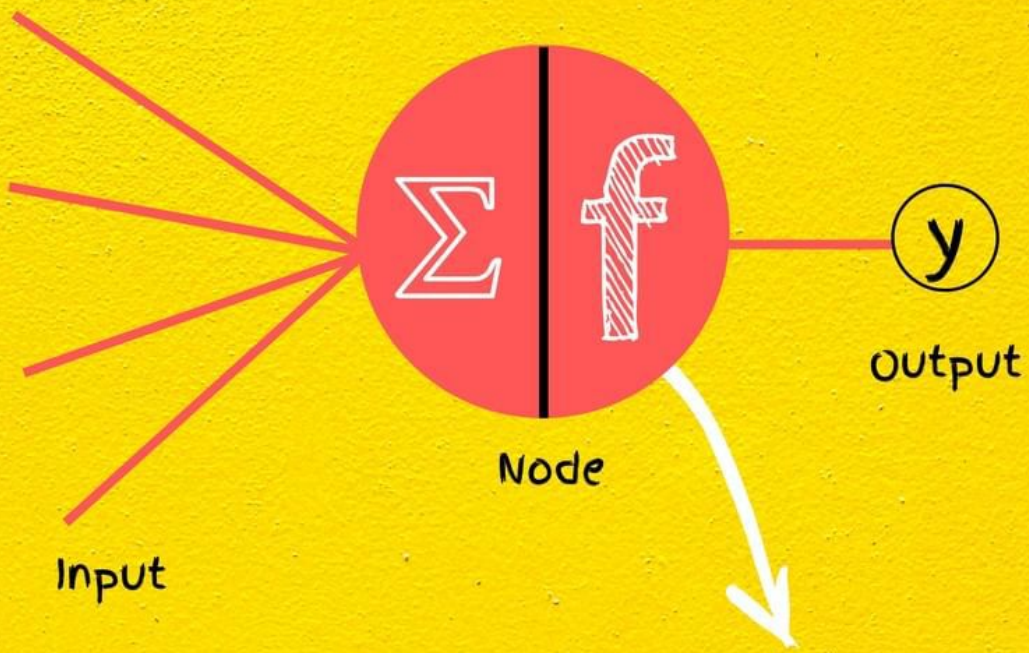
01

Binary, Linear, Sigmoid, TanH and ReLu

5 ACTIVATION FUNCTIONS FOR DEEP LEARNING

@learn.machinelearning

02



WHAT IS ACTIVATION FUNCTION?

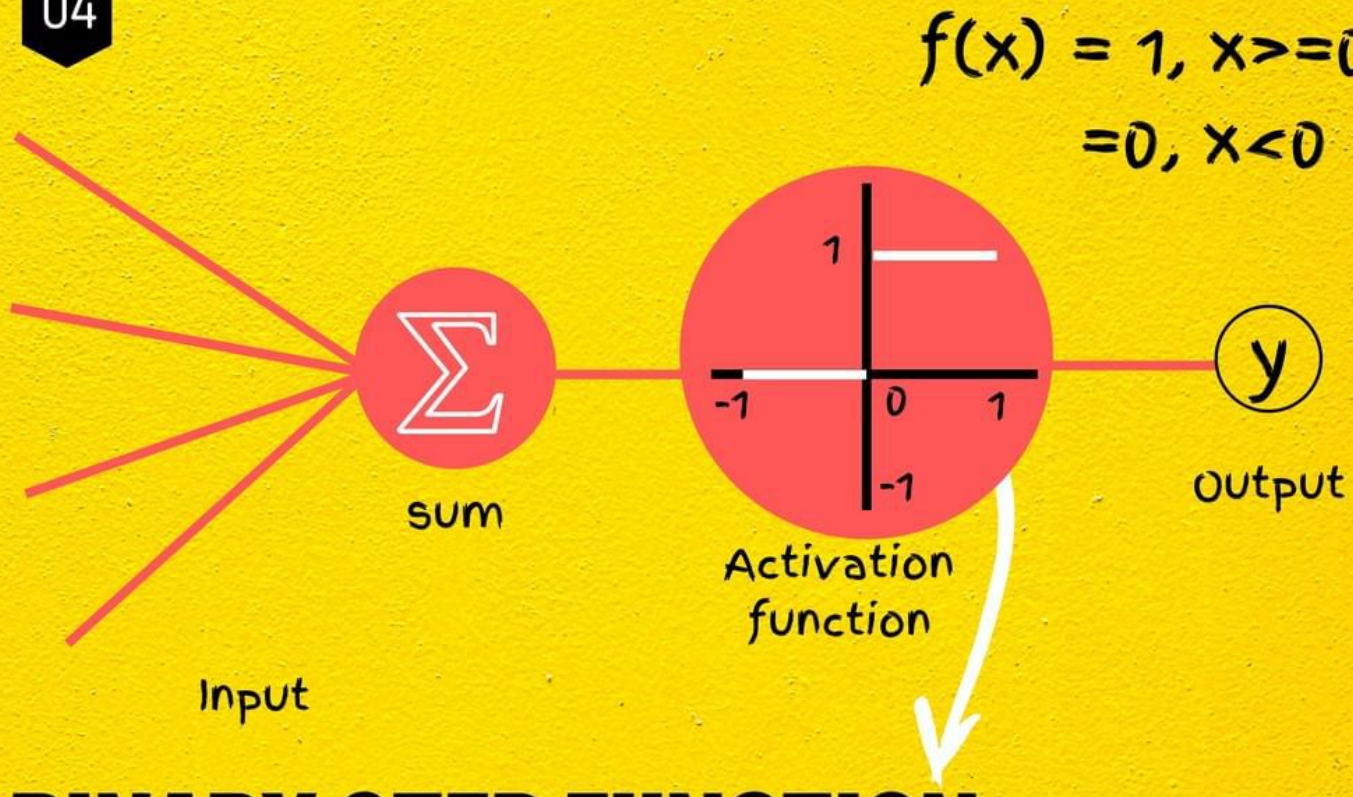
- The ultimate goal of an Activation function is to convert the input and generate the output which will be used as input for the next layer.
- They decide whether to activate a neuron or not.
- It also helps to normalize the output of any input in the range between 1 to -1 or 0 to 1 or >=1

