E0 259 DATA ANALYTICS ASSIGNMENT 1: DL METHOD

Name: - Samarth Atulkumar Patel

Sr. No.: - 05-08-02-10-51-21-1-19794

SUMMARY:

DATA ANALYSIS

The data taken for analysis is the first inning score. Total no. of unique matches for first innings is equal to 1423 out of which 990 were fully completed and 1300 matches have been played above 40 overs in the first innings. So, all the matches were consider in building the model as only 9 percent of matches were played less than 40 overs (135 Matches).

DATA CLEANING

The data in the csv file was incorrect. Total.Runs, Runs, Remaining Column has some wrong values so to take care of it Total.Runs columns was cleaned by taking cumulative sum of Runs Column for given match. Now the data for 0th over used was not in the data so a function call add row was written which adds row that contains situation of match when zero overs were used for a given match.

PARAMETERS INITIALIZATION

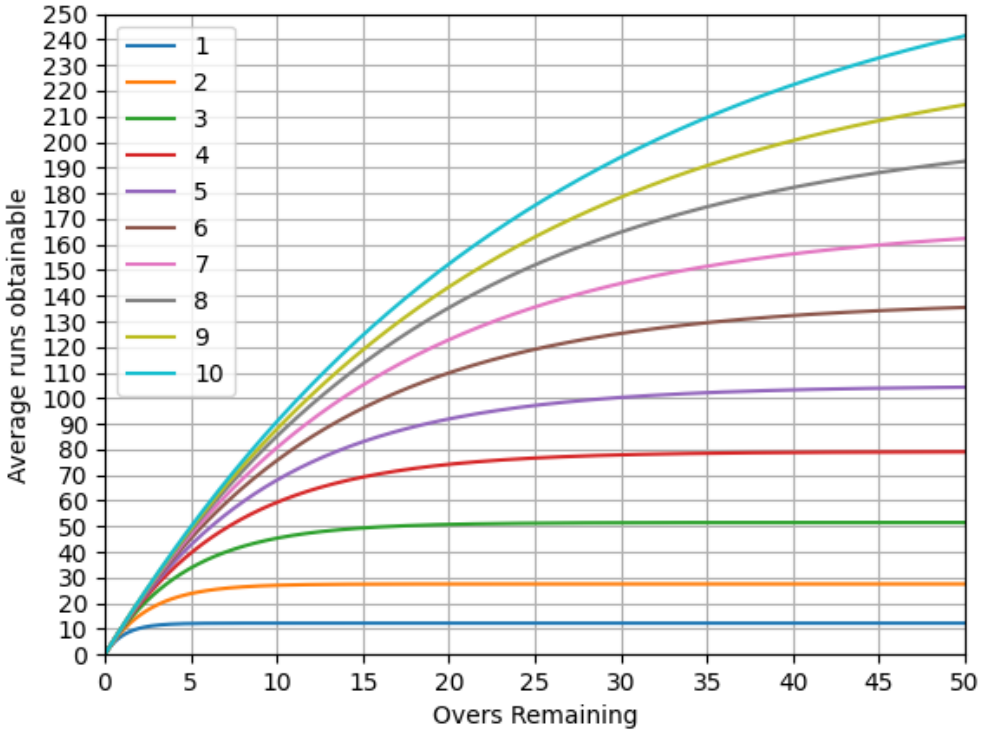
Data was grouped by for given wickets and again each group were further grouped by with match and max score for given match were taken and median of that value was used as a initialization for a given wicket. Median was taken as to make the initialization outlier prone. L was initialized as 9.

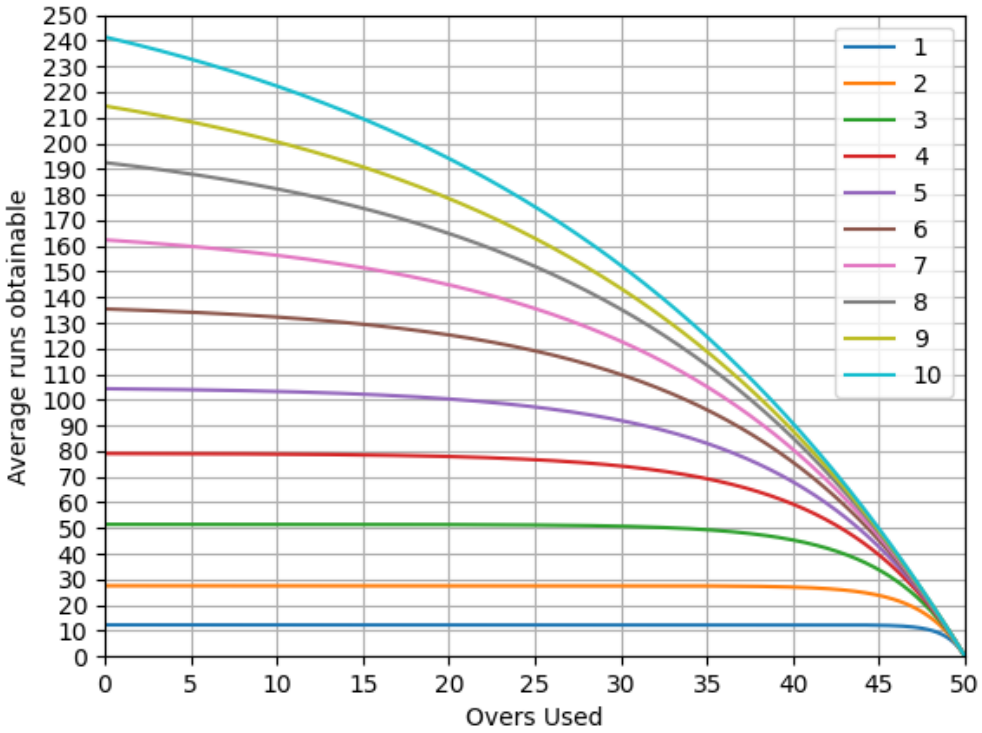
OPTIMIZING

Mean square error was used as a loss function and scipy optimization was used and method used for optimization was L-BFGS-B as it ended to be fastest.

RESULTS:

```
L = 10.96702470119143
mse = 1605.7913443793025
      Wicket_remaining          Z0
0   Wicket_remaining-1   12.268316
1   Wicket_remaining-2   27.515045
2   Wicket_remaining-3   51.451287
3   Wicket_remaining-4   79.143964
4   Wicket_remaining-5  104.877712
5   Wicket_remaining-6  138.008916
6   Wicket_remaining-7  168.905690
7   Wicket_remaining-8  207.133142
8   Wicket_remaining-9  238.450376
9  Wicket_remaining-10  281.619181
```


Average runs obtainable
Overs Remaining
1
2
3
4
5
6
7
8
9
10


Average runs obtainable
Overs Used
1
2
3
4
5
6
7
8
9
10