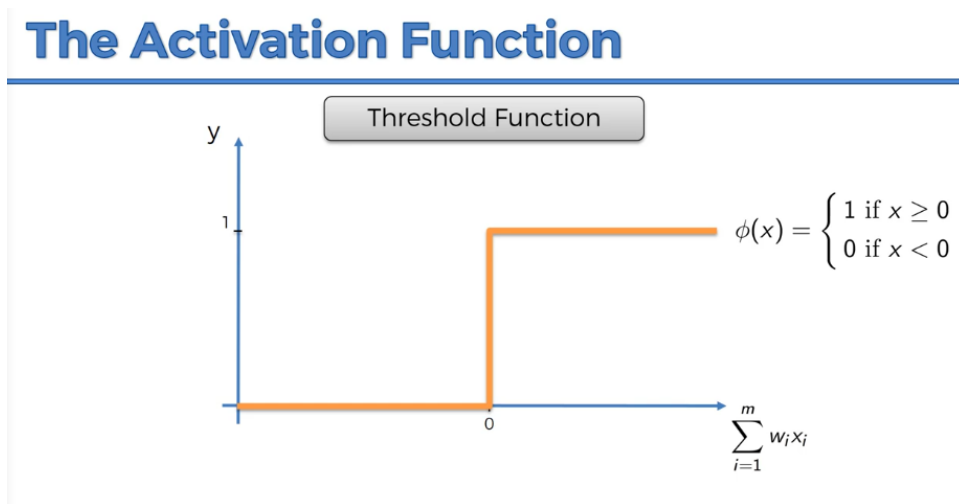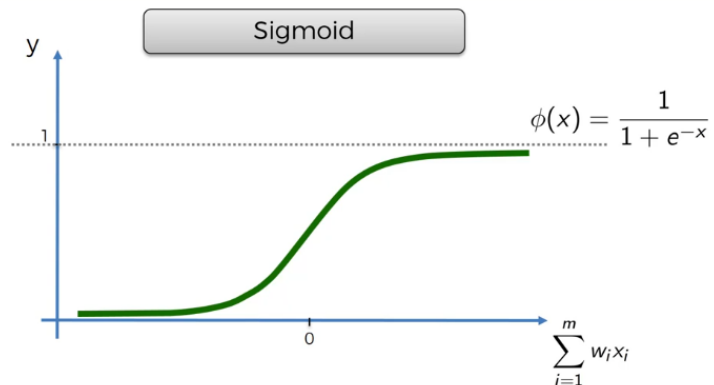# Activation Functions

1. Threshold Function

- It is a straightforward process. It can be called a Yes/No function. If the value is lower than 0, then the threshold function passes 0. On the other hand, if the value is greater than 0, then the function passes 1.

The Activation Function

Threshold Function

$$\phi(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$$

$$\sum_{i=1}^{m} w_i x_i$$

2. Sigmoid Function

- The Sigmoid function is also used in Logistic Regression.

- It is a smoother function when compared with the Threshold function. Sigmoid function is extremely useful when predicting probabilities in the output layer.
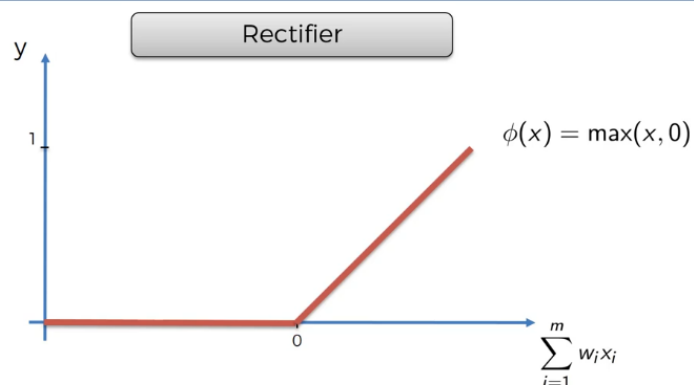
The Activation Function

Sigmoid

$$\phi(x) = \frac{1}{1 + e^{-x}}$$

3. Rectifier Function

- One of the most popular activation in ANN.

