DA Assignment 4

Shreeman Agrawal

October 2024

1 Implementation Summary

1.1 Data Loading and Preprocessing

The script first loads cricket match data from 1999 to 2011, removing rows with missing values and filtering out erroneous data. The relevant columns such as `Runs.Remaining`, `Overs`, and `Wickets.in.Hand` are selected for further processing. The number of overs is converted from overs bowled to overs remaining using the transformation:

$$\text{Over} = 50 - \text{Over}$$

1.2 Model Definition

The `DLModel` class defines the mathematical formulation of the Duckworth-Lewis (DL) method. The primary function, which predicts the runs remaining (Z), is expressed as:

$$Z = Z_0 \cdot \left(1 - e^{-\frac{L \cdot X}{Z_0}}\right)$$

where Z_0 is the initial possible score, L is the decay factor, and X is the number of overs remaining. The class also includes a custom loss function based on logarithmic divergence to compare predicted and actual runs:

$$\text{Loss} = (y_{\text{pred}} + 1) \log\left(\frac{y_{\text{pred}} + 1}{y_{\text{actual}} + 1}\right) - y_{\text{pred}} + y_{\text{actual}}$$

1.3 Training and Evaluation

The `train_model` function runs the optimization process to fit the model to the dataset. It refines the Z_0 and L parameters and calculates the final model loss. It first minimizes Z_w and L_w for each w, then minimizes all the equations together with Z_0 initialised as Z_w and L initialised as the the weighted average of L_w.

2 Results

2.1 Plots

The plot for the initial 10 runs is shows below 2, and the plot for the final run is also shown 1

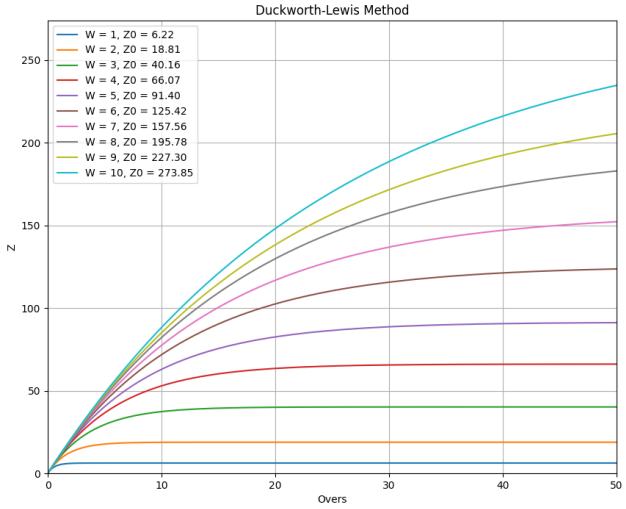

Figure 1: Plot for the final Run

2.2 Average Loss

The loss values for each wicket, are as follows:

- Wicket 1: 3.949
- Wicket 2: 4.914
- Wicket 3: 5.758
- Wicket 4: 5.930
- Wicket 5: 5.979
- Wicket 6: 5.984
- Wicket 7: 5.876
- Wicket 8: 6.509

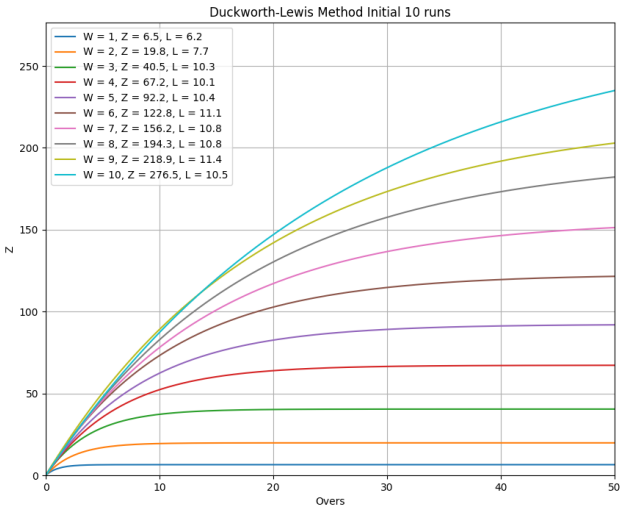

Figure 2: Plot for the initial 10 runs

- Wicket 9: 7.098
- Wicket 10: 6.798

Average loss for the final run : 58.891

2.3 Value of Model Parameters

For the first 10 runs:

$Z_w = [6.497, 19.815, 40.455, 67.233, 92.242, 122.849, 156.184, 194.256, 218.928, 276.508]$
$L_w = [6.237, 7.701, 10.285, 10.118, 10.407, 11.119, 10.824, 10.793, 11.446, 10.486]$

Final run:

$Z_0 = [6.222, 18.811, 40.156, 66.074, 91.403, 125.421, 157.555, 195.784, 227.299, 273.848]$
$L = 10.647$