Tutorial 2

Overview

- Matrix and Vector Product
- Matrix Dimensions
- Perceptron and XOR
- Activation Functions
- Loss Functions
- Partial Derivatives

Matrix and Vector Product

Can you multiply these?

$$W = \begin{bmatrix} 3 & 4 \\ 1 & 2 \\ 5 & 6 \end{bmatrix} \qquad x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

What are their dimensions?

$W \sim 3 \times 2$

$x \sim 2 \times 1$

What are their dimensions?

$W \sim 3 \times 2$

$x \sim 2 \times 1$

$Wx \sim ?$

What are their dimensions?

$W \sim 3 \times 2$

$x \sim 2 \times 1$

$Wx \sim 3 \times 1$

Can you multiply them?

$$W = \begin{bmatrix} 3 & 4 & 5 & 1 \\ 1 & 7 & 7 & 3 \end{bmatrix} \qquad x = \begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}$$

In other words: addition of columns of W with weights of x

$$Wx = 3 \begin{bmatrix} 3 \\ 1 \end{bmatrix} + 5 \begin{bmatrix} 4 \\ 7 \end{bmatrix} + 4 \begin{bmatrix} 5 \\ 7 \end{bmatrix} + ?? \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Matrix Dimensions

Describe the NN

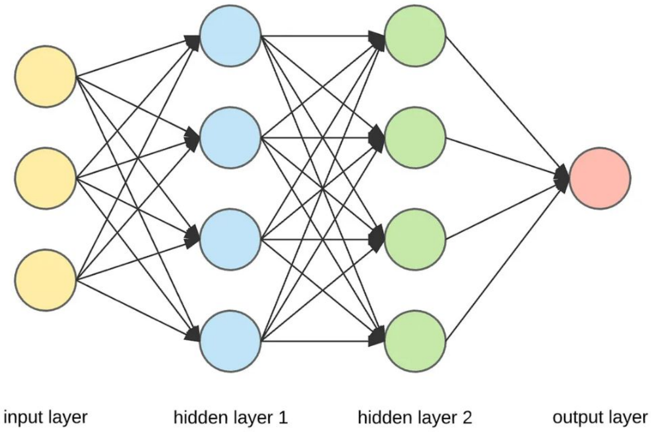


Figure 1

Given:

- Input layer and hidden layers are row vectors (a horizontal vectors)
- Every node is calculated like this:

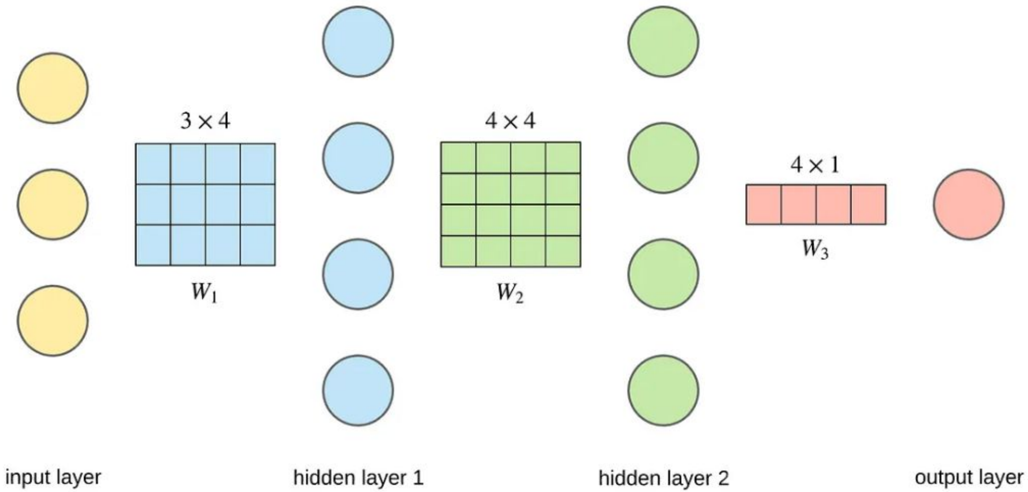
$$z = f(x \cdot w) = f\left(\sum_{i=1}^{n} x_i w_i\right)$$

Solve:

- Define the dimensions of all layers
- Define the dimensions of the weight matrices

