## horizontal line

Backpropagation (class 10)

20.01.2025

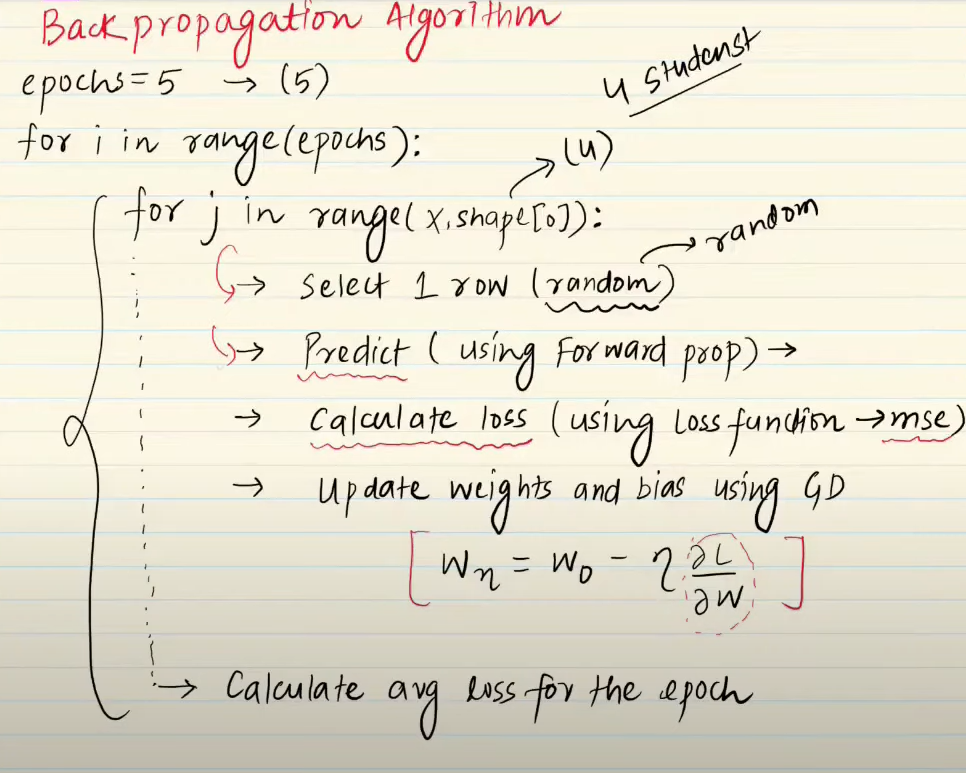
Backpropagation is an algorithm for supervised learning of ANN . given an ANN and an error function , the method calculates the gradient of the error function with respect to its weights.

[Backpropagation in regression](https://colab.research.google.com/drive/1S8VD1Y2a3calU1_1eiwYckkmZwJTCNmM)

[Backpropagation in regression using keras](https://colab.research.google.com/drive/1hskMbyqQ6TlIXJo5SM7JB_W6_gcy1fno)

[Backpropagation in classification](https://colab.research.google.com/drive/1WYcPSe2kPTTiIGbzruMtr2ajv4uSHvaI)

[Backpropagation in classification using keras](https://colab.research.google.com/drive/1B9Uqu6d6VBnyYdJt5EAngz_2VoVoin6C)

