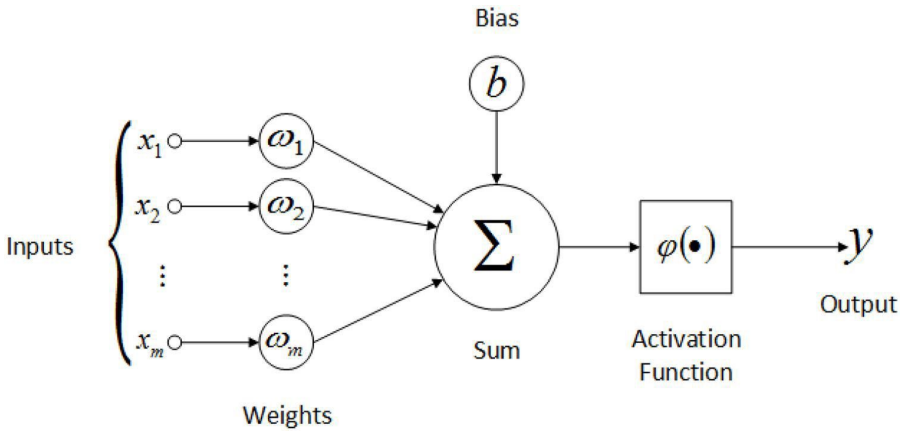
Activation Functions

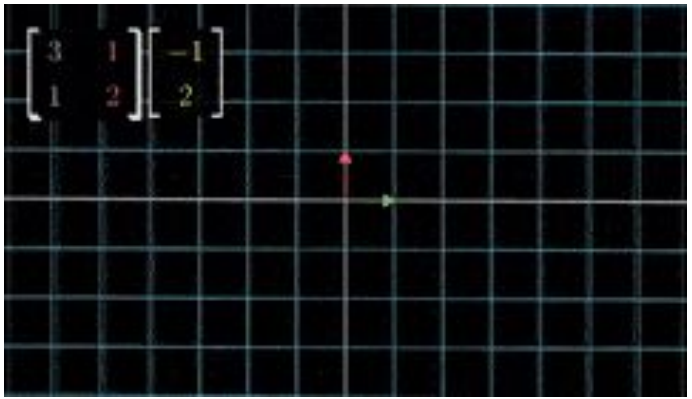
What are they?

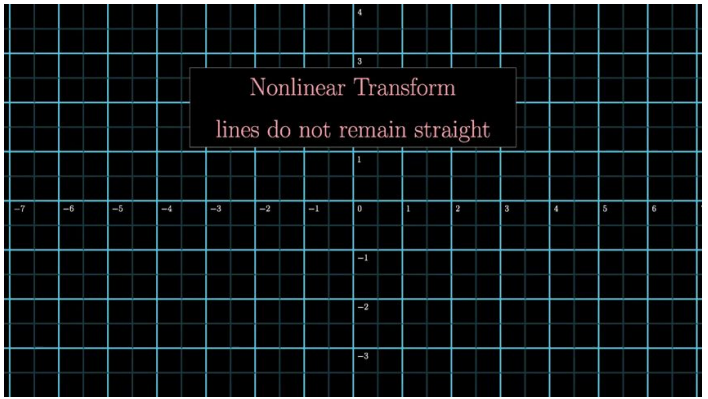
- Gate between the current neuron and the output going to the next layer



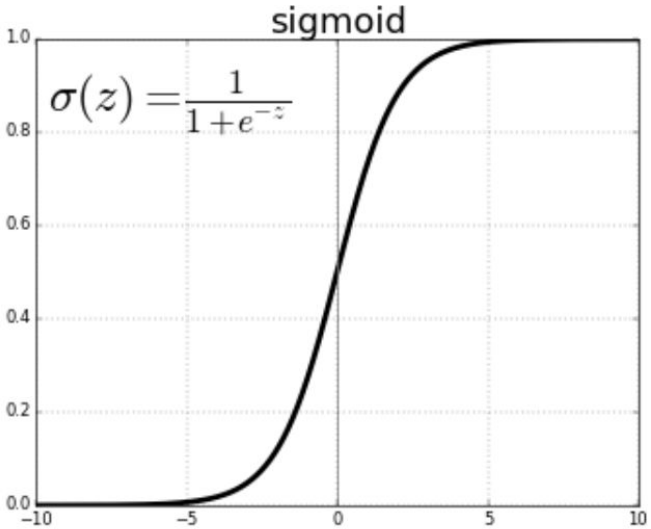
Why use Non-Linear Activation Functions?

- Enables more complex mappings between inputs and outputs





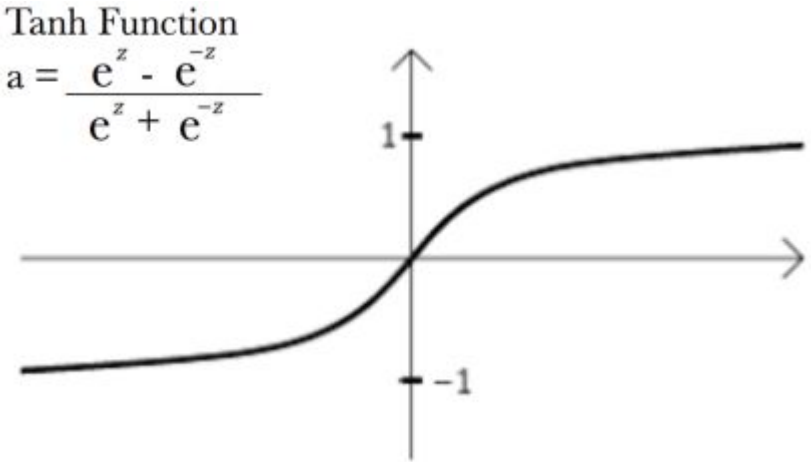
Sigmoid



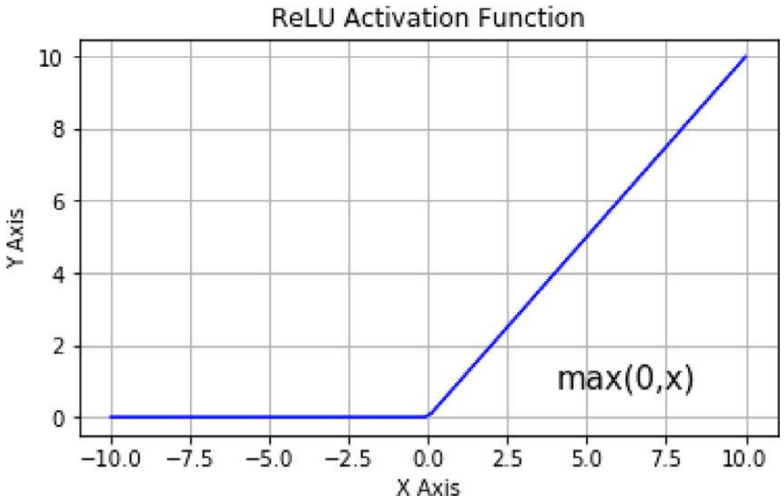
- Adv.
  - Smooth Gradient
  - Bounded 0 to 1
  - Clear predictions
- Dis.
  - Vanishing Gradient
  - Not zero-centered

Hyperbolic Tangent

- Adv.
  - Zero centered
  - Bounded -1 to 1
- Dis.
  - Vanishing Gradient



Rectified Linear Unit



- Adv.
  - Efficient
  - Non-linear
- Dis.
  - No gradient on negative values

Leaky ReLU

- Adv.
  - Negative values have slope
- Dis.
  - Negative values do not give consistent predictions