https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6#106c

Describe the NN: Solution

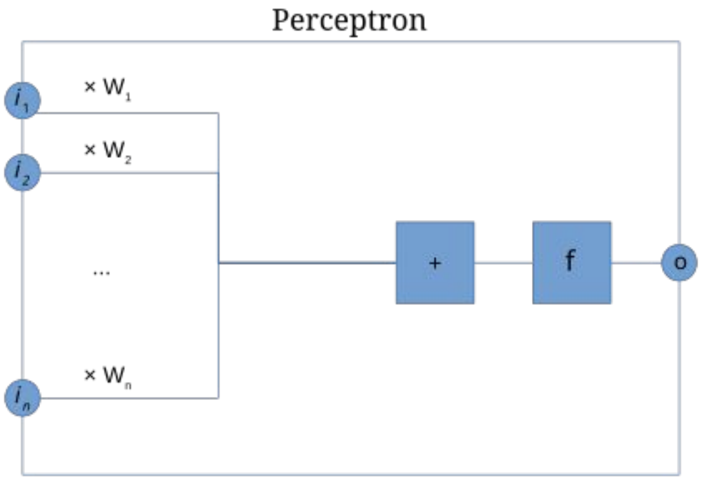


- Input: 1 x 3
- Hidden 1: 1 x 4
- Hidden 2: 1 x 4
- Output: 1 x 1

https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6#106c

Perceptron and XOR

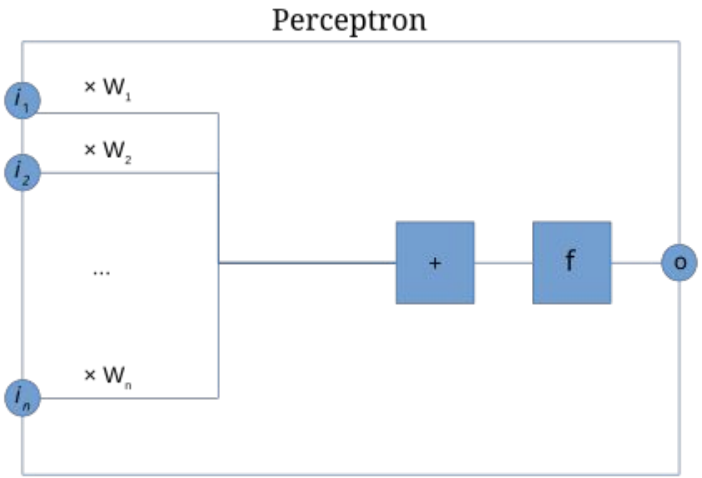
What is the perceptron?



Source: https://en.wikipedia.org/wiki/Perceptron

What is the perceptron?

$$f(x) = h(Wx + b)$$



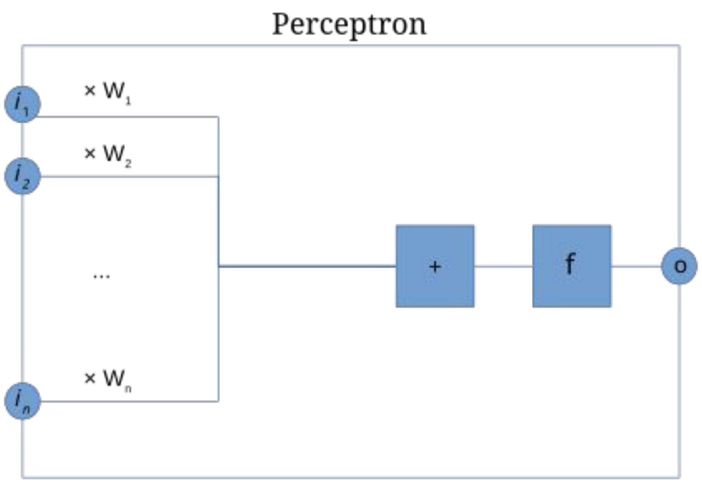
Source: https://en.wikipedia.org/wiki/Perceptron

What is the perceptron?

$$f(x) = h(Wx + b)$$

Where h = Heaviside step function

$$H(x) := \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



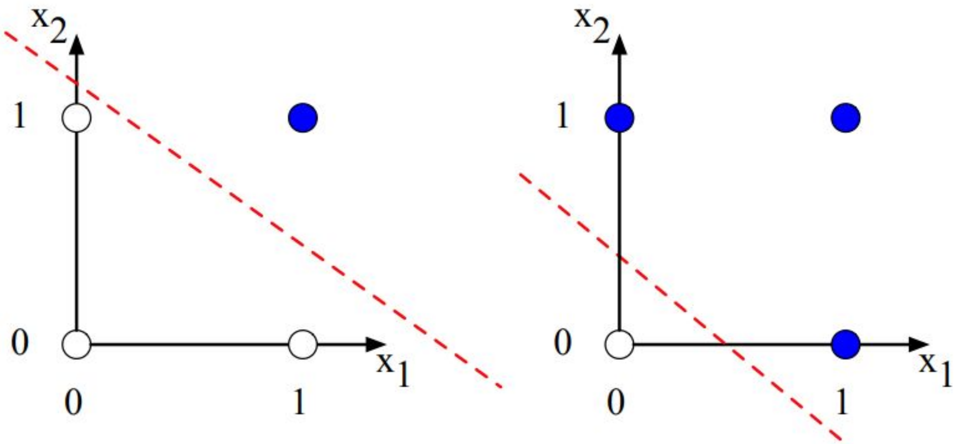
Source: https://en.wikipedia.org/wiki/Perceptron

Logical operators

AND			OR			XOR		
x1	x2	y	x1	x2	y	x1	x2	y
0	0	0	0	0	0	0	0	0
0	1	0	0	1	1	0	1	1
1	0	0	1	0	1	1	0	1
1	1	1	1	1	1	1	1	0

