Neural Network

1. Component of a neural network and discuss the similarities of neural network.

A neural network is a computation model inspired by human brain. And its component are (i) Input layer (ii) Hidden layer (iii) Output layer (iv) Neuron Node (v) weights (vi) Bias (vii) Activation function (viii) Loss function (ix) optimizer

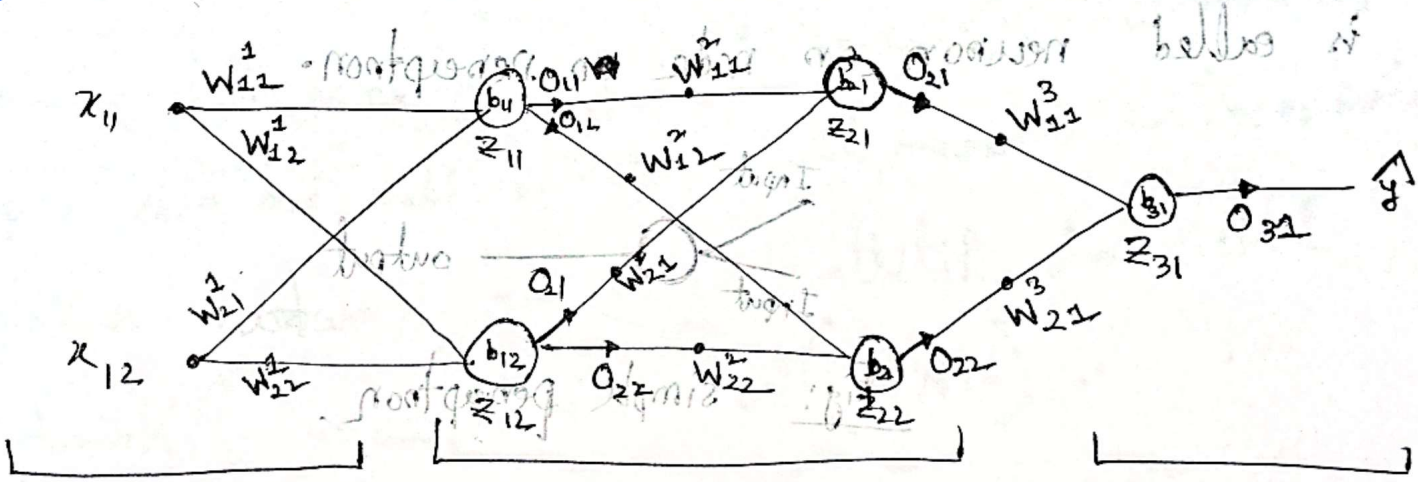


Fig: simple artificial neural network

Backpropagation:

(Loss $\rightarrow \hat{y} \rightarrow Z_{31} \rightarrow O_{21} \rightarrow Z_{21} \rightarrow O_{11} \rightarrow Z_{11} \rightarrow W_{11}$)

(computational graph for updating W_{11})

$$\frac{\partial L}{\partial W_{11}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial Z_{31}} \times \frac{\partial Z_{31}}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial W_{11}}$$

$$z_{11} = x_{11} w_{11}^1 + x_{21} w_{21}^1 + b_{11}$$

$$O_{11} = \text{Activation function } (z_{11})$$

$$z_{21} = O_{11} w_{11}^2 + O_{12} w_{21}^2 + b_{21}$$

$$O_{21} = \text{Activation function } (z_{21})$$

~~Here, the ANN made up multiple neuron or node of its~~ ~~call per~~. The smallest unit of a neural network is called neuron or node or perceptron.

