Data Analytics Assignment 2

Anshul Shivhare (Sr no. 17828)

September 12, 2021

1 Score Prediction using DLS Method

So the task is to find the best fit 'run production functions' in terms of wickets-in-hand w and overs-to-go u. Using the sum of squared errors loss function, summed across overs and wickets.

2 Data Set

Our data set contains data of ODI matches from 1999 to 2011. Data include total of 38 columns and 126768 rows. Every row contain data, over by over.

Initially some preprocessing is done, Which includes:

1. As we have to consider only first inning score, hence removed the second innings data. The total number of rows left is 67794.
2. Dropping of rows having value 1 in "Error.In.Data" column, so total number of rows left 67288.
3. One match had data from 16Th over, hence removed those rows.
4. Irrelevant columns are removed, in our case only 7 columns were useful.
5. Note: Many of the matches details are from wicket 9th , 8th or 7th, But i am not removing those rows, as it will lead to data loss. And in this case we can still use the data when others wickets are in hand.

3 Procedure

1. Calculated number of overs left, using total overs and over columns. So this is our value of u.
2. We don't have to calculate w, as it is directly given in the data. i.e. the wickets in hand column.
3. Now using the above information, we can calculate the runs scored when u overs were left and w wickets in hand using total runs and innings total column.

4. So finally we have our 3 main columns using which we will do prediction, The columns are "Remaining overs(u)", "Wickets.in.Hand(w)", "Runs scored".

5. Output for the below function is:

3.1 20 parameters model

$$Z(u, w) = Z_0(w)(1 - e^{-b(w)u})$$

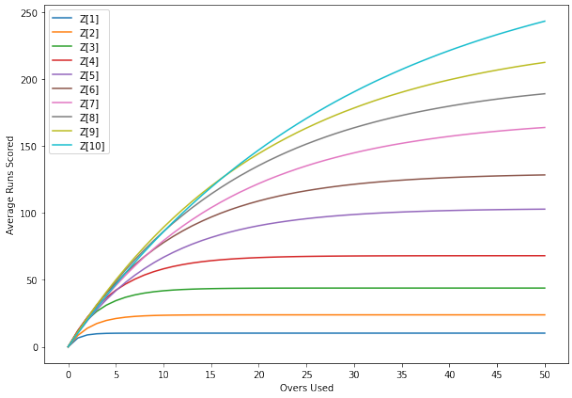

Figure 1: Plot for first model (20 parameters)

$Z_0 = [10.20, 23.88, 43.89, 68.10, 103.50, 129.88, 171.75, 201.69, 233.75, 297.33]$
$b = [1.01, 0.43, 0.30, 0.19, 0.10, 0.09, 0.06, 0.05, 0.04, 0.03]$

3.2 11 parameters model

$$Z(u, w) = Z_0(w)(1 - e^{\frac{-Lu}{Z_0(w)}})$$

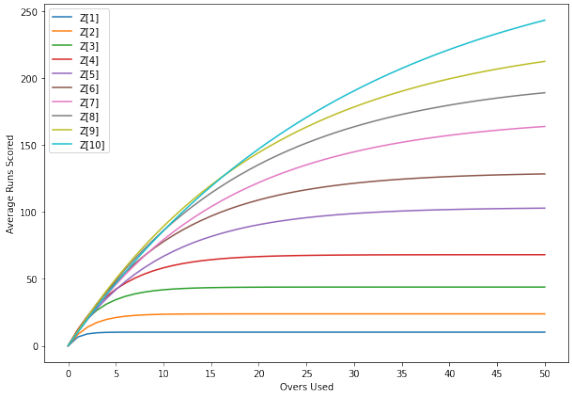

Figure 2: Plot for second model (11 parameters)

$Z_0 = [10.20, 23.88, 43.89, 68.10, 103.50, 129.88, 171.75, 201.69, 233.75, 297.33]$, $L = 15.4$

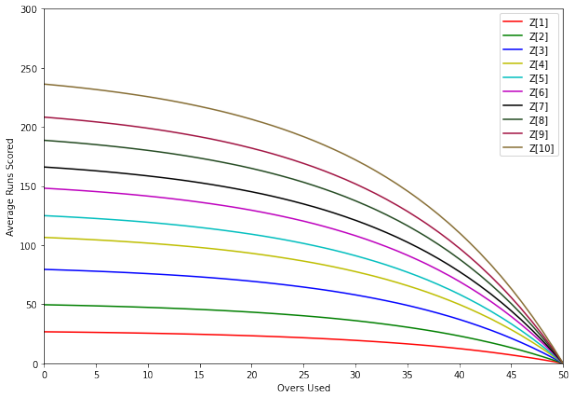

Figure 3: Overs used plot for second model (11 parameters)