

The background of the slide is a dark gray to black gradient, overlaid with a complex, white, abstract network pattern. This pattern consists of numerous small circles (nodes) of varying sizes, connected by thin, white, irregular lines (edges). The nodes and lines are distributed across the entire frame, creating a sense of a vast, interconnected digital or neural network. The density of the connections is higher in some areas, particularly towards the right side of the slide.

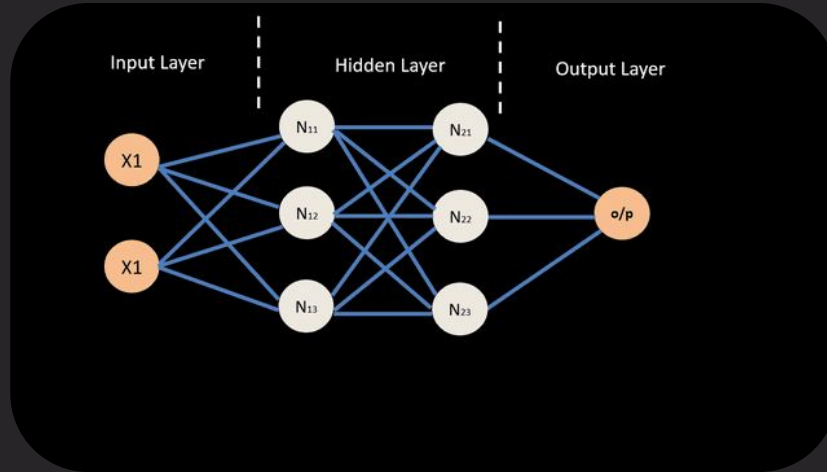
Understanding workings of Neural Networks

Video 4: What is Backpropagation?

