Chain Rule in Back Propagation

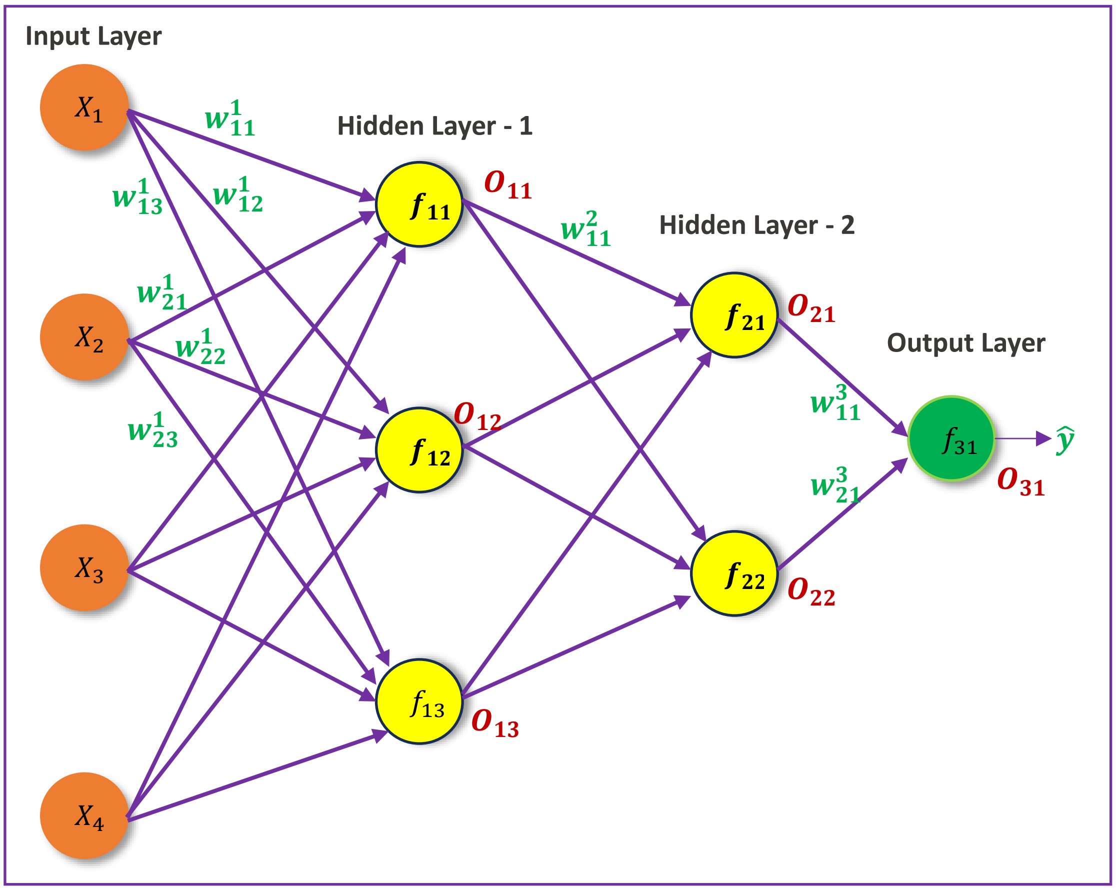

Figure 1: Multilayer neural network

Let's take gradient descent wide updating formula for reference:

$$\omega_{new} = \omega_{old} - \alpha \frac{\partial L}{\partial \omega_{old}} \dots\dots\dots\dots \quad (1)$$

The chain rule formula for update w^3_{11} at first step from output side is:

$$w^3_{11new} = w^3_{11} - \alpha \frac{\partial L}{\partial w^3_{11}} \dots\dots\dots\dots \quad (2)$$

$$\frac{\partial L}{\partial w_{11}^3} = \frac{\partial L}{\partial O_{31}} \cdot \frac{\partial O_{31}}{\partial w_{11}^3}$$

We should calculate $\frac{\partial L}{\partial w_{11}^3}$ by using above formula and substitute in w_{11new}^3 (equation- 2)

Similarly, the formula for update the w_{21}^3 is:

$$w_{21new}^3 = w_{21}^3 - \alpha \frac{\partial L}{\partial w_{21}^3} \ldots\ldots\ldots\ldots \quad (3)$$

$$\frac{\partial L}{\partial w_{21}^3} = \frac{\partial L}{\partial O_{31}} \cdot \frac{\partial O_{31}}{\partial w_{21}^3}$$

The formula for update the w_{12}^2 is:

$$w_{12new}^2 = w_{12}^2 - \alpha \frac{\partial L}{\partial w_{12}^2} \ldots\ldots\ldots\ldots \quad (2)$$

$$\frac{\partial L}{\partial w_{12}^2} = \left[\frac{\partial L}{\partial O_{31}} \cdot \frac{\partial O_{31}}{\partial O_{21}} \cdot \frac{\partial O_{21}}{\partial w_{11}^2}\right] + \left[\frac{\partial L}{\partial O_{31}} \cdot \frac{\partial O_{31}}{\partial O_{22}} \cdot \frac{\partial O_{22}}{\partial w_{12}^2}\right]$$

It will absorb the above equation The second term is extra added after + Because when we want to change the weight w_{12}^2 the output of O_{31} affected in two paths are $w_{11}^2 \to O_{21} \to O_{31}$ and $w_{12}^2 \to O_{22} \to O_{31}$. That's why second chain added in above equation.

Note: This is just a basic chain calculation, if you assume 100 connections of neurons are existed, then the formula will be bigger and many differentiations will be performed in that formula.