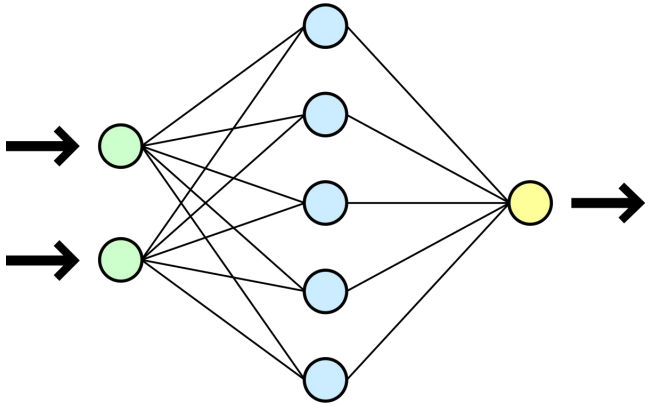


THE CODING LIBRARY

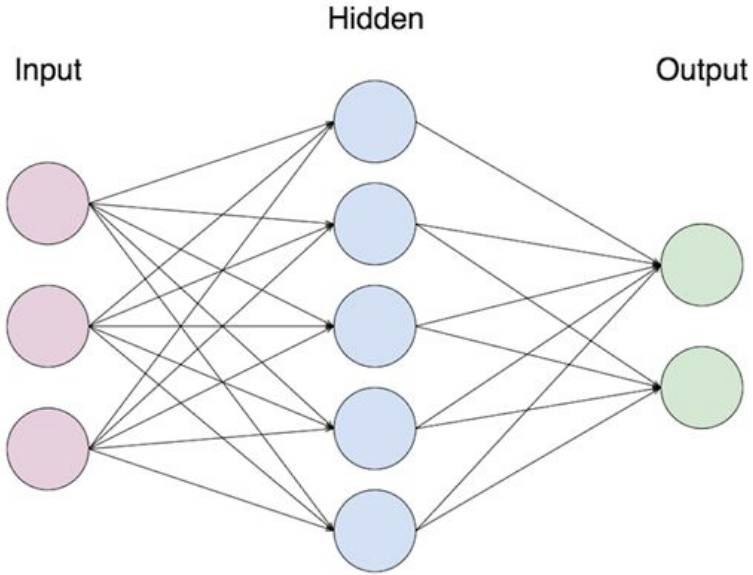
Neural Networks

Artificial Neural Networks

- Layers and Activation Functions

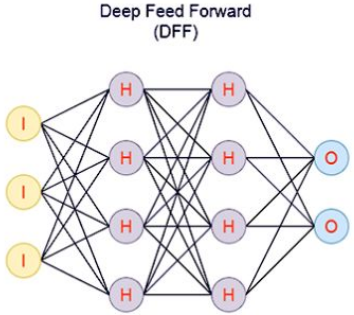
Artificial Neural Network

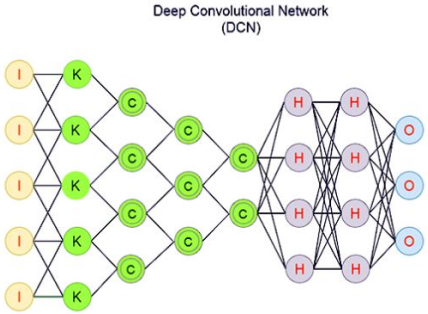
- Neurons or Nodes in a ANN
- Numerical Data
- Layers
- Weights
- Sum of Weights
- Feedforward
- Activation Functions

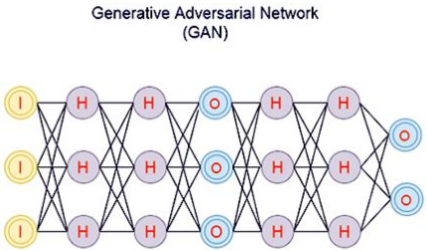


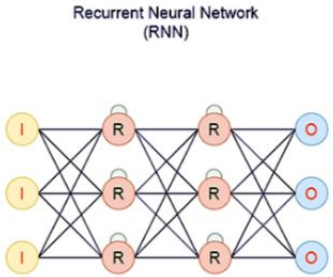
Layers in an Neural Network

- Dense Layers
- Convolutional Layers
- Deconvolutional Layers
- Pooling Layers
- Recurrent Layers
- Normalization Layers

