Lecture - 2

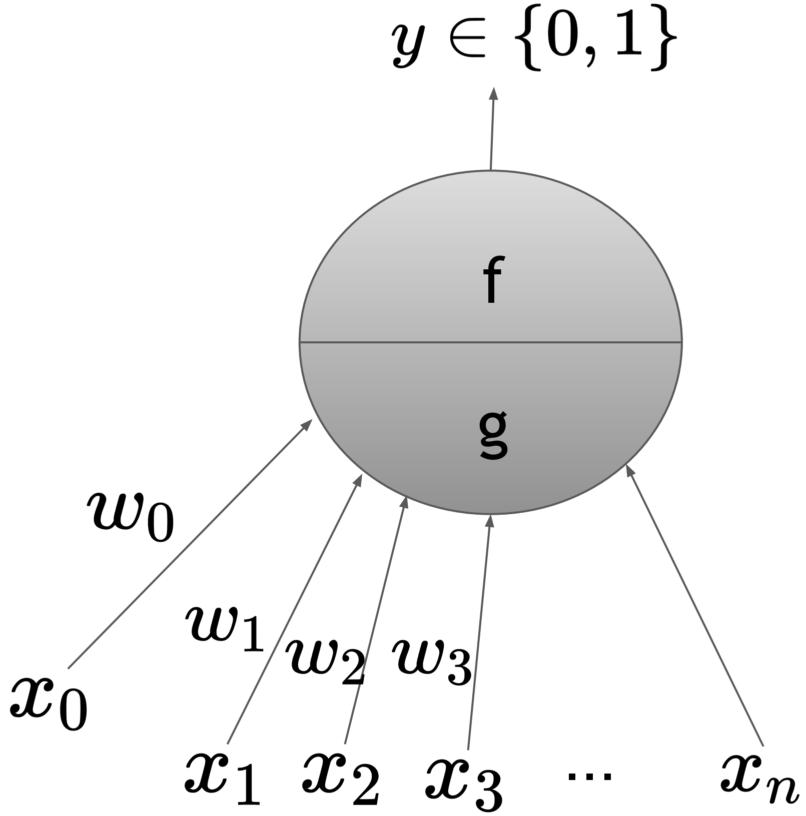
Neural Networks

Outline (today)

- Module 1: Perceptron Learning Algorithm
- Module 2: Network of Perceptron
- Module 3: Sigmoid neuron

Module - 1: Perceptron Learning Algorithm

Perceptron



$$y = 1 \ \ if \ \ \sum_1^n w_i x_i - \theta >= 0$$

$$y = 0 \ \ if \ \ \sum_1^n w_i x_i - \theta < 0$$

$$\boxed{w_0 = -\theta \ \ and \ \ x_0 = 1}$$

$$y = 1 \, if \, \sum_0^n w_i x_i >= 0$$

$$y = 0 \, if \, \sum_0^n w_i x_i < 0$$

Example:

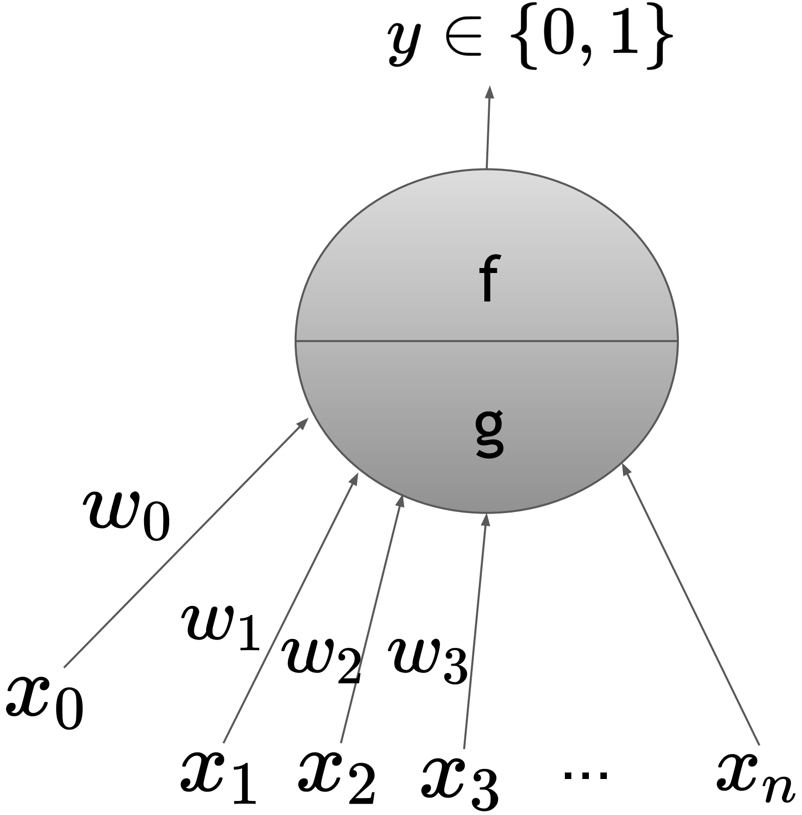
x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 + w_1 x_1 + w_2 x_2 < 0$
0	1	0	$w_0 + w_1 x_1 + w_2 x_2 < 0$
1	0	0	$w_0 + w_1 x_1 + w_2 x_2 < 0$
1	1	1	$w_0 + w_1 x_1 + w_2 x_2 >= 0$

Example:

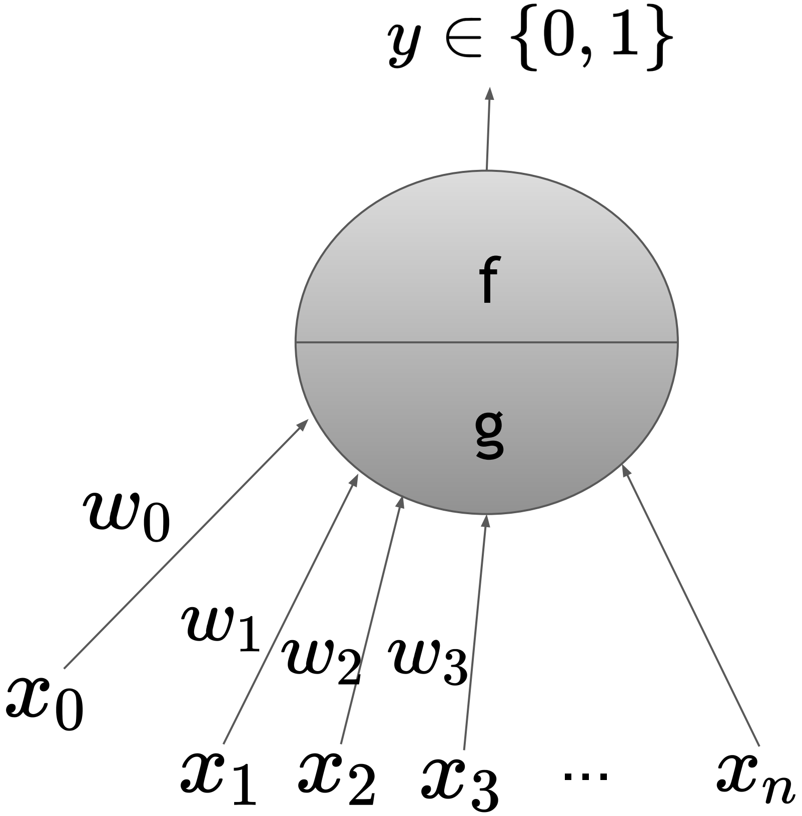
x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

What to learn?

What to learn?

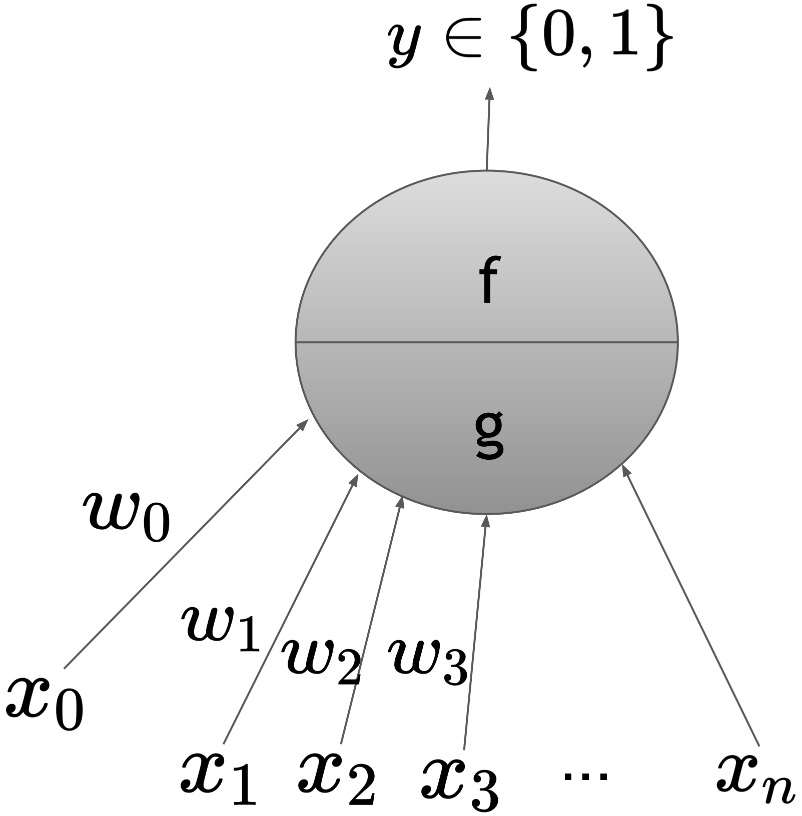


What to learn?



Right w's

What to learn?



Right w's

What is right w?

Example:

x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

If I choose $w_0 = 0, w_1 = 0, w_2 = 0$, how many samples I misclassify?

Example:

x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

If I choose $w_0 = 0, w_1 = 0, w_2 = 0$, how many samples I misclassify? 3

Example:

x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

If I choose $w_0 = -1, w_1 = -1, w_2 = 0$, how many samples I misclassify?

Example:

x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

If I choose $w_0 = -1, w_1 = -1, w_2 = 0$, how many samples I misclassify? 1

Example:

x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

If I choose $w_0 = -1, w_1 = 0.9, w_2 = 0.9$, how many samples I misclassify?

Example:

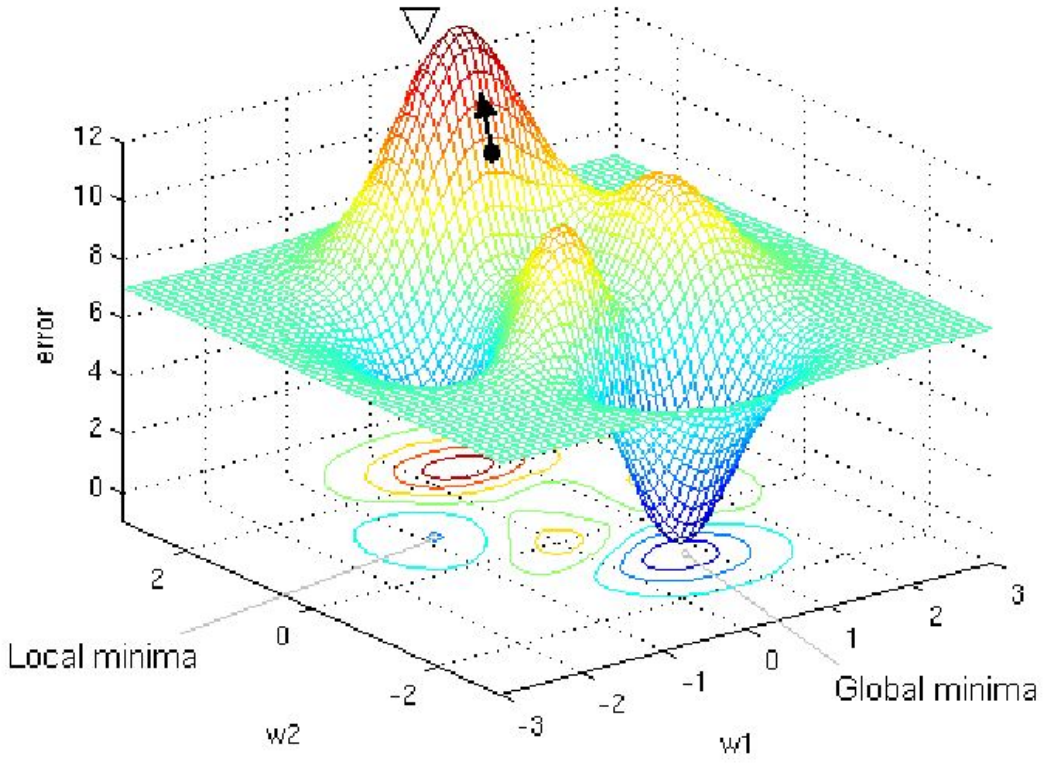
x_1	x_2	y = x_1 AND x_2	Conditions
0	0	0	$w_0 < 0$
0	1	0	$w_0 + w_2 < 0$
1	0	0	$w_0 + w_1 < 0$
1	1	1	$w_0 + w_1 + w_2 >= 0$

If I choose $w_0 = -1, w_1 = 0.9, w_2 = 0.9$, how many samples I misclassify? 0

Error surface

w_0	w_1	w_2	#misclassification
0	0	0	3
-1	-1	0	1
-1	0.9	0.9	0
.	.	.	.
.	.	.	.

