Deep Learning: Activation Functions
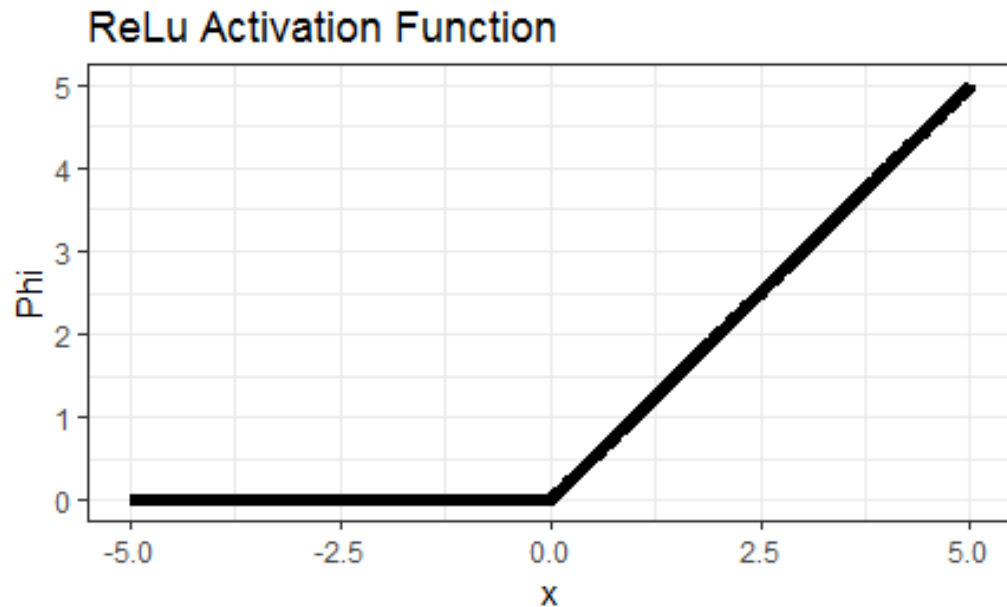
Bert Gollnick

Deep Learning: Activation Functions

There are different activation functions.

Rectified Linear Unit (ReLu)
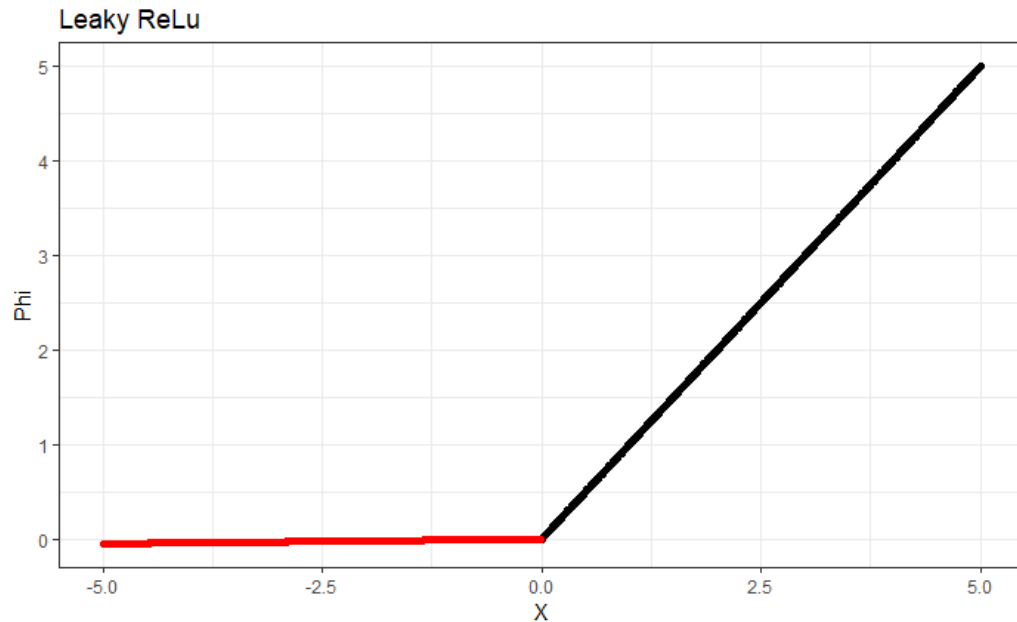
- Phi = max(0, x)
- Most common
- Non-linear

ReLu Activation Function

Deep Learning: Activation Functions

Leaky Rectified Linear Unit (Leaky ReLu)