Jurafsky & Martin, chapter 7:
https://web.stanford.edu/~jurafsky/slp3/7.pdf

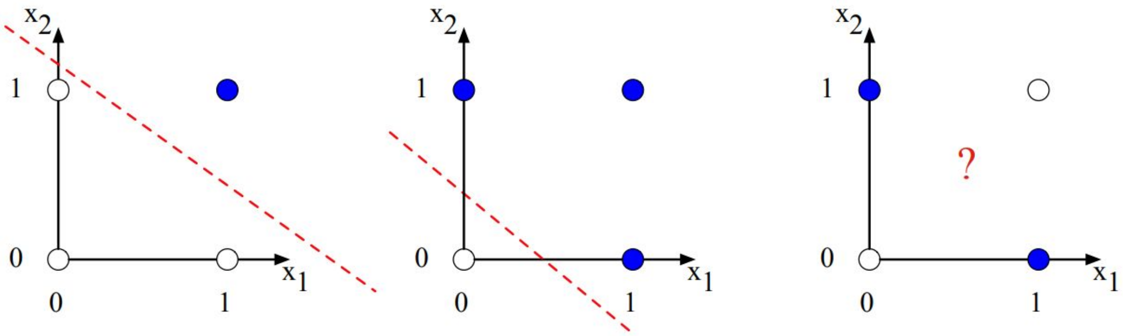
AND, OR



a) x_1 AND x_2 b) x_1 OR x_2

Jurafsky & Martin, chapter 7:
https://web.stanford.edu/~jurafsky/slp3/7.pdf

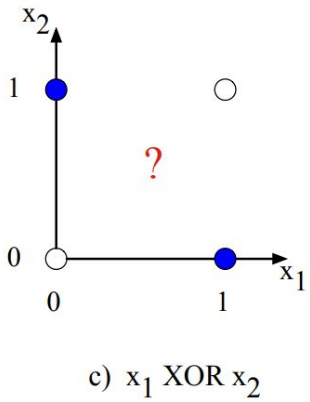
AND, OR, XOR



a) x_1 AND x_2 b) x_1 OR x_2 c) x_1 XOR x_2

Jurafsky & Martin, chapter 7:
https://web.stanford.edu/~jurafsky/slp3/7.pdf

AND, OR, XOR



Jurafsky & Martin, chapter 7:
https://web.stanford.edu/~jurafsky/slp3/7.pdf

How do you solve the XOR problem?

How do you solve the XOR problem?

Nonlinearity + Hidden layers!

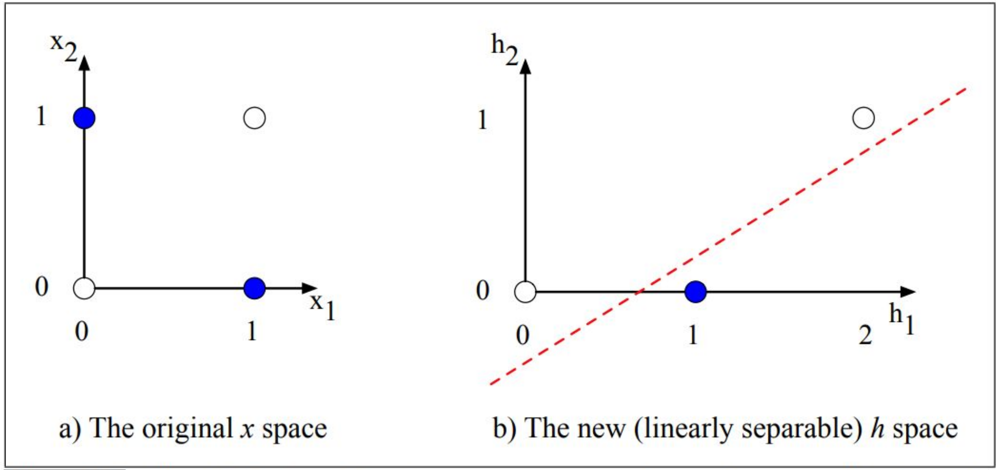


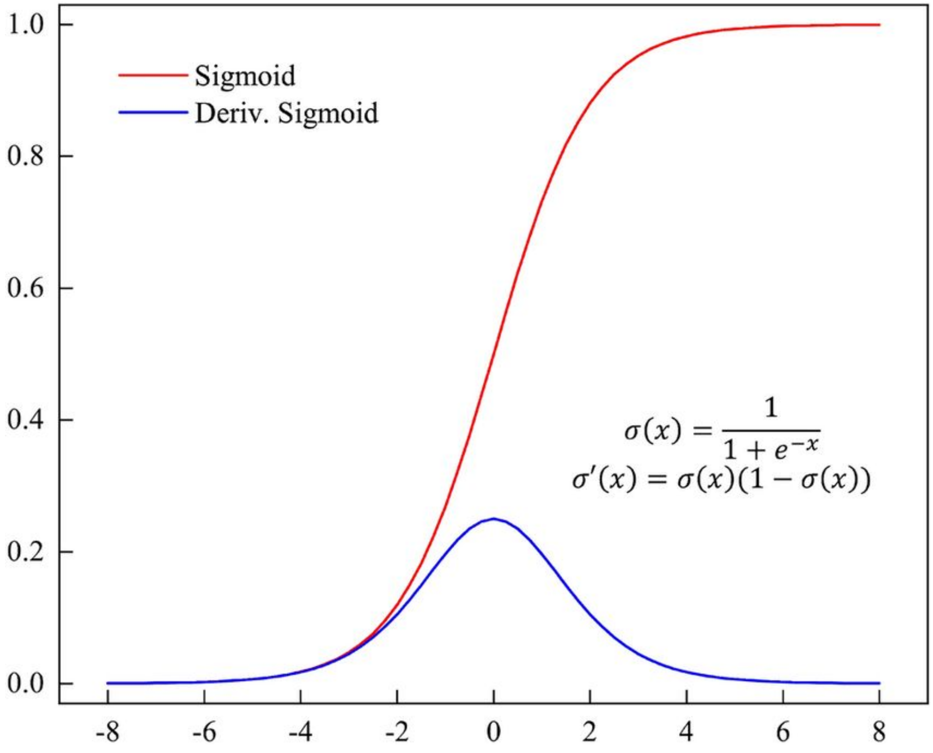
Figure 7.7 The hidden layer forming a new representation of the input. (b) shows the representation of the hidden layer, **h**, compared to the original input representation **x** in (a). Notice that the input point [0, 1] has been collapsed with the input point [1, 0], making it possible to linearly separate the positive and negative cases of XOR. After Goodfellow et al. (2016).