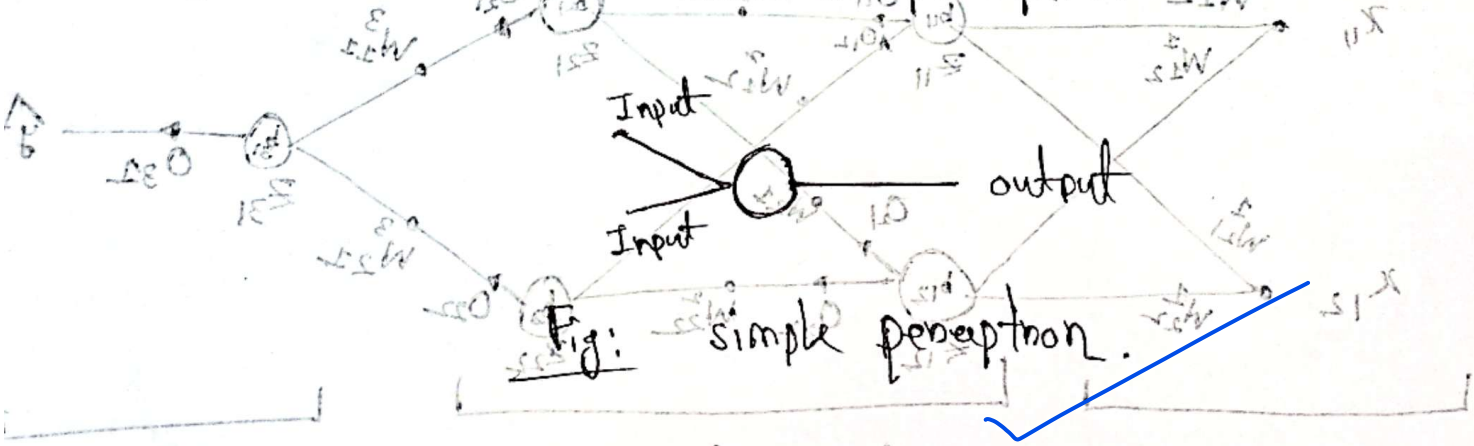
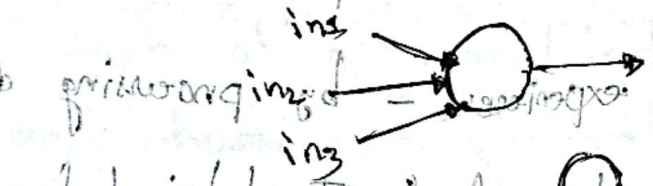


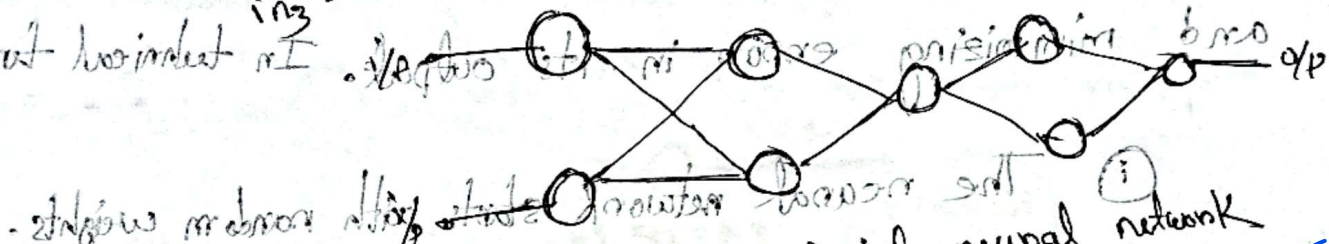
Fig: simple perceptron.