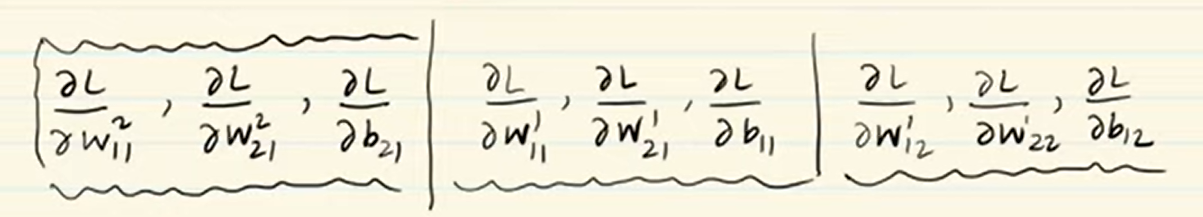


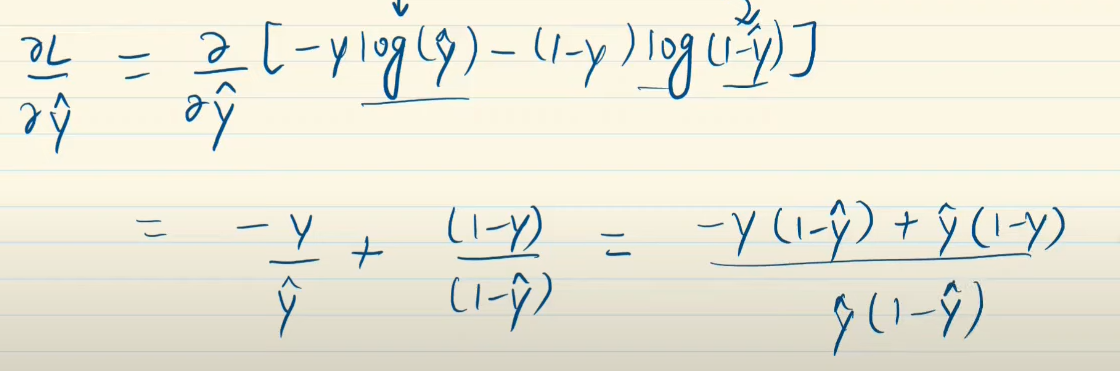
L = (yi - yi\_hat)^2



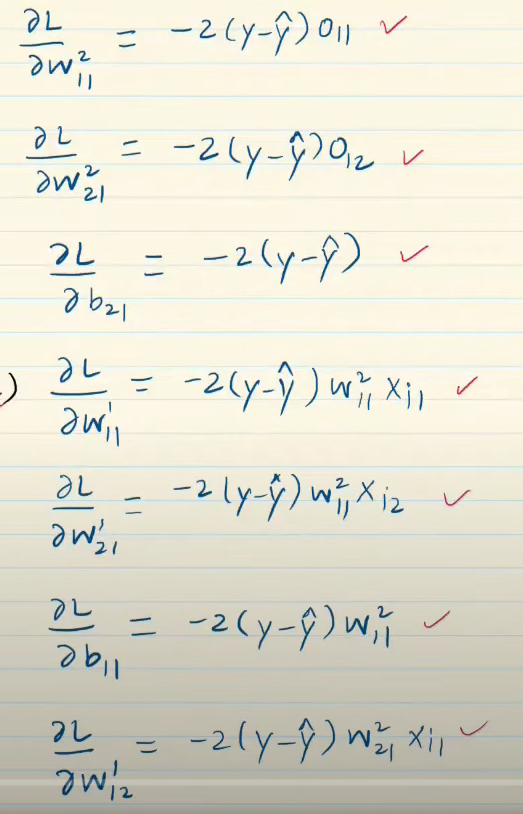
L depends on yi\_hat . yi\_hat depends on O11 ,O12 , W211 , W221 , b21

Similarly O11 depends on W111 , Xi1 , W121 , Xi2 , b11 and O12 depends on W112 , Xi1 , W122 , Xi2 , b12.