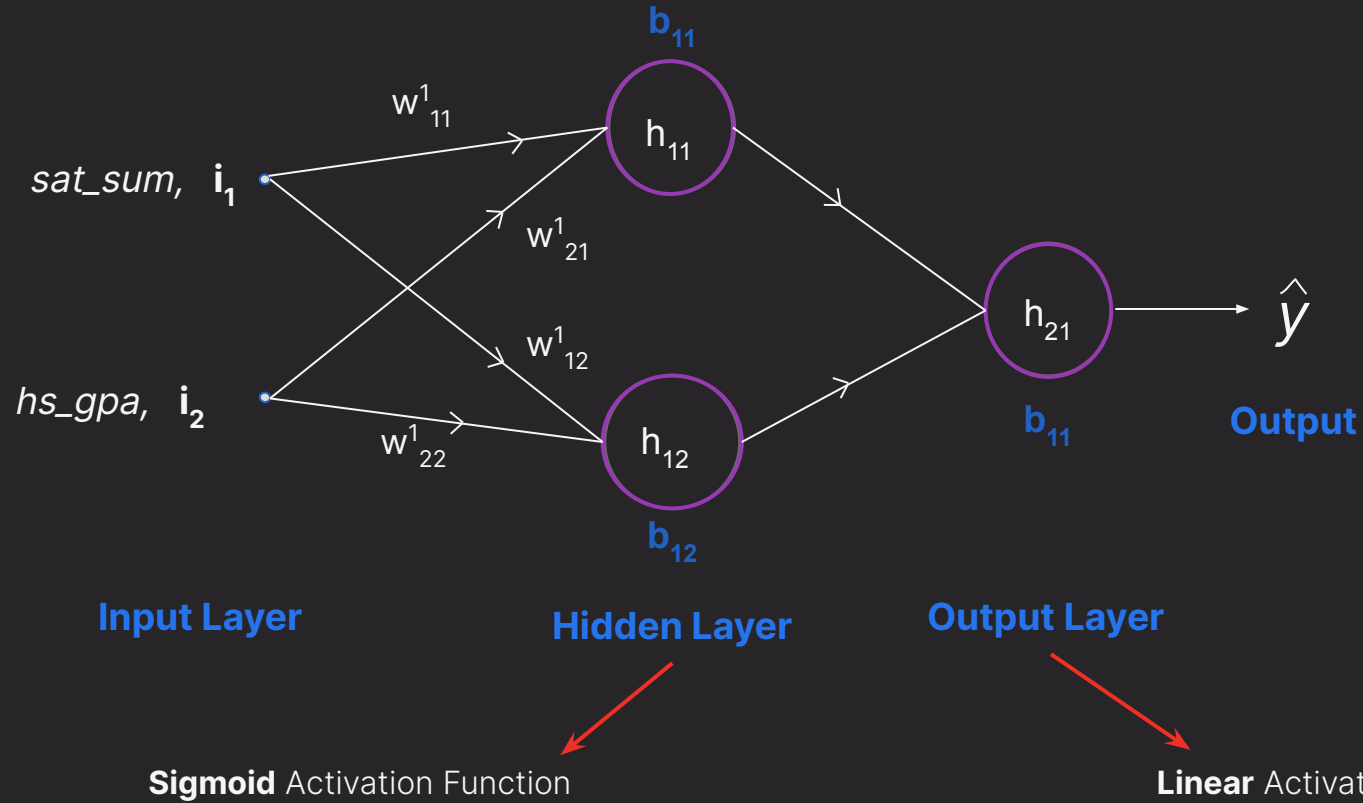
In air

Backpropagation

Computes gradient of loss function at output & distributes gradient backward through all layers.



- Results in updated neuron weights & biases
- Leads to improved neural outputs



$$L = (1/n) \sum |\hat{y} - y|^2$$

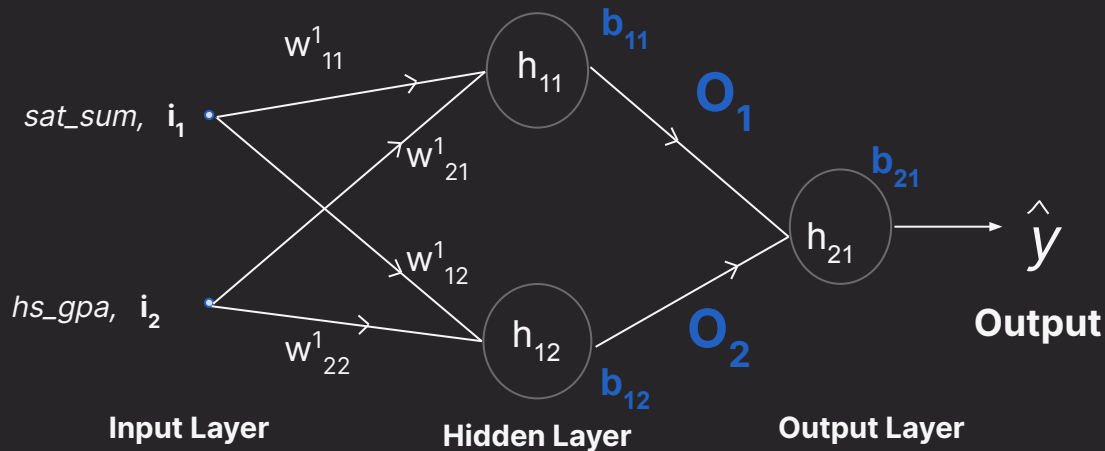


Predicted values
can be changed

Labels
cannot be changed

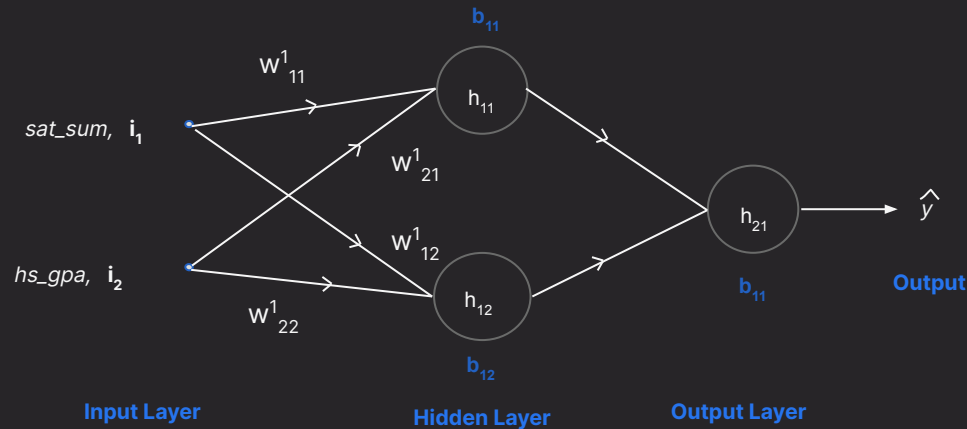
Optimizing Neural Networks

$$\hat{y} = W_{21}^2 * O_1 + W_{22}^2 * O_2 + b_{21}$$



Optimizing Neural Networks

Which parameters does \hat{y} depend on ?