Neuron1 Synapse Neuron2

A human brain has billon of neurons. All the neur

interconnected with each other making a big network called neural network. "Similarly, a perceptron on it's own is called artificial neuron



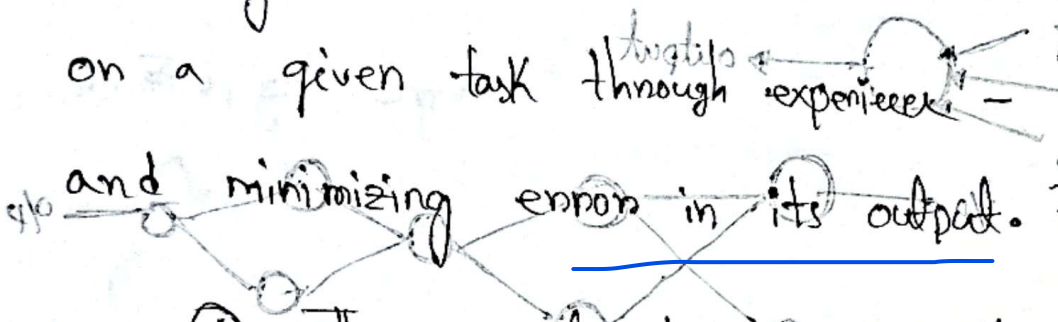


simple artificial neural network

Similarly, thousand of artificial neuron (numbers) based on input data and model complexity) interconnected with eachother mark a network and we call it Artificial Neural Network (ANN).

2. What does "learning" mean in relation to neural nets.

Learning means that neural network improves its performance on a given task through experience by processing data and minimizing error in its output. In technical terms,



(i) The neural network starts with random weights.

(ii) It proces the data through it's layer (called forward propagation)

(iii) The output is compared with the expected or actual output

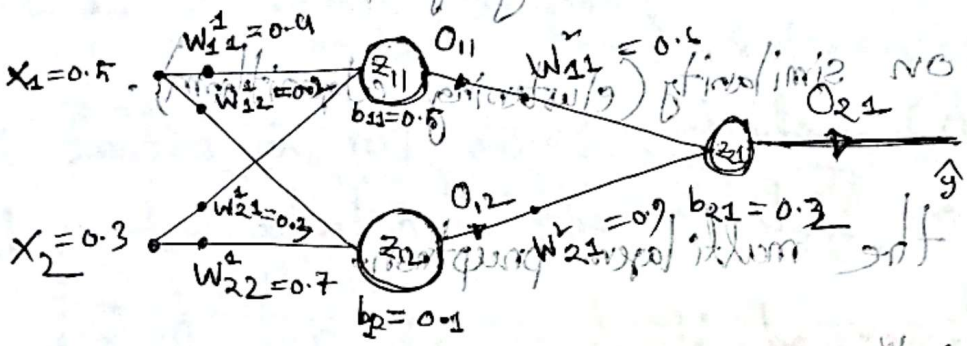
(iv) The difference between them is error or loss.

(v) Using the backpropagation and gradient descent the neural network adjusts its weights.

(vi) Repeating this process many times called (training).

3. Explain forward propagation and self-organization.

Forward propagation is the process of moving inputs through the layers of a neural network to get an output. Let's take an example below



Dataset

x_1	x_2	Tanget(x)
0.5	0.3	1

and, Lets take activation function as sigmoid.

$$O_{11} = \frac{1}{1+e^{-z_{11}}} = \frac{1}{1+e^{-(0.5\times0.4 + 0.3\times0.2 + 0.5)}} = 0.68$$

$$O_{12} = \frac{1}{1+e^{-z_{12}}} = \frac{1}{1+e^{-(0.5\times0.2 + 0.3\times0.7 + 0.1)}} = 0.601$$

$$O_{21} = \frac{1}{1+e^{-z_{21}}} = \frac{1}{1+e^{-(0.68\times0.6 + 0.601\times0.9 + 0.2)}} = 0.76$$