Error surface



Perceptron Learning Algorithm

P $\leftarrow$ Set of Positive Samples (y=1)

N $\leftarrow$ Set of Negative Samples (y=0)

$\mathbf{w} \leftarrow [w_0, w_1, w_2, \cdots, w_n]$ (randomly)

While !Convergence

Do

………….

Done

Perceptron Learning Algorithm

P ← Set of Positive Samples (y=1)

$\boxed{\text{N} \leftarrow \text{Set of Negative Samples (y=0)}}$

$\mathbf{w} \leftarrow [w_0, w_1, w_2, \cdots, w_n]$ (randomly)

While !Convergence

Do

....

Done

Perceptron Learning Algorithm

P $\leftarrow$ Set of Positive Samples (y=1)

N $\leftarrow$ Set of Negative Samples (y=0)

w $\leftarrow$ $\left[w_0, w_1, w_2, \cdots, w_n\right]$ (randomly)

While !Convergence

Do

…………

Done

Perceptron Learning Algorithm

While !Convergence

Do

$for\ x \in P \cup N\{$

If $x \in P\ and\ \sum_{i=0}^{n} w_i x_i < 0\ then$//Positive misclassified as negative

$$w = w + x$$

If $x \in N\ and\ \sum_{i=0}^{n} w_i x_i >= 0\ then$ //Negative misclassified as positive

$$w = w - x$$

}

Done

Linear Algebraic Interpretation

$$\sum_{i=0}^{n} w_i x_i = \mathbf{w}^T \mathbf{x}$$

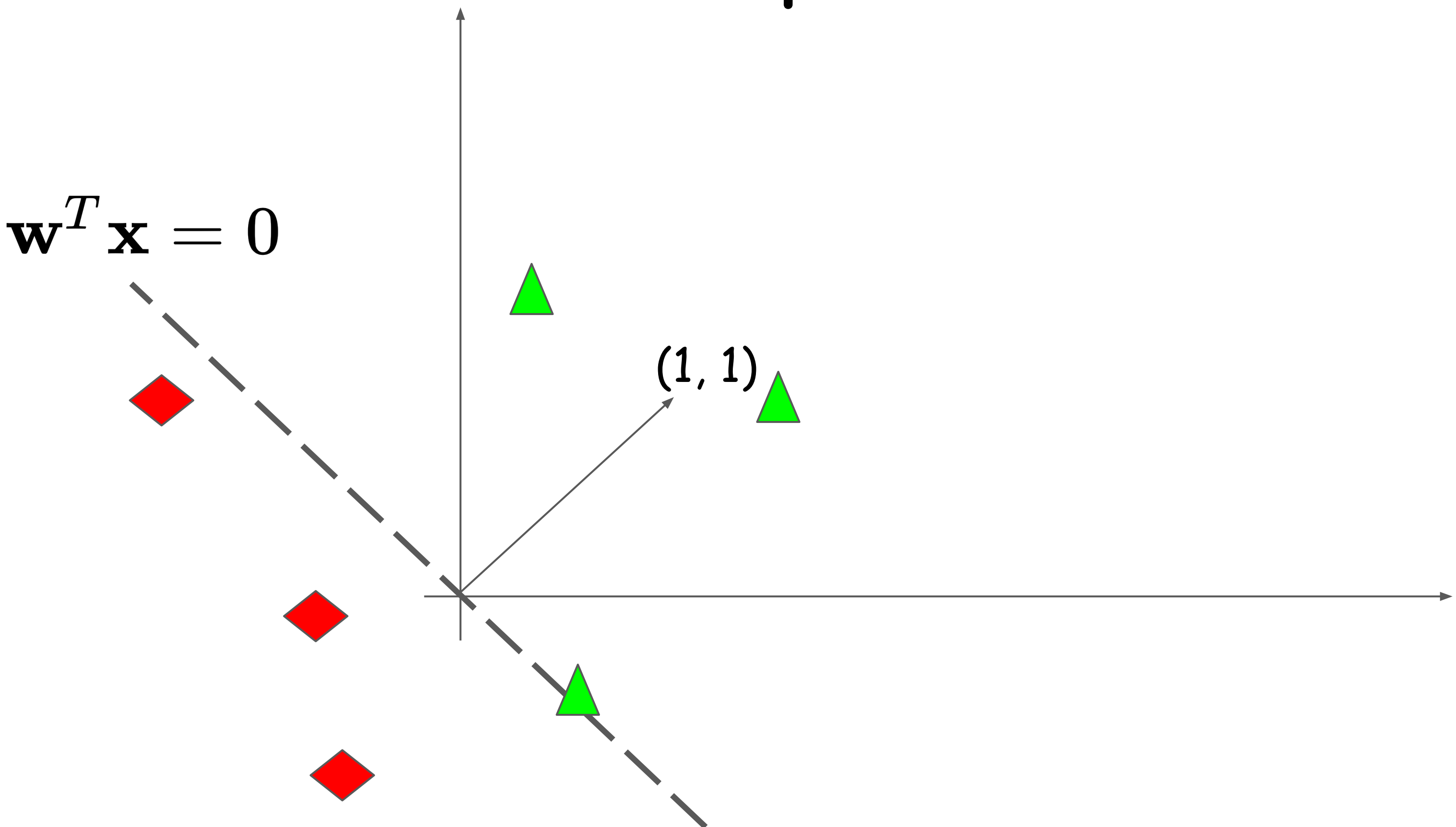
Where

$$\mathbf{w} = \begin{bmatrix} 1 & w_1 & w_2 & \cdots & w_n \end{bmatrix}$$

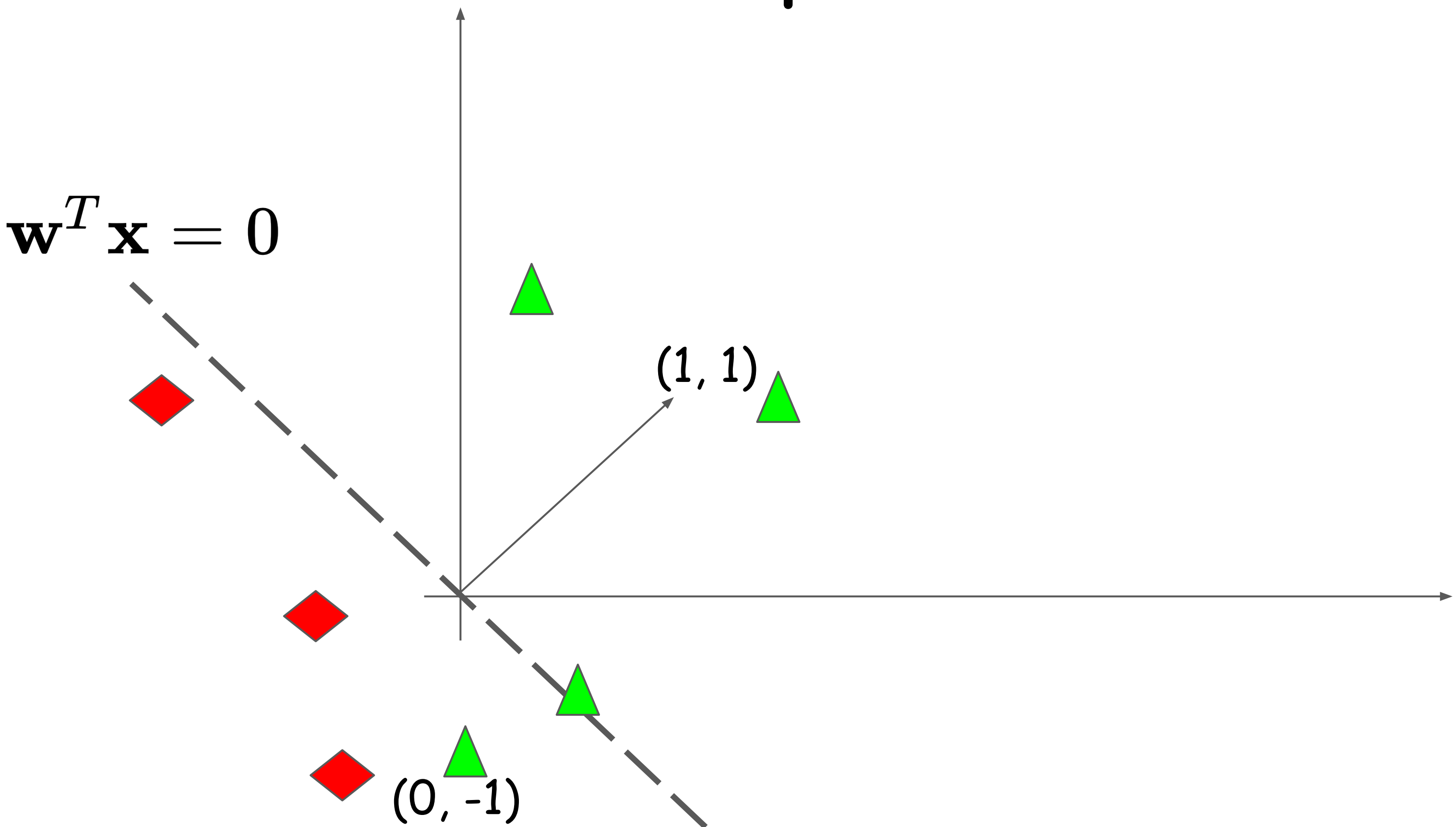
$$\mathbf{x} = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Decision boundary: $\mathbf{w}^T \mathbf{x} = 0$

