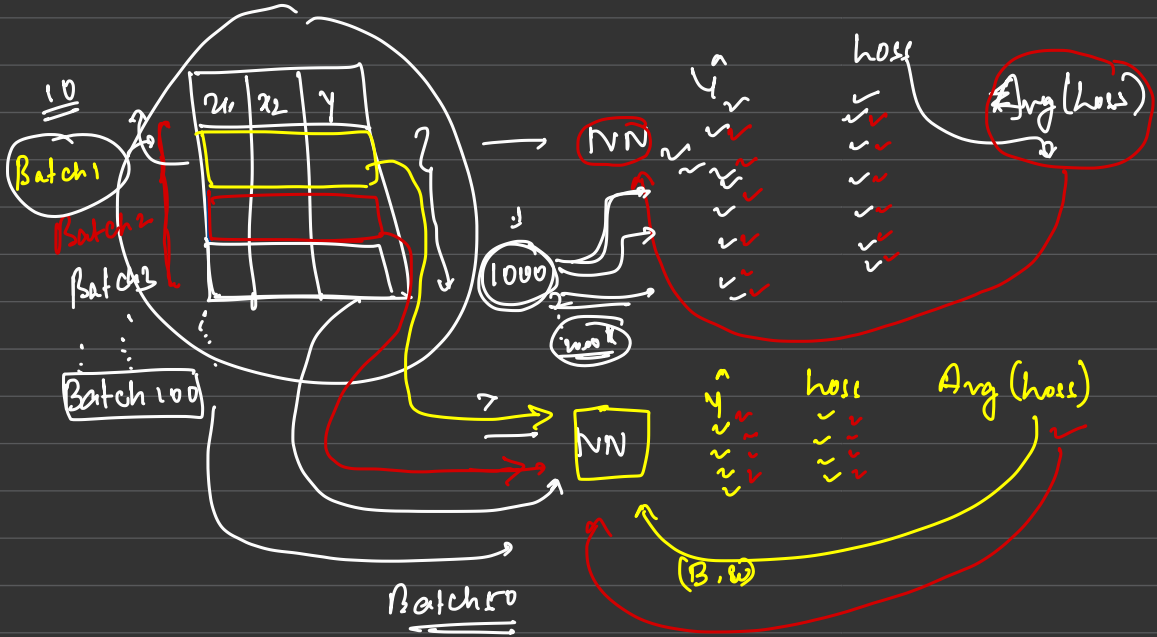


Batch Normalization



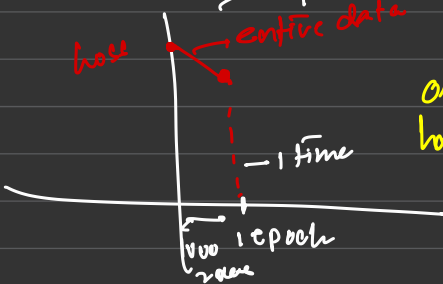
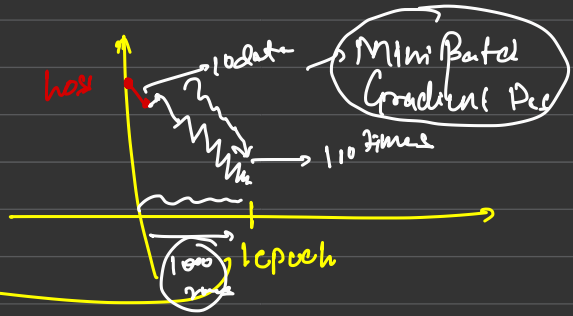
49 times

