Data Analytics Assignment1-Report Duckworth-Lewis Method

Alice Laguri

21275

August 21, 2023

Implementation

AS Duckworth-Lewis-Stern(DLS) method is used to calculate a revised target score for overs to go and wicket in hand. To begin with data pre-processing, taking samples when innings=1, this reduced the dataset from 126768 samples to 67794 samples. Removed all unnecessary columns and considered only ['Over','Runs', 'Wickets.in.Hand','Total.Runs', 'Runs.Remaining'].

Total of 600 cells were not in format(DD-MM-YYYY) and it can either be changed to pd.datetime format or can be removed as for few innings, few data were repeating with match held on that particular day like 29-30 May, 1999 with 29/05/1999. Also there was 'Error.In.Data' column with value 0 or 1, considered 'Error.In.Data' = 0 and got 67288 samples. Further, I added 'overs.remaining' column from overs as we had data for 'wickets.in.hand', **hence DLS method was applied on 67288 x 6 data after data pre-processing.**

Run Production Function(Z):

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

where, $Z_0(w)$ = average runs scored
w = wickets in hand
$L = b(w)Z_0(w)$ (Slope)
$b(w)$ = growth rate for the corresponding wicket.

There are two methods were used to find parameters of production function :
Scipy.optimize.Curve_fit and Scipy.optimize.minimize. Both were giving nearby same values but to calculate loss, we need Scipy.optimize.minimize as Curve_Fit is better choice for fitting curve and not for optimization. Thus I used Scipy.optimize.minimize to calculate parameters and mean squared loss.

Few Observations can be made by output below:
Parameter table Z0 satisfies Z0(1)<Z0(2)<Z0(3)<.....<Z0(10). Also value increment for wicket 9 to 10, i.e., Z0(9)<Z0(10) has increased drastically but was not the case for others. The Reason being, 10 wicket may be reached if no interruption occurred, but for interruption 10 wicket may not complete. We are calculating based on remaining wicket and remaining overs while overs could be less than 50, thus the wicket remaining for smaller value is comparatively less.
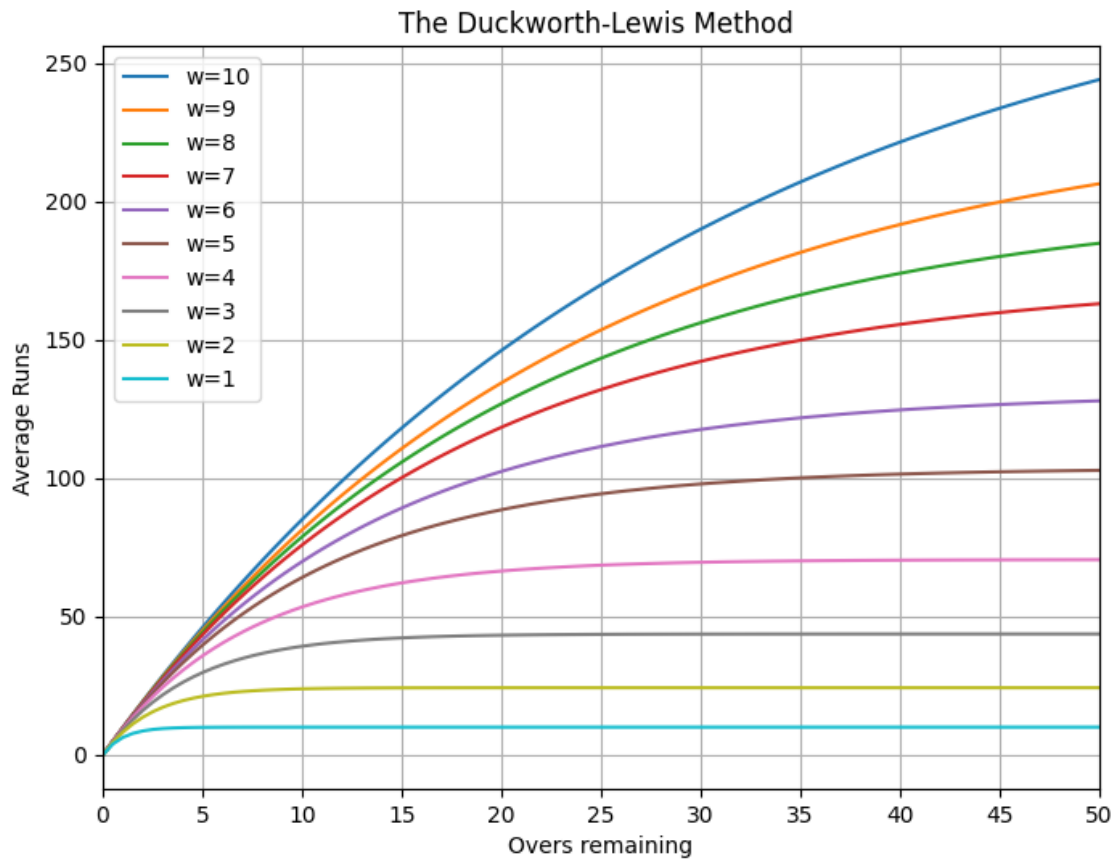
Results

Figure 1: Average Run vs Overs remaining plot after estimating 11 unknown parameters

Model Parameters observation:

w	1	2	3	4	5	6	7	8	9	10
Z0	11.982	28.365	50.772	79.993	106.081	136.194	169.794	206.991	234.886	294.80

Table 1: Different values of Z0 for different wicket

L	10.30829

Table 2: Slope L

Average Loss:

MSE: 1569.2217