Geometric Interpretation



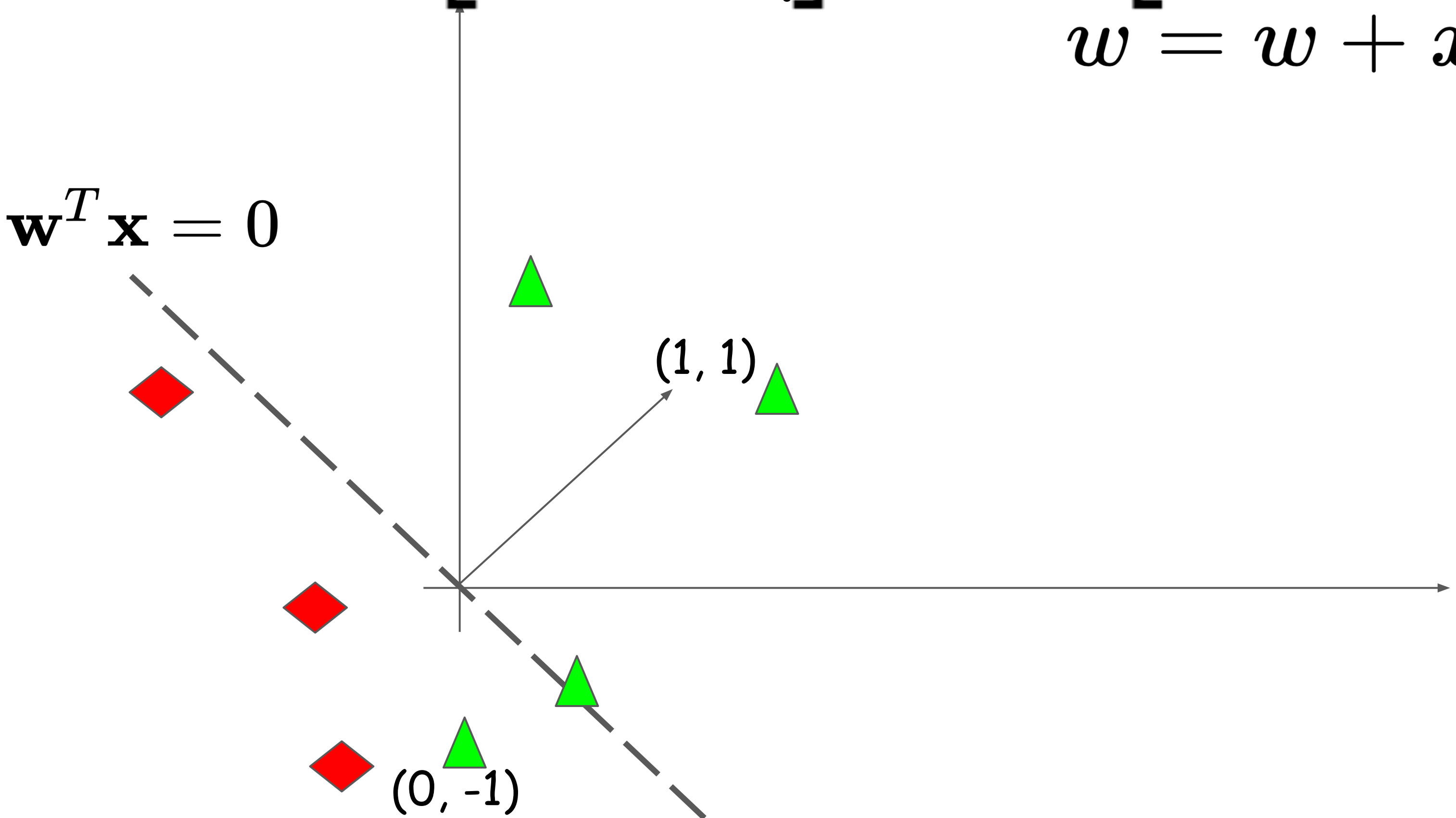
Geometric Interpretation



$$w = [1, 1] + [0, -1]$$

Geometric Interpretation

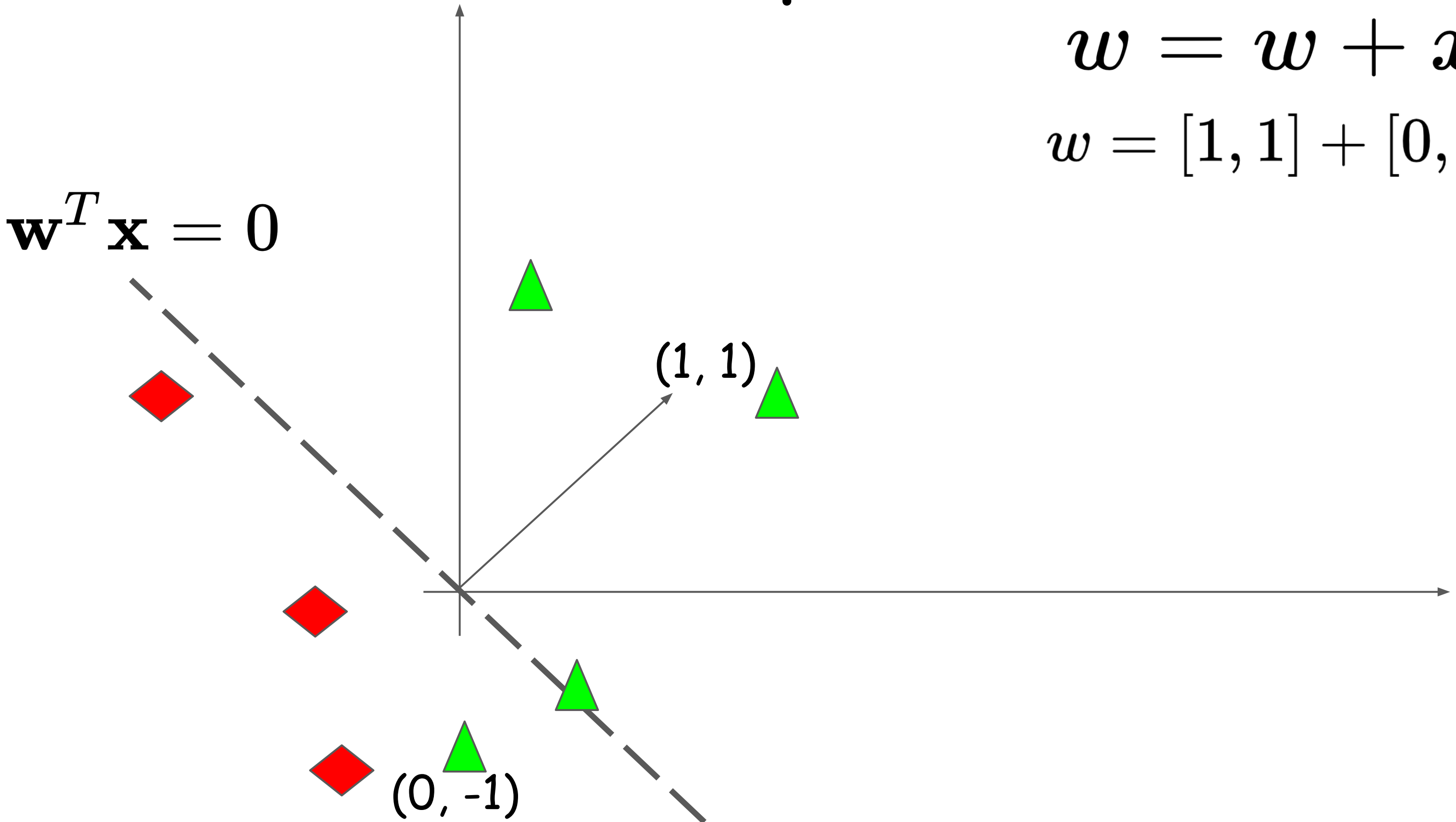
$$w = w + x$$



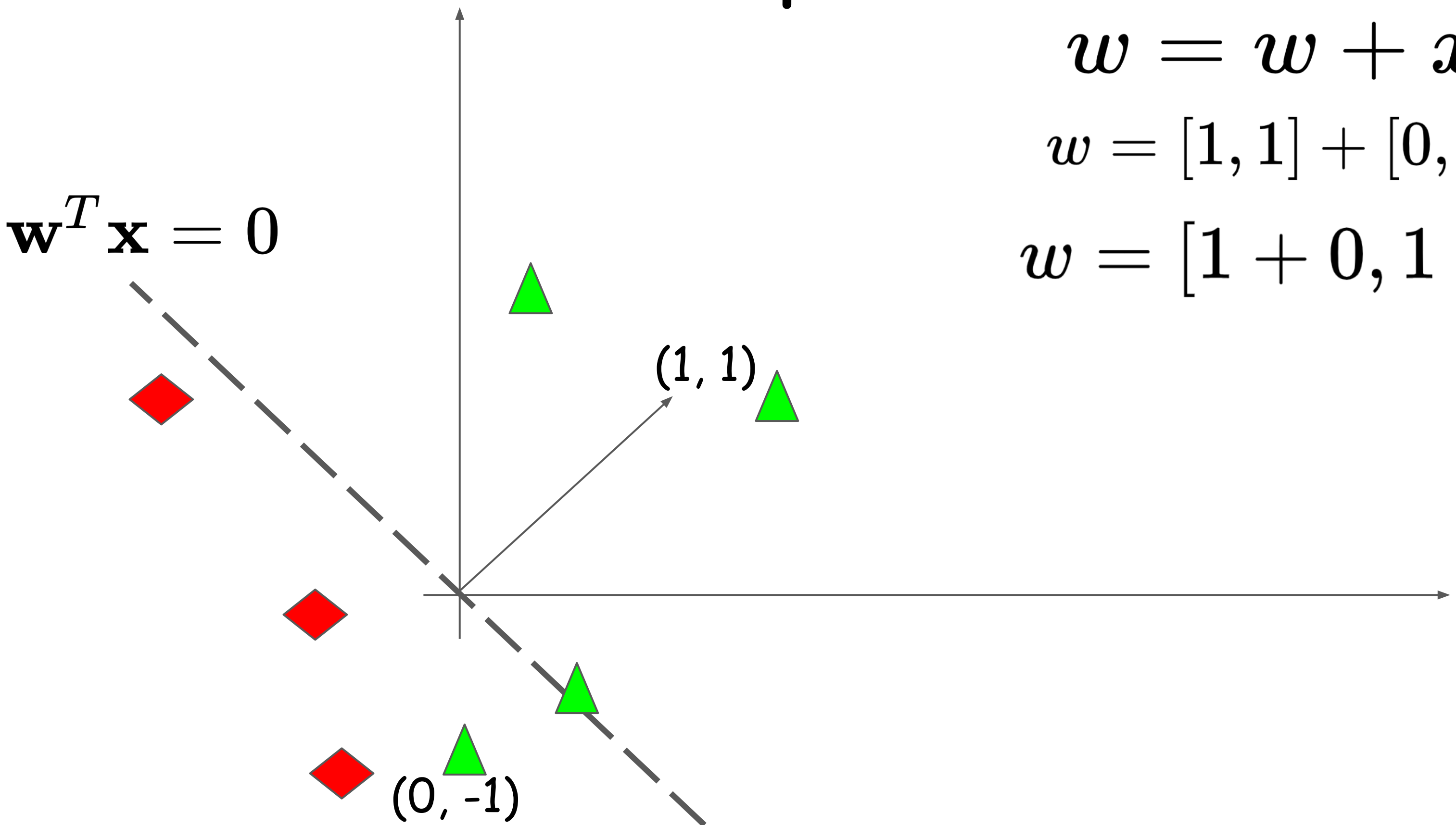
Geometric Interpretation

$$w = w + x$$

$$w = [1, 1] + [0, -1]$$



Geometric Interpretation



$$w = w + x$$

$$w = [1, 1] + [0, -1]$$

$$w = [1 + 0, 1 - 1]$$

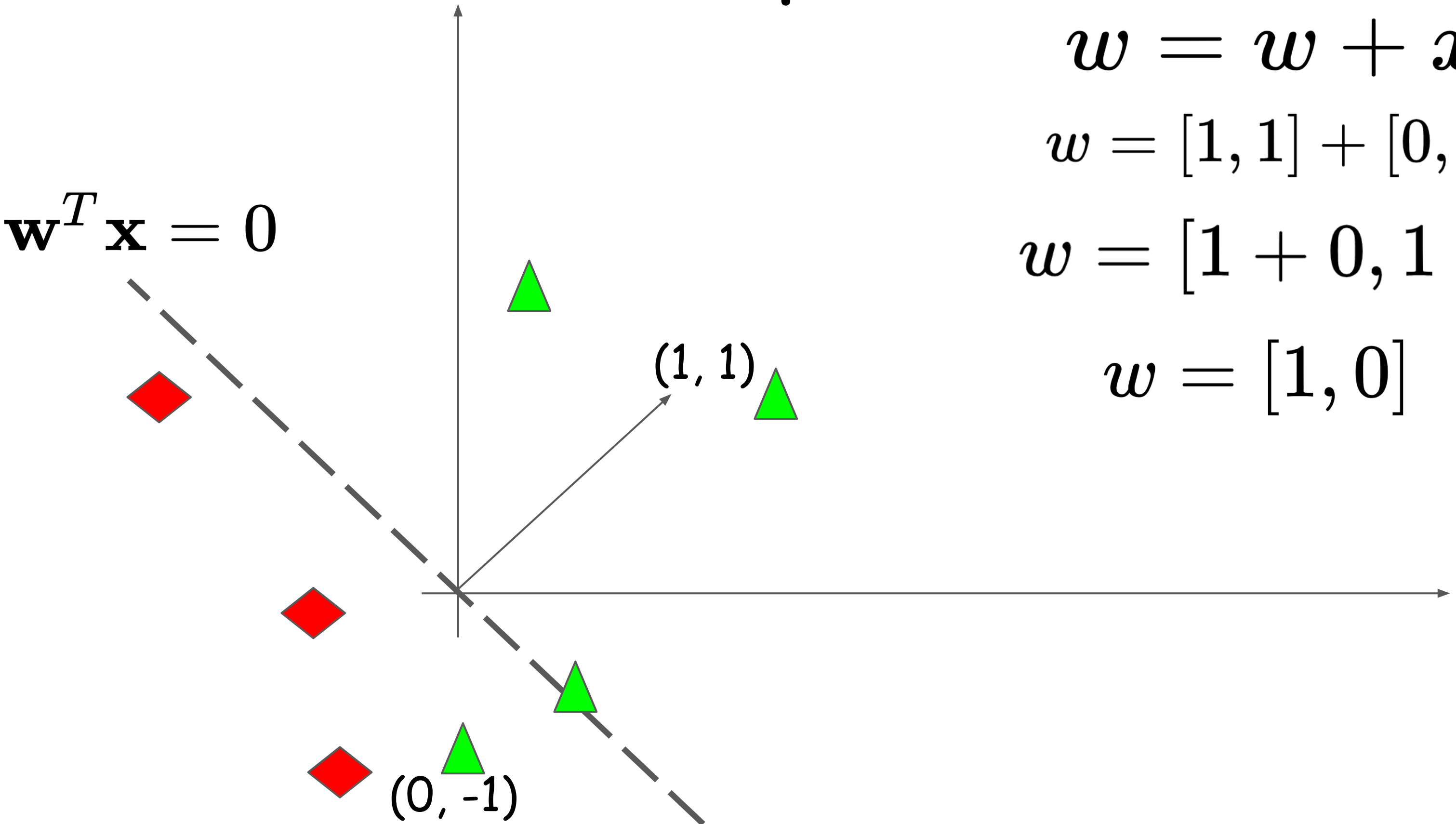
Geometric Interpretation

$$w = w + x$$

$$w = [1, 1] + [0, -1]$$

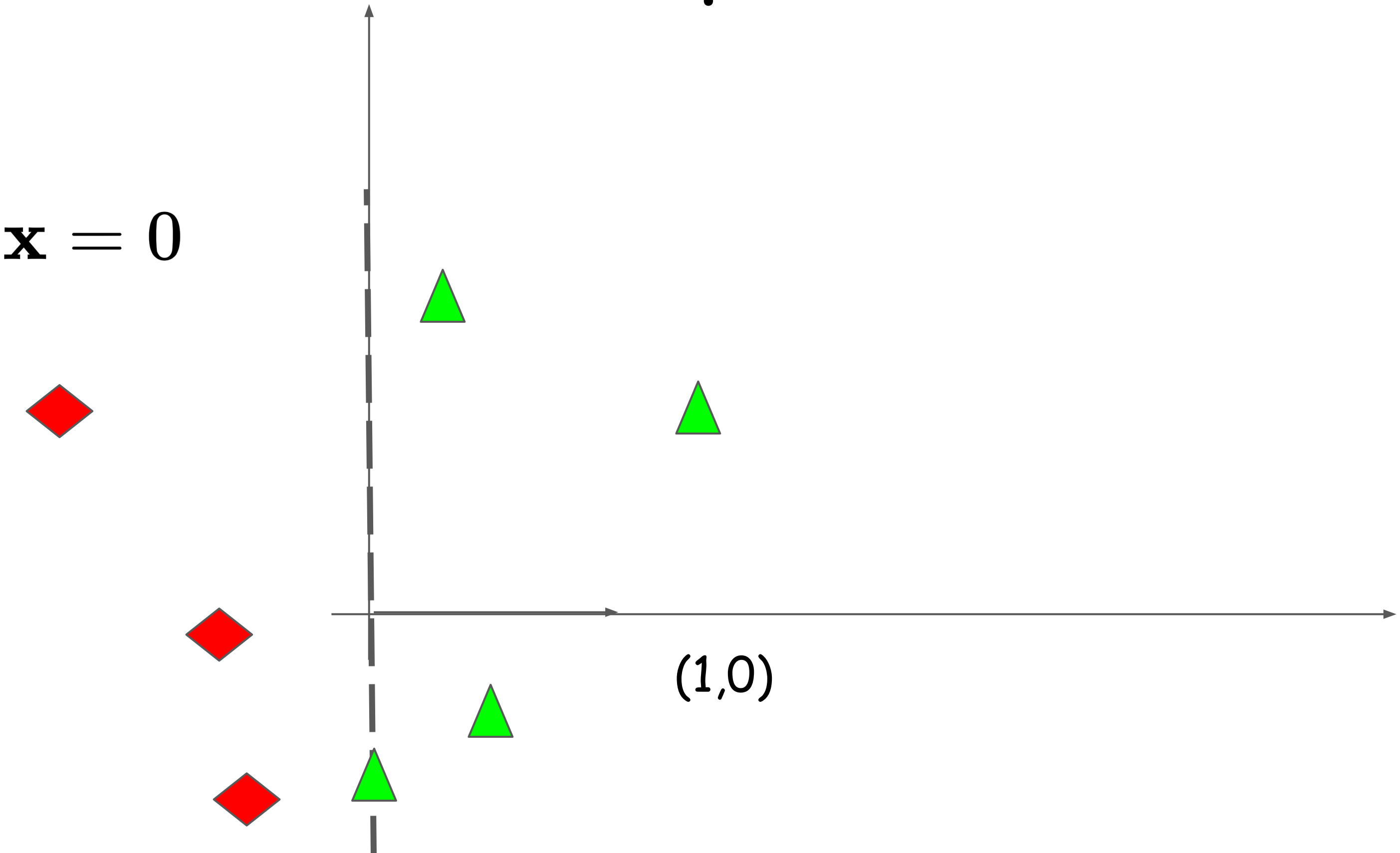
$$w = [1 + 0, 1 - 1]$$

$$w = [1, 0]$$



Geometric Interpretation

$$\mathbf{w}^T\mathbf{x} = 0$$

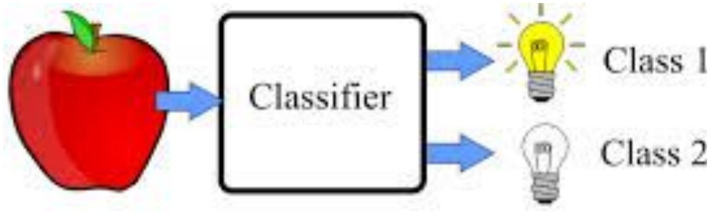


Perceptron Learning algorithm

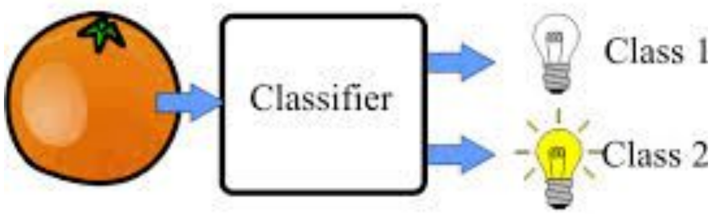
Demo:
https://www.cs.utexas.edu/~teammco/misc/perceptron/

Let us code!

Task: Apple vs Orange classification



$$y = 0$$
$$y = 1$$

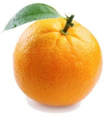


$$y = 0$$
$$y = 1$$

Let us code!

Task: Apple vs Orange classification

Sample	Redness	Weight