Jurafsky & Martin, chapter 7:
https://web.stanford.edu/~jurafsky/slp3/7.pdf

Activation functions

Sigmoid

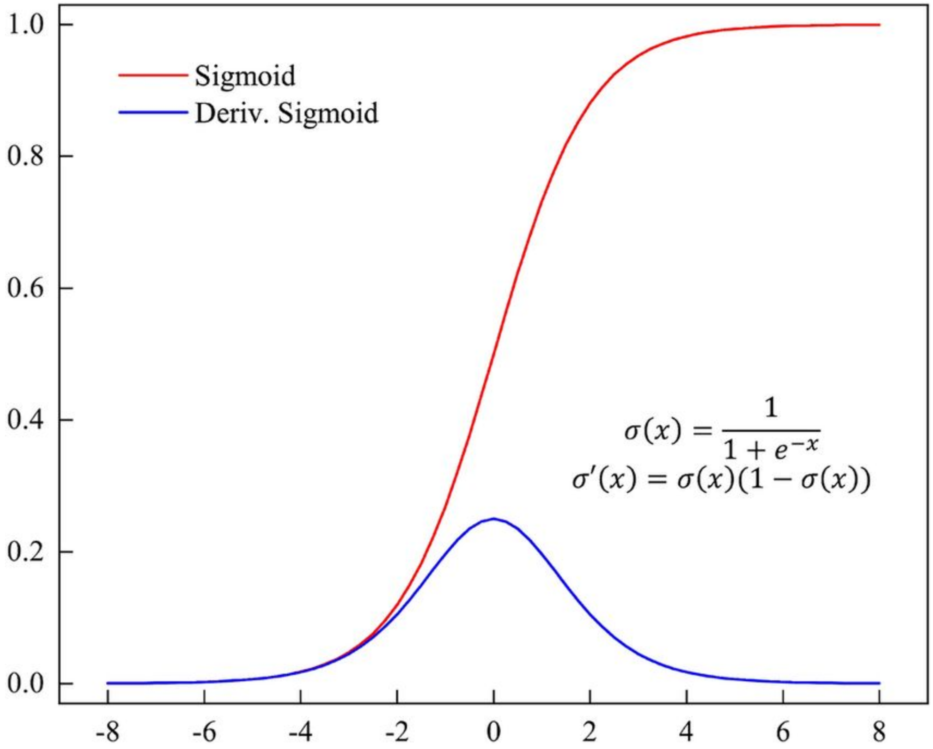
Properties of Sigmoid



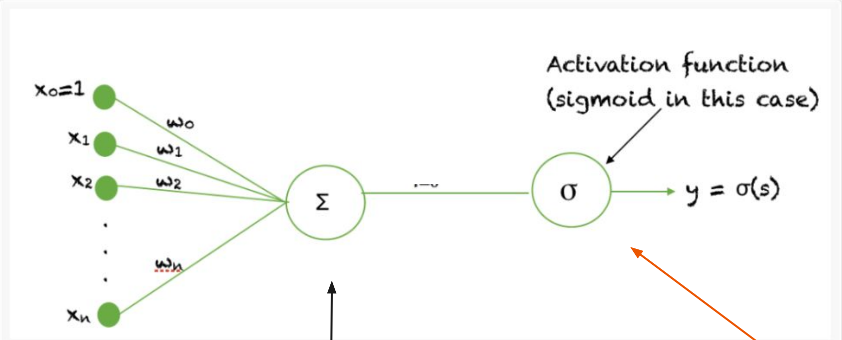
Sigmoid

Properties of Sigmoid

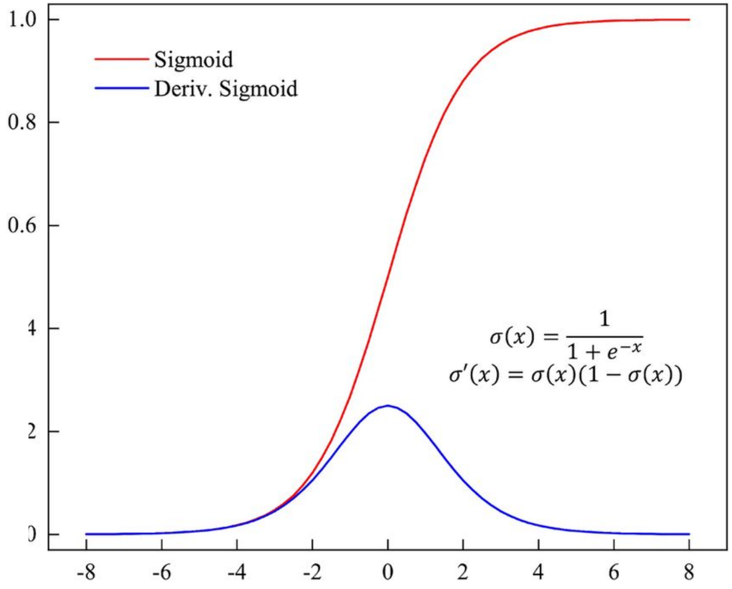
- Range (0, 1)
- Linear around 0
- Outliers squashed to 0 and 1
- ...



Sigmoid: calculation



A sigmoid unit in a neural network



So now you have a zero here

Your goal: calculate this!

