. https://medium.com
3. https://statisticsolutions.com