$\sum_{i=1}$

Lecture-3

Stochastic Gradient Descent.

Problem with Batch GD -

$$\frac{\partial L}{\partial B_1} = \frac{-2}{n} \sum_{i=1}^{n} \cdots$$

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

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 - n updates.

hence we reach soln faster.

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

- Time comparison -

e = 100

batch
↳ it will do
100 updates

Stochastic
↳ 100 × (n) → no. of rows

→ 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 soln then also it fluctuates.
So we use learning schedule which changes the value of lr as we move closer to soln

Page No.
Date

Lecture-3

- Mini-Batch Gradient Descent.
  - ↳ we make batches.

$n = 1000$

Suppose batch_size = 10

So $\Rightarrow$ 100 batches

So in each epoch we do 100 updates