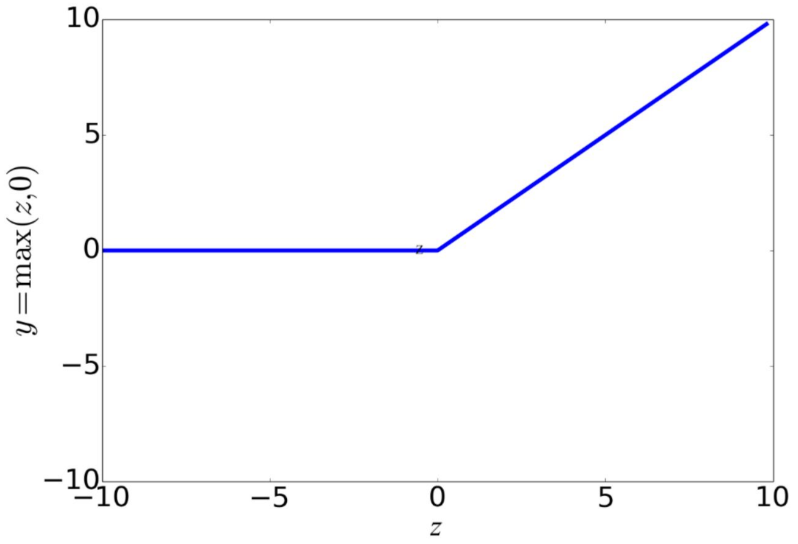
image source(left): https://machinelearningmastery.com/a-gentle-introduction-to-sigmoid-function/

ReLU

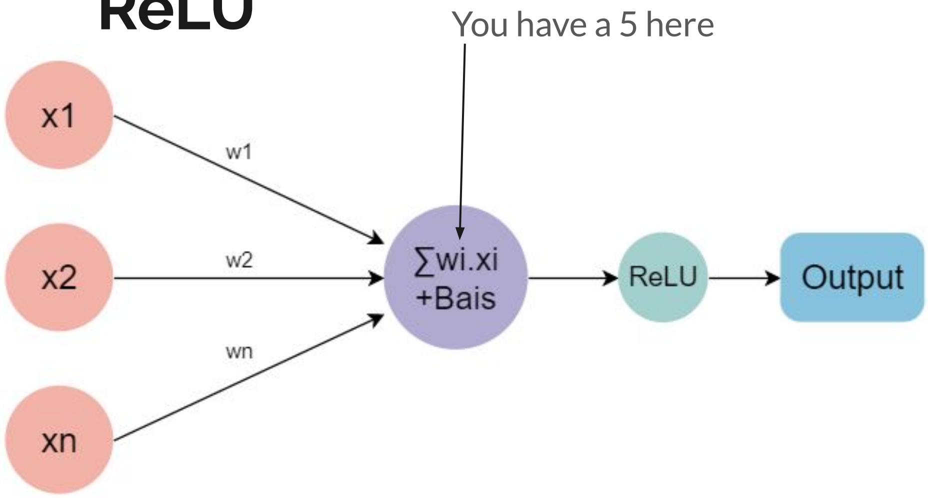
Properties of ReLU

- Range [0, +inf)

ReLU: Rectified Linear Unit



ReLU



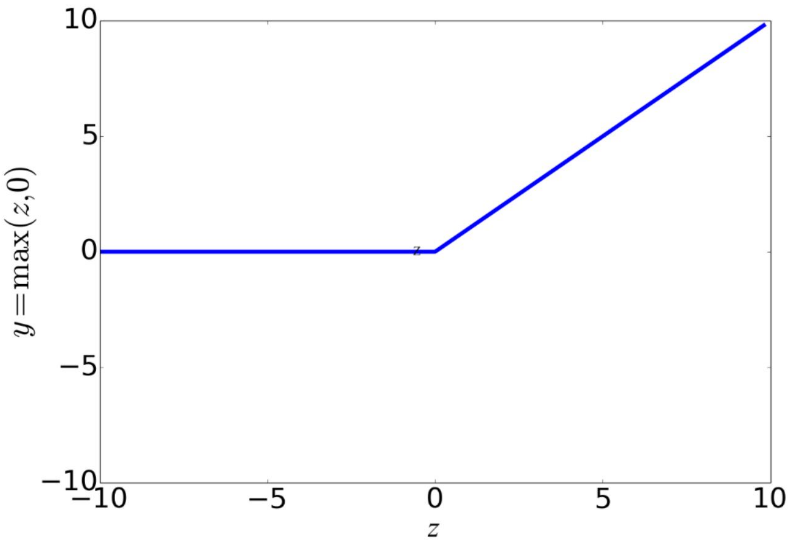
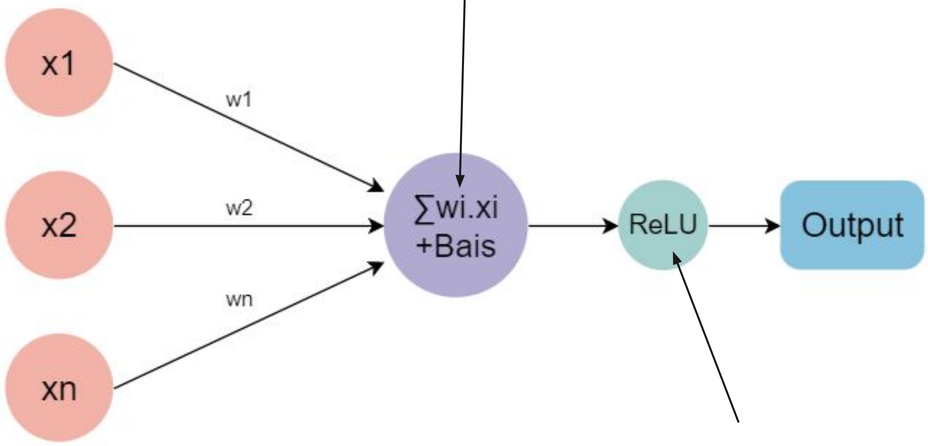


Image source(left): https://medium.com/@MrBam44/the-dying-relu-problem-causes-and-solutions-1a970f177c

ReLU

You have a 5 here



What is the output here?