03

WHY ACTIVATION FUNCTION

- Imagine a neural network without the activation functions, the output signal would simply be a simple linear function.
- These linear equations are limited in complexity and less power to find the complex mappings between input and output.
- Activation functions introduce non-linearity into the model which can help in finding the complex pattern in data like images, videos, audio, etc.

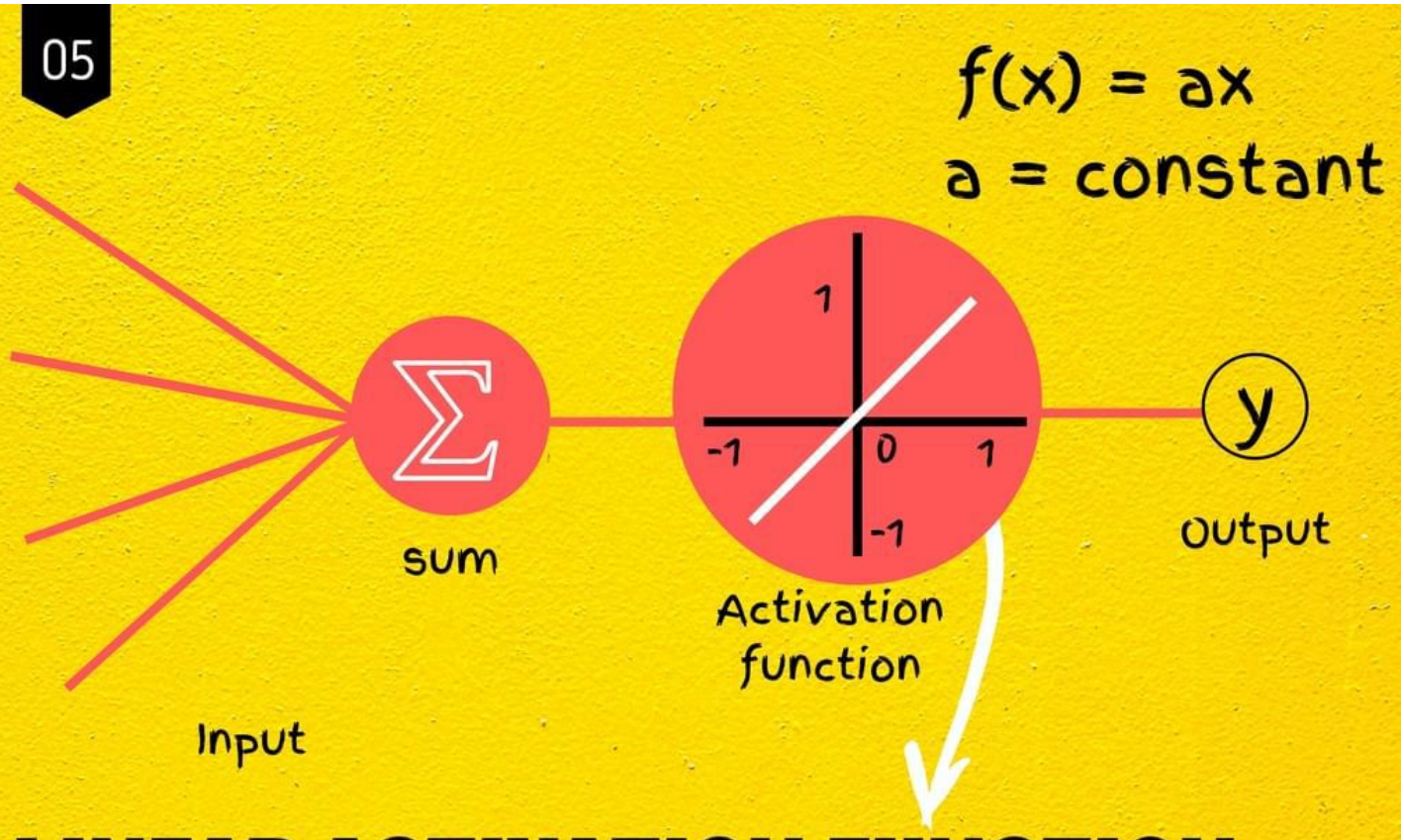
04



BINARY STEP FUNCTION

- It is a threshold-based activation function.
- If x value is >= 0 then y = 1, neuron activated else not activated.
- The gradient of the step function is zero which causes a hindrance in the backpropagation process.
- Gradients are calculated to update the weights and biases during the backprop process. Since the gradient of the function is zero, the weights and biases don't update.

