Data Analytics - Assignment #1

Nishant Kumar (Sr.NO. - 21495)

August 2023

Run Prediction using DLS Method

1 Implementation Summary

1. I initially selected the key columns in the dataset "04 cricket 1999to2011.csv," which contains information about One Day International (ODI) cricket matches played from 1999 to 2011. During the data processing phase, I extracted four crucial aspects for each match: **innings number, remaining runs, remaining overs, wickets remain.**

2. Start by setting up the parameters, and then use the optimization function from the Scipy package, specifically the "minimize" function. This process aims to minimize the squared error loss. The outcome of this optimization will be the most suitable values for the parameters. The squared error loss is calculated in the following manner:

$$\text{Squared Error Loss} = \Sigma_{i=1}^{n}(y_i - y_i')^2 \tag{1}$$

where:
n represents the number of data points or samples.
y_i is the actual or observed value for the ith data point.
y_i' is the predicted or estimated value for the ith data point.

3. The method I used is **L-BFGS-B** that is present in the library **scipy.optimize**, **L-BFGS-B** is an optimization algorithm for finding a function's minimum. It's particularly suitable for solving optimization problems where the objective function is smooth and bounded, and it doesn't require the computation of the Hessian matrix, making it memory-efficient.

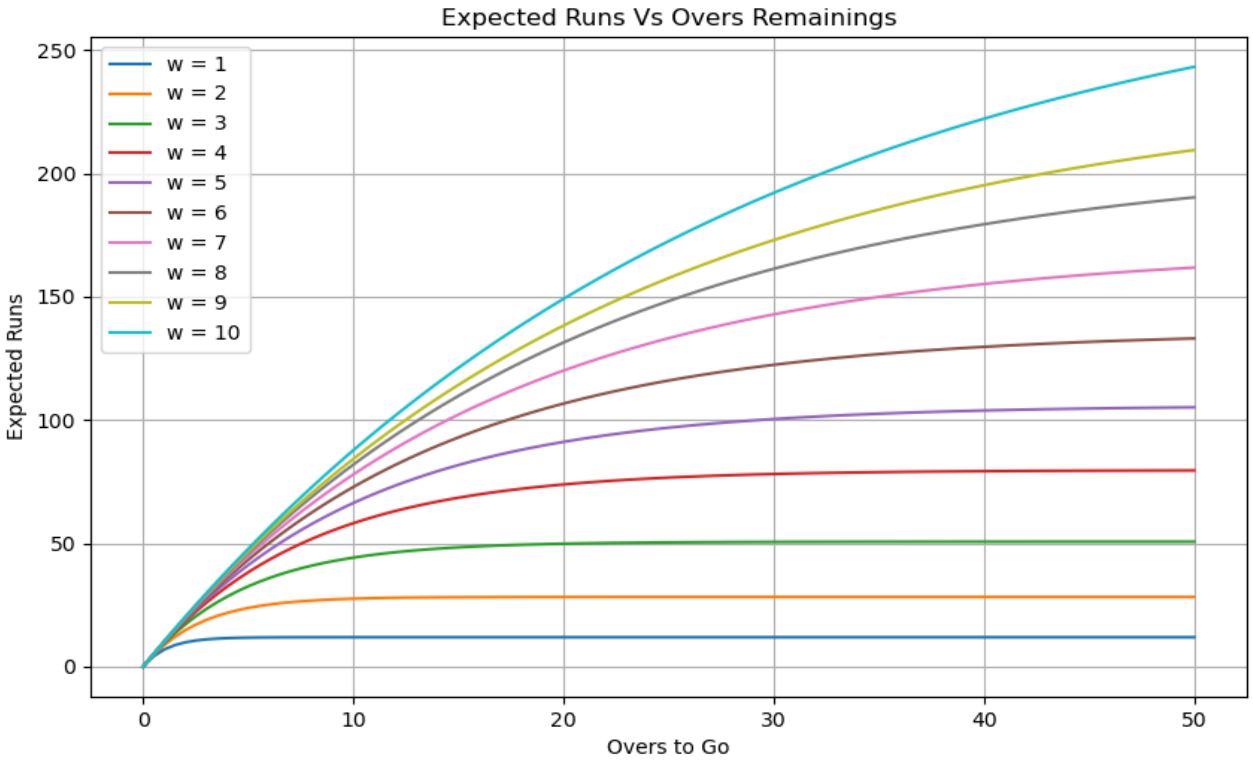
In the program, **L-BFGS-B** is applied within the **train_model** function to optimize the parameters for predicting run production based on the given data. The minimize function iteratively adjusts the parameters to minimize the loss function, leading to better predictions.

2 Results

2.1 The plot with 10 curves

2.2 Average Loss

Average Loss : 0.0860642705928258



2.3 Value of model Parameters

Optimized values of all Z and L Parameters:

Z[1]	Z[2]	Z[3]	Z[4]	Z[5]	Z[6]	Z[7]	Z[8]	Z[9]	Z[10]	L
11.959	28.288	50.702	79.700	105.942	136.022	169.648	206.924	235.008	292.255	10.433