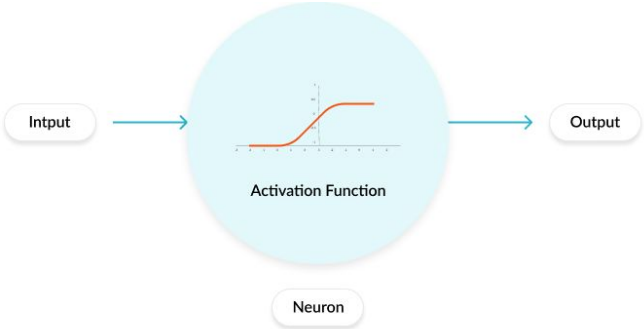


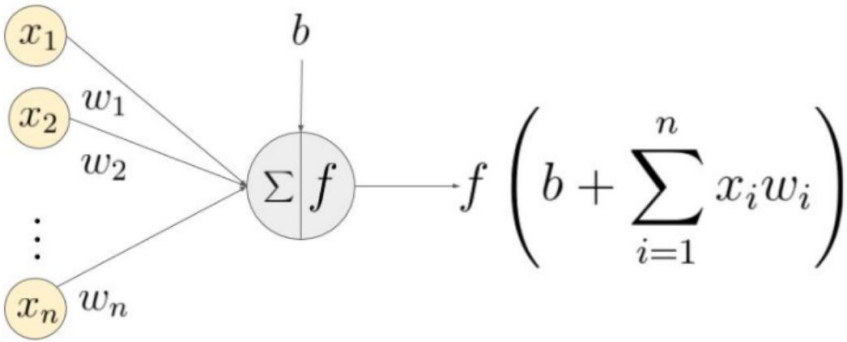






Activation Function in a Neural Network





How Activation Functions affect Neurons

THE CODING LIBRARY

- Activation Functions do non-linear transformations to the inputs which makes it capable to learn and perform more complex tasks.
- Makes the Neural Network able to work with videos, audio, speech etc.
- Linear Activation Function
  - Output is proportional to the input
  - When we backpropagate the gradient will be the same
    - Constant when we differentiate a linear function
- Non-linear Activation Functions
  - Most used ones
  - Easier for the model to generalize or adapt to the variety of data and to differentiate between the output
  - Used to back-propagation and updating the weights during training

Activation Functions

THE CODING LIBRARY

- Activation Functions should be differential and monotonic
  - Makes back-propagation possible which updates the weights and biases during training

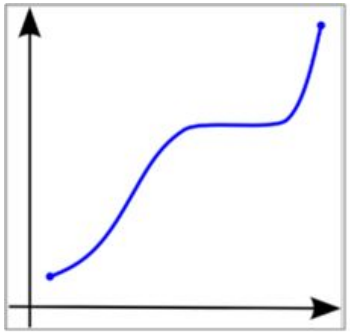


Figure 1 - A monotonically increasing function

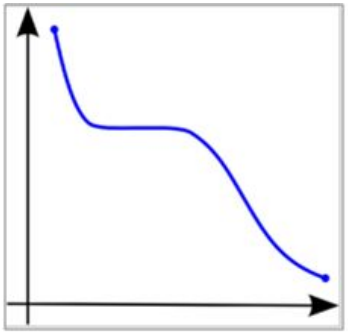


Figure 2 - A monotonically decreasing function

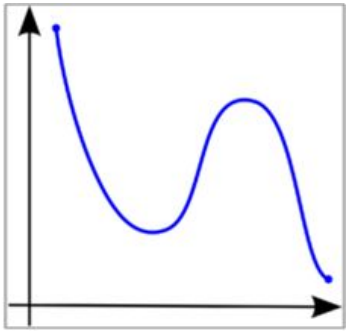
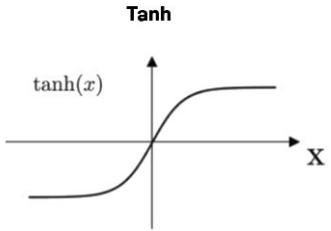
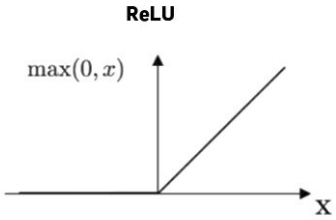


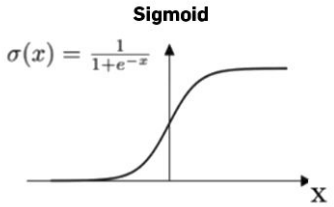
Figure 3 - A function that is not monotonic

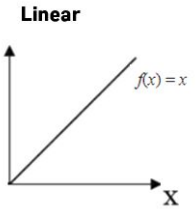
Types of Activation Functions

- Linear
  - Output is the same as the input
- Sigmoid
  - Can cause vanishing gradients
  - Output is not zero centered
    - Makes optimization harder
- ReLU
  - Not all neurons are activated at the same time
  - Efficient and easy for computation
  - Output is not zero centered
  - Dead neuron if gradient hits zero when backpropagating
- TanH
  - Output is zero centered so optimization is easier
  - Vanishing gradient problem. Can saturate and kill gradients









When to use which Activation Function

THE CODING LIBRARY

- ReLU is the most used one because of general better results
- In case of dead neurons with ReLU, Leaky ReLU or ELU can be used instead
- ReLU should only be used inside of hidden layers
- Sigmoid works good for classification problems
- Sigmoid and tanh is often avoided if possible
    - Vanishing gradient problem
- Tanh is avoided most of the time because of dead neurons