

# Multilayer NN Working

1. Multilayer NN working.
  - a. Forward propagation
  - b. Loss calculation
  - c. Backward Propagation

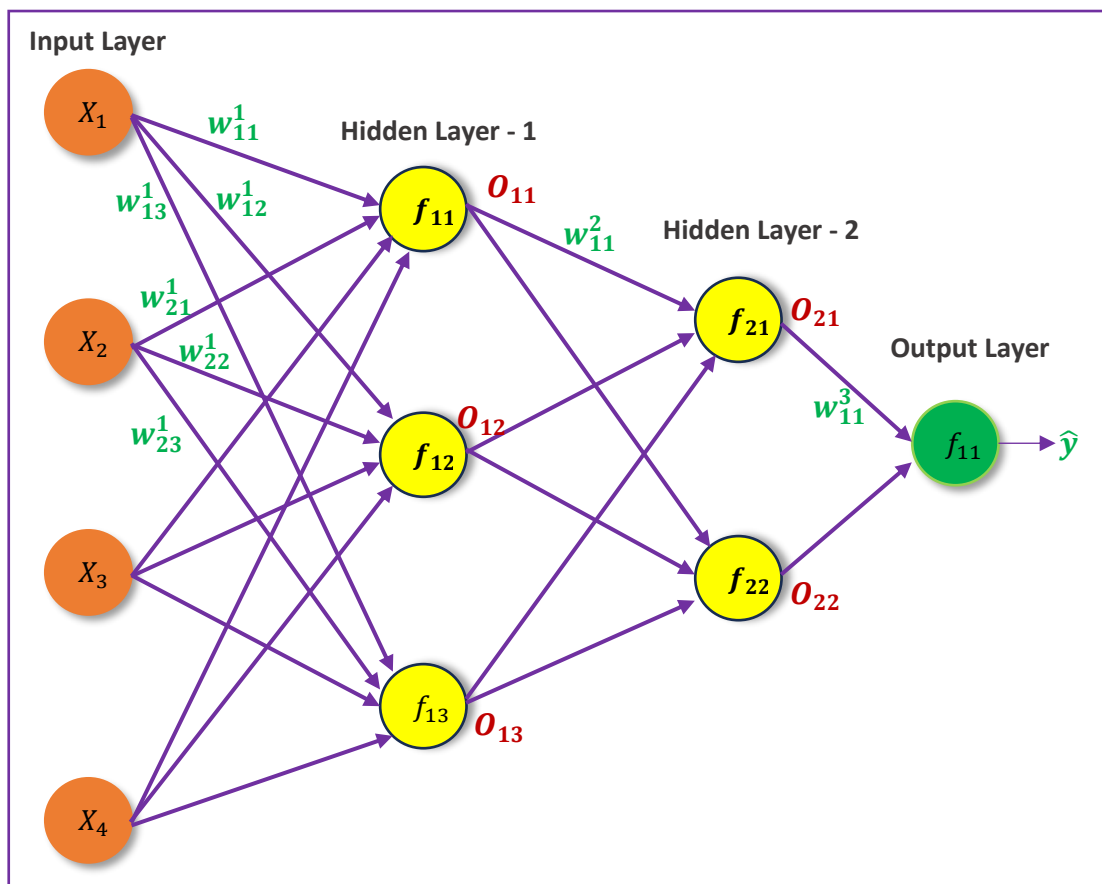


Figure 1: Multilayer neural network

The denotation of weight:  $W_{(input\ node\ number)(output\ node\ number)}^{layer\ number}$

The weights matrix at first layer connections:

$$\begin{bmatrix} w_{11}^1 & w_{12}^1 & w_{13}^1 \\ w_{21}^1 & w_{22}^1 & w_{23}^1 \\ w_{31}^1 & w_{32}^1 & w_{33}^1 \\ w_{41}^1 & w_{42}^1 & w_{43}^1 \end{bmatrix}$$

The weights matrix at second layer connections:

$$\begin{bmatrix} w_{11}^2 & w_{12}^2 \\ w_{21}^2 & w_{22}^2 \\ w_{31}^2 & w_{32}^2 \end{bmatrix}$$

The weights matrix at third layer connections:

$$\begin{bmatrix} w_{11}^3 \\ w_{21}^3 \end{bmatrix}$$

Let's understand this multilayer neural network working in three stages show in below. Observe the above multilayer neural network diagram and link with below explanation for better understanding:

1. Forward propagation
2. Loss calculation
3. Backward propagation

### 1. Forward Propagation:

- First it will take the input features  $(x_1, x_2, x_3, x_4)$  and pass them to each neuron  $(f_{11}, f_{12}, f_{13})$  in the first hidden layer and weights also passed through input connection.
- Once the neuron receives the weights and inputs then it will start the *Hypothesis function*, and *Activation function* calculation internally. To understand it better let's consider the first neuron  $(f_{11})$  working in the first hidden layer work.

Step-1: Calculating the hypothesis value:

$$f_{11} = w_{11}^1 x_1 + w_{21}^1 x_2 + w_{31}^1 x_3 + w_{41}^1 x_4 + b$$

Step-2: Calculating the output value by applying the activation function:

$$O_{11} = \text{act}(f_{11})$$

- Next the generated output  $(O_{11})$  Passing as an input of next layer neurons.
- Again, the same neuron operations will happen inside the second layer neurons, as explained in the 3<sup>rd</sup> point above.
- Finally, the predicted y value will be generated at output layer.

### 2. Loss Calculation:

- The cost/ loss function value will calculate after complete the one round of forward training.
- The cost function is:  $L = (y_i - \hat{y}_i)^2$

### 3. Backward Propagation:

- Updating the weights is the main work of backward propagation.

- We need to select a suitable optimizer (like Gradient descent, stochastic gradient descent etc..) for update the weights. Let's consider *Gradient Descent* for simple understand and consider  $w_{11}^3$  weight updating process. The resultant equation is:

$$w_{11}^3_{new} = w_{11}^3 - \alpha \frac{\partial L}{\partial w_{11}^3}$$

- Once all the weights got updated then it will again start the forward propagation. This cycle repeats until to get the global minimum point of the loss function.

There are four steps involved for update weights:

Step-1: Find out the predicted value  $\hat{y}$ .

Step-2: Pass the predicted  $\hat{y}$  value to find the loss function L.

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)$$

Step-3: Calculate the loss function.

Step-4: Substitute the loss function in the weight formula and update the weight by back propagation process.

- These four steps are continued to update all the weights and this back propagation process continues till old  $\hat{y}$  predicted is similar to new  $\hat{y}$  predicted.