Understanding workings of Neural Networks

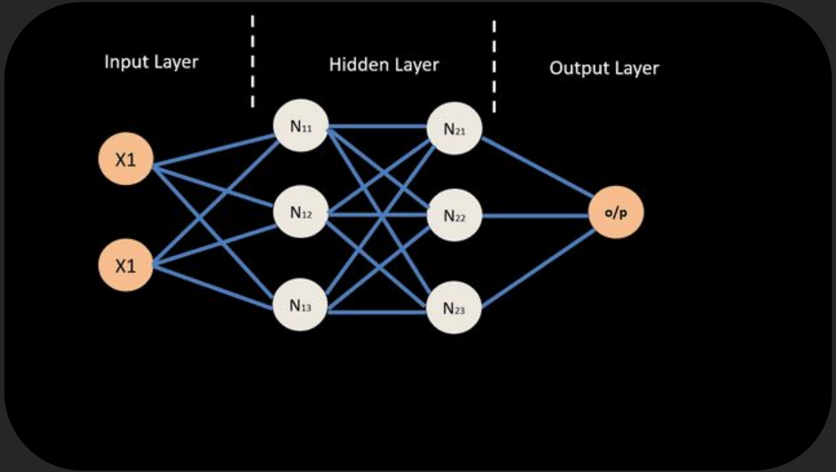
Video 4: What is Backpropagation?

