## horizontal line

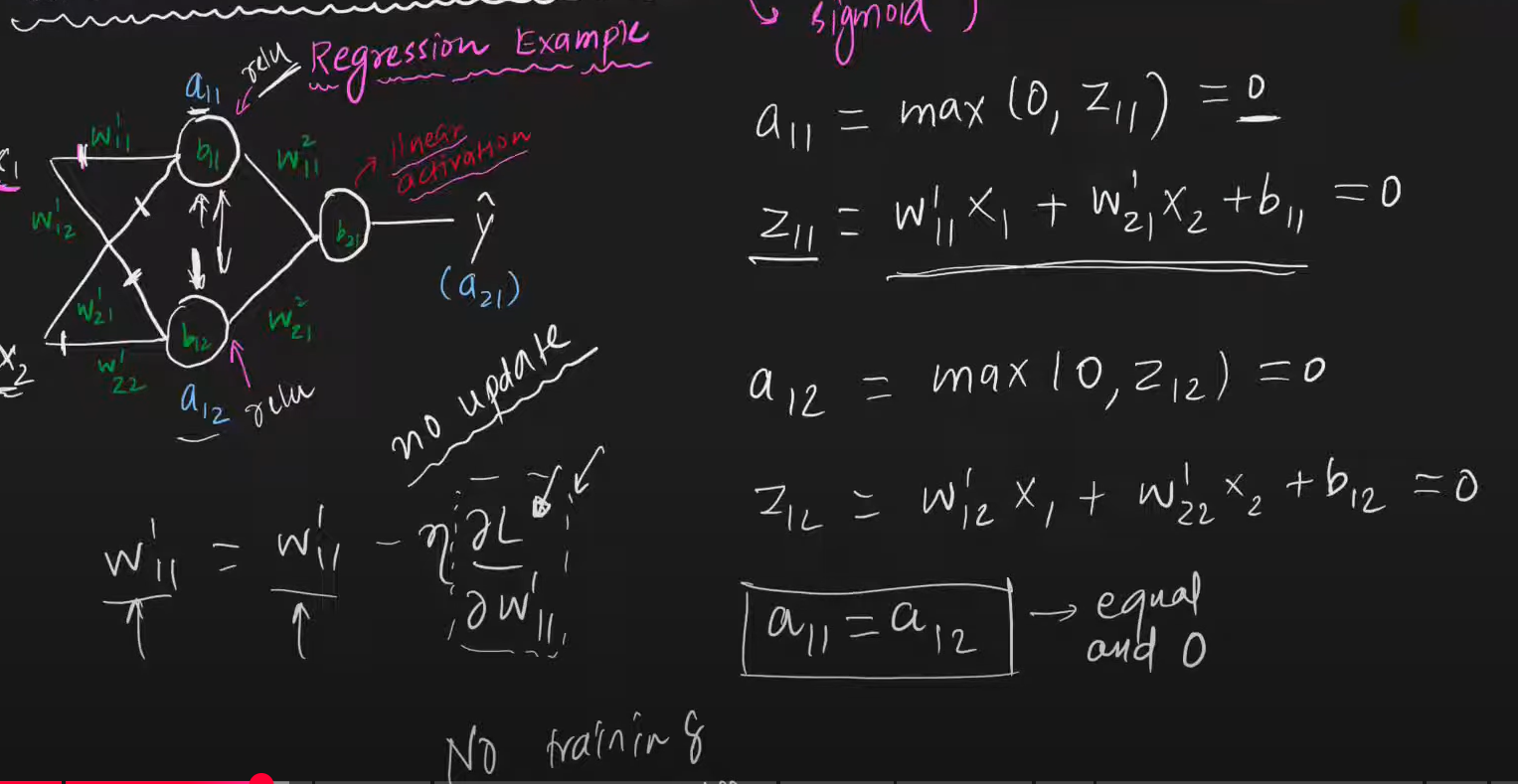
Weight Initialization

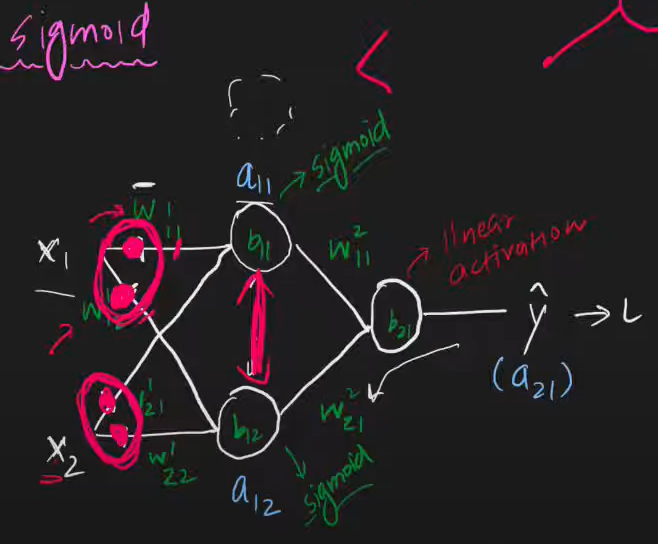
20.03.2025

[Xavier and He Weight Initialization](https://colab.research.google.com/drive/1dXFTBebGAyAEPdgr4SJNSZaphkIP8i5w#scrollTo=Uv_S02e00qaL)

# Case 1 : Zero Initialization

For tanh and relu , no training will take place .