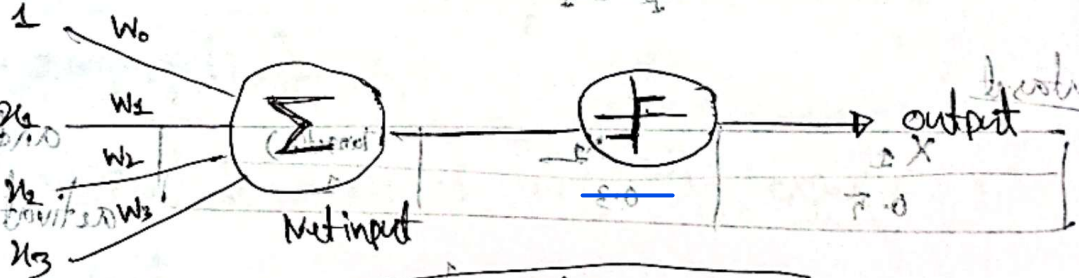
$$\hat{y} = O_{21} = 0.76$$

The process of getting output from input is called forward propagation.

Self-Organization

Self organization is a learning process without supervision (unsupervised learning). The network organizes data itself and finds a pattern of data. The system automatically groups similar inputs together based on similarity (clustering algorithm).

4. Describe the multilayer perceptron.



a single neuron

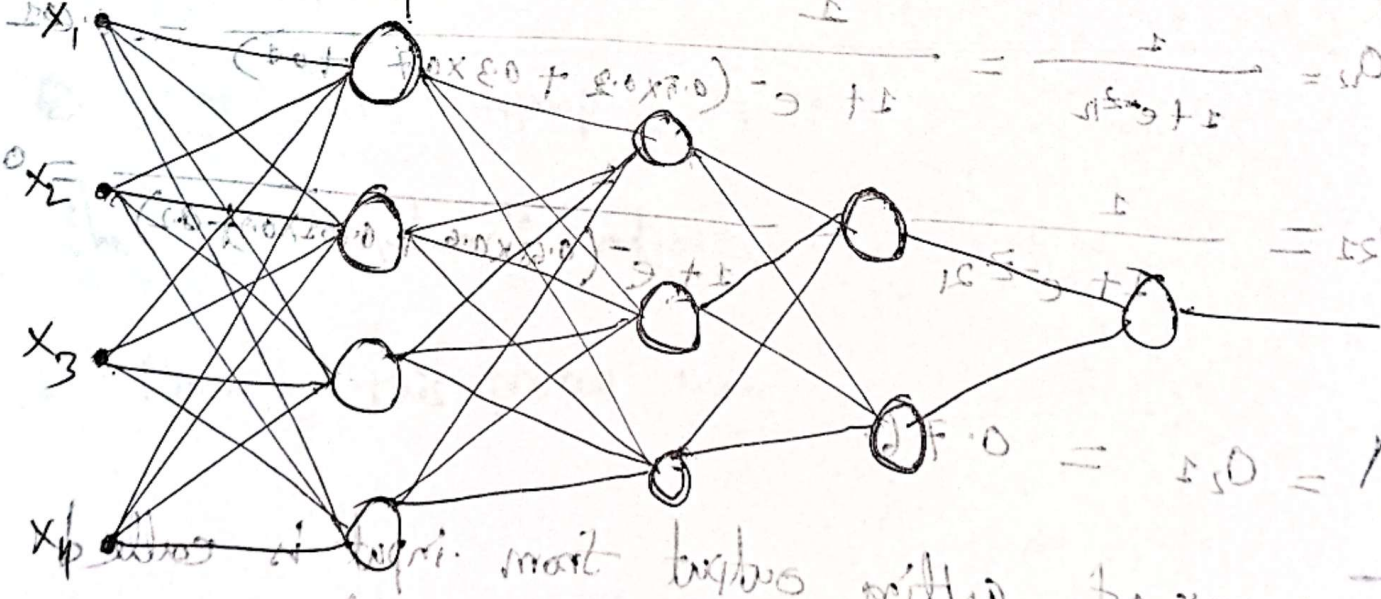


Fig: Multi-layer Perceptron (MLP)

5. Illustrate and describe the standard activation function.

In artificial neural networks each weighted sum of input passes through a function called activation function.

