

Mini Batch Gradient Descent

Q Why we use Mini Batch Gradient Descent, when we have the Stochastic Gradient Descent?

Ans • In Stochastic Gradient Descent we get the results [somewhat ^{heaver} to our best results]. But as we know that Stochastic Gradient Descent is random in nature. So, its results are fluctuating. and sometimes it cannot find the best results.

• So, to overcome this problem, Mini Batch Gradient Descent selects some random rows and this collection of rows is termed as Batch. In that particular Batch, 1 time the values of the coefficients are updated only. This leads to less randomness in the data.

Visualization of Types of Gradient Descent

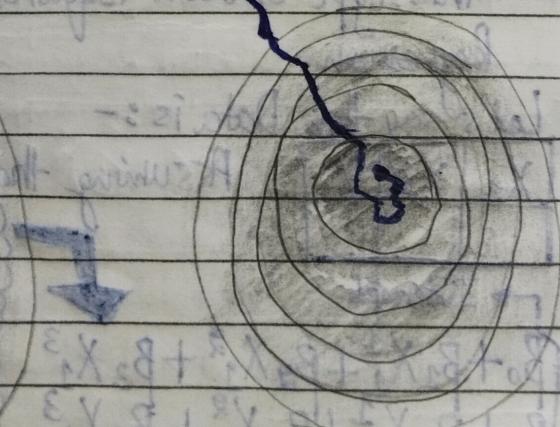
"No Randomness"

"Too much Randomness"

"Batch is selected"



Batch Gradient Descent



Stochastic Gradient Descent



Mini Batch Gradient Descent Spiral

Date.....

Note:- The Formulas are nearly same as in the case of Batch Gradient Descent and also the Algorithm or class building is also same.

[To calculate the slope \Rightarrow Derivative of Intercept]
$$-2 * \text{hp}.treat(Y_{\text{train}}[\text{idx}] - Y_{\text{predicted}})$$

[To calculate the Coefficients]
$$\Rightarrow -2 * \text{hp}.mean(Y_{\text{train}}[\text{idx}] - Y_{\text{predicted}}) * X_{\text{train}}[\text{idx}]$$

Where $\text{idx} = \text{hp}.random.sample(\text{range}(X_{\text{train}}.shape[0]), 10)$

It means we are selecting a Batch which includes 10 Random Rows.