

# Data Analytics Assignment 1

Kalyan Reddy, 21361

## 1 Implementation summary

### 1.1 Pre processing data

- **Column Selection:** Selected a specific set of columns from the input dataset, removing unnecessary columns.
- **Date Formatting:** It then processes the 'Date' column, handling different date formats such as 'DD/MM/YYYY' and 'Month-DD-YYYY'. The dates are converted into 'DD-MM-YYYY' format.
- **Data Cleaning:** If any missing values(NaN or null values) are found in the columns, the respective rows are dropped.
- **Training Data Extraction:** Extracted training data by considering different 'Wickets in Hand' (ranging from 1 to 10). For each 'Wickets in Hand', relevant data points are extracted from the dataset. Special handling is done for the case of 10 wickets and 50 overs remaining, where additional data points are collected.

### 1.2 Training model

- **Loss Function:** A custom loss function, 'func', is defined. It computes the loss for model parameters and training data. The loss formula involves parameters, input values, and a specific equation.
- **Optimization and Model Update:** Using an optimization library (sciPy.optimize.minimize), the function minimizes the loss. It adjusts trainable parameters ('*Z\_list*') (contains  $Z_0(1)....Z_0(10)$ ) and '*L*' to minimize mean squared loss across all data points. The optimized parameters update the model, which is returned.

The model is defined as:

$$Z(u, w) = Z_0(w) \cdot \left(1 - \exp\left(-Lu \cdot \frac{1}{Z_0(w)}\right)\right)$$

where  $u$  is the number of overs remaining,  $w$  is number of wickets in hand.

$L$  and  $Z_0(1)....Z_0(10)$  are optimizable parameters.

## 2 Results

### 2.1 Plot with 10 curves

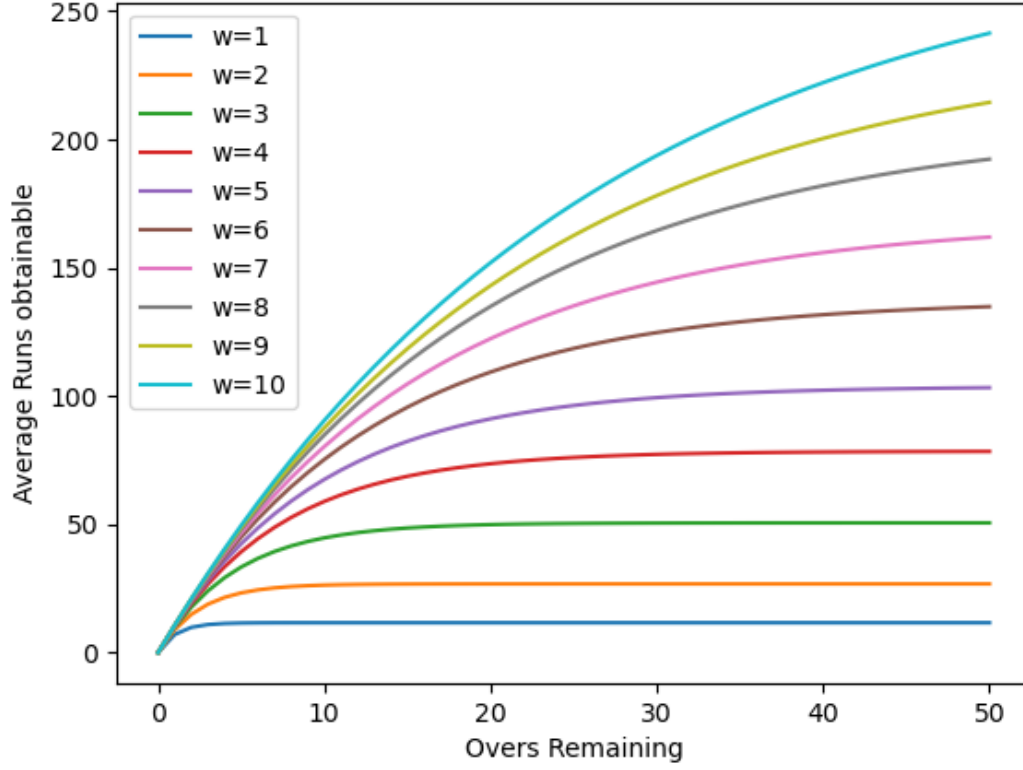


Figure 1: Overs Remaining vs Average Runs Obtainable for wickets in hand from 1 to 10

### 2.2 Average Loss

Normalised Squared Error Loss for all the training data points is: 1609.545297842099

### 2.3 Value of model parameters

Model Parameter  $L$ : 10.91423734759565

Wickets in Hand( $w$ )	Model Parameter $Z_0(w)$
$Z_0(1)$	11.663332803721149
$Z_0(2)$	26.794473796996286
$Z_0(3)$	50.58482608782695
$Z_0(4)$	78.50059076041828
$Z_0(5)$	103.82409447482873
$Z_0(6)$	137.4533446502598
$Z_0(7)$	168.57394246025524
$Z_0(8)$	207.21747295038145
$Z_0(9)$	238.7335827349905
$Z_0(10)$	282.2706942680824

Table 1: Model Parameters  $Z_0$