	171	80
	175	78
	180	90

Sample	Redness	Weight
	100	60
	99	65
	102	59

Let us code!

In Google Colab ...

Definitions

Linearly separable: Two sets of points P and N in an n-D space are called linearly separable if $\exists\ (w_0, w_1, w_2, \cdots, w_n) \in \mathcal{R}^n$ such that

$$\sum_{i=1}^{n} w_i x_i - w_0 >= 0 \ \ \forall (x_0, x_1, x_2, \cdots, x_n) \in P$$

$$\sum_{i=1}^{n} w_i x_i - w_0 < 0 \ \ \forall (x_0, x_1, x_2, \cdots, x_n) \in N$$

Convergence

If P and N are finite and linearly separable then the perceptron learning algorithm updates the weight vector a finite number of times.

Proof: on paper and pen.

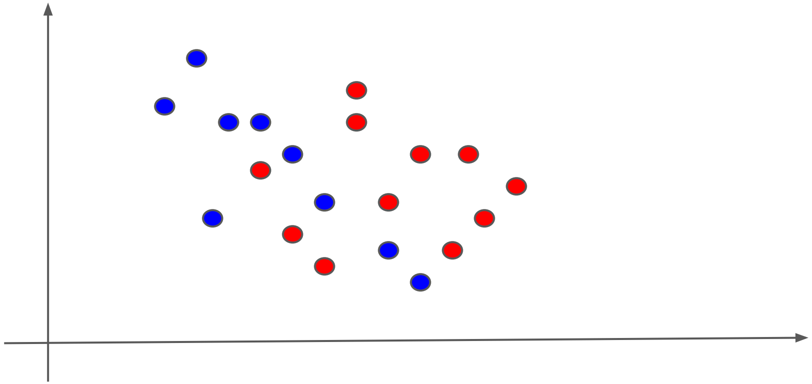
Summary so far...

- We can weight inputs.
- We can classify linearly separable samples.

What about non-linearly separable samples?

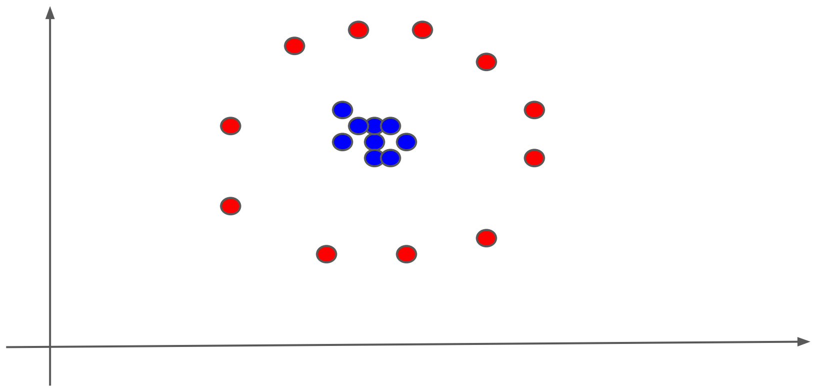
Summary so far...

- We can weight inputs.
- We can classify linearly separable samples.



Summary so far...

- We can weight inputs.
- We can classify linearly separable samples.



