

Deep Learning: Activation Functions

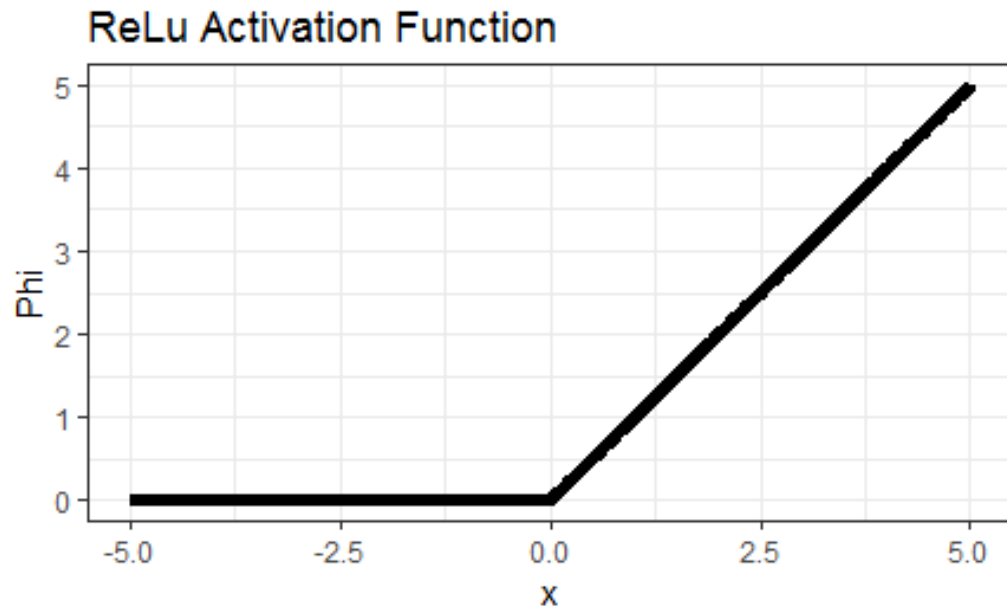
Deep Learning: Activation Functions

Activation Functions

There are different activation functions.

Rectified Linear Unit (ReLU)

- $\Phi = \max(0, x)$
- Most common
- Non-linear



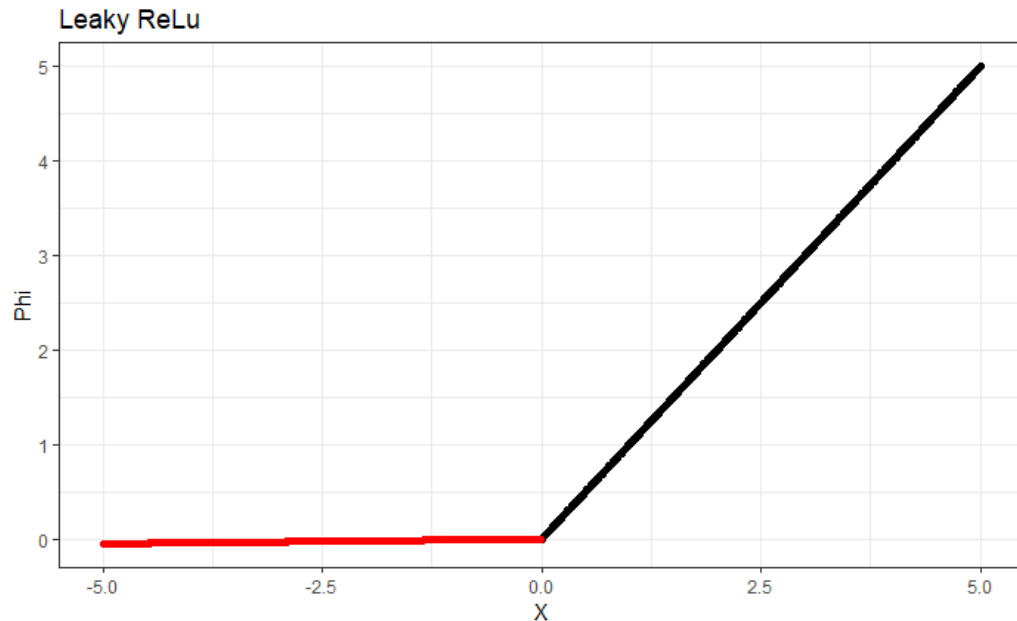
Deep Learning: Activation Functions

Activation Functions

Leaky Rectified Linear Unit (Leaky ReLU)

- $\Phi(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha * x & \text{otherwise} \end{cases}$

- α typically 0.01
- Instead zero for negative inputs, small gradient
- Gradient never zero

