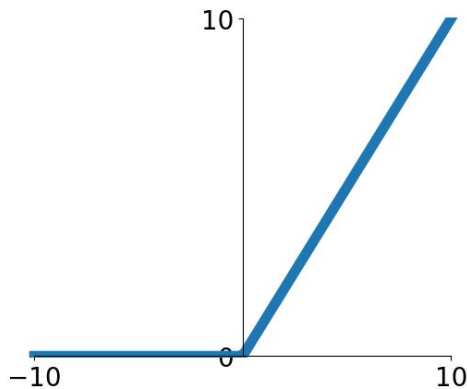


# Activation Functions



## ReLU (Rectified Linear Unit)

- Computes  $f(x) = \max(0, x)$
  - Does not saturate (in +region)
  - Very computationally efficient
  - Converges much faster than sigmoid/tanh in practice (e.g. 6x)
  - Actually more biologically plausible than sigmoid
- 
- Not zero-centered output
  - An annoyance:

hint: what is the gradient when  $x < 0$ ?