- An example for the Rectifier Function may be coal. When coal is coal, there is no value for it. However, when it turns into diamond, then its value increases a lot.
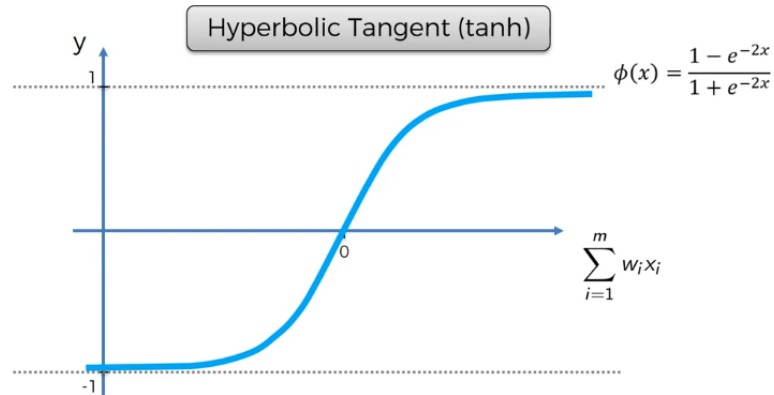
The Activation Function

Rectifier

$$\phi(x) = \max(x, 0)$$

4. Hyperbolic Tangent

- It is similar to Sigmoid function but the values goes below 0.

The Activation Function

Hyperbolic Tangent (tanh)
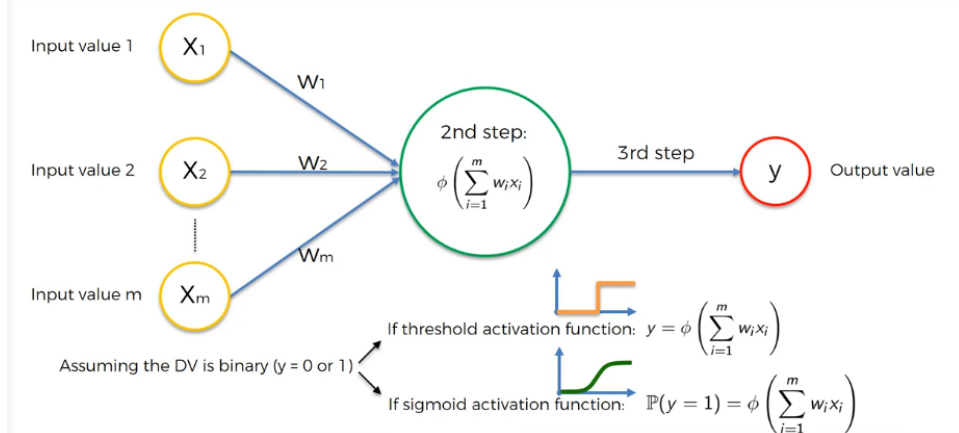
$$\phi(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

$$\sum_{i=1}^{m} w_i x_i$$

Remarks:

- In a binary classification, it is possible to use either Threshold function or Sigmoid function as activation function.

The Activation Function

Input value 1 X_1 W_1

Input value 2 X_2 W_2 2nd step: $\phi\left(\sum_{i=1}^{m} w_i x_i\right)$ 3rd step y Output value

Input value m X_m W_m

Assuming the DV is binary (y = 0 or 1)

If threshold activation function: $y = \phi\left(\sum_{i=1}^{m} w_i x_i\right)$

If sigmoid activation function: $\mathbb{P}(y = 1) = \phi\left(\sum_{i=1}^{m} w_i x_i\right)$

Resources:

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/3ec212ad-4768-46d0-b576-28d60ad2192e/glorot11a.pdf