05

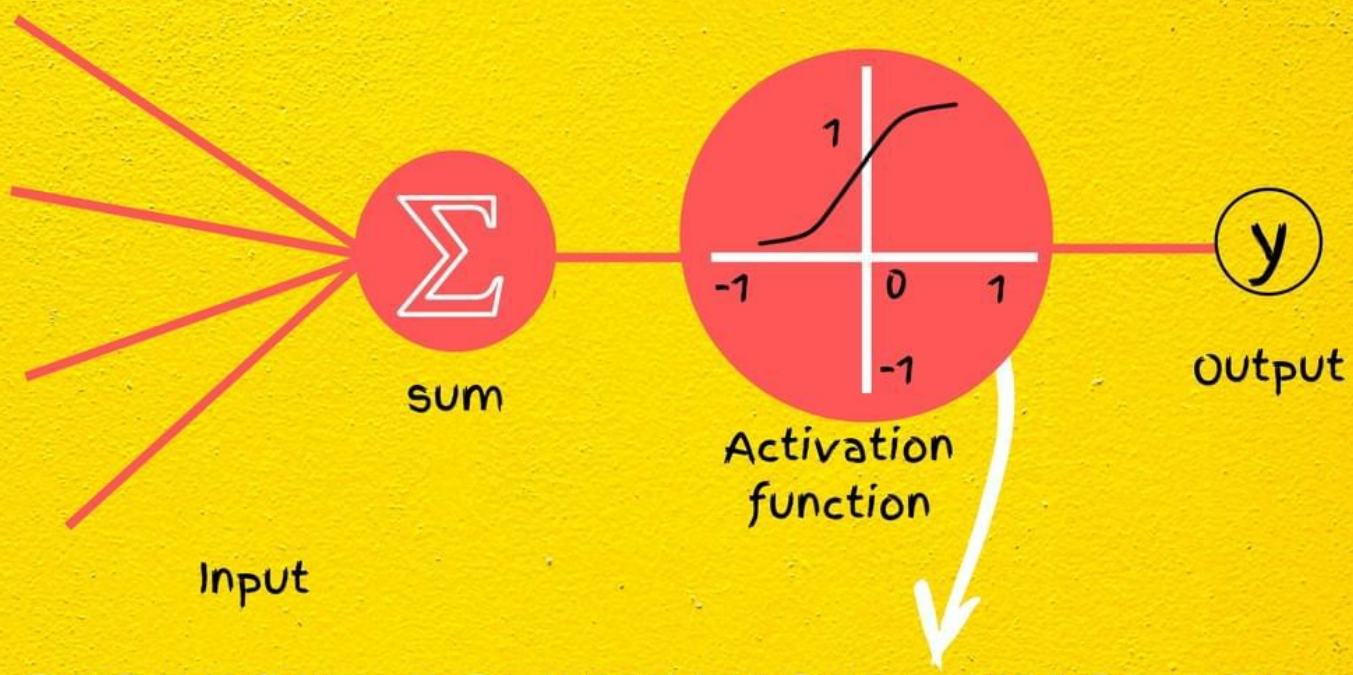


LINEAR ACTIVATION FUNCTION

- It is better than step function as it can have multiple values as we multiple with constant.
- One disadvantage of this function is gradient is constant as it doesn't depend on the input x.
- No matter how many layers we have, if all are linear in nature, the final activation function of the last layer is nothing but just a linear function of the input of the first layer

06

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