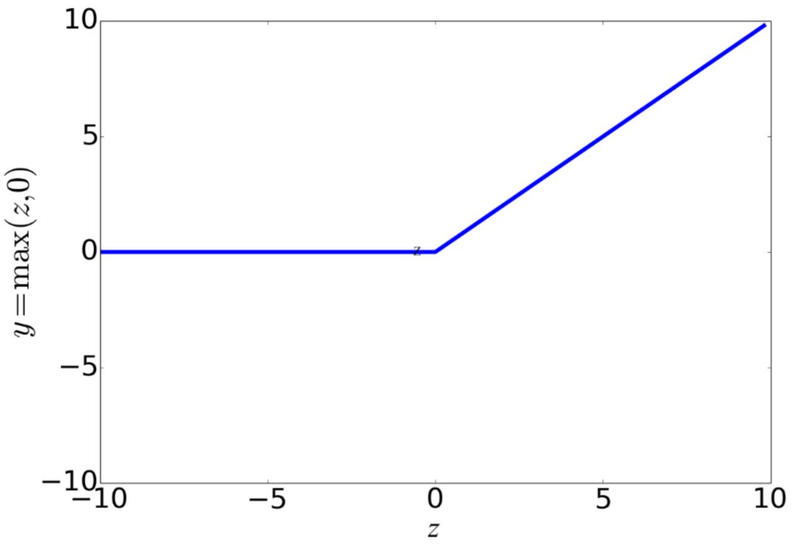
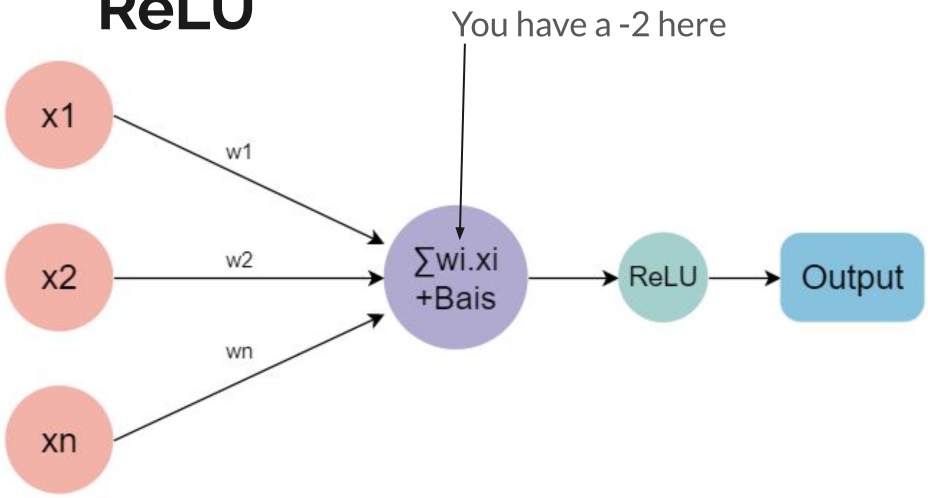


Image source(left): https://medium.com/@MrBam44/the-dying-relu-problem-causes-and-solutions-1a970f177c

ReLU



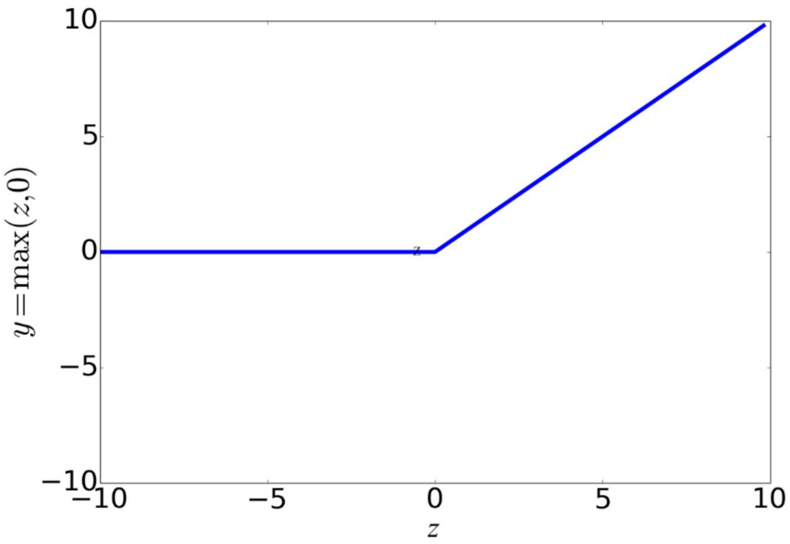
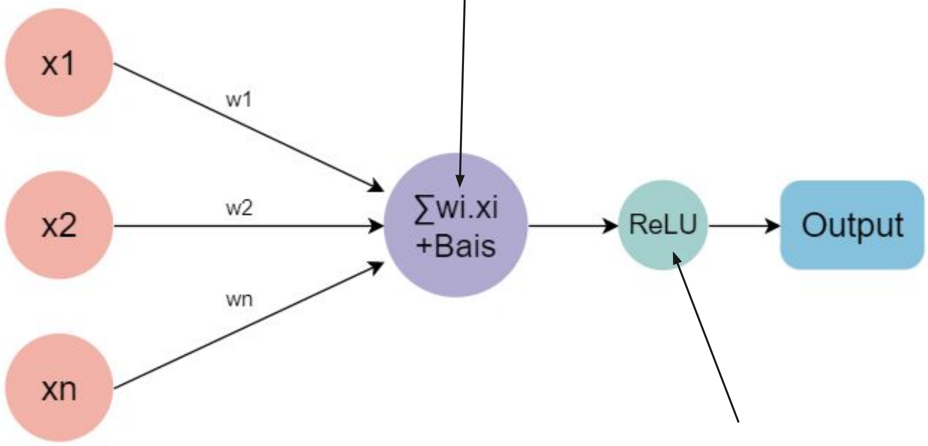


