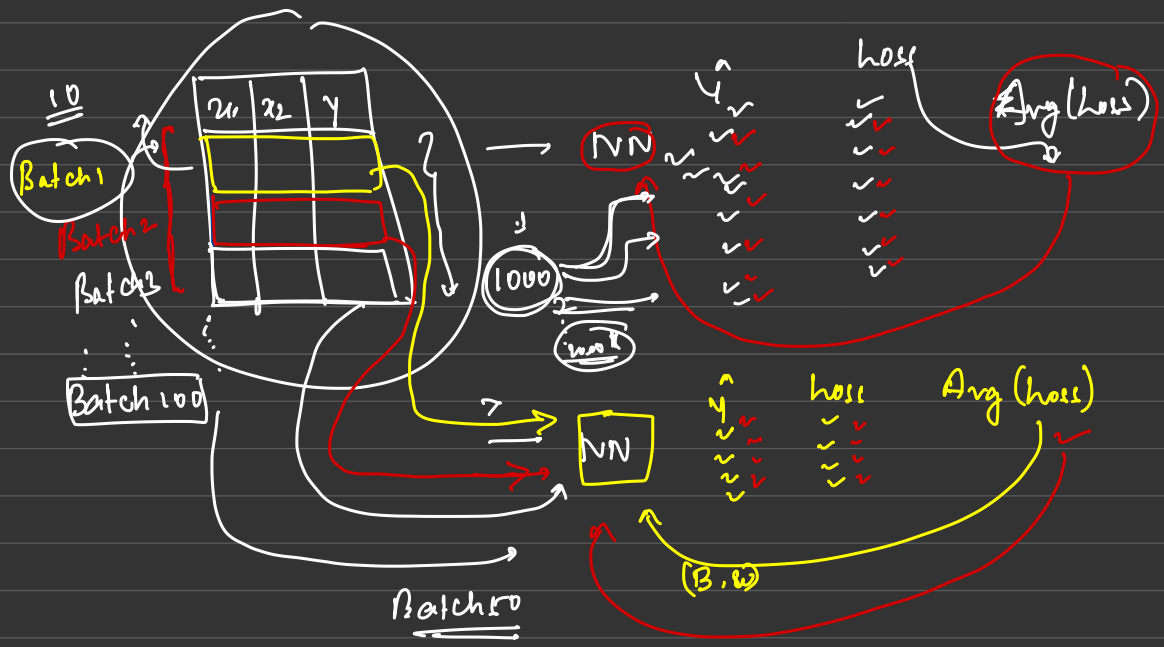
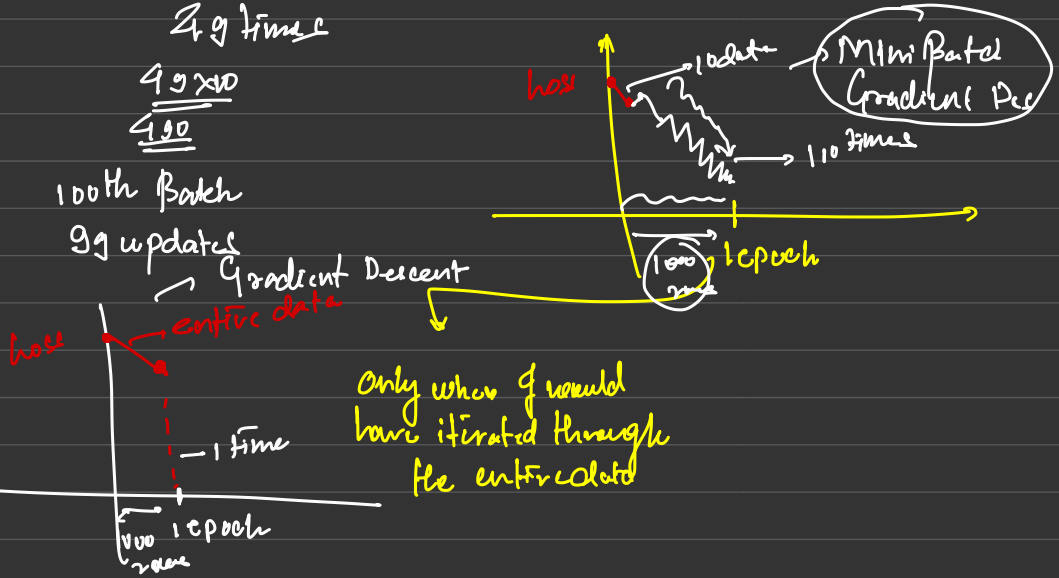
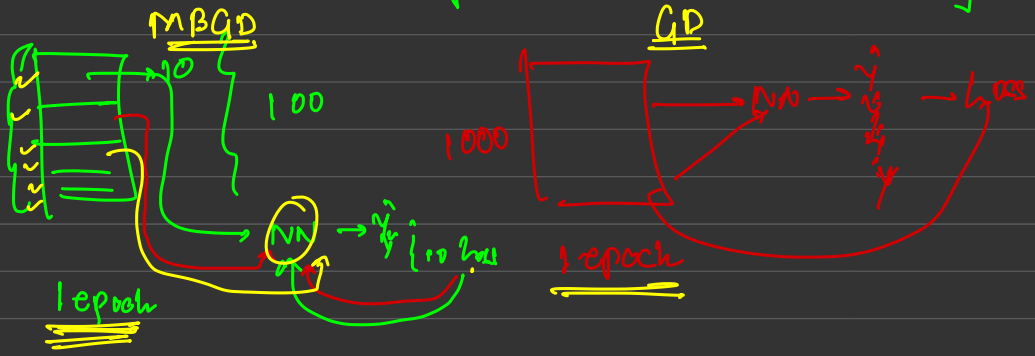
Batch Normalization