- w_{21}^2
- O_1
- w_{22}^2
- O_2
- b_{21}

Optimizing Neural Networks

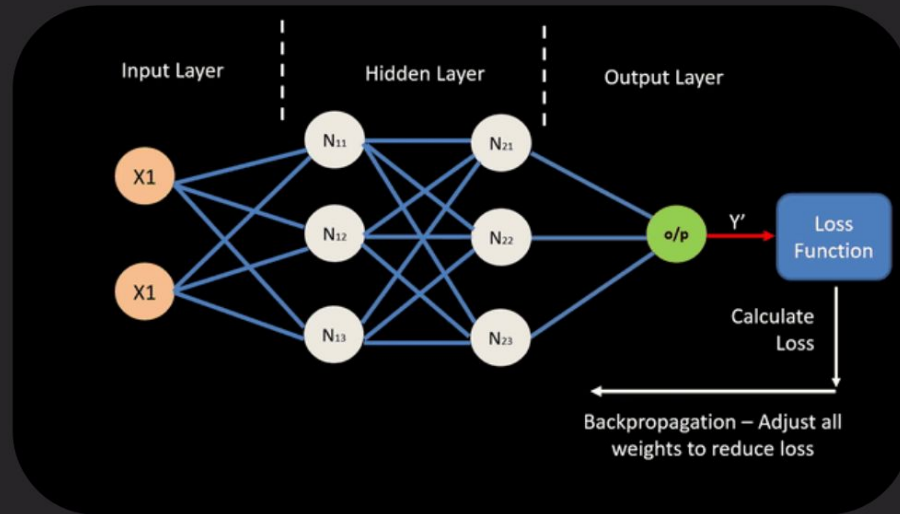
$$O_1 = \text{sigmoid} (w_{11}^1 * i_1 + w_{21}^1 * i_2 + b_{11})$$

$$O_2 = \text{sigmoid} (w_{12}^1 * i_1 + w_{22}^1 * i_2 + b_{12})$$

Can not update w_{11}^1, w_{12}^1 etc, directly due to their dependencies

Backpropagation

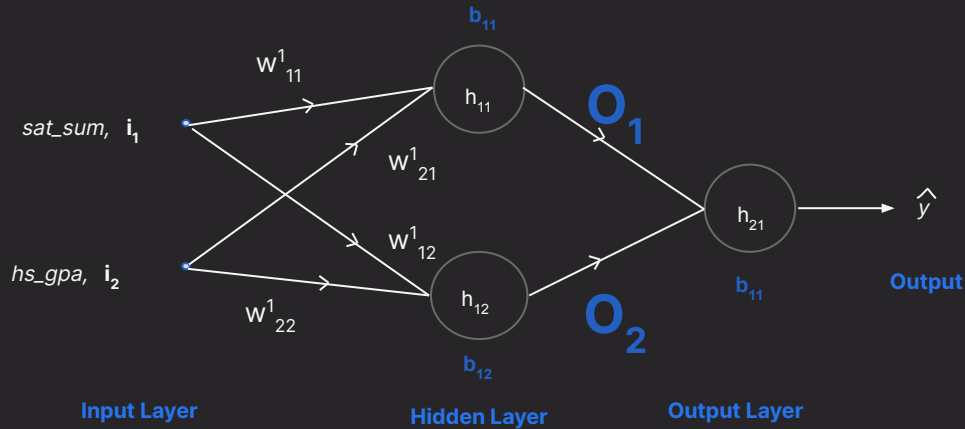
Back Propagation: A systematic way of updating weights and biases across all layers to minimize the loss function.



Backpropagation

$$y = W_{21}^2 * O_1 + W_{22}^2 * O_2 + b_{21}$$

$$\frac{dL}{d\hat{y}} = 2(\hat{y} - y)$$



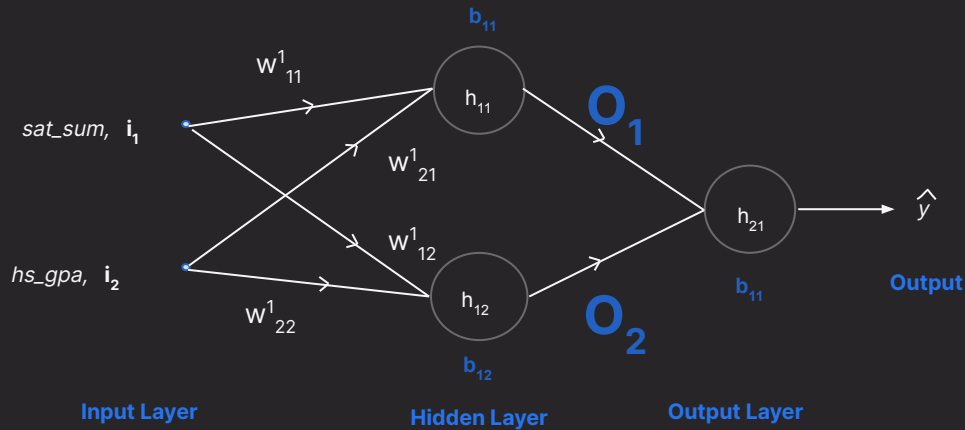
$$\frac{dL}{dw_{21}^2} = \frac{dL}{d\hat{y}} * \frac{d\hat{y}}{dw_{21}^2} = 2(\hat{y} - y) * O_1$$

$$\frac{dL}{dw_{22}^2} = \frac{dL}{d\hat{y}} * \frac{d\hat{y}}{dw_{22}^2} = 2(\hat{y} - y) * O_2$$

$$\frac{dL}{db_{21}} = \frac{dL}{d\hat{y}} * \frac{d\hat{y}}{db_{21}} = 2(\hat{y} - y)$$

Backpropagation

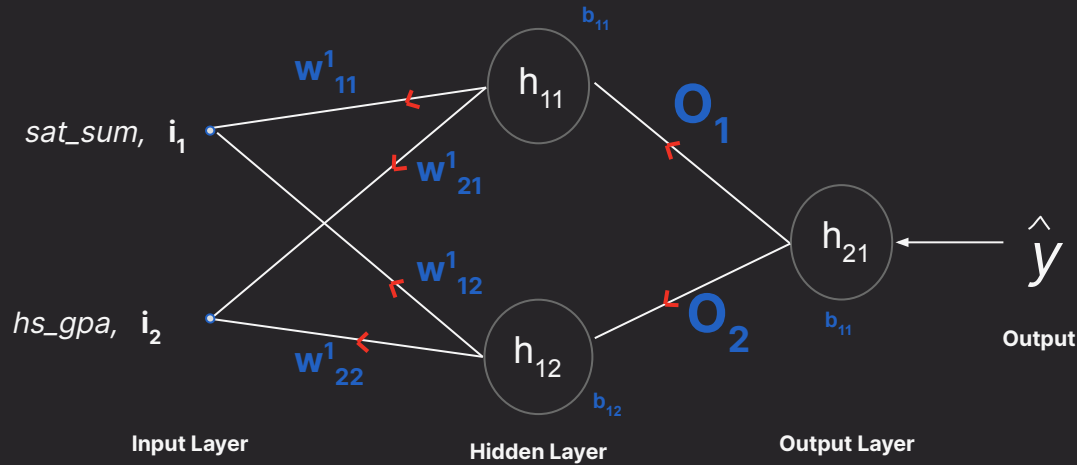
Calculate how Loss get affected by O_1 & O_2



$$\frac{dL}{dO_1} = \frac{dL}{d\hat{y}} * \frac{d\hat{y}}{dO_1}$$

$$\frac{dL}{dO_2} = \frac{dL}{d\hat{y}} * \frac{d\hat{y}}{dO_2}$$

Backpropagation



$$O_1 = \text{sigmoid} (w^1_{11} * i_1 + w^1_{21} * i_2 + b_{11})$$

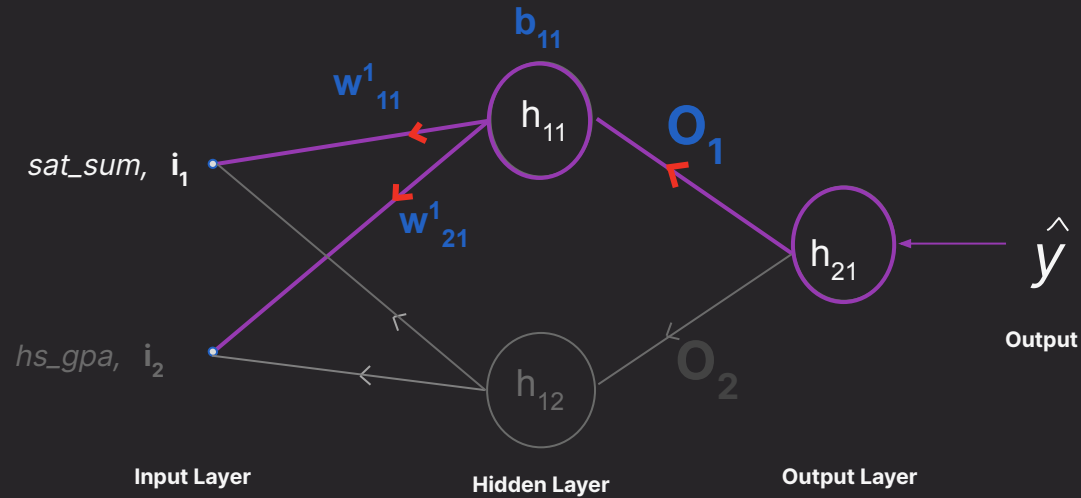
$$O_2 = \text{sigmoid}(w^1_{12} * i_1 + w^1_{22} * i_2 + b_{12})$$

Optimizing parameters with Chain Rule

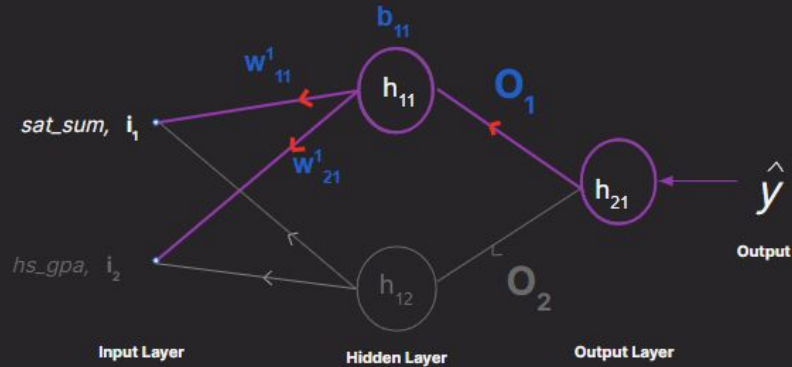
Determine how \mathbf{w}_{11}^1 , \mathbf{w}_{21}^1 , \mathbf{b}_{11} , \mathbf{w}_{12}^1 , \mathbf{w}_{22}^1 and \mathbf{b}_{12} are affecting loss function