Analytics Vidhya

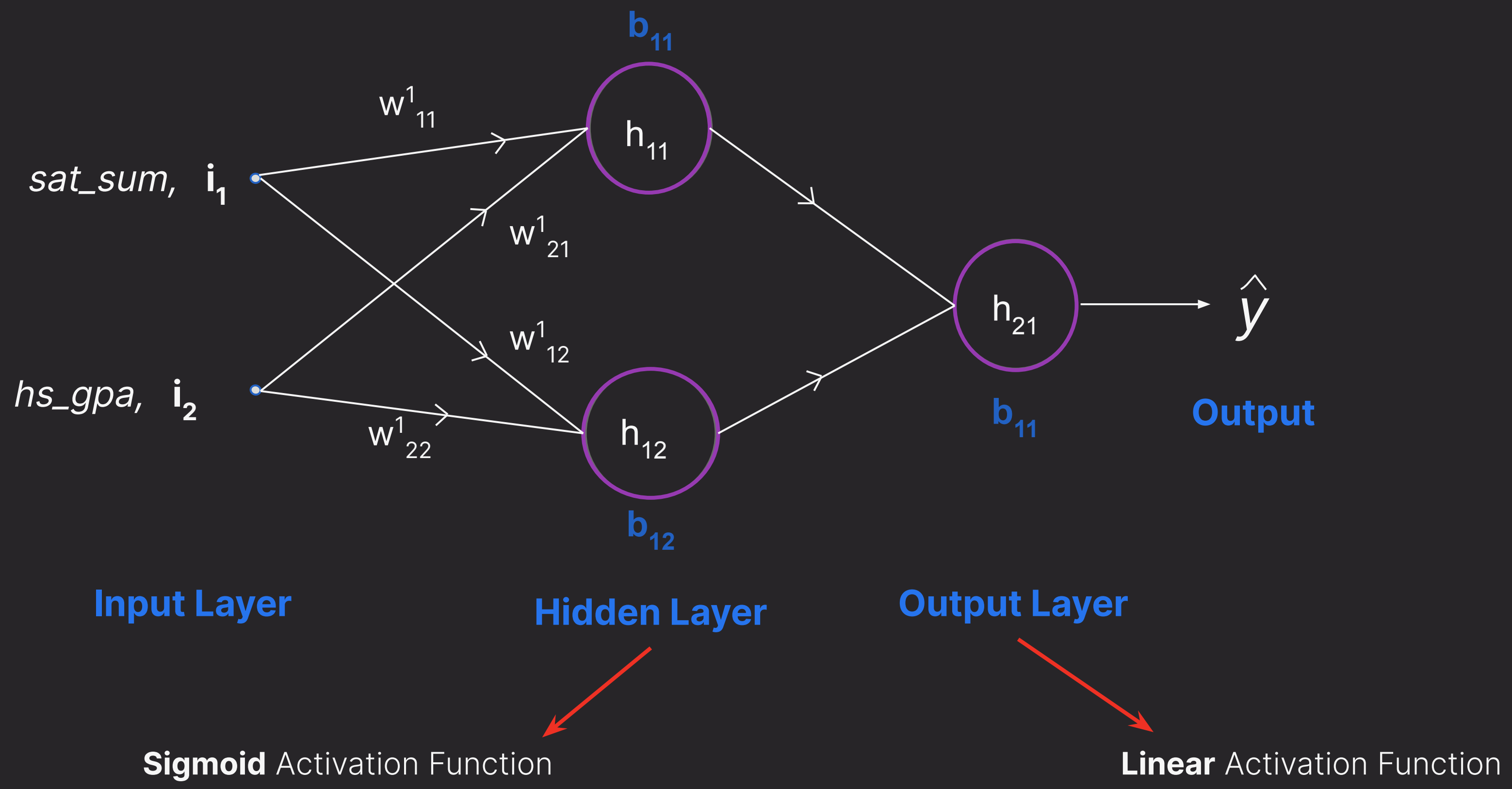
In air

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



Analytics Vidhya

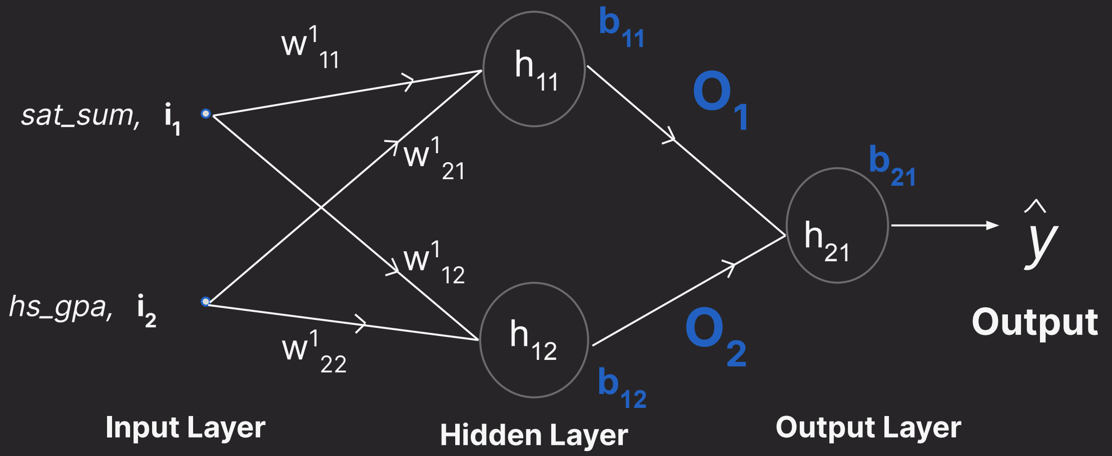
$$L = (1/n)\,\Sigma\left|\hat{y} - y\right|^2$$

Predicted values
can be changed

Labels
cannot be changed

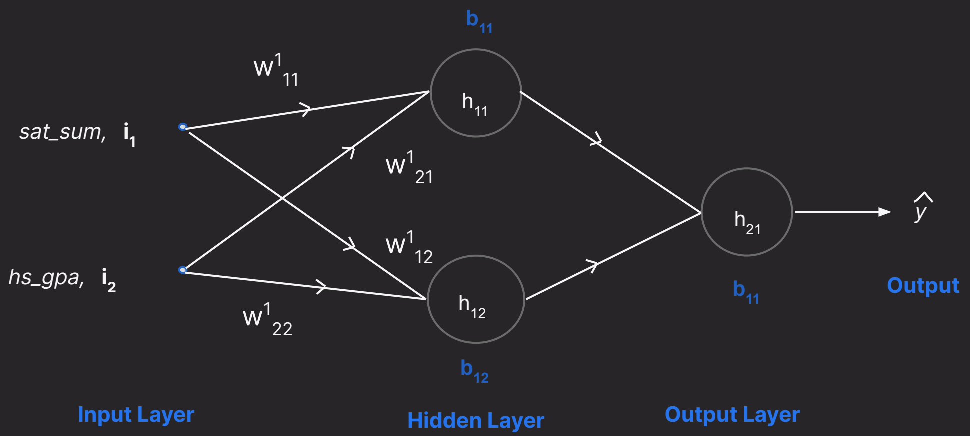
Optimizing Neural Networks

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



Optimizing Neural Networks

Which parameters does ŷ depend on ?