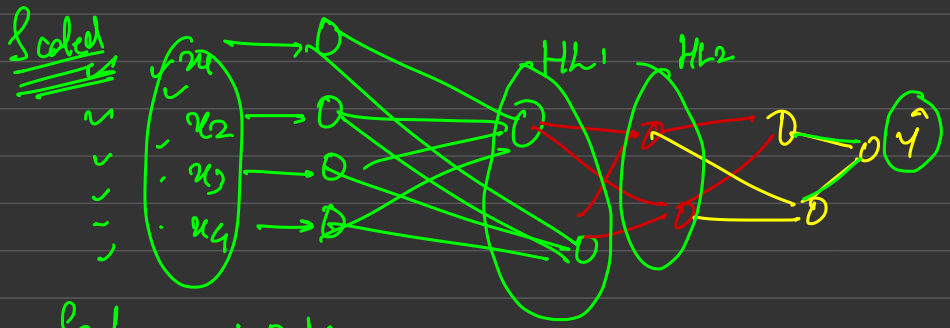




In a mini batch gradient descent, we divide the total data set into mini batches - for eg if my total no of samples is lets say 1000, and I choose the batch size = 10, then I will have 100 batches at my hand.



Batch Normalization



Scale our inputs

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

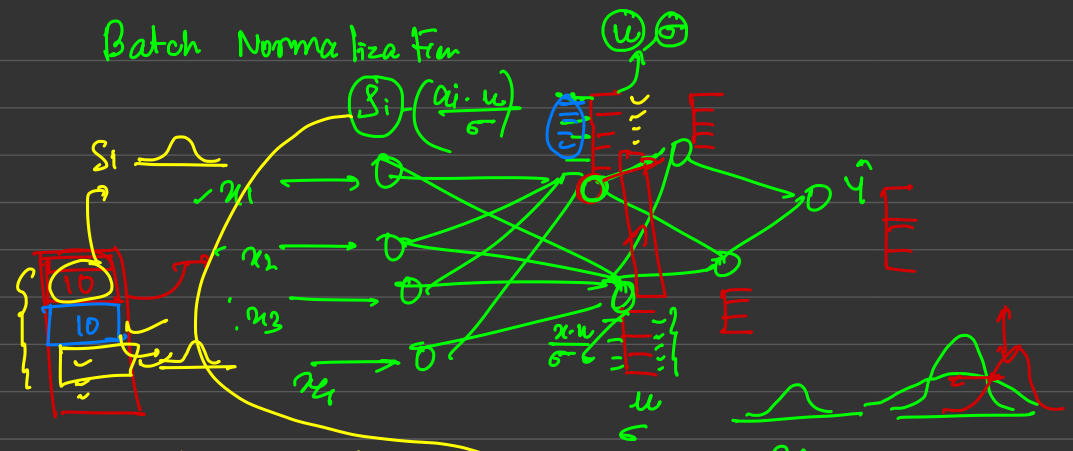
β_2

Ram
Honda City
50,000 kms
3rd Owner

Shyam
Honda City
52,000 kms
2000
1st Owner

Batch Normalization

Batch Normalization



putting too much of a constraint to my output

Ni = γ Si + β

Gamma

Beta

learnable parameters