

Data Analytics Assignment - 1

G Ramesh Babu
SR No. 20950

Duckworth Lewis Method

Abstract

This Report shows the analysis done on Duckworth Lewis method as part of Data Analytics Assignment - 1

1 Implementation Summary

1.1 Data Preprocessing

- The given .csv file is loaded using pandas.
- For this problem, the columns that are required are: 'Date', 'Innings', 'Over', 'Runs', 'Total.Runs', 'Innings.Total.Runs', 'Runs.Remaining', 'Wickets.in.Hand' and 'Total.Overs'. Rest of the columns are removed using *iloc()* method of pandas dataframes.
- The date format is converted to 'dd-mm-yyyy' by looping through the whole column.
- *dropna()* method is used to remove any rows with empty cells.
- Rows related to the 2nd Innings are removed.
- Finally data is arranged in a numpy array in such a way that, the first 10 indices have lists of (runs, overs) tuples pertaining to respective wickets in hand and the last index has the list of runs scored in the 50th over in all the matches.

1.2 Training

- Non linear Regression is applied using `scipy.optimize.minimize()`.
- Loss function and initializations are to be given as arguments to this method. *totalLoss(Params)* is used as loss function which is defined in *train_model()*. *totalLoss(Params)* takes the model parameters and calculates the normalized squared error loss over all the datapoints using the *calculate_loss()* methods that are already defined.
- Model Parameters are initialized with *model.Z0* set to a list of all 1s and *model.L* set to the approximate value of *L* which is calculated using the runs scored in the 50th Over.
- Once training is done, Model Parameters are finally set to the obtained values

1.3 Plots and Loss

- Normalized squared error loss is calculated by finding the predicted runs using the final model parameters and the *calculate_loss()* method.
- Plots are drawn using the final model parameters and then using linearly spaced overs between 1 – 50

2 Results

2.1 The plot with 10 Curves

The following plot shows the Run Production Curves for each 'Wickets.in.Hand' value

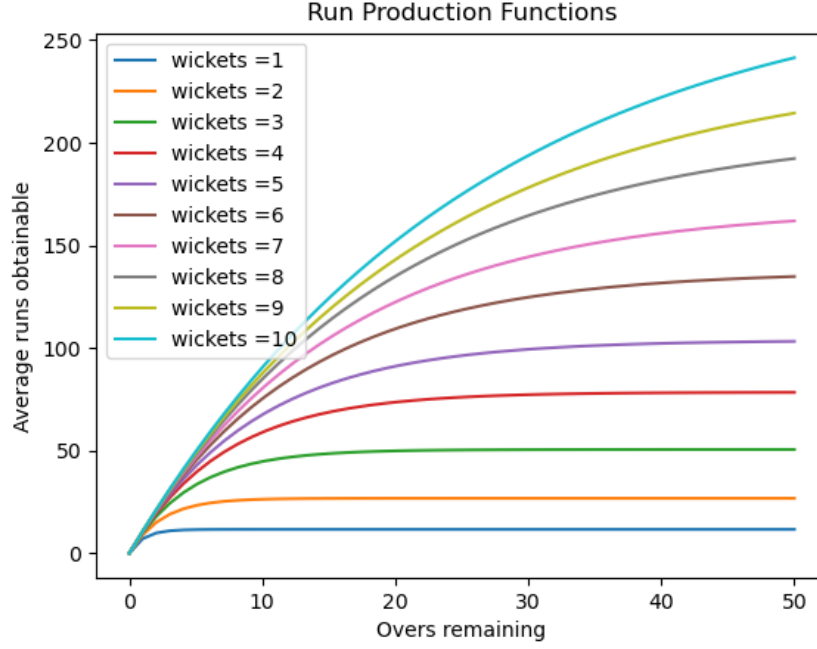


Figure 1: Run Production Curves

2.2 Normalized Squared Error Loss

The Normalized Squared Error Loss obtained is: **1609.5452968488664**

2.3 Values of Model Parameters

Model Parameters Obtained are as follows:

Wickets in Hand	Z0	L
1	11.663121444031065	10.91458299143258
2	26.794659873752533	10.91458299143258
3	50.58495531538822	10.91458299143258
4	78.49999091514236	10.91458299143258
5	103.82252940799044	10.91458299143258
6	137.45176417176683	10.91458299143258
7	168.56996331341938	10.91458299143258
8	207.21186470646617	10.91458299143258
9	238.72675759297806	10.91458299143258
10	282.2656439983367	10.91458299143258

Table 1: Model Parameters