49 x 20
= 980

100th Batch

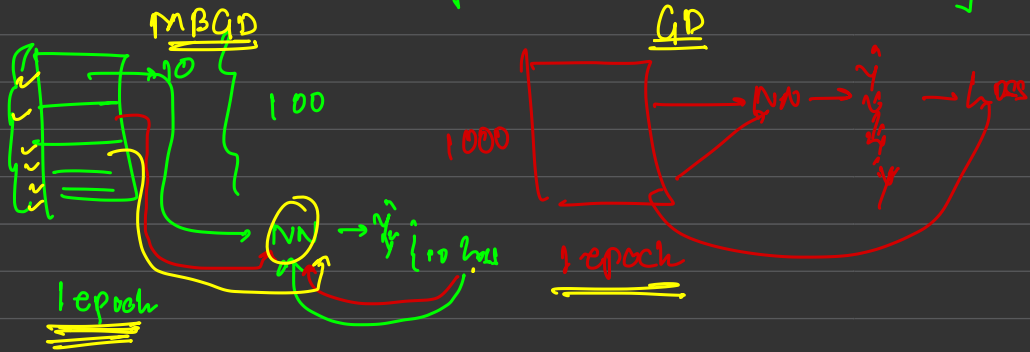
99 updates

Gradient Descent

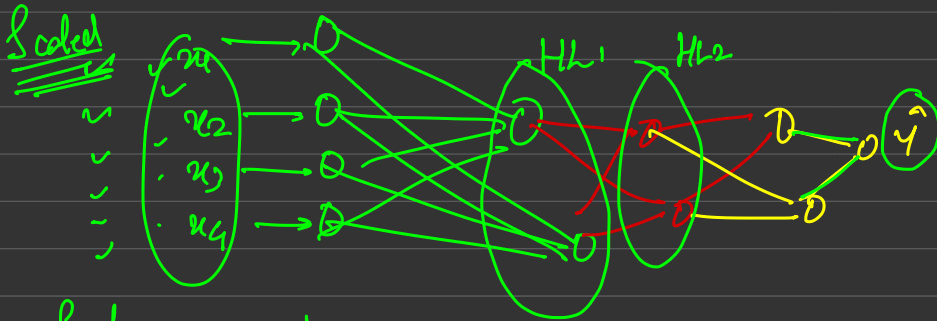


Only when I would have iterated through the entire data

In a mini batch gradient descent, we divide the total data set into mini batches - for eg if my total no of samples is lets say 1000, and I choose the batch-size = 10, then I will have 100 batches at my hand.



Batch Normalization



Scale our inputs

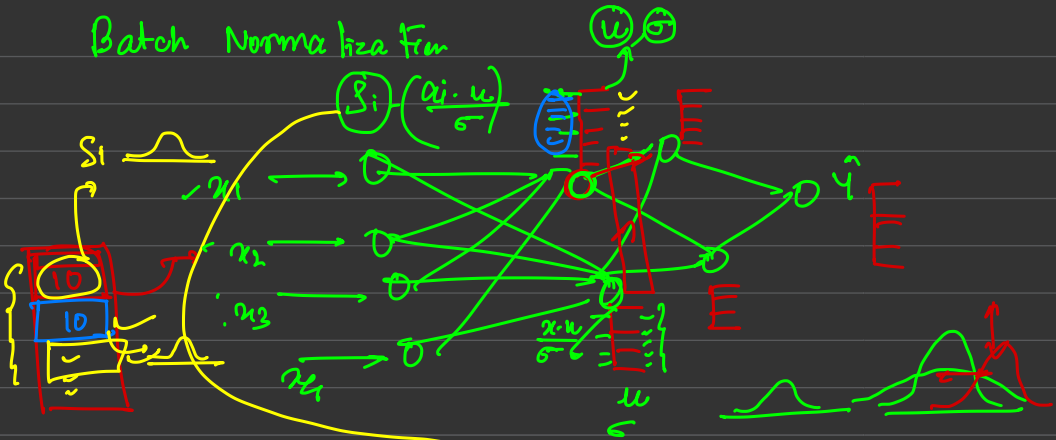
$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

(b_2)

Ram	Shyam
Honda City	Honda City
50,000 km	52,000 km
3rd owner	1st owner
2777	2111

Batch Normalization

Batch Normalization



Putting too much of a constraint
to my output

$$N_i = \gamma S_i + \beta$$

γ is labeled as **Gamma** and β is labeled as **beta**.

learnable parameters