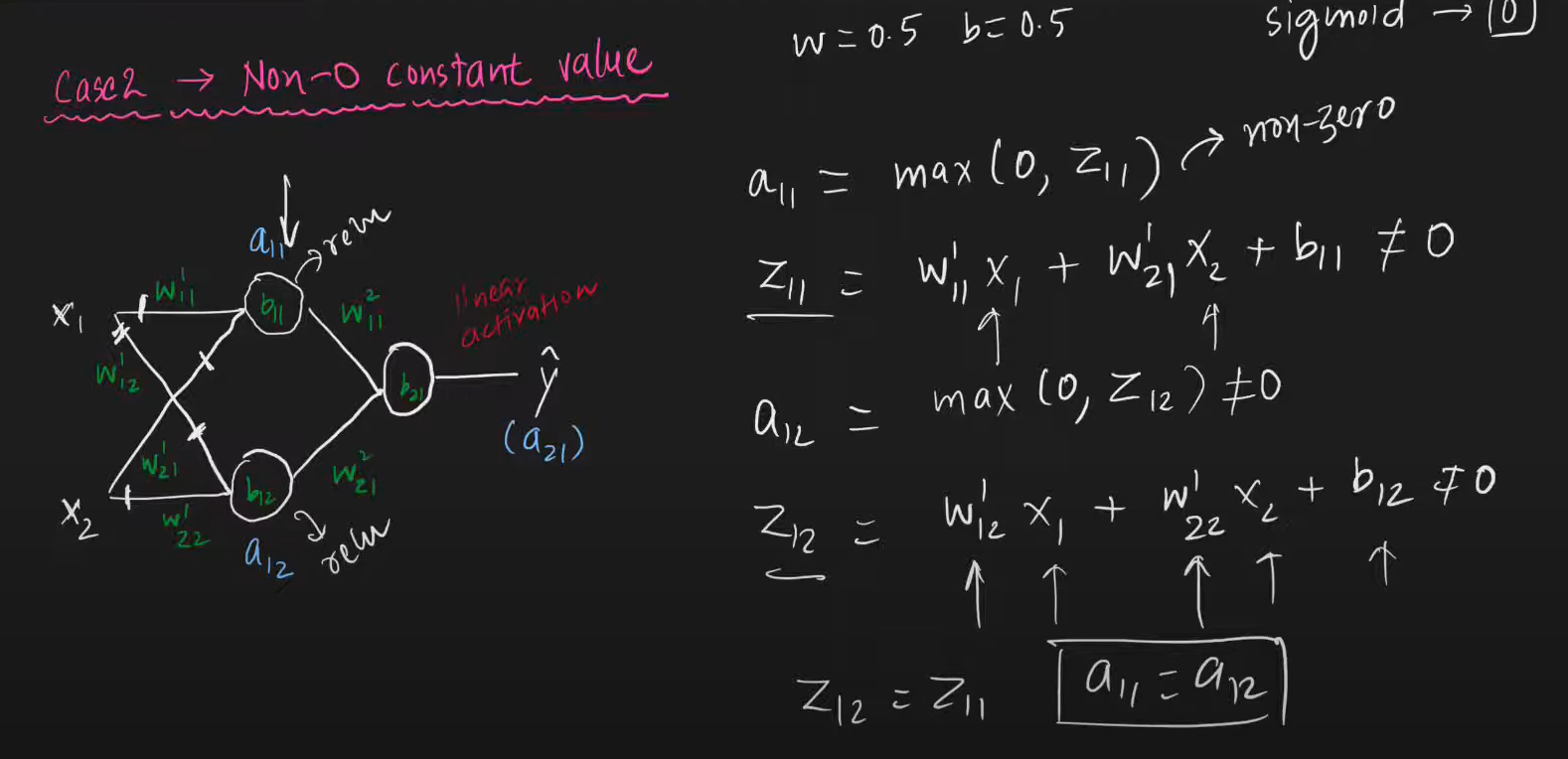


But in case of sigmoid , it will act as a linear model perceptron .

Here all nodes in the hidden layer act as one single layer.

