

Stochastic Gradient Descent

Q Why we are using Stochastic Gradient Descent? When we have the Batch Gradient Descent? What are the problems behind the Batch Gradient Descent?

Ans [Problems of Batch Gradient Descent]

- 1> When we have Big Dataset i.e., 1000 Rows and 5 features
- So for that, we will calculate 6 Coefficients $\{B_0, B_1, B_2, B_3, B_4, B_5\}$
- for 1 Coefficient the computation will be done as $\Rightarrow B_0 \times 1000 = 1000$ times..
for 6 Coefficients $\Rightarrow 6000$ values..

→ It leads to high Computation Problem.

- 2> As, per the Hard Code which we have done, there we load the Complete X-train while calculating the "Y_Predicted". When the "X-train" is too much big so it leads to Hardware Problem..

Q What is Stochastic Gradient Descent?

Ans 1> Stochastic Gradient Descent is faster because as per the mechanism of Stochastic Gradient \Rightarrow For a Big Dataset i.e., 1000 Rows & 5 features..
So every Coefficients [6 Coefficients] the epochs required will be less. because in 1 epoch.
["t" updates will be there in the value of coefficients..]

2> No Hardware Problem because In Stochastic Gradient Descent we select 1 Row at a time and updates all the coefficient values..

3> No Steady Solution \Rightarrow It means as per the Stochastic Gradient Descent selects any random row and update the values of the coefficient. It means the result will vary or will not same at every execution of the model.

Date.....

① To calculate the value of Intercept $\Rightarrow -2(Y_i - \hat{Y}_i)$

② To calculate the values of Coefficients $\Rightarrow -2(Y_i[\text{idx}] - \hat{Y}_i) / X_i[\text{idx}]$