Module 2: A network of Perceptron

Theorem

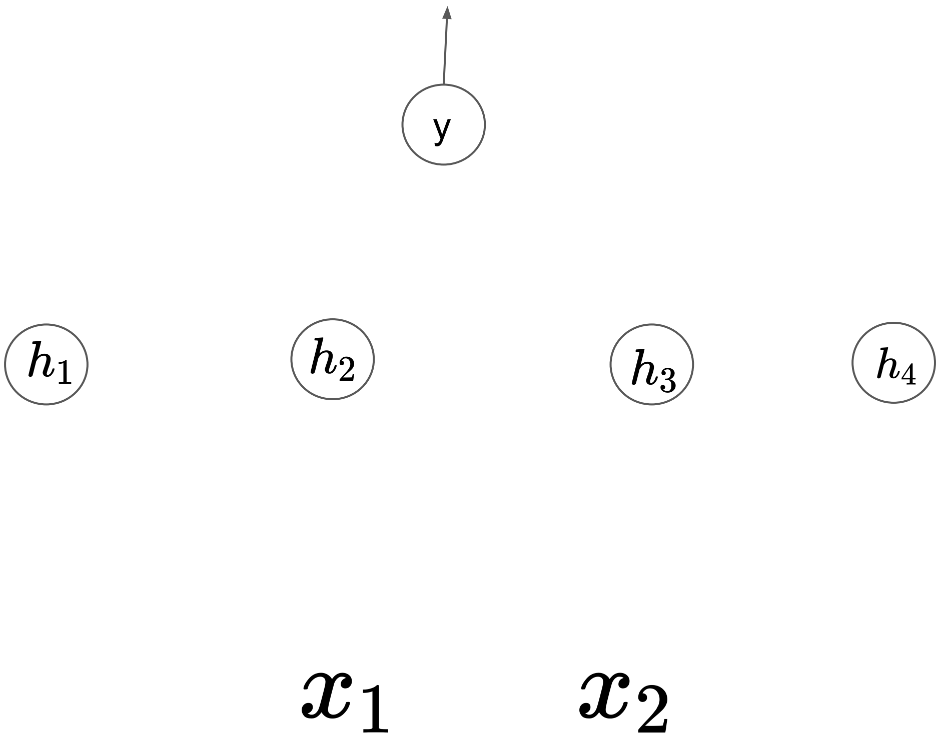
Any Boolean function of n inputs can be represented by a network of perceptron containing 1 hidden layer with 2^n perceptron and one output layer containing one perceptron.

Network of Perceptron

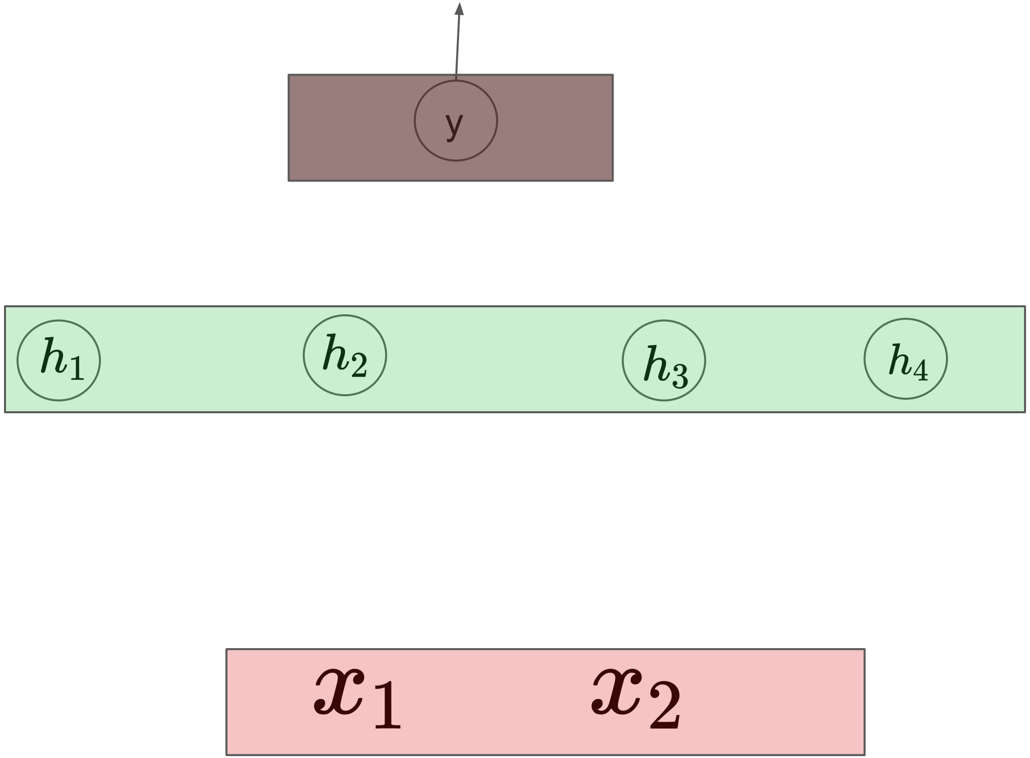
$x_1 \in \{-1, 1\}$
$x_2 \in \{-1, 1\}$

$x_1 \quad x_2$

Network of Perceptron



Network of Perceptron



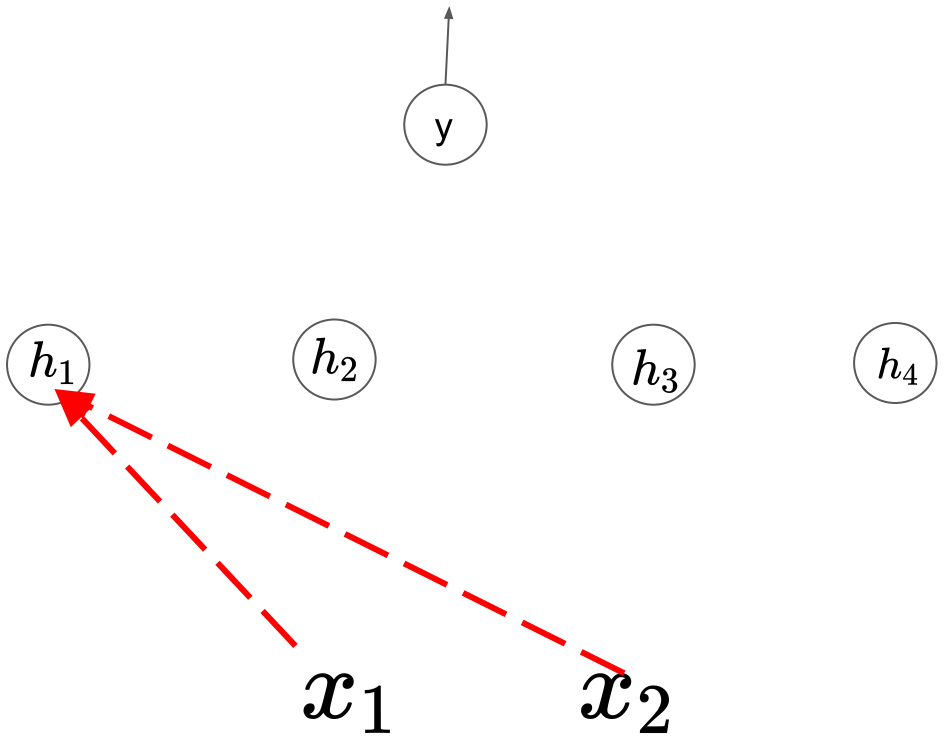
Output Layer

Hidden Layer

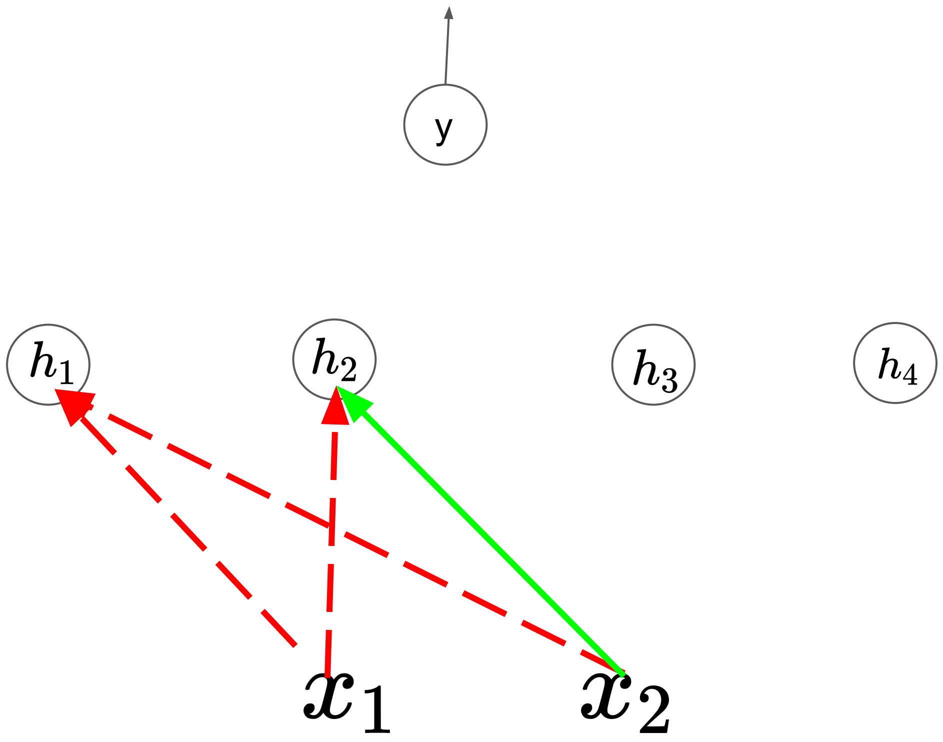
Input Layer

Network of Perceptron

$$h_k = f(wx_1 + wx_2 - 2)$$
$$w = -1 \;\; for \; red \; dotted \; arrow$$
$$w = 1 \;\; for \; green \; arrow$$



Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$$w = -1 \;\; \mathit{for\ red\ dotted\ arrow}$$

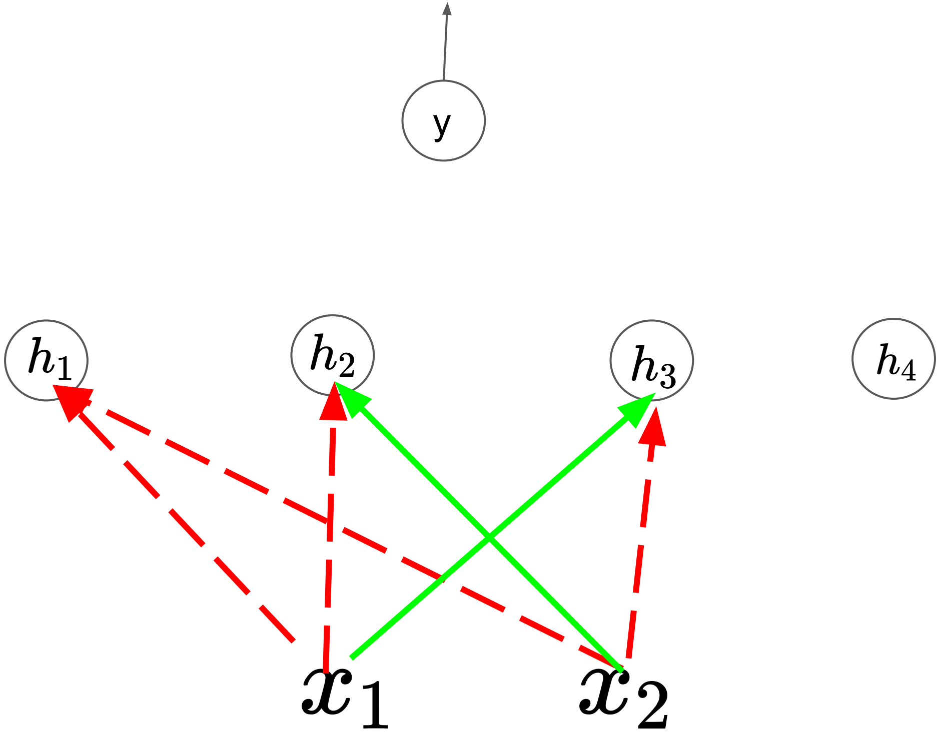
$$w = 1 \;\; \mathit{for\ green\ arrow}$$

