Name: Dapse Vidyesh Rajendra

SR.NO. 06-18-01-10-51-21-1-19705

Duckworth-Lewis-Stern method

1. **Problem statement:**

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

   Assume the model $Z(u,w) = Z0(w)[1 - exp\{-Lu/Z0(w)\}]$.

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

2. **Data:**
   Data file is "04_cricket_1999to2011.csv". This file contains data on ODI matches from 1999 to 2011. Data consists of 1423 matches data. For each match, details are provided for each over. So overall data has 126768 data samples and 38 features.
3. **Implementation summary:**
   a. **Data pre-processing:**
      i. As we do not require second inning data, so I remove the second inning data samples.
      ii. I selected 8 features: Match, Over, Runs, Total.Runs, Innings.Total.Runs, Runs.Remaining, Innings.Total.Out, Wickets.in.Hand. At this stage, we have 67794*8 data shape.
      iii. By grouping the data according to the Wickets.in.Hand and Over. I created a new data for optimizing cost function. This new data contains Wickets.in.Hand, Remain_Over and Runs.Remaining.
   b. **Optimization using scipy library**
      i. Use **scipy.optimize.least_squares** function for finding parameters. These parameters consist of z_o(w) and growth rate (L).
      ii. **Optimization function:**

      Cost function = $\sum_i (\boldsymbol{yi} - Z(u,w))^2$

      iii. For this function, initialise the parameters which we want to optimise.
      iv. Define the loss function as per requirement of **scipy library**.
4. **Results:**

4.1 Parameters:

L	$Z_o(1)$	$Z_o(2)$	$Z_o(3)$	$Z_o(4)$	$Z_o(5)$	$Z_o(6)$	$Z_o(7)$	$Z_o(8)$	$Z_o(9)$	$Z_o(10)$
10.88	11.66	26.80	50.61	78.57	103.94	137.65	168.84	207.57	239.137	284.21

4.2 Normalised square error:

Normalised square error is **1557.34.**

4.3 Plot:

Curve for optimized parameters: