1. The Numerical Differentiation

Definition:

It's a numerical value of a derivative of a given function at a given point.

Formula:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}.$$

● What is a main Problem we have to Face?

The main problem is "Numerical Instability". Typically, we use numerical differentiation for approximation of results. When we have to use the smaller value of h which is less then 1. If we use 0 then the result goes to infinity this is numerical instability.

● How can we remove this Numerical Instability?

For removing the Numerical Instability we must have to use that value of h which is >0.

2. The Chain Rule

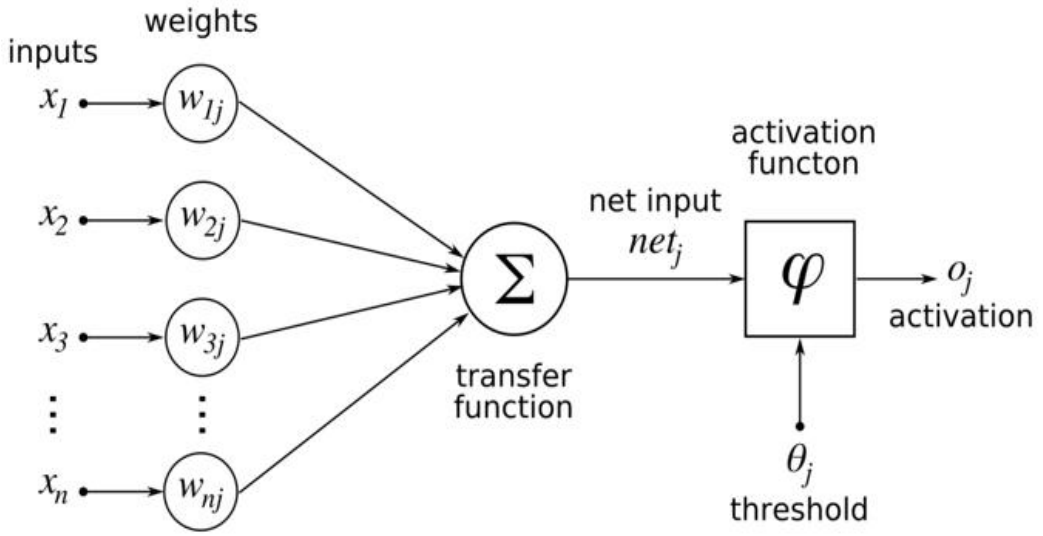
Definition:

The Rate of change of Error with respect to weight or we can say that it's a partial derivative of error with respect to its weight.

Formula:

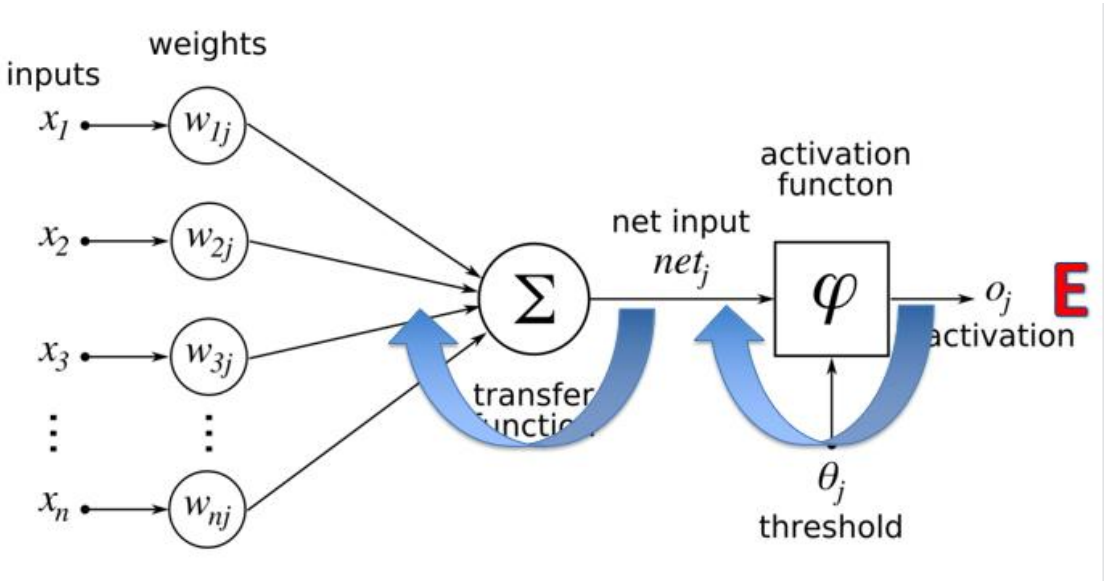
$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial \text{net}_j}\frac{\partial \text{net}_j}{\partial w_{ij}}$$

- **First we need to understand how neural network works?**



E

We have Inputs, weights, Transfer Function , Activation Function and then the Output. From output we can calculate the error.



$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial \text{net}_j}\frac{\partial \text{net}_j}{\partial w_{ij}}$$

From above diagram and the formula now you understand how it works. It consist on multiple layers each layer connected with the previous layer that's why it's a back propagation. First we have Error then take partial deferential with out put O_j then O_j and net_j and so on until you get the rate change of error with respect to weights.

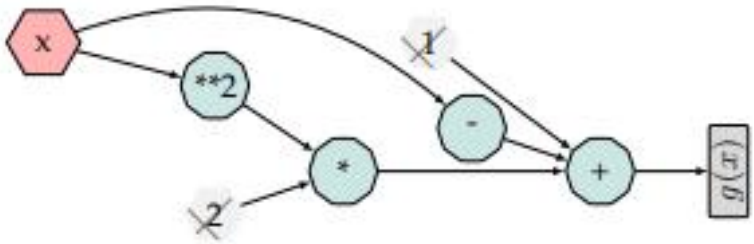
Usage:

1. The chain Rule is used to find the derivatives of a composite function.The Composite Function is the combination of two or more Function.
2. The chain rule is widely applied by back-propagation algorithms to calculate the error of the loss function with respect to each weight.

3. Directed Acyclic Graph (DAG)

Such graphs which have no loop or cycles connecting with the other edges and each edge have some direction is called Directed Acyclic Graph. This is only one way graph means we can't traverse back to previous edge. The Directed Acyclic Graph is a topological sorting, means each node is in a specific order.

Example:



f(1) = 2
f(2) = x^2
f(3) = -x
g(x) = f(1) * f(2) - f(3) + f(4)
SO,

$g(x) = 2x^2 - x + 1$

● Application of DAG

1. Routing in computer networks
2. Data processing