Network of Perceptron

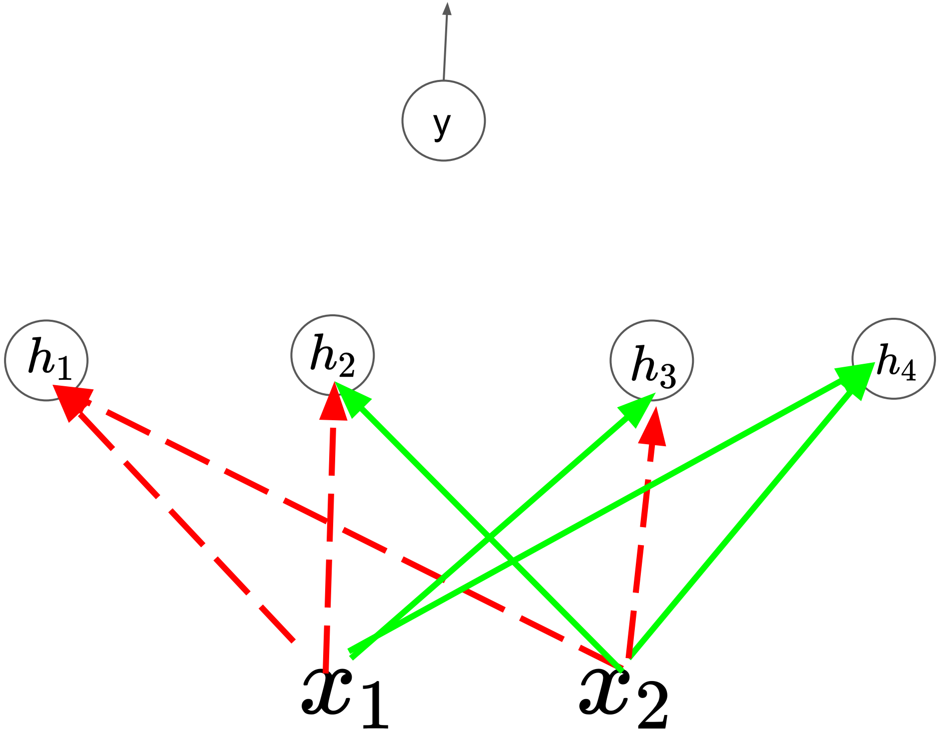
$$h_k = f(wx_1 + wx_2 - 2)$$

$$w = -1 \;\; for \; \textcolor{red}{red \; dotted \; arrow}$$

$$w = 1 \;\; for \; \textcolor{green}{green \; arrow}$$



Network of Perceptron

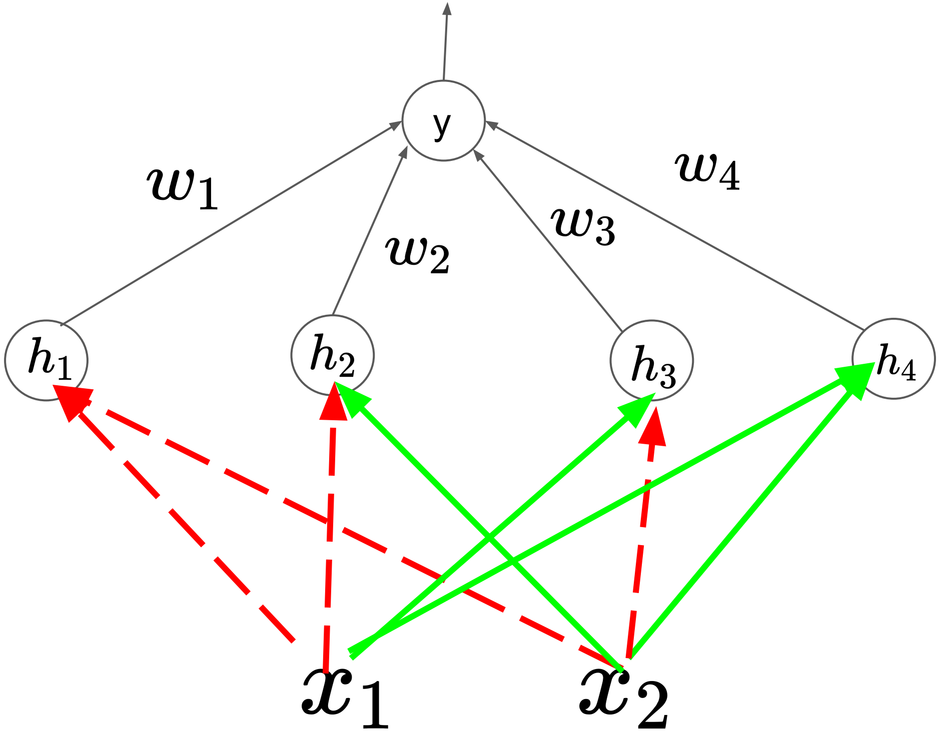


$$h_k = f(wx_1 + wx_2 - 2)$$

$$w = -1 \;\; for \; red \; dotted \; arrow$$

$$w = 1 \;\; for \; green \; arrow$$

Network of Perceptron



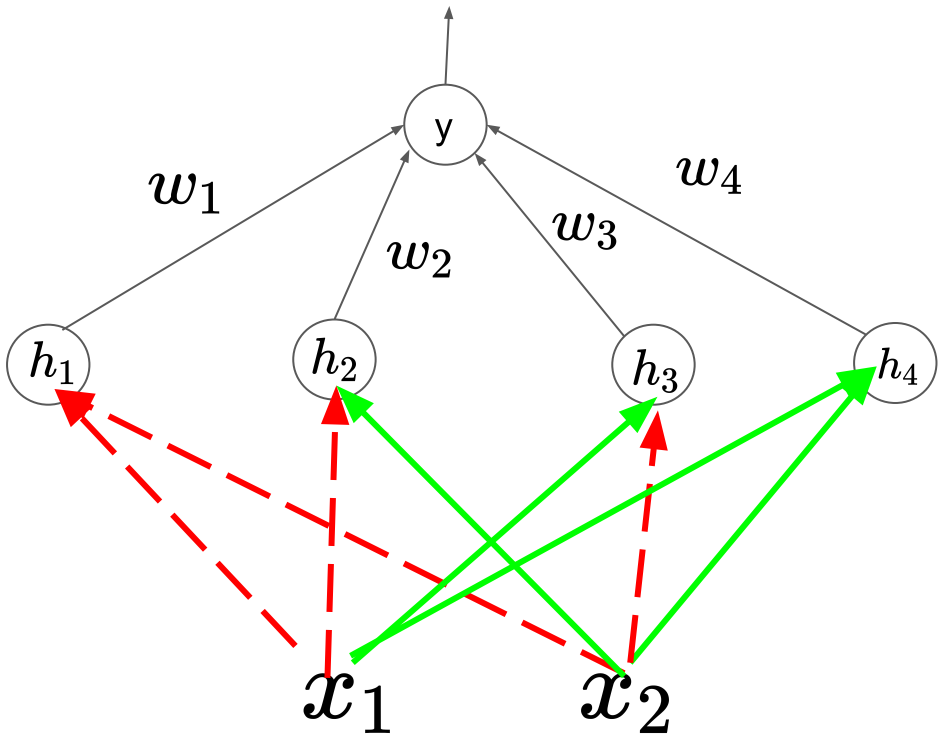
$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

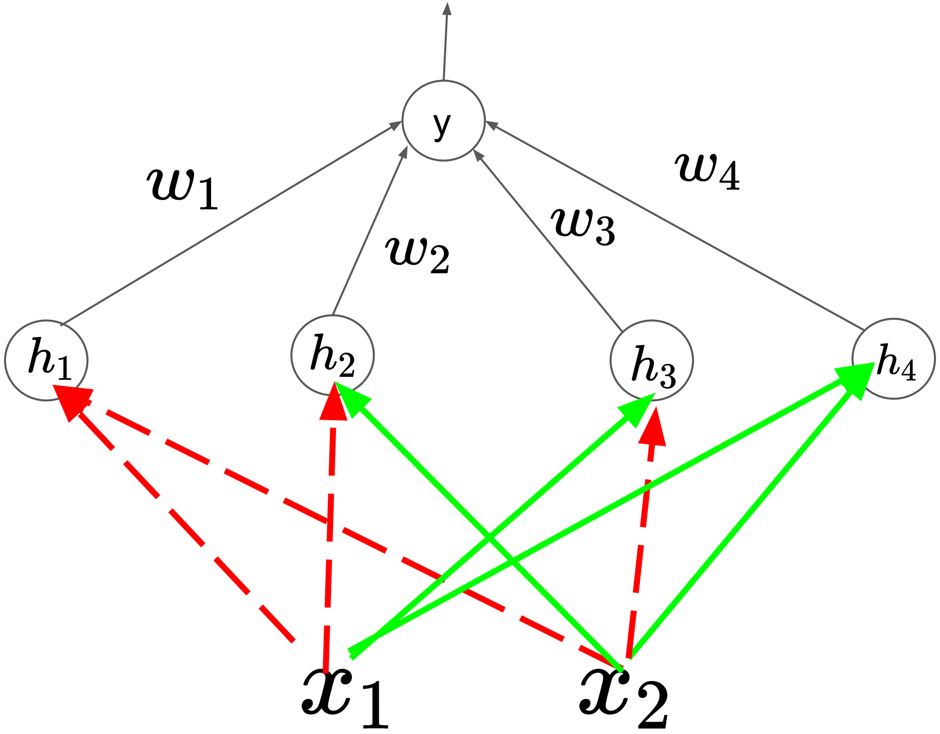
$w = -1$ *for red dotted arrow*

$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

$$f(p) = 0 \; if \; p < 0$$
$$f(p) = 1 \; if \; p >= 0$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

