

Lecture-3

Stochastic Gradient Descent.

Problem with Batch GD -

$$\frac{\partial L}{\partial \beta_1} = -\frac{2}{n} \sum_{i=1}^n \dots$$

~~for~~ we have to do lots of derivations.

- i) Algorithm becomes slow
- ii) we need more space.

- In Stochastic we see 1 row and then do update

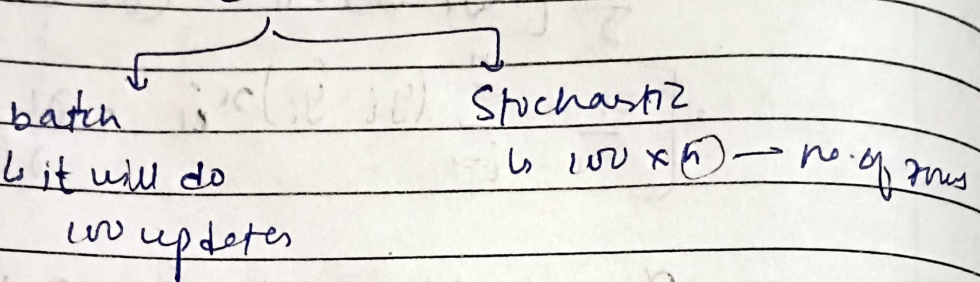
So if n rows then in 1 epoch - ~~need~~ n updates.

hence we reach Sol^n faster.

→ In this we select the rows randomly
So every time we get diff result

• Time comparison -

$e = 100$



→ But in large dataset stochastic don't need 100 epochs

• When to use SGD -

i) when we have Big data

ii) when we don't have convex fn

• There is one problem in SGD →

when we move closer to Sol^n then also it fluctuates.

So we use learning schedule which changes the values of α as we move closer to Sol^n

Lecture-3

• mini-Batch Gradient Descent

↳ we make batches

$$n = 10000$$

Suppose batch size = 100

So \Rightarrow 100 batches

So in each epoch we do 100 updates