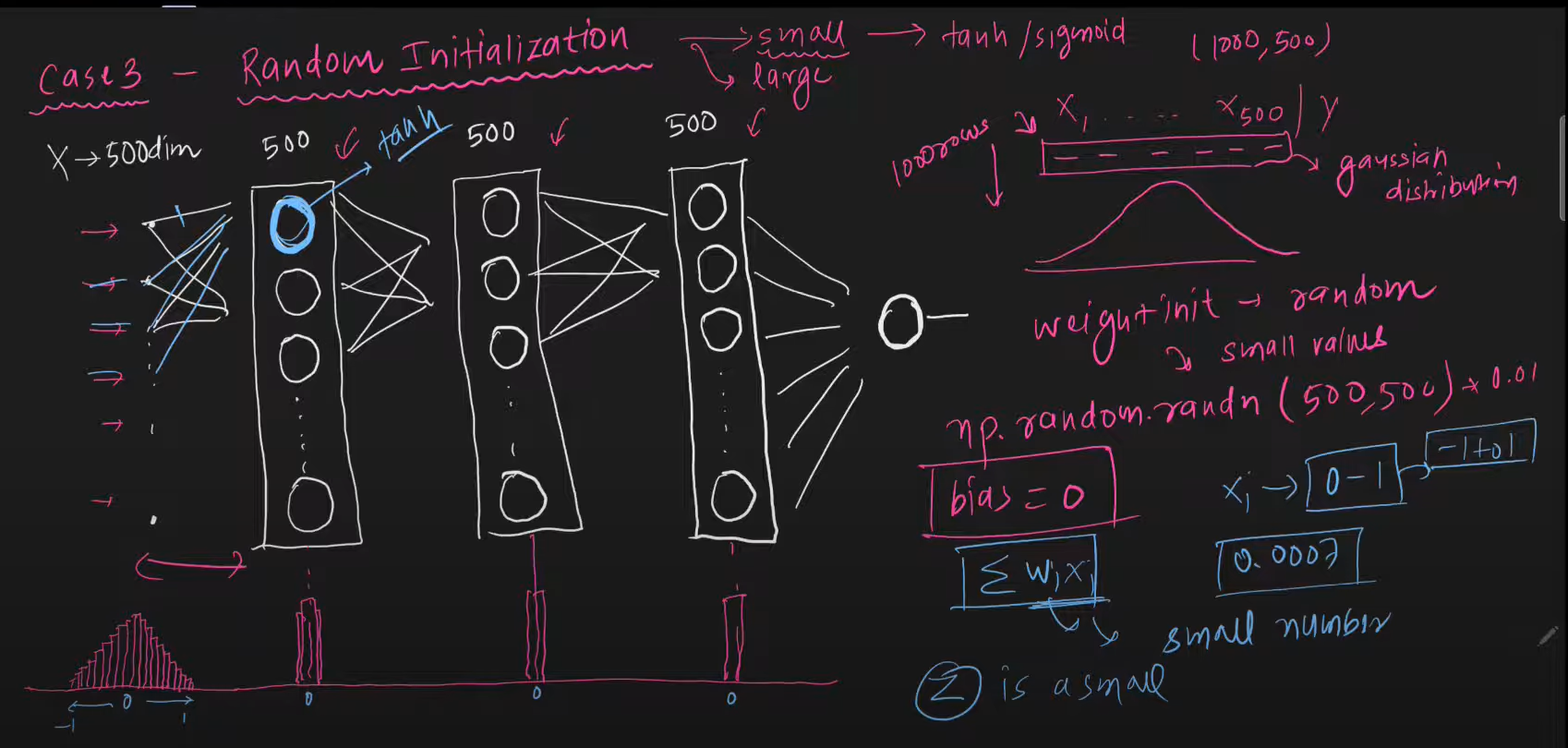
# Case 2 : Non Zero Constant Value

In case of all , it takes all the linear model perceptrons.



# Case 3 : Random Initialization

## Small values



Z is very small , close to 0.

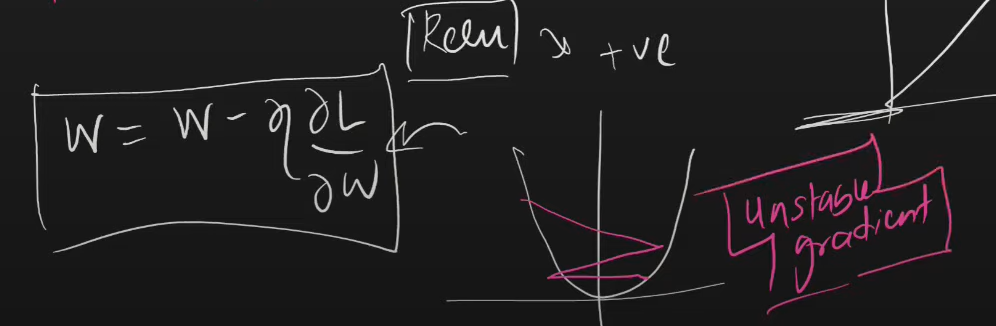
Thus in case of tanh, sigmoid VGP occurs while for Relu its a slow training.

## Large values

# 

# 

For relu there will be unstable gradients.



# Things not to do :

1. Zero initialization
2. Non zero constant initialization
3. Random initialization with small weights
4. Random initialization with large weights

# Xavier/Glorat and He Weight Initialization

Here we take variance as 1/n . If n is very very large weights would be divided by 1/n resulting in a large no. , and if n is very small , 1/n would be large increasing the sum WiXi.

1. Formula for Xavier Normal : and may be

This should act like np.random.randn(2,2)\*

1. Formula for He Normal :
2. Xavier Uniform : generate no.s between [-limit , limit]

Limit =

1. He Uniform : generate no.s between [-limit , limit]

Limit =