Image source(left): https://medium.com/@MrBam44/the-dying-relu-problem-causes-and-solutions-1a970f177c

ReLU

You have a -2 here



What is the output here?

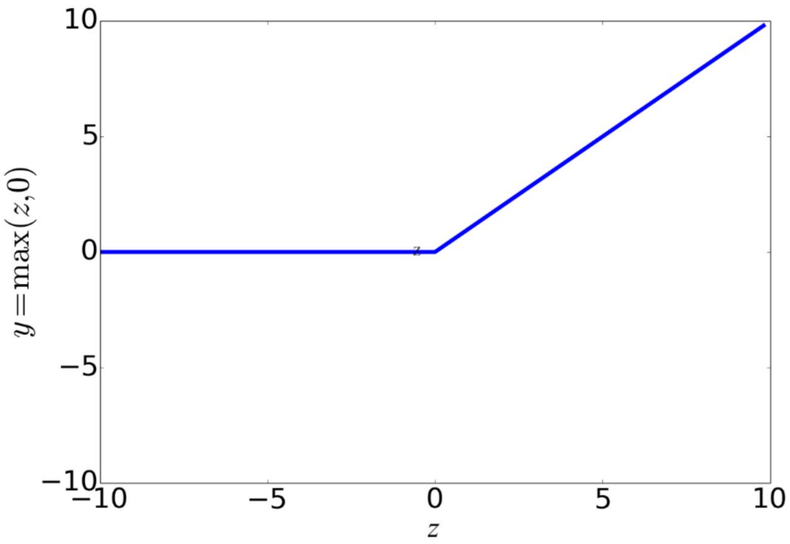
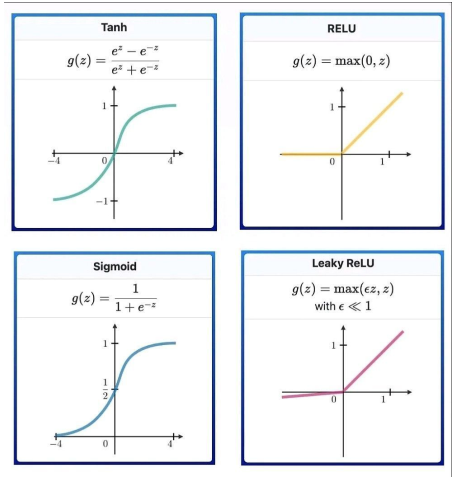


Image source(left): https://medium.com/@MrBam44/the-dying-relu-problem-causes-and-solutions-1a970f177c

More



And a lot more...

Image source: https://ai-artificial-intelligence.webyes.com.br/most-used-activation-functions-in-neural-networks/

Loss function

Why is it needed?

The training procedure:

- we use our model to make predictions on some input data,
- compare those predictions to the actual observed outcomes (or labels, targets, ground truths)
- use a loss (or cost) function to measure how "wrong" our model's predictions are
- parameter update: Based on the loss, we adjust the model's parameters to try to reduce this loss in future predictions
- repeat the process for certain repetition / some result has been achieved / model no longer improves

Cross-Entropy Loss

For binary classification $L_{CE}(\hat{y}, y) = -\log p(y|x) \;=\; -[y \log \hat{y} + (1-y)\log(1-\hat{y})]$

More: https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html

Calculation exercise

For binary classification $L_{CE}(\hat{y}, y) = -\log p(y|x) = \boxed{-[y \log \hat{y} + (1-y)\log(1-\hat{y})]}$

Now your label is 1, and your prediction is 0.9, what should the above formula be?

(just plug in the numbers)

Cross-Entropy Loss

For binary classification $L_{CE}(\hat{y}, y) = -\log p(y|x) \;=\; -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$

Now your label is 1, and your prediction is 0.9, what should the above formula be?

(just plug in the numbers)

Cross-Entropy Loss

For binary classification $L_{CE}(\hat{y}, y) = -\log p(y|x) \;=\; \boxed{-[y \log \hat{y} + (1-y)\log(1-\hat{y})]}$

Now your label is 1, and your prediction is 0.9, what should the above formula be?

-1*log0.9 + (1-1)*log(1-0.9) = -1 * log0.9 + 0 = -log0.9

Partial Derivatives

Calculate the following derivatives...

$f(x) = 3x^2$

$f(x) = ax^2$, where a is a constant

$f(y) = ay$, where a is a constant

$f(y) = a^2y$, where a is a constant

Calculate the following derivatives...