All function is not activation function. Criterion for a function belong to activation or step function

(i) Activation function should be non-linear.

(ii) Activation function should be differentiable.

(iii) numerically inexpensive.

(iv) zero centred

(v) non-saturating

means bound within range. Like, saturating function sigmoid output range (0~1) and tanh output rang (-1, 1). For very very higher input sigmoid give always 1 same for tanh that why it's faces vanishing gradient problem.

activation function with formula

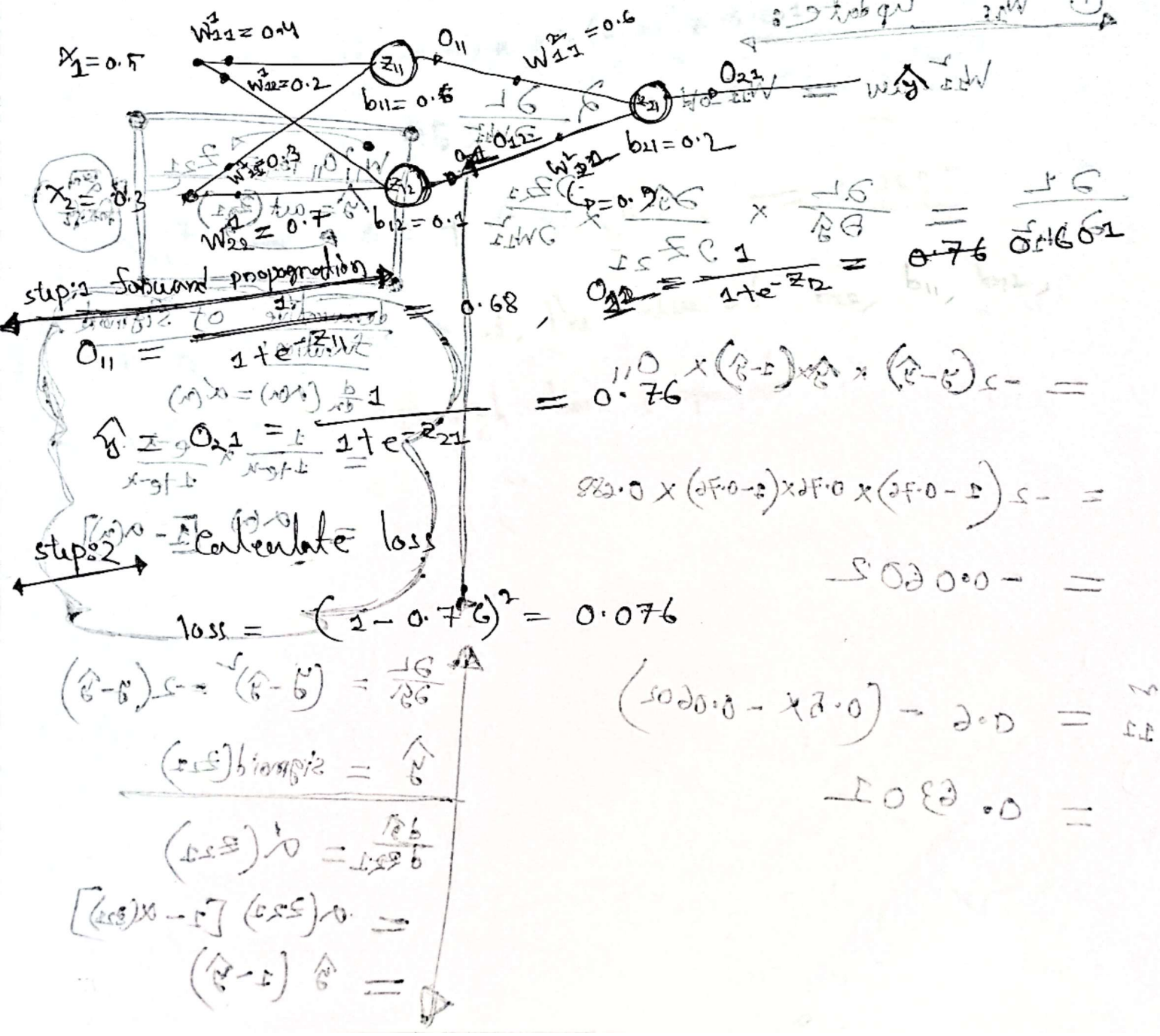
Function	Formula	Range	Remarks
sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	(0, 1)	prone to vanishing gradient
ReLu	$f(x) = \max(0, x)$	$(0, \infty)$	may cause dead neuron when input $x \leq 0$
Leaky ReLu	$f(x) = \begin{cases} x, & x > 0 \\ 0.01x, & x \leq 0 \end{cases}$	$(-\infty, \infty)$	solve the dead neuron problem
tanh	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1, 1)	zero-centered (easier to normalize) fast convergence than sigmoid
softmax	$f(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	(0, 1)	Converts output into probability