- Phi(x) = { x if x>0
  alpha * x otherwise

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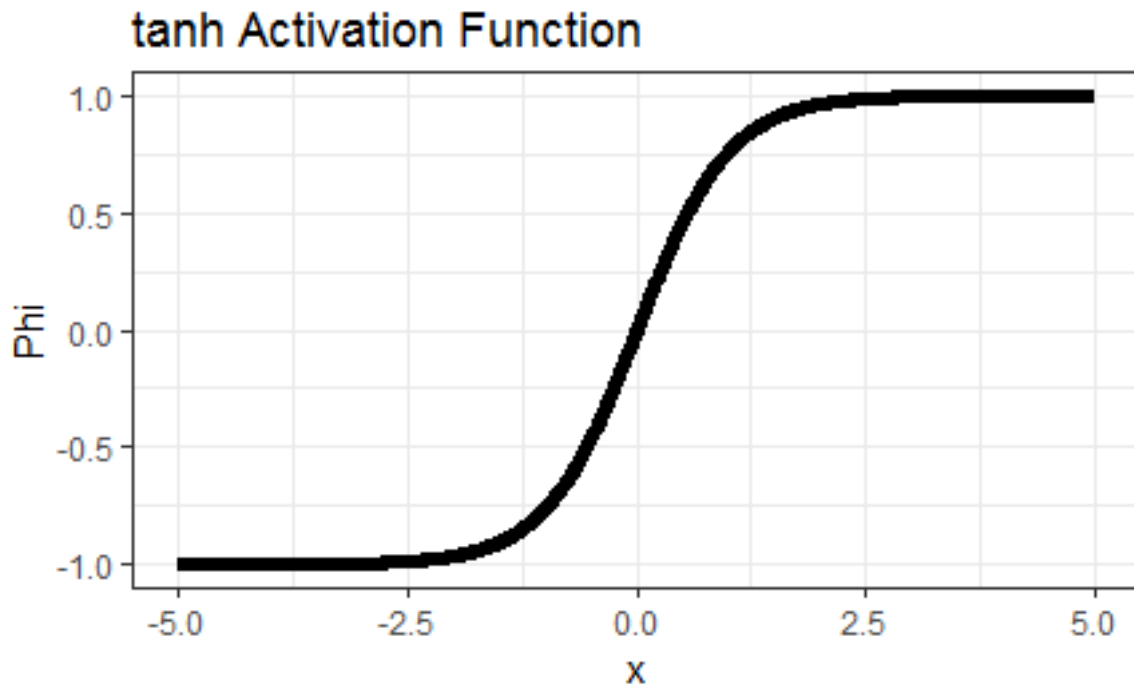- Instead zero for negative inputs, small gradient
- Gradient never zero

Leaky ReLu

Hyperbolic Tangent (tanh)

- $Phi(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

- Non-linear
- Relatively flat, except for small range
- Derivative small except for small range
- Might suffer vanishing gradient problem

tanh Activation Function

Sigmoid

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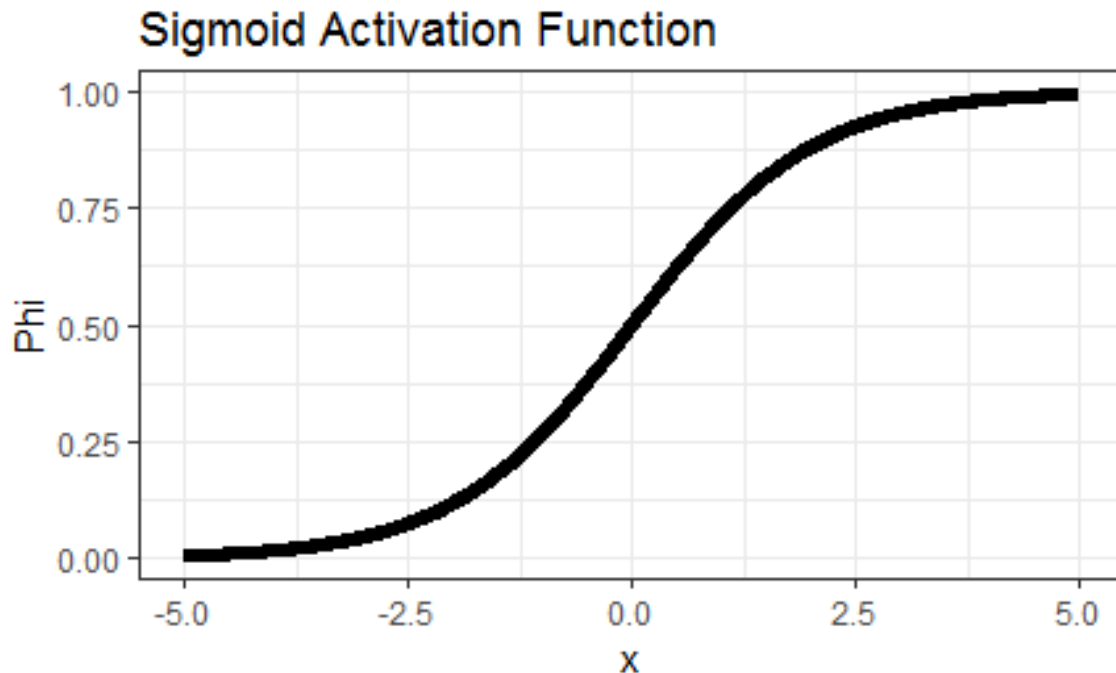- Might suffer vanishing gradient problem
- Result range 0 to 1

Sigmoid Activation Function

Deep Learning: Activation Functions

Softmax

- Used for multi-class prediction

1 (x1)

4 (x2)

2 (x3)

3 (x4)

Softmax

$$Phi(x) = \frac{e^{x_i}}{\sum_i e^{x_i}}$$

p(x1)

p(x2)

p(x3)

p(x4)