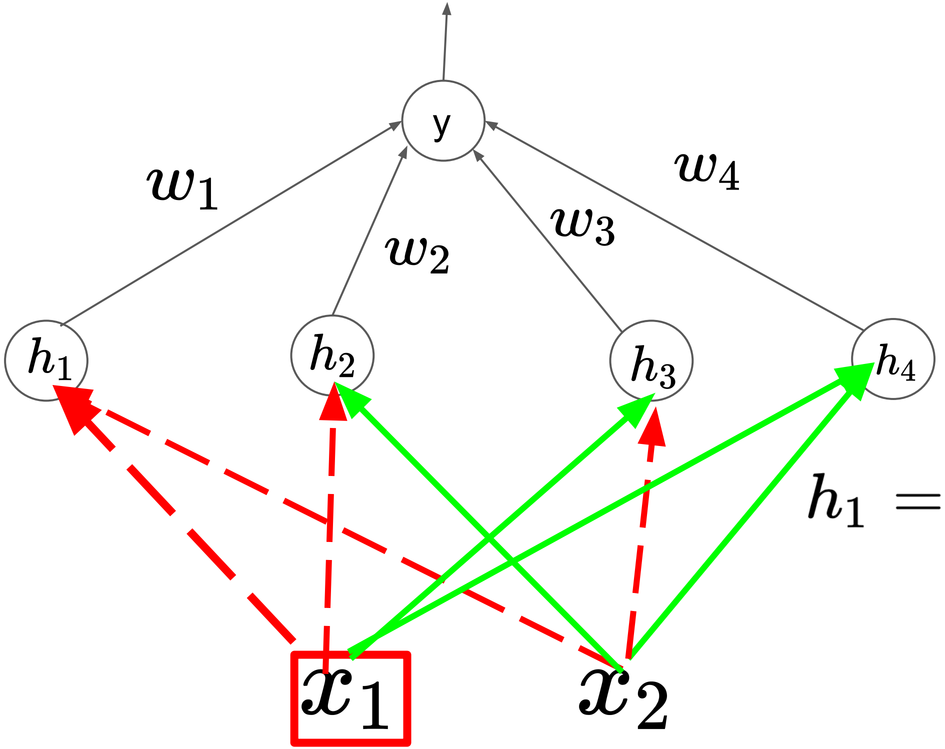
$w = -1$ *for red dotted arrow*

$w = 1$ *for green arrow*

$$y = f\left(\sum_{i=1}^{4} w_i h_i\right)$$

$x_1 = -1$ *and* $x_2 = -1$, *compute* h_1

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

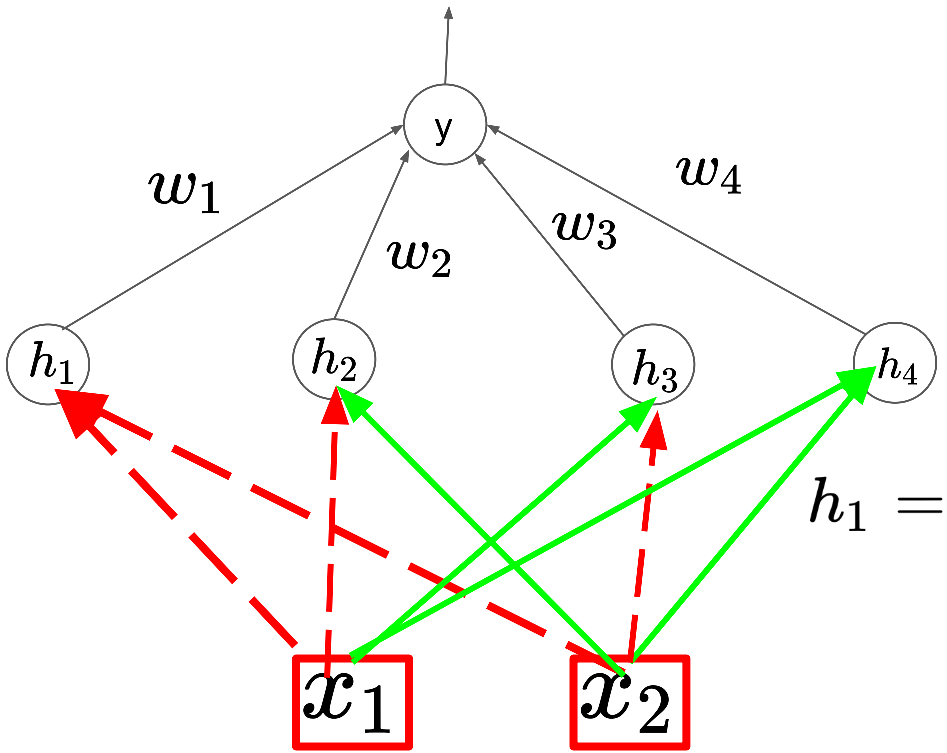
$w = 1$ *for green arrow*

$$y = f\left(\sum_{i=1}^{4} w_i h_i\right)$$

$x_1 = -1$ *and* $x_2 = -1$, *compute* h_1

$$h_1 = f(-1 \times -1 +$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

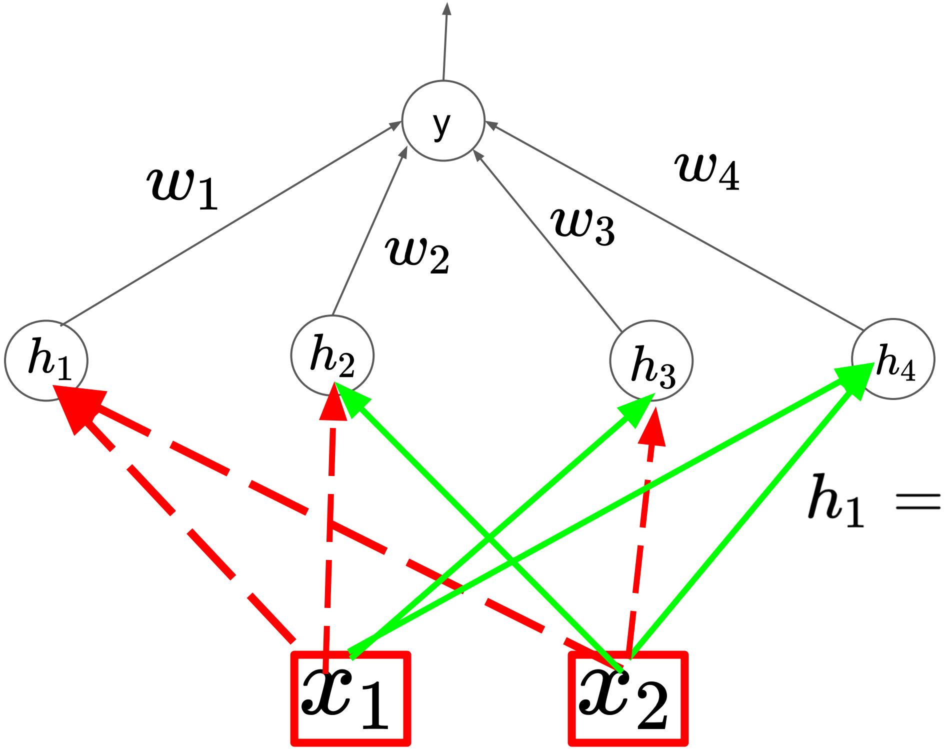
$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = -1$, *compute* h_1

$$h_1 = f(-1 \times -1 + -1 \times -1$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

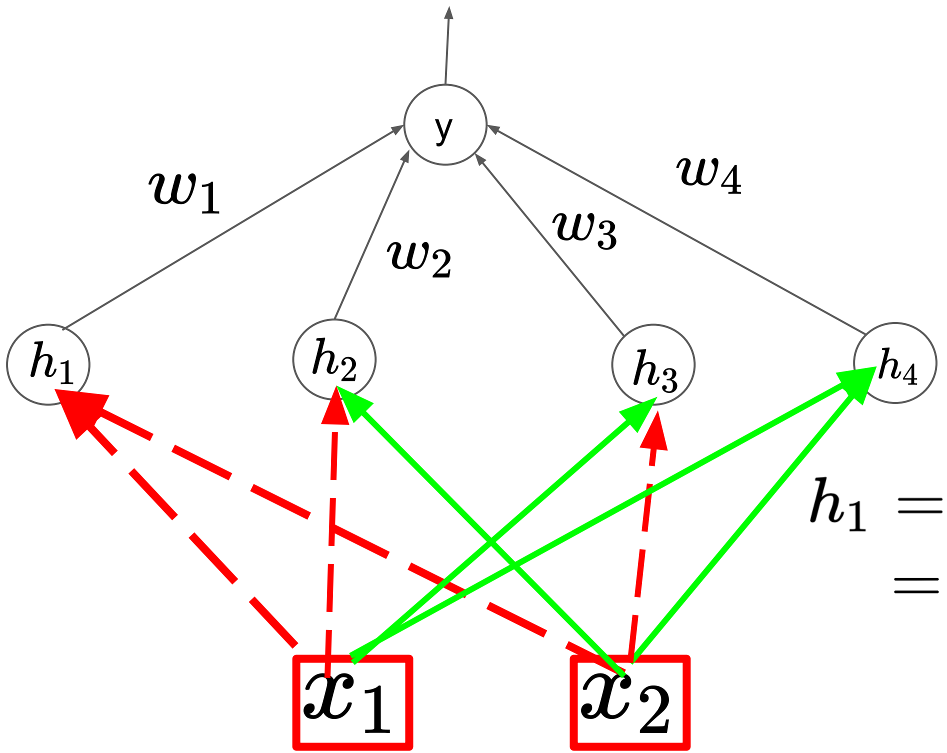
$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = -1$, *compute* h_1

$$h_1 = f(-1 \times -1 + -1 \times -1 - 2)$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

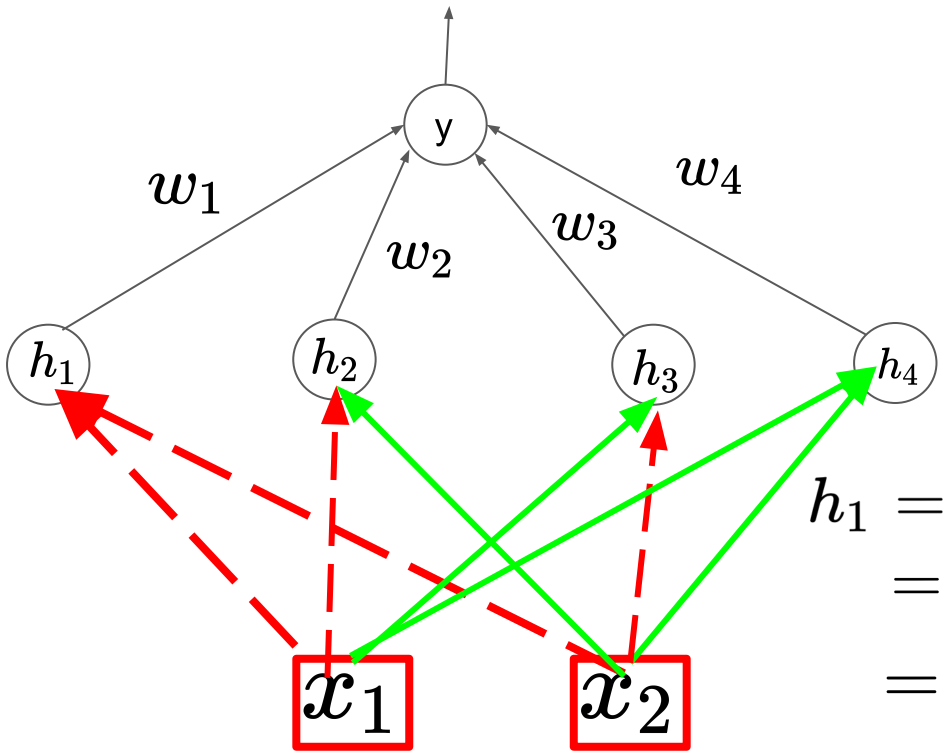
$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = -1$, *compute* h_1

$$h_1 = f(-1 \times -1 + -1 \times -1 - 2)$$
$$= f(0)$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

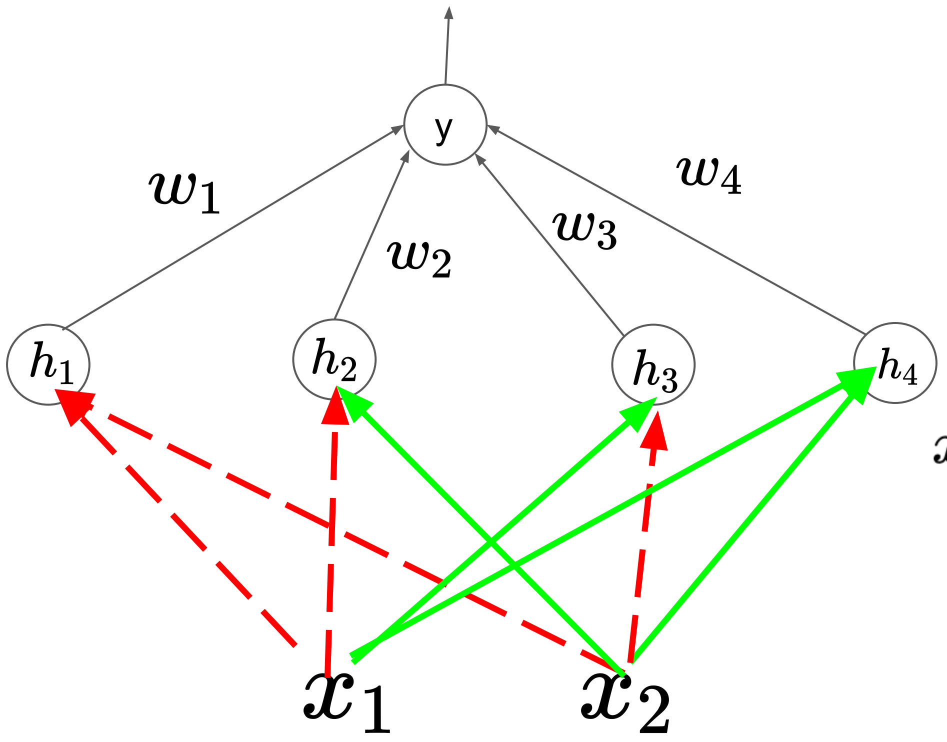
$w = 1$ *for green arrow*

$$y = f(\textstyle\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = -1$, *compute* h_1

$$\begin{aligned} h_1 &= f(-1 \times -1 + -1 \times -1 - 2) \\ &= f(0) \\ &= 1 \end{aligned}$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