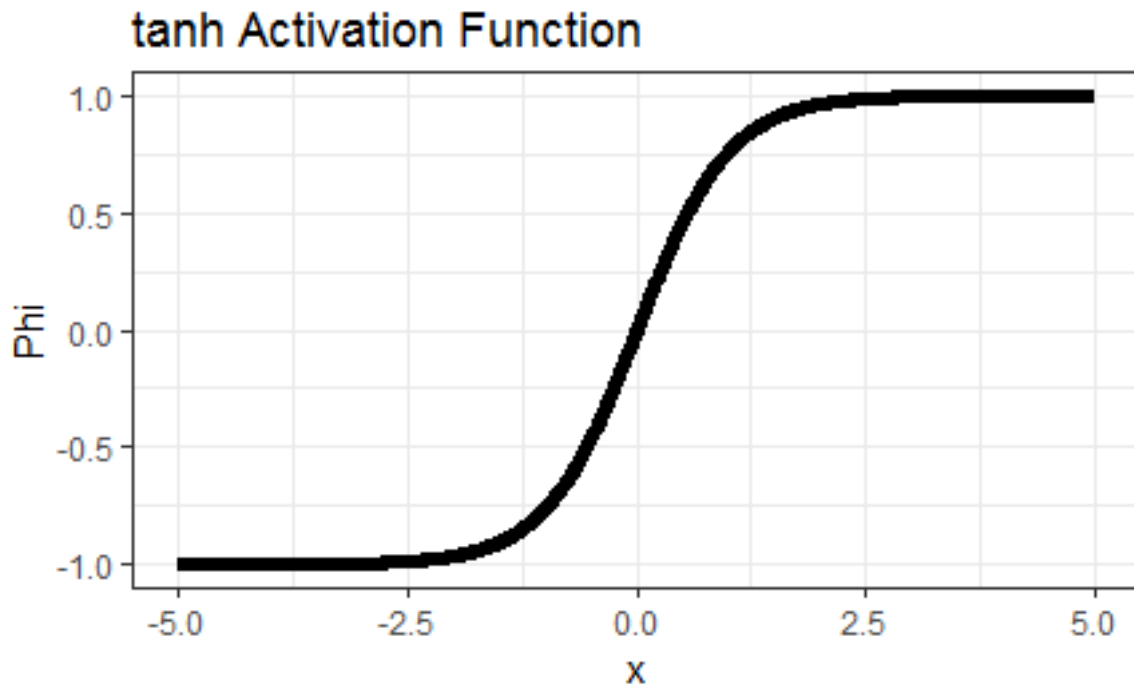


Deep Learning: Activation Functions

Activation Functions

Hyperbolic Tangent (tanh)

- $\Phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- Non-linear
- Relatively flat, except for small range
- Derivative small except for small range
- Might suffer vanishing gradient problem

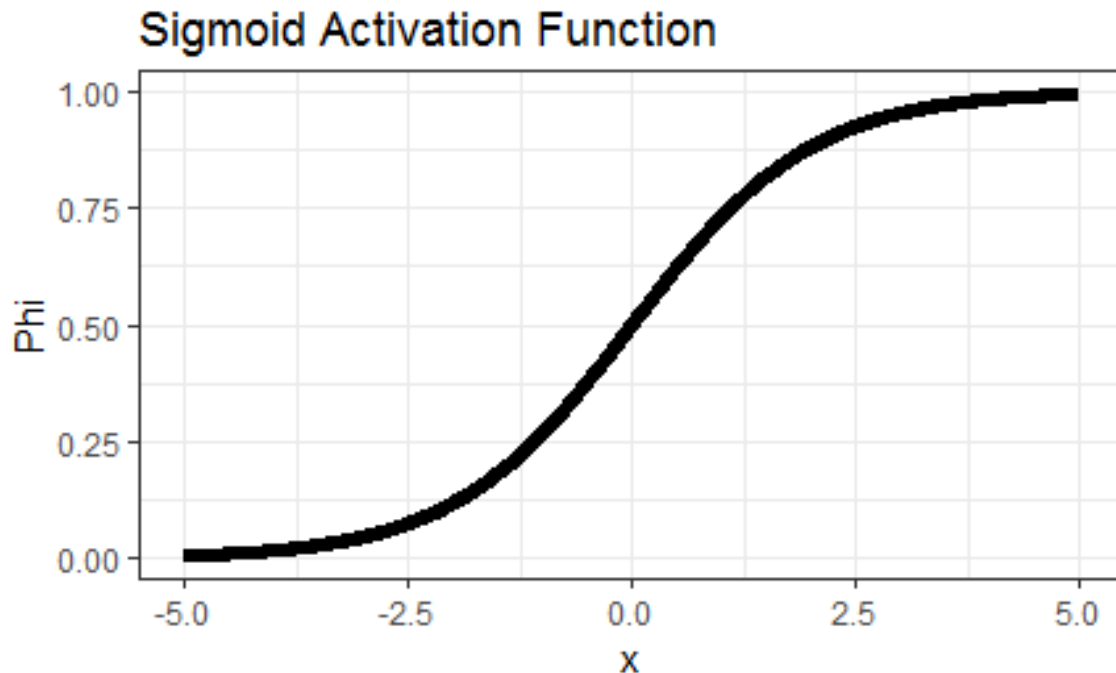


Deep Learning: Activation Functions

Activation Functions

Sigmoid

- $\Phi(x) = 1 / (1 + e^{-x})$
- Non-linear
- Relatively flat, except for small range
- Derivative small except for small range
- Might suffer vanishing gradient problem
- Result range 0 to 1



Deep Learning: Activation Functions

Activation Functions

Softmax

- Used for multi-class prediction