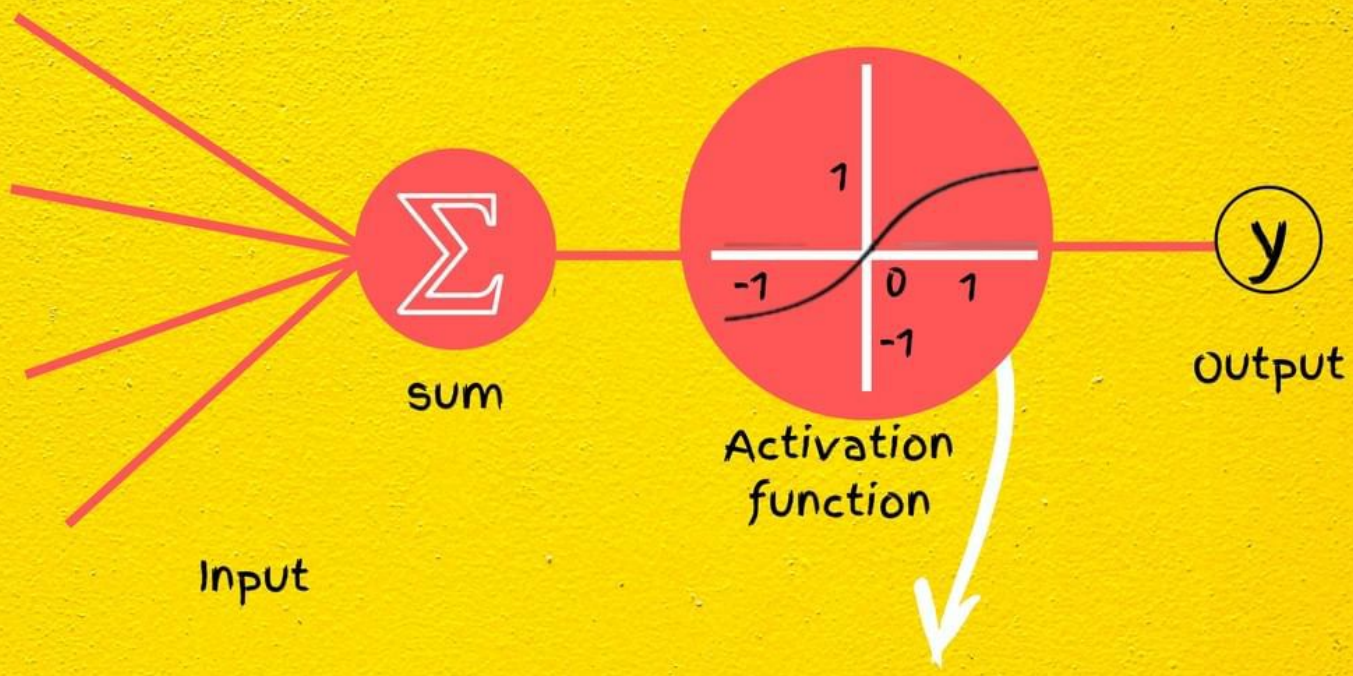


SIGMOID ACTIVATION FUNCTION

- If you look, it is an s-shaped smooth curve and bounds its values between o and i.
- It is a non-linear function, so we can add multiple non-linear functions to find complex patterns and it is also differentiable.
- That less steep values(small gradient values) causes vanishing gradient problem. The network refuses to learn further or slow convergence.
- It's output isn't zero centred.