- $W^2{}_{21}$
- O_1
- $W^2{}_{22}$
- O_2
- b_{21}

Optimizing Neural Networks

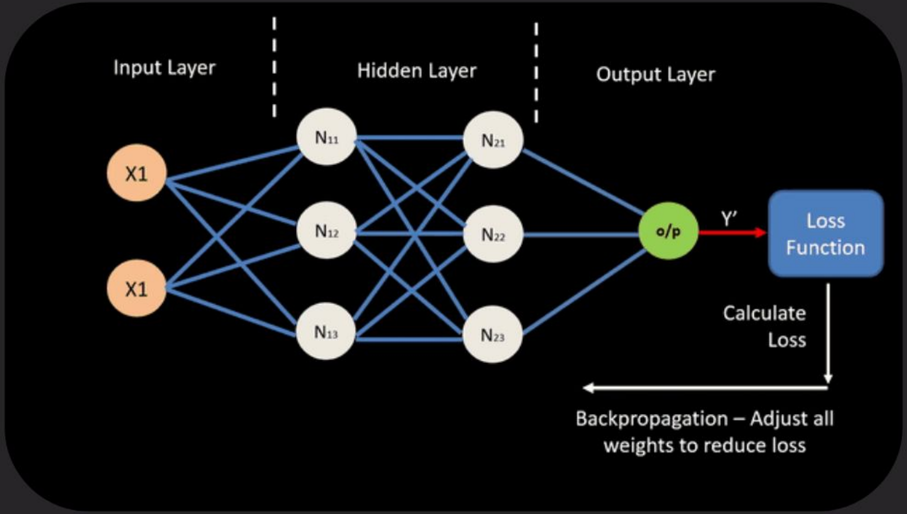
$O_1 = \text{sigmoid}\ (w^1_{11} * i_1 + w^1_{21} * i_2 + b_{11})$

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

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

Backpropagation

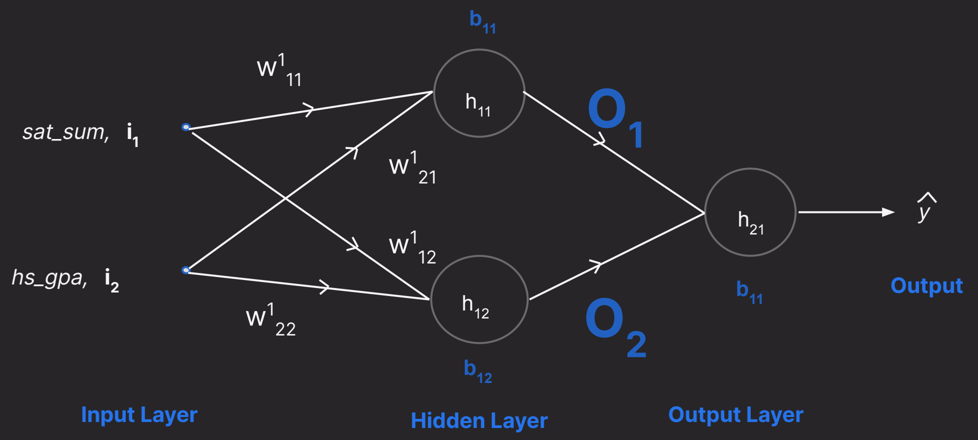
Back Propagation: A systematic way of updating weights and biases across all layers to minimize the loss function.



Backpropagation

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

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



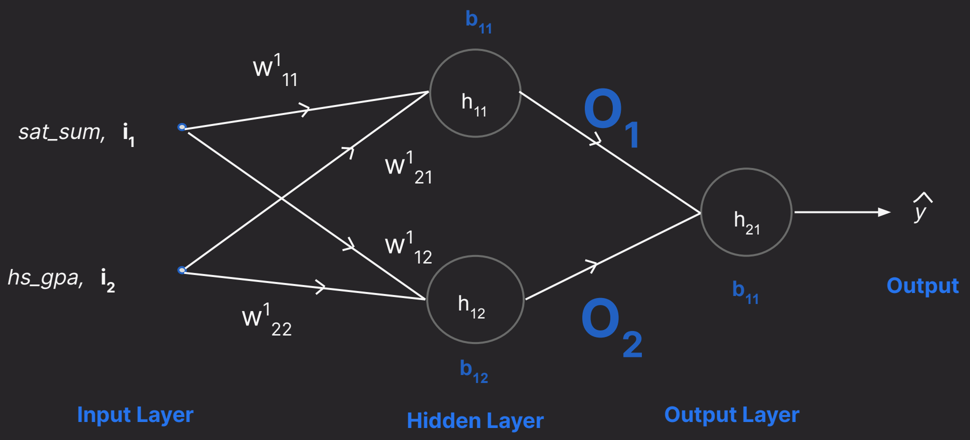
$$\frac{dL}{dw^2_{21}} = \frac{dL}{d\hat{y}} * \frac{d\hat{y}}{dw^2_{21}} = 2(\hat{y} - y) * O_1$$

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

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

Backpropagation

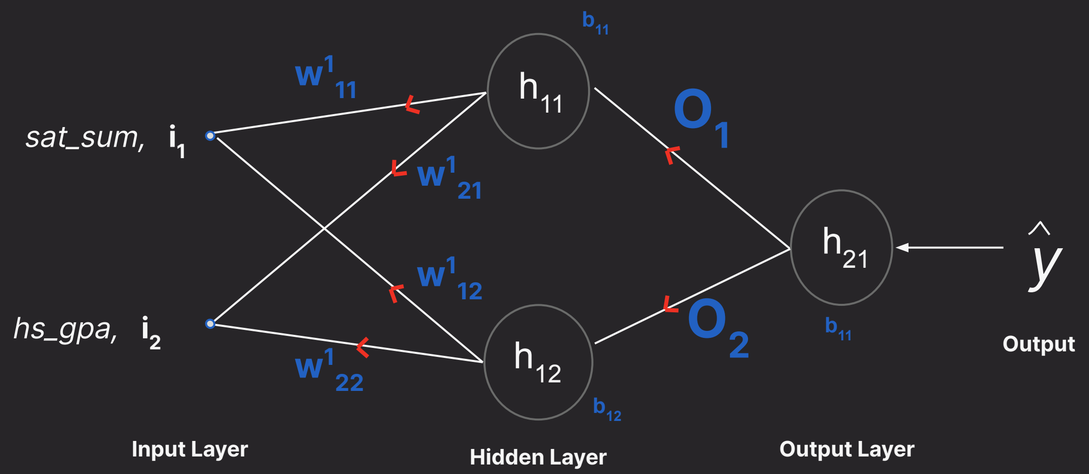
Calculate how Loss get affected by O_1 & O_2



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

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

Backpropagation



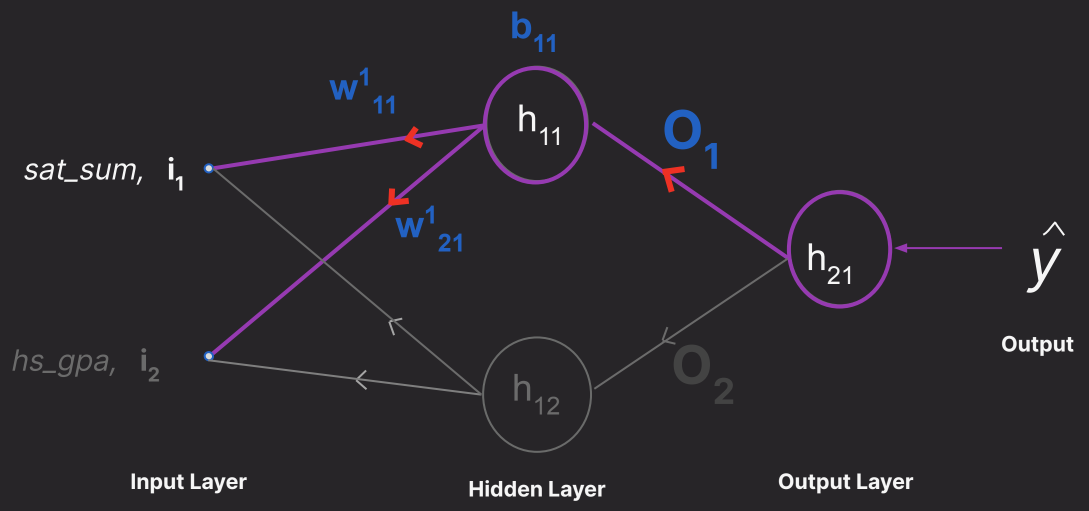
O_1 = sigmoid ($w^1_{11} * i_1 + w^1_{21} * i_2 + b_{11}$)

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

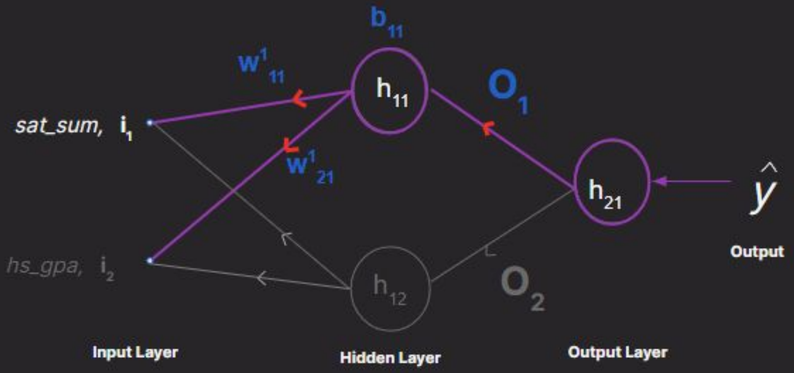
Optimizing parameters with Chain Rule

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

Focus on parameters associated with $\mathbf{O_1}$: $\mathbf{w^1_{11}}$, $\mathbf{w^1_{21}}$, $\mathbf{b_{11}}$



Optimizing parameters with Chain Rule

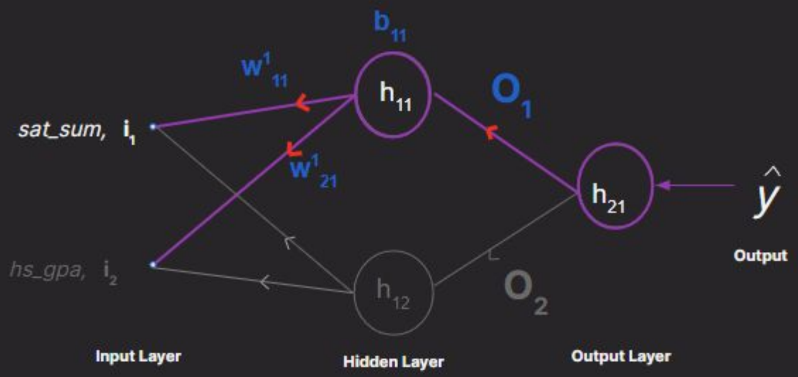


$$\frac{dL}{dw_{11}^{1}} = \frac{dL}{dO_1} * \frac{dO_1}{dw_{11}^{1}}$$

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

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

Optimizing parameters with Chain Rule

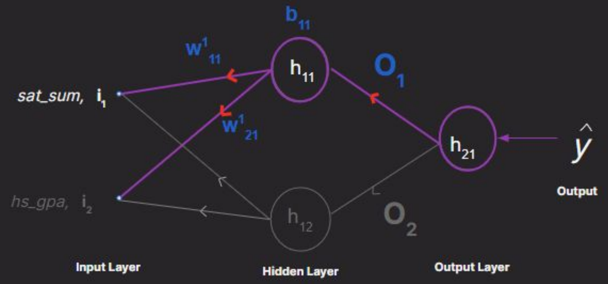


O_1 = sigmoid ($w^1_{11} * i_1 + w^1_{21} * i_2 + b_{11}$)

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

Analytics Vidhya

Optimizing parameters with Chain Rule



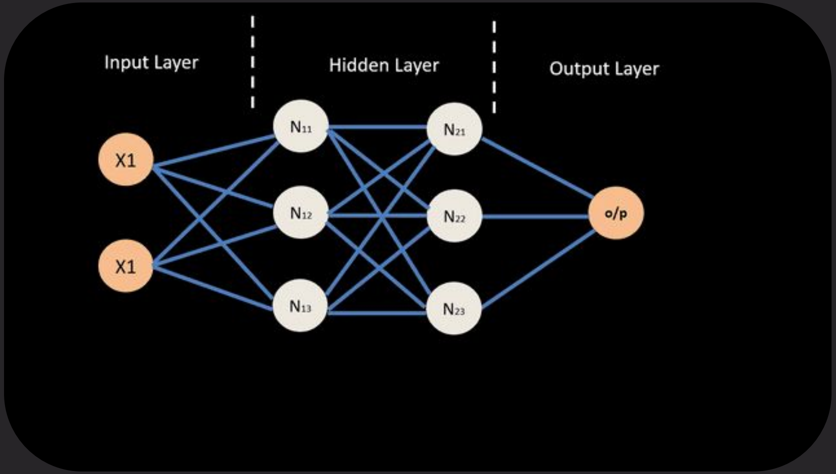
$$\sigma'' = \frac{\mathrm{d}}{\mathrm{d}x}\sigma(x) = \sigma(x)\cdot\left(1-\sigma(x)\right)$$

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

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

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

Finalizing Neural Network Training



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

Analytics Vidhya

Updation Formulas

Weights: Input Layer

$$\mathrm{w}_{11}^{1}(\mathrm{new}) = \mathrm{w}_{11}^{1} - \eta \frac{\partial L}{\partial w_{11}^{I}}$$

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

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

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

Weights: From Hidden Layer

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

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

Bias Terms

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

$$b_{12}(\mathrm{new}) = \mathrm{b}_{12} - \eta \frac{\partial L}{\partial b_{12}}$$

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

In air