

# Data Analytics - Assignment I

Anup Patel (Sr.No. - 15474)  
M.tech CSA

August 29, 2019

## Run Prediction using DLS Method

**Problem Statement:** Using the first innings data alone in the above data set, find the best fit run production functions in terms of wickets-in-hand  $w$  and overs-to-go  $u$ . Assume the model is:

$$Z(u, w) = Z_0(w)[1 - \exp\frac{-Lu}{Z_0(w)}]$$

Here  $u$  represent remaining overs and  $w$  represent wickets in hand.

Use the sum of squared errors loss function, summed across overs and wickets.

## Data Preprocessing

Given data file contains data for ODI matches between 1999 - 2011. It consist of 38 columns and 126768 data instances. Out of 38 Columns we just required 7 columns (*Innings*, *Innings.Total.Runs*, *Total.Runs*, *Total.Overs*, *Over*, *Wickets.in.Hand*) for this assignment. There are 67794 instances of first innings. I had computed **Runs Remaining** and **over remaining** as follows:

**Run Remaining = Innings Total Runs - Total Runs**

**Overs Remaining = Total Overs - Over**

## Approach

Initialise parameter and then minimize squared error loss by using *scipy* package *minimize* function. This will return optimized value of parameters. Squared error loss is computed as follows:

$$\text{Squared Error Loss} = \sum_{n=1}^{67794} (A_n - P_n)^2$$

Here,  $A_n$  represent actual remaining Run when  $u$  overs are remaining and  $w$  wickets in hand. Whereas  $P_n$  represent predicted remaining run when  $u$  overs are remaining and  $w$  wickets in hand.

## Results

Parameter  $L = 10.89$

$Z_1 = 13.51$

$Z_2 = 27.36$

$Z_3 = 51.17$

$Z_4 = 78.82$

$Z_5 = 104.04$

$Z_6 = 137.75$

$Z_7 = 168.78$

$Z_8 = 207.49$

$Z_9 = 239.04$

$Z_{10} = 284.09$

**Total Squared Error Loss = 104818087.11**

**Root Mean Squared Error Loss = 28.75**

$$Z_{10} > Z_9 > Z_8 > Z_7 > Z_6 > Z_5 > Z_4 > Z_3 > Z_2 > Z_1$$