$f(x) = 3x^2$ — $f'(x) = 3 \cdot 2x = 6x$

$f(x) = ax^2$, where a is a constant — $f'(x) = a \cdot 2x = 2ax$

$f(y) = ay$, where a is a constant — $f'(y) = a$

$f(y) = a^2y$, where a is a constant — $f'(y) = a^2$

Calculate the following derivatives…

$f(x) = 3x^2$	$f'(x) = 3 \cdot 2x = 6x$
$f(x) = ax^2$, where a is a constant	$f'(x) = a \cdot 2x = 2ax$
$f(y) = ay$, where a is a constant	$f'(y) = a$
$f(y) = a^2y$, where a is a constant	$f'(y) = a^2$
$f(x, y) = 3x^2y$, w.r.t. x and y	?

Partial derivatives

How to write the partial derivatives for: $f(x, y) = 3x^2y$

The partial derivative with respect to x

The partial derivative with respect to y

Partial derivatives

$f(x, y) = 3x^2y$

The partial derivative with respect to x is written $\frac{\partial}{\partial x}3x^2y$

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

Partial derivatives walkthrough

f(x, y) = $3x^2y$

The partial derivative with respect to x is written $\frac{\partial}{\partial x}3x^2y$ What are the constants?

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

Partial derivatives walkthrough

f(x, y) = $3x^2y$

The partial derivative with respect to x is written $\frac{\partial}{\partial x}3x^2y$

What are the constants?

3, y

$$\frac{\partial}{\partial x}3x^2y = 3y\frac{\partial}{\partial x}x^2$$

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

Partial derivatives walkthrough

f(x, y) = $3x^2y$

The partial derivative with respect to x is written $\frac{\partial}{\partial x}3x^2y$

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

$$\frac{\partial}{\partial x}3x^2y = 3y\frac{\partial}{\partial x}x^2$$

$$\frac{\partial}{\partial x}3x^2y = 3y\frac{\partial}{\partial x}x^2 = 3y \times 2x = 6xy$$

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is $6xy$

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is 6xy

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

What are the constants here?

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is 6xy

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

What are the constants here?
$3, x^2$

$$\frac{\partial}{\partial y}3x^2y = 3x^2\frac{\partial}{\partial y}y$$

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is 6xy

The partial derivative with respect to y is written $\frac{\partial}{\partial y}3x^2y$

$$\frac{\partial}{\partial y}3x^2y = 3x^2\frac{\partial}{\partial y}y$$

$$\frac{\partial}{\partial y}3x^2y = 3x^2\frac{\partial}{\partial y}y = 3x^2$$

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is 6xy

The partial derivative with respect to y is $3x^2$

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is $6xy$

The partial derivative with respect to y is $3x^2$

Rewrite them!

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is 6xy

The partial derivative with respect to y is $3x^2$

$$\frac{\partial}{\partial x}3x^2y = 6xy$$
$$\frac{\partial}{\partial y}3x^2y = 3x^2$$

Next task: put them together

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

The partial derivative with respect to x is 6xy

The partial derivative with respect to y is $3x^2$

$$\boxed{\begin{aligned}\frac{\partial}{\partial x}3x^2y &= 6xy\\ \frac{\partial}{\partial y}3x^2y &= 3x^2\end{aligned}}$$

$$\left[\frac{\partial}{\partial x}3x^2y \quad \frac{\partial}{\partial y}3x^2y\right]$$

Partial derivatives walkthrough

$f(x, y) = 3x^2y$

~~The partial derivative with respect to x is~~ 6xy

~~The partial derivative with respect to y is~~ $3x^2$

Plug it back...

$$\frac{\partial}{\partial x}3x^2y = 6xy$$
$$\frac{\partial}{\partial y}3x^2y = 3x^2$$

$$\left[\frac{\partial}{\partial x}3x^2y \quad \frac{\partial}{\partial y}3x^2y\right]$$

Partial derivatives walkthrough

$$\left[\tfrac{\partial}{\partial x}3x^2y \quad \tfrac{\partial}{\partial y}3x^2y\right] = \left[6xy \quad 3x^2\right]$$

Plug it back...

$$\left[\frac{\partial}{\partial x}3x^2y \quad \frac{\partial}{\partial y}3x^2y\right]$$

Partial derivatives walkthrough

$$\begin{bmatrix} \frac{\partial}{\partial x}\boxed{3x^2y} & \frac{\partial}{\partial y}\boxed{3x^2y} \end{bmatrix} = \begin{bmatrix} 6xy & 3x^2 \end{bmatrix}$$

Now let's write this $3x^2y$ as f(x, y)

Partial derivatives walkthrough

$$\nabla f(x,y) = \left[\frac{\partial}{\partial x}\boxed{f(x,y)} \quad \frac{\partial}{\partial y}\boxed{f(x,y)}\right] = \begin{bmatrix} 6xy & 3x^2 \end{bmatrix}$$

Done! We represented it by f(x,y)

Partial derivatives walkthrough

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) & \frac{\partial}{\partial y} f(x,y) \end{bmatrix} = \begin{bmatrix} 6xy & 3x^2 \end{bmatrix}$$

nabla symbol

Partial derivatives walkthrough

$$\boxed{\nabla f(x,y)} = \left[\tfrac{\partial}{\partial x} f(x,y) \quad \tfrac{\partial}{\partial y} f(x,y)\right] = \begin{bmatrix} 6xy & 3x^2 \end{bmatrix}$$

Denotes the gradient
(the gradient in "gradient descent")