6. How to adjust the weights of perceptron in nn? or state the back-propagation algorithm for learning in a multilayer network.

Lets take the below mlp

Dataset

x_1	x_2	output
0.5	0.3	1



Given:

$x_{11} = 0.5 \quad x_{12} = 0.3, \quad y = 1$ (target)

$O_{11} = 0.688 \quad O_{12} = 0.601, \quad \hat{y} = O_{21} = 0.76$

$W_{11}^2 = 0.6, \quad W_{21}^2 = 0.9$

$b_{21} = 0.2$, Let's $\lambda = 0.5$

① W_{11}^2 update:

$$W_{11}^{2\,new} = W_{11}^{2\,old} - \lambda \frac{\partial L}{\partial W_{11}^2}$$

$$\frac{\partial L}{\partial W_{11}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_{21}} \times \frac{\partial z_{21}}{\partial W_{11}^2}$$

$$= -2(y-\hat{y}) \times \hat{y}(1-\hat{y}) \times O_{11}$$

$$= -2(1-0.76) \times 0.76 \times (1-0.76) \times 0.688$$

$$= -0.0602$$

$$W_{11}^2 = 0.6 - (0.5 \times -0.0602)$$

$$= 0.6301$$

$W_{11}^2 O_{11} + b_{21} = z_{21}$

$\hat{y} = \text{out}(z_{21})$

derivative of Sigmoid function:

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

$= \frac{1}{1+e^{-x}} \times \frac{e^{-x}}{1+e^{-x}}$

$= \sigma(x)[1-\sigma(x)]$

$\frac{\partial L}{\partial \hat{y}} = (y-\hat{y})^2 = -2(y-\hat{y})$

$\hat{y} = \text{sigmoid}(z_{21})$

$\frac{d\hat{y}}{dz_{21}} = \sigma'(z_{21})$

$= \sigma(z_{21})[1-\sigma(z_{21})]$

$= \hat{y}(1-\hat{y})$

(ii) For, W_{21}^2,

$$z_{21} = o_{11} W_{11}^2 + o_{12} W_{21}^2 + b_{21} \quad \text{—(i)}$$

$$\hat{y} = \text{sigmoid}(z_{21}) \quad \text{—(ii)}$$

$$\frac{\partial L}{\partial W_{21}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_{21}} \times \frac{\partial z_{21}}{\partial W_{21}}$$

$$= -2(y - \hat{y}) \times \hat{y} \times (1 - \hat{y}) \times o_{12}$$

$$= -2(1 - 0.76) \times 0.76 (1 - 0.76) * 0.601$$

$$= -0.053$$

$$W_{12}^2 = 0.0 + (0.5)(0.53) = 1.165$$

Similarly, we well adjust the value of b_{21}, b_{11}, b_{12}, W_{21}^1, W_{12}^1 through back propagation.