

# Recap

---

## What is the purpose of the activation function?

There are two reasons for the use of an activation function. The first is *modelling the behavior of a natural neurons spiking activity* and the second is, to *introduce nonlinearity* into the network, since otherwise all layers would collapse to a simple affine transformation.

## Activation functions

### Sigmoid

Function:

$$\sigma(x) \triangleq \frac{1}{1 + e^{-x}}$$

Derivative:

$$\frac{\partial \sigma}{\partial x} = \sigma(x)(1 - \sigma(x))$$

### Rectified Linear Unit

Function:

$$\text{ReLU}(x) \triangleq \max(x, 0)$$

Derivative:

$$\frac{\partial \text{ReLU}}{\partial x} = \text{Heaviside}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Unlike the stepfunction both sigmoid and ReLU have derivatives that are information preserving i.e. their derivatives do change w.r.t. the differentiated variable. If the step function would be used, all partial derivative running through the stepfunction according to the chainrule would be zeroed out during backpropagation and no information from the loss could reach the previous weights.