We can also calculate it when

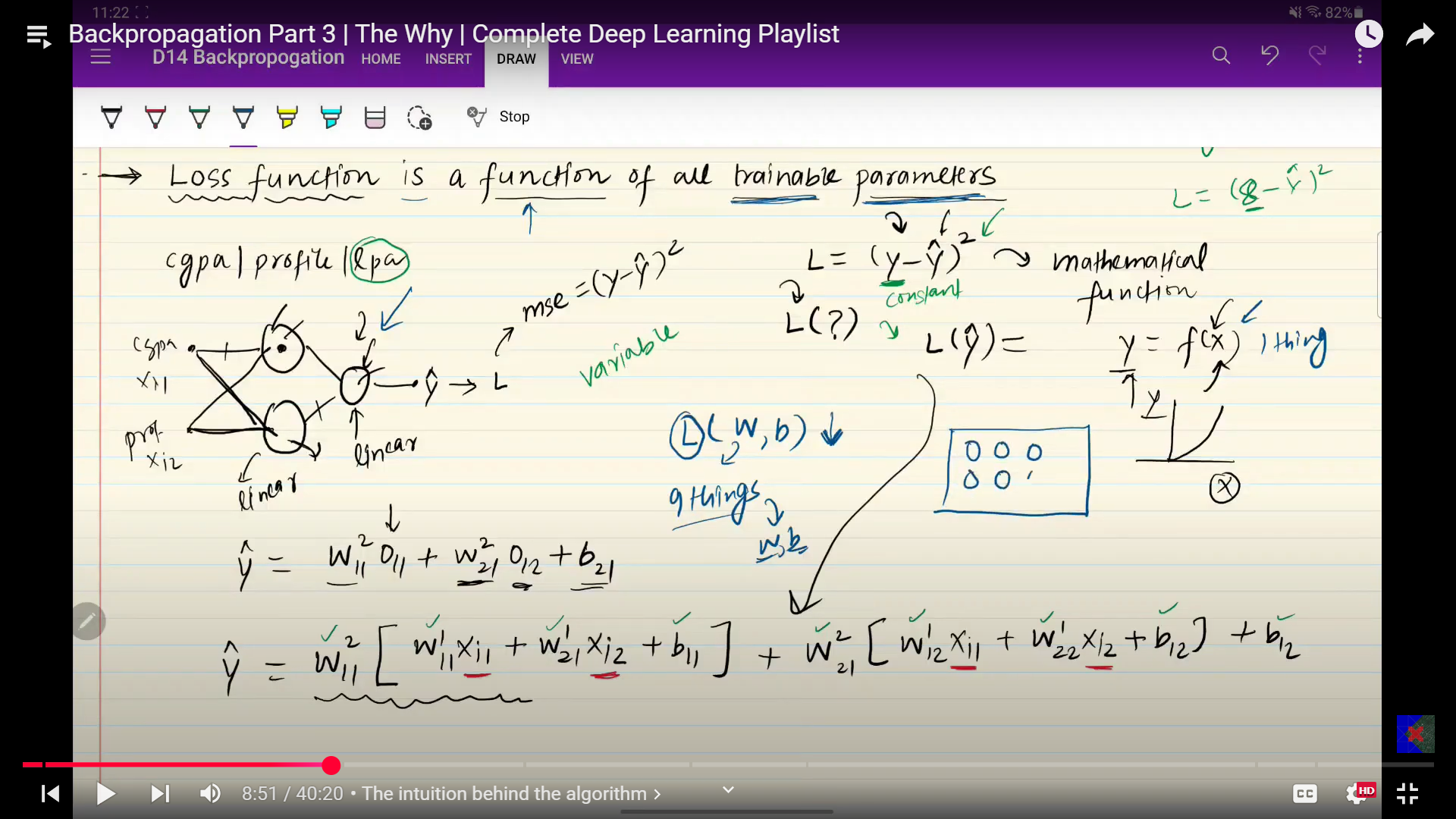


# Why Backpropagation work like this ?

We need to see intuition behind this formula:

Wnew = Wold - L/W

1. Loss function is a function of all trainable parameters.



1. Gradient is a partial derivative of any function that depends on 2 or more variables.

/ , /

While differentiation is derivative of the function that depends on only one variable.

1. If dy/dx = +ve so , increase in x increases y .
2. If dy/dx = -ve so, increase in x decreases y .
3. If dy/dx = 0 at any point means value is minimum here . Same as we need to find where Loss is minimum .
4. Let's take L now only depends on b21 .

If dL/db21 = +ve it means we need to reduce the value of b21 from its previous one and by the amount of dL/db.

If dL/db21 = -ve it means we need to increase the value of b21 .

Wnew = Wold - L/W

So, the minus sign here works well for both cases.

1. Learning rate is used to set the step size neither too small nor too high .