Data Analytics: DLS Assignment

Narendra Thiramdas (SR.No: 19306)

August 22, 2022

1 Task

Using the first innings data alone in the above data set, find the best fit 'run production functions' in terms of wickets-in-hand w and overs-to-go u. Assume the model

$$Z(u, w) = Z_0(w)[1 - exp(-Lu/Z_0(w))] \quad (1)$$

Use the sum of squared errors loss function, summed across overs, wickets, and data points for those overs and wickets.

Provide a plot of the ten functions, and report the (11) parameters associated with the (10) production functions, and the normalised squared error (total squared error summed across overs, wickets, and data points for those overs and wickets, and normalised by the total number of data points across all overs and wickets).

2 Approach

1. Extract four important features(Innings, Runs Remaining, Overs Remaining, Wickets in Hand) from the given csv file

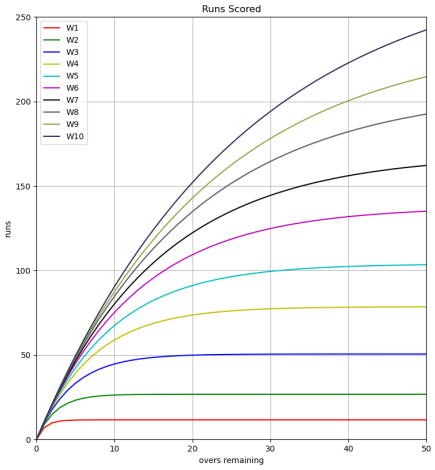
2. Take all the data points of first innings and define the loss function as sum of squared errors loss function, summed across overs and wickets.

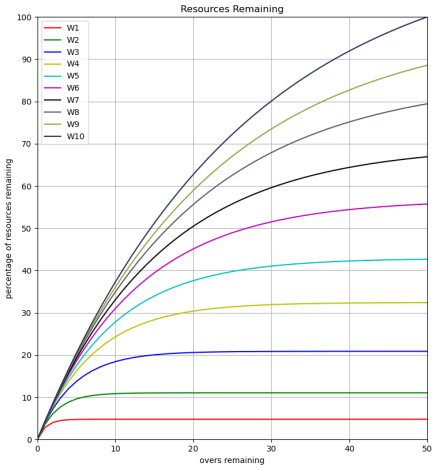
3. Minimize the objective function using scipy.optimize.minimize function.

3 Results

MSE = 1549.9182553374153
Optimized parameters for avg score = [11.66921239, 26.81265025, 50.6237097, 78.5979276, 103.9751508, 137.71205471, 168.9074295, 207.66220961, 239.23836829, 284.34198406]
Optimized parameter(L) = 10.874600352509741

4 Plots


Runs Scored
W1
W2
W3
W4
W5
W6
W7
W8
W9
W10
runs
overs remaining
0
50
100
150
200
250
0
10
20
30
40
50


Resources Remaining
W1
W2
W3
W4
W5
W6
W7
W8
W9
W10
percentage of resources remaining
overs remaining
0
10
20
30
40
50
60
70
80
90
100
0
10
20
30
40
50