Harshit Samani

M.Tech Artificial Intelligence

SR No. 17862

Sketch of code:

- **Pre-processing:**
  1. Removed all the 2nd innings data.
  2. Total.Runs and Runs fields are not consistent for some matches. I checked those matches on ESPN website and observed that Total.Runs is correct one(at least for those matches which I checked). So, I considered Total.Runs as reference to calculate Runs.Remaining for given resources.
  3. I observed that few matches are shortened possibly due to rain. It is possible that rain started before the match, or started in after few overs of first innings. Batting teams scoring approach depends on no. of overs they get to play. For example, If batting team knows about short match beforehand they would have scored more aggressively. In general, I found this data quite noisy due to lack of information. So, I decided to remove short match data. For that I removed all the matches for which wicket.in.hand > 0 and overs remaining<50.

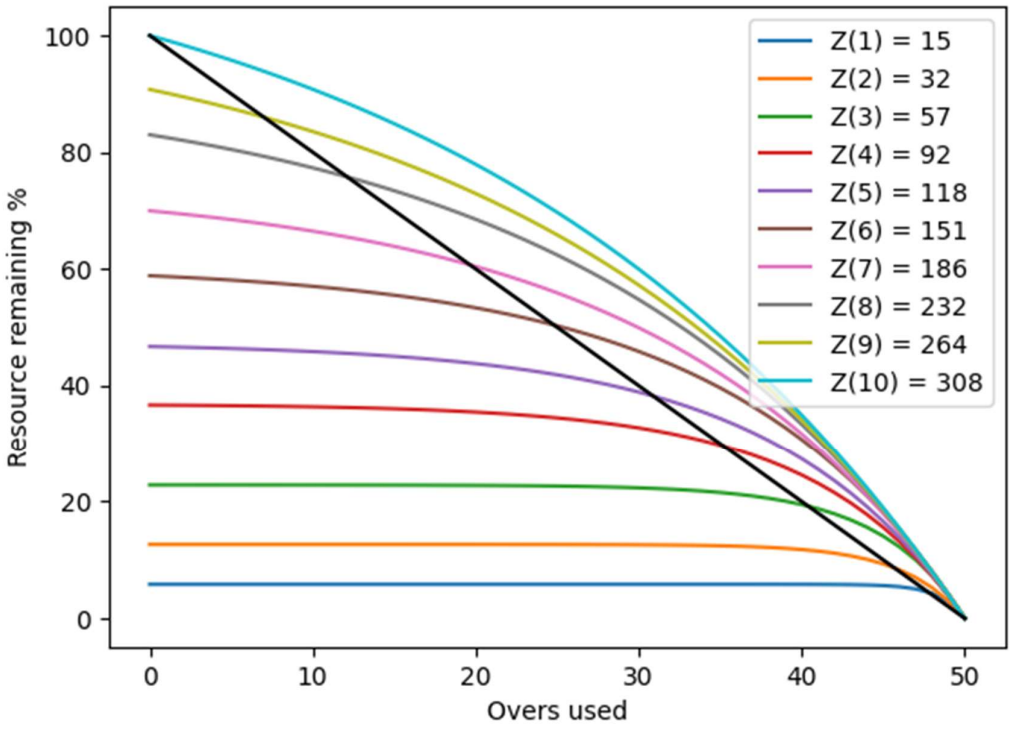
- **Regression:**
  1. Used scipy.optimize.minimize with L-BFGS -B method.
  2. Z0(w) is initialized with average runs for w wickets in hand.
  3. L is initialized with random value.

Results:

Model	Parameters		Slope = b*Z0	Error per point
20 parameters	Z0(1) = 14.55	b(1) = 0.662	9.64	1406.54
	Z0(2) = 31.59	b(2) = 0.267	8.42	
	Z0(3) = 56.99	b(3) = 0.192	10.91	
	Z0(4) = 91.93	b(4) = 0.110	10.10	
	Z0(5) = 118.16	b(5) = 0.086	10.19	
	Z0(6) = 151.25	b(6) = 0.071	10.69	
	Z0(7) = 186.40	b(7) = 0.055	10.29	
	Z0(8) = 232.44	b(8) = 0.044	10.30	
	Z0(9) = 264.36	b(9) = 0.039	10.27	
	Z0(10) = 307.68	b(10) = 0.033	10.25	
Model	**Parameters**		**Slope = L**	**Error per point**
11 Parameters	Z0(1) = 14.42	L = 10.24	10.24	1428.22
	Z0(2) = 30.18			
	Z0(3) = 58.15			
	Z0(4) = 91.94			
	Z0(5) = 118.16			
	Z0(6) = 155.57			
	Z0(7) = 186.40			
	Z0(8) = 232.44			
	Z0(9) = 264.36			
	Z0(10) = 307.68			

Slope values for both the models remains almost similar. For model1, since no. of parameters are more compared to no. of parameters in model2, so model1 can better fit the given data and we can achieve less training error. However, error doesn't reduce significantly. Hence, common initial slope is fair assumption for this model.

Question 1



Question 2