07

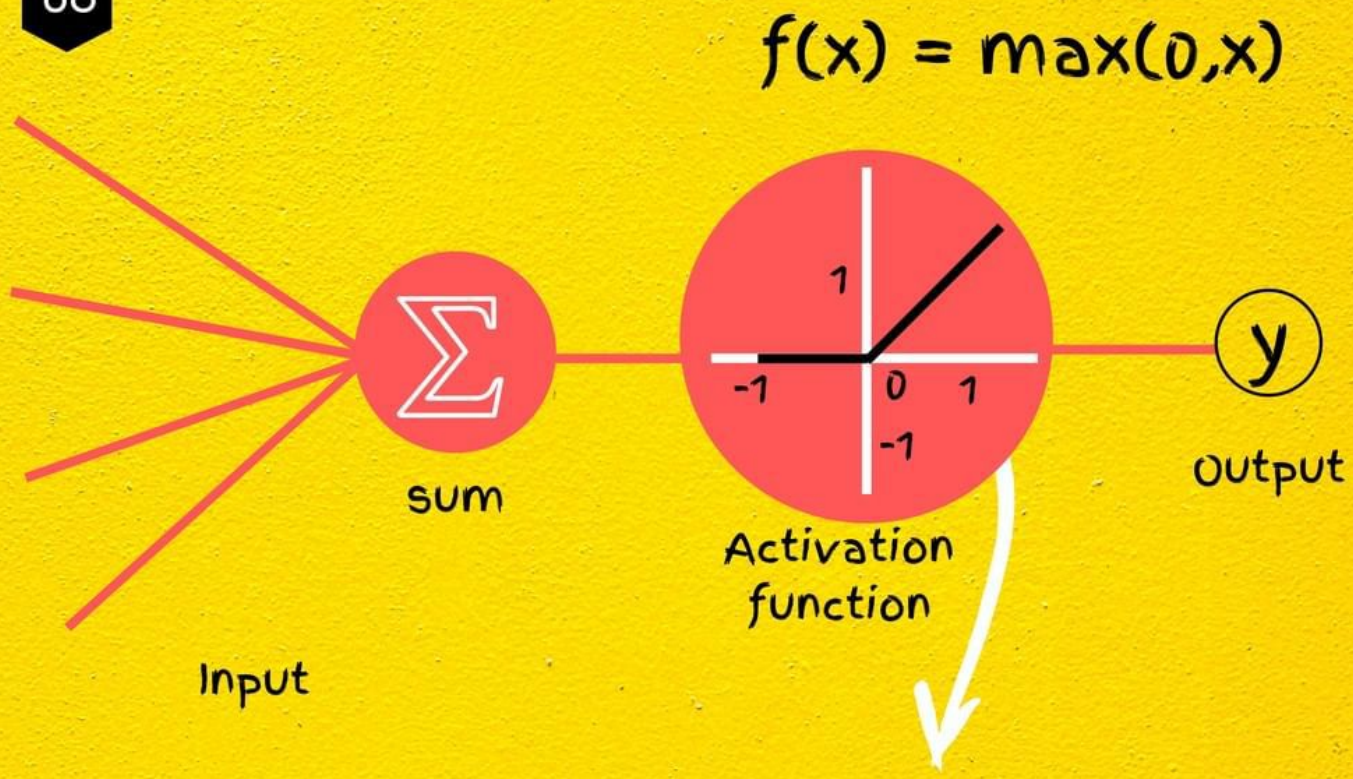
$$f(x) = (2/(1+e^{-2x})) - 1$$



TANH ACTIVATION FUNCTION

- It's mathematically shifted version of the sigmoid function.
- The values bound between -1 and 1, so the mean will be zero or very close to zero.
- Convergence is faster than sigmoid and it has the same disadvantages as sigmoid
- Usually, tanh is preferred over the sigmoid function since it is zero centred.

08



RELU ACTIVATION FUNCTION

- Values can range between [0, x) and this is a non-linear function.
- Big advantages of ReLu function is it doesn't activate all neurons at the same time which introduces sparsity in the model, computationally efficient and converges very fast
- And the disadvantage of ReLu is (dead neurons or dying ReLu) which is because of gradients going towards zero when X is negative and the weights will not be adjusted during backprop

09

RECAP

01 WHAT IS ACTIVATION FUNCTION

02 WHY ACTIVATION FUNCTION

03 BINARY STEP FUNCTION

04 LINEAR ACTIVATION FUNCTION

05 SIGMOID ACTIVATION FUNCTION

06 TANH ACTIVATION FUNCTION

07 RELU ACTIVATION FUNCTION