NAME: K UGENDER
SR.NO: 21008
ugenderk@iisc.ac.in

Duckworth-Lewis Method
Data Analytics Assignment 1

2023-08-21

Task

Using the first innings data alone in the given data set, finding the best fit 'run production functions' in terms of wickets-in-hand W and overs-to-go u. Assuming the model $Z_{(u,w)} = Z0_{(w)}[1 - exp{-}Lu/Z0_{(w)}]$. Using the sum of squared errors loss function, summed across overs, wickets, and data points for those overs and wickets. And also plotting the ten functions and reporting the (11) parameters associated with the (10) production functions and the normalized squared error

1 Implementation

First, I loaded the data. Then I cleaned it by removing erroneous rows, selecting relevant columns, filtering by innings, calculating derived features, and handling missing values. This cleaned dataset than fed into the model to make accurate predictions.

1.1 DLModel

In The DLModel class, I essentially encapsulate the functionality of training, predicting, calculating loss, saving, and loading the model for predicting scores based on the given parameters L, Z0, wickets in hand, and number of overs remaining. I have predicted scores based on the given input parameters using the equation

$$Z_{(u,w)} = Z0_{(w)}[1 - exp{-}Lu/Z0_{(w)}]$$

Then I calculated the loss using Squared Error loss for the model's predictions compared to the actual average score values. using the equation

$$\mathbf{E} = \sum_{i}^{n}(Y_i - X_i)^2$$

where Y is the actual average score value and X is the predicted score.

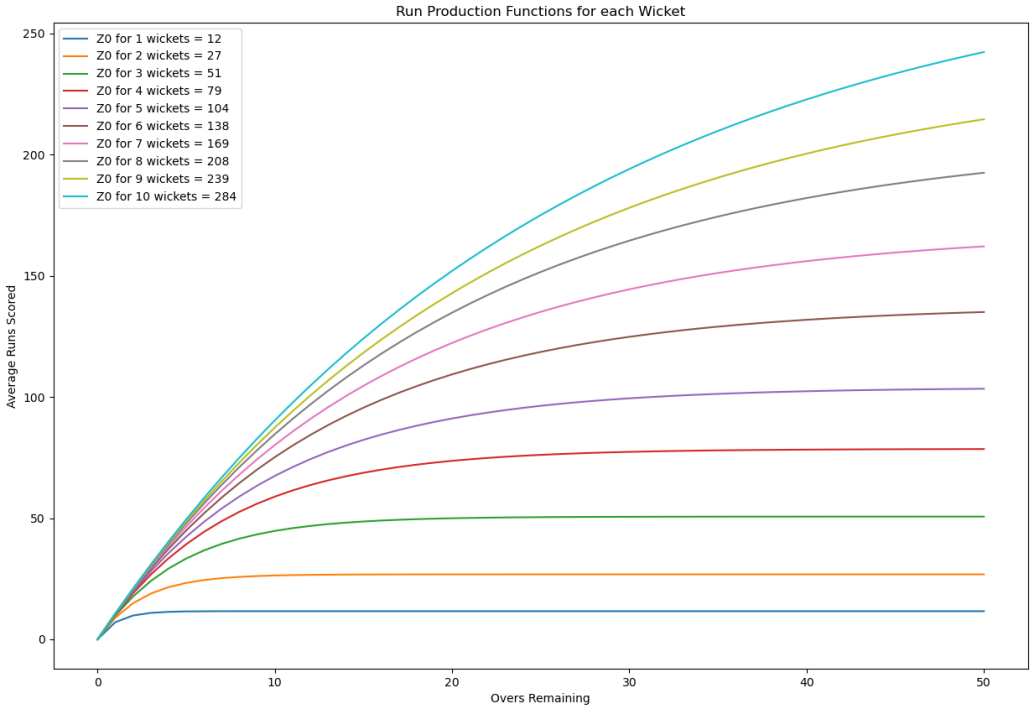
Then I used scipy.optimize.minimize to minimize the loss function. Then I got the optimal values for L and $Z0_{(w)}$ for all wickets.

1.2 Plotting

Then using the optimal values for L and $Z0_{(w)}$ for all wickets, I plotted the run production functions for all the values in a single plot. And saved the plot and model.

2 Results

2.1 Plot



2.2 Average Loss

I computed the normalized loss as squared error loss using Mean Squared Error for the trained model using the equation

$$\textbf{MSE} = \frac{1}{n}\sum_{i}^{n}(Y_i - X_i)^2$$

where Y is the actual average score value and X is the predicted score.

The minimized normalized squared error: 1549.9182899049802

2.3 Value of Model Parameters

The optimal values of 10 parameters $Z0_{(w)}$ associated with the production functions:

$Z0$ for 1 wicket: 11.62009182
$Z0$ for 2 wickets: 26.81656214
$Z0$ for 3 wickets: 50.65436987
$Z0$ for 4 wickets: 78.59012888
$Z0$ for 5 wickets: 103.95402227
$Z0$ for 6 wickets: 137.70281732
$Z0$ for 7 wickets: 168.89300626
$Z0$ for 8 wickets: 207.6501864
$Z0$ for 9 wickets: 239.22287589
$Z0$ for 10 wickets: 284.3184486

The optimal value of L: 10.875765419269804