Focus on parameters associated with O_1 : $w_{11}^1, w_{21}^1, b_{11}$



Optimizing parameters with Chain Rule

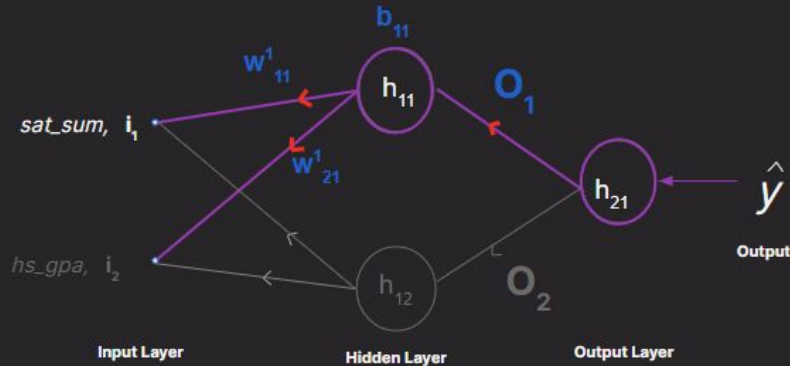


$$\frac{dL}{dw^1_{11}} = \frac{dL}{dO_1} * \frac{dO_1}{dw^1_{11}}$$

$$\frac{dL}{dw^1_{21}} = \frac{dL}{dO_1} * \frac{dO_1}{dw^1_{21}}$$

$$\frac{dL}{db_{11}} = \frac{dL}{dO_1} * \frac{dO_1}{db_{11}}$$

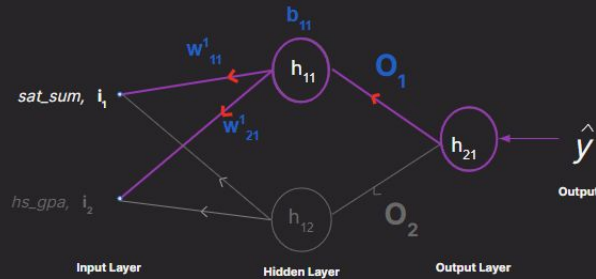
Optimizing parameters with Chain Rule



$$O_1 = \text{sigmoid} (w_{111}^1 * i_1 + w_{121}^1 * i_2 + b_{11})$$

$$\frac{d}{dx} \sigma(x) = \sigma(x) \cdot (1 - \sigma(x))$$

Optimizing parameters with Chain Rule



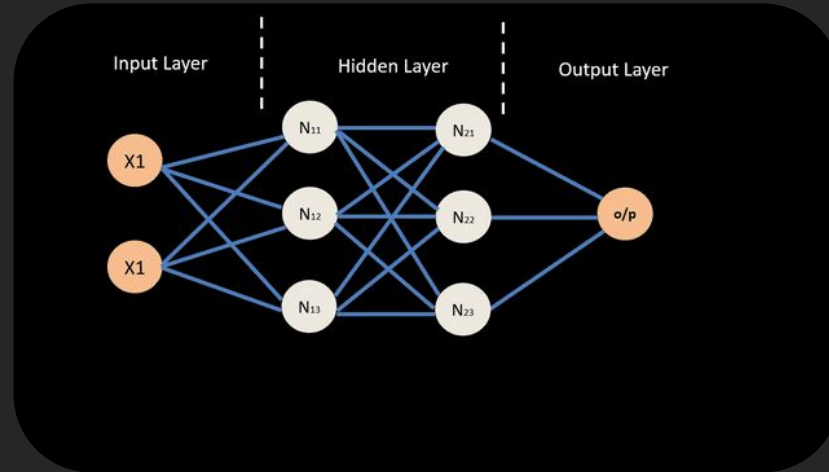
$$\sigma'' = \frac{d}{dx} \sigma(x) = \sigma(x) \cdot (1 - \sigma(x))$$

$$\frac{dl}{dw_{11}^1} = 2(\hat{y} - y) \cdot w_{21}^2 \cdot \sigma''(w_{11}^1 i_1 + w_{21}^1 i_2 + b_{11}) \cdot i_1$$

$$\frac{dl}{dw_{21}^1} = 2(\hat{y} - y) \cdot w_{21}^2 \cdot \sigma''(w_{11}^1 i_1 + w_{21}^1 i_2 + b_{11}) \cdot i_2$$

$$\frac{dl}{db_{11}} = 2(\hat{y} - y) \cdot w_{21}^2 \cdot \sigma''(w_{11}^1 i_1 + w_{21}^1 i_2 + b_{11})$$

Finalizing Neural Network Training



Use the **gradients** to **update weights & biases** across the network.

Updation Formulas

Weights: Input Layer

$$w_{11}^1(\text{new}) = w_{11}^1 - \eta \frac{\partial L}{\partial w_{11}^1}$$

$$w_{12}^1(\text{new}) = w_{12}^1 - \eta \frac{\partial L}{\partial w_{12}^1}$$

$$w_{21}^1(\text{new}) = w_{21}^1 - \eta \frac{\partial L}{\partial w_{21}^1}$$

$$w_{22}^1(\text{new}) = w_{22}^1 - \eta \frac{\partial L}{\partial w_{22}^1}$$

Weights: From Hidden Layer

$$w_{11}^2(\text{new}) = w_{11}^2 - \eta \frac{\partial L}{\partial w_{11}^2}$$

$$w_{21}^2(\text{new}) = w_{21}^2 - \eta \frac{\partial L}{\partial w_{21}^2}$$

Bias Terms

$$b_{11}(\text{new}) = b_{11} - \eta \frac{\partial L}{\partial b_{11}}$$

$$b_{12}(\text{new}) = b_{12} - \eta \frac{\partial L}{\partial b_{12}}$$

$$b_{21}(\text{new}) = b_{21} - \eta \frac{\partial L}{\partial b_{21}}$$

In air