Activation Function

Activation Function

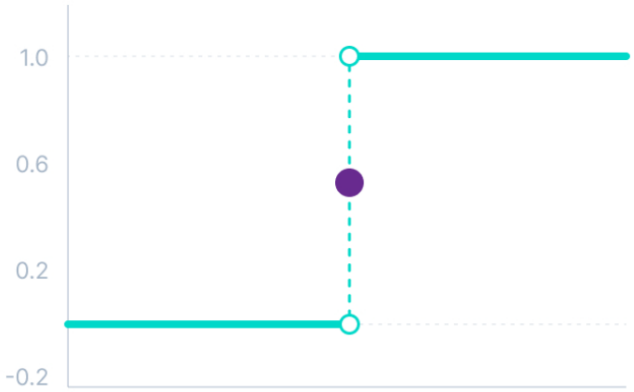
- Decides whether a neuron should be activated or not.
- *used to determine the output of neural network*
- basically divided into 2 types
  - Linear Activation Function
  - Non-linear Activation Functions

Binary Step Function

- Depends on a threshold value that decides whether a neuron should be activated or not.
- The input fed to the activation function is compared to a certain threshold;
  - if the input is greater than it, then the neuron is activated, else it is deactivated

$$\textit{Binary step}$$

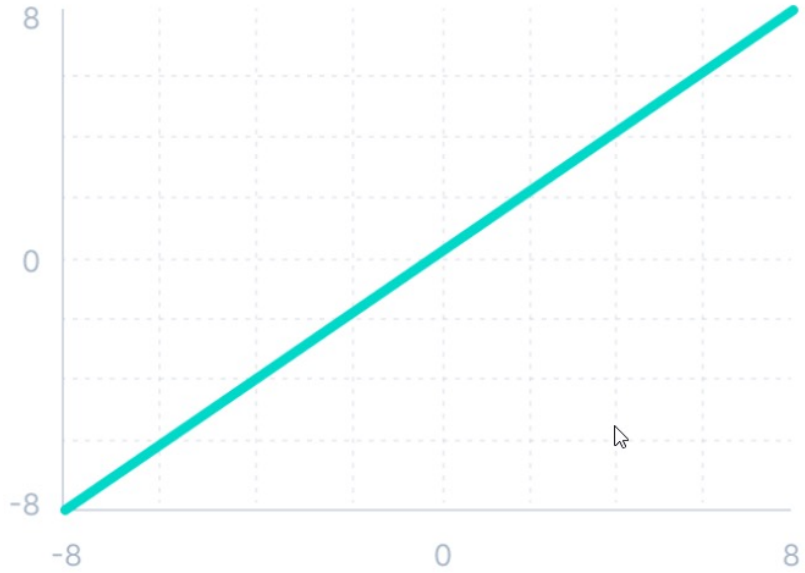
$$f(x) = \begin{cases} 0 & for\ x < 0 \\ 1 & for\ x \geqslant 0 \end{cases}$$



- limitations of binary step function:
  - It cannot provide multi-value outputs—for example, it cannot be used for multi-class classification problems.
  - The gradient of the step function is zero, which causes a hindrance in the backpropagation process.

Linear Activation Function

- The linear activation function, also known as "no activation," or "identity function" (multiplied x1.0), is where the activation is proportional to the input.



Linear

$$f(x) = x$$

Non-Linear Activation Functions

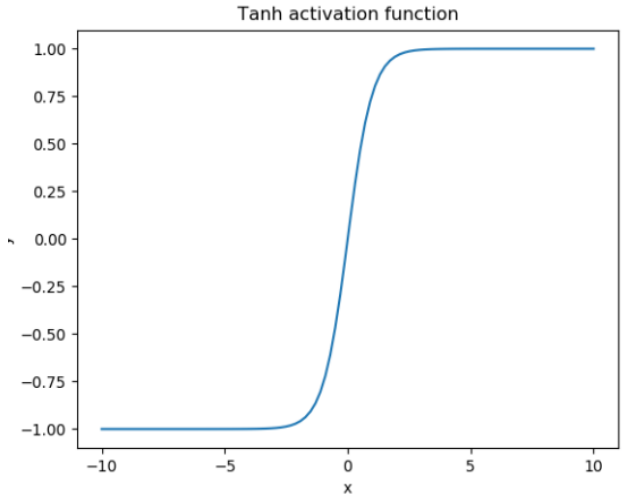
- **Sigmoid / Logistic Activation Function**
  - This function takes any real value as input and outputs values in the range of 0 to 1.
  - The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to 0.0

Sigmoid / Logistic

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

- **When the output value is close to 1, the neuron is active and enables the flow of information, while a value close to 0 corresponds to an inactive neuron.**
- **the output of the sigmoid activation function can be interpreted as a probability** since it lies in the range . That's why it is also used in the output neurons of a **prediction task**.
- most widely used functions:
  - used for models where we have to predict the probability as an output

- **Tanh Function (Hyperbolic Tangent)**
- Tanh function is very similar to the sigmoid/logistic activation function, and even has the same S-shape with the difference in output range of **-1 to 1.**
- The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to -1.0.

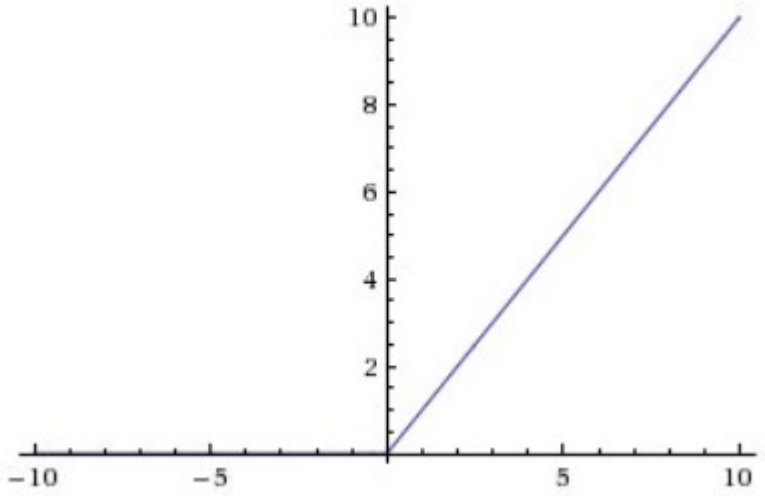

Tanh activation function
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−10
−5
0
5
10
x

Rectified Linear Unit (ReLU)

- A linear function that will output the input directly if it is positive, otherwise, it will output zero.
- The function returns 0 if it receives any negative input, but for any positive value x it returns that value back.

$$f(x)=max(0,x)$$

Rectified Linear Unit (ReLU)



- It acts as a linear function for positive values and as a non-linear activation function for negative values.

Vanishing gradient problem

- The sigmoid function is one of the most popular activations functions used for developing deep neural networks.
- The use of sigmoid function restricted the training of deep neural networks because it caused the vanishing gradient problem
- This causes the neural network to learn at a slower pace or in some cases no learning at all.

Vanishing gradient problem