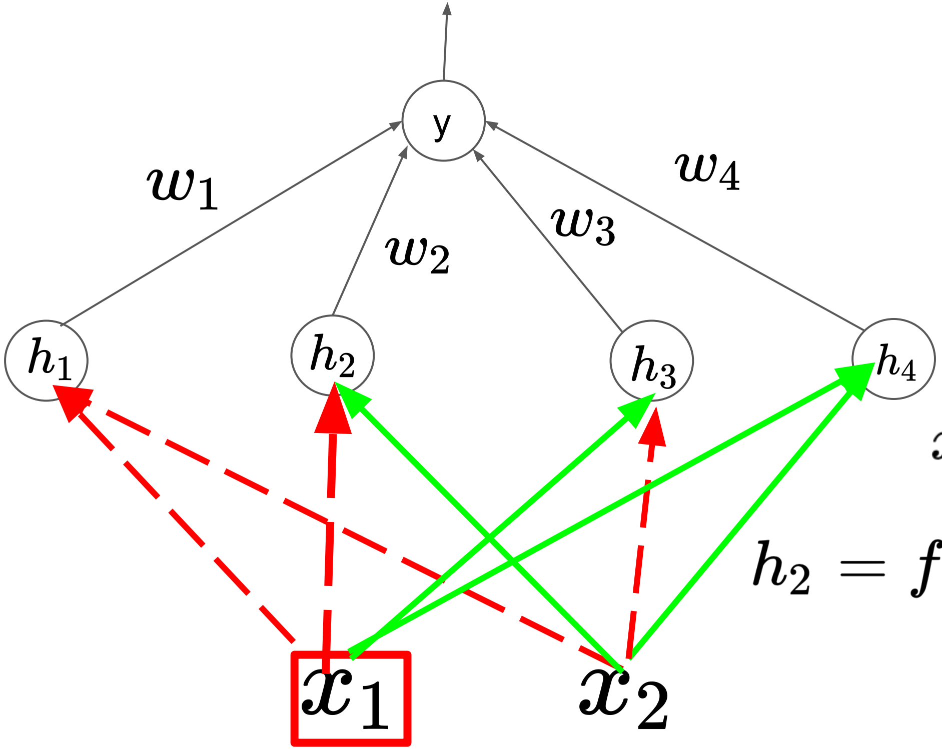
$w = -1$ *for red dotted arrow*

$w = 1$ *for green arrow*

$$y = f\left(\textstyle\sum_{i=1}^{4} w_i h_i\right)$$

$x_1 = -1$ *and* $x_2 = 1,$ *compute* h_2

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

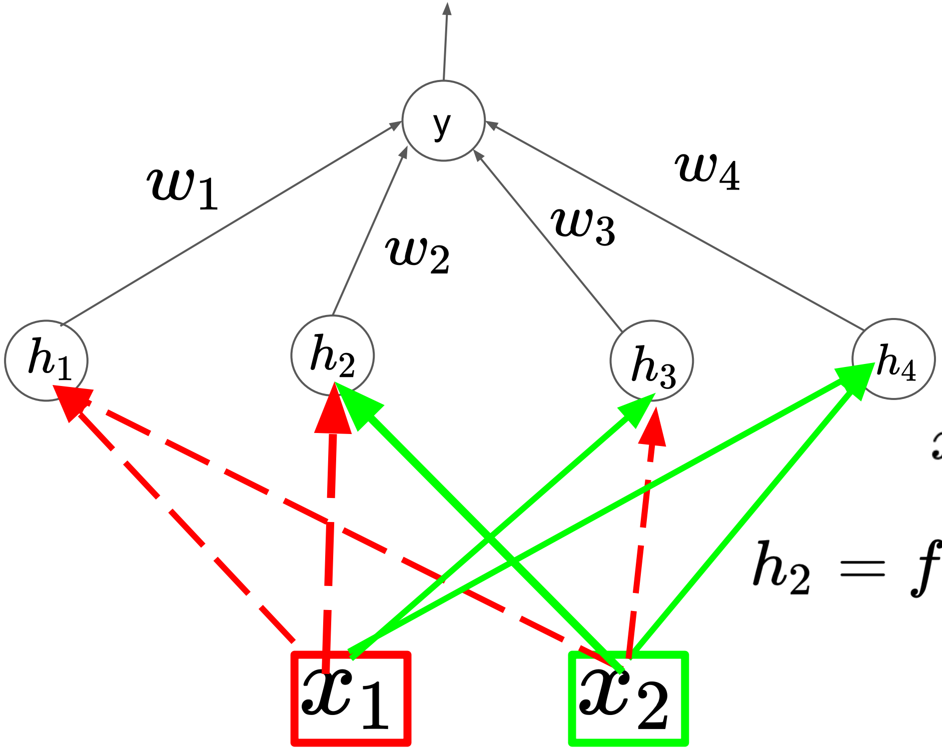
$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = 1,$ *compute* h_2

$$h_2 = f(-1 \times -1$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

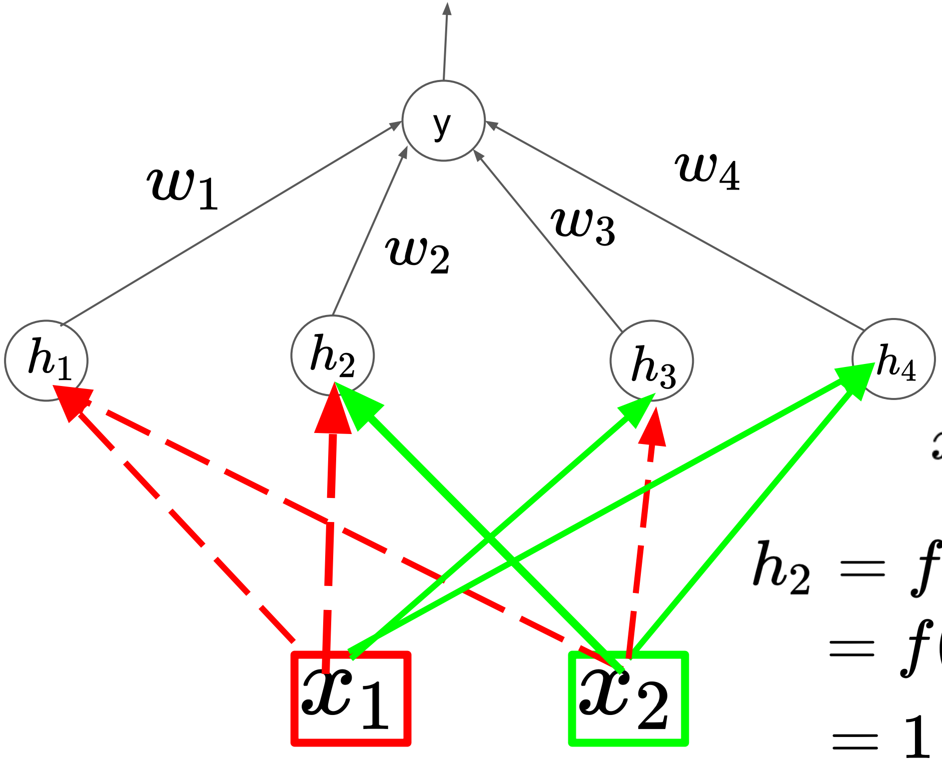
$w = 1$ *for green arrow*

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = 1,$ *compute* h_2

$$h_2 = f(-1 \times -1 + 1 \times 1$$

Network of Perceptron



$$h_k = f(wx_1 + wx_2 - 2)$$

$w = -1$ *for red dotted arrow*

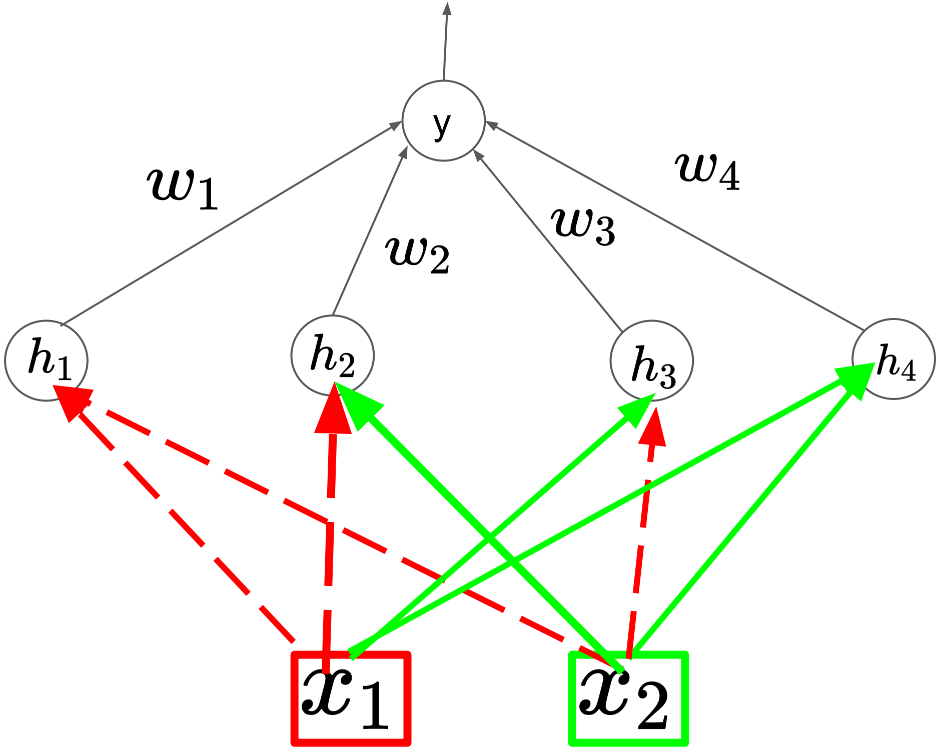
$w = 1$ *for green arrow*

$$y = f(\textstyle\sum_{i=1}^{4} w_i h_i)$$

$x_1 = -1$ *and* $x_2 = 1,$ *compute* h_2

$$\begin{aligned} h_2 &= f(-1 \times -1 + 1 \times 1 - 2) \\ &= f(0) \\ &= 1 \end{aligned}$$

Network of Perceptron

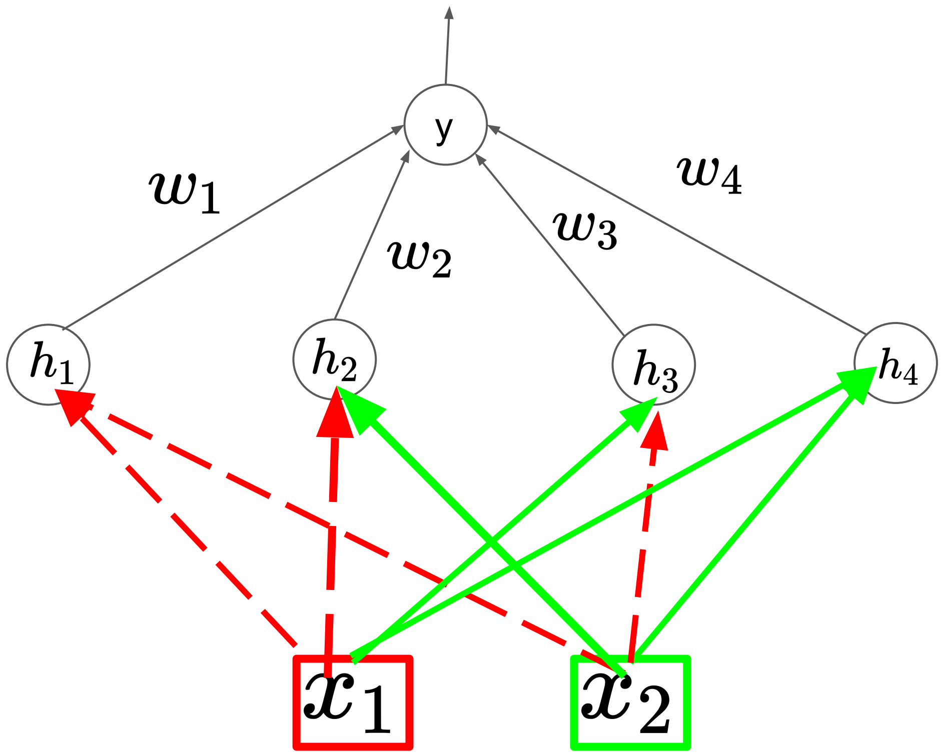


$$h_k = f(wx_1 + wx_2 - 2)$$

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

x_1	x_2	h_1	h_2	h_3	h_4	y
-1	-1					
-1	1					
1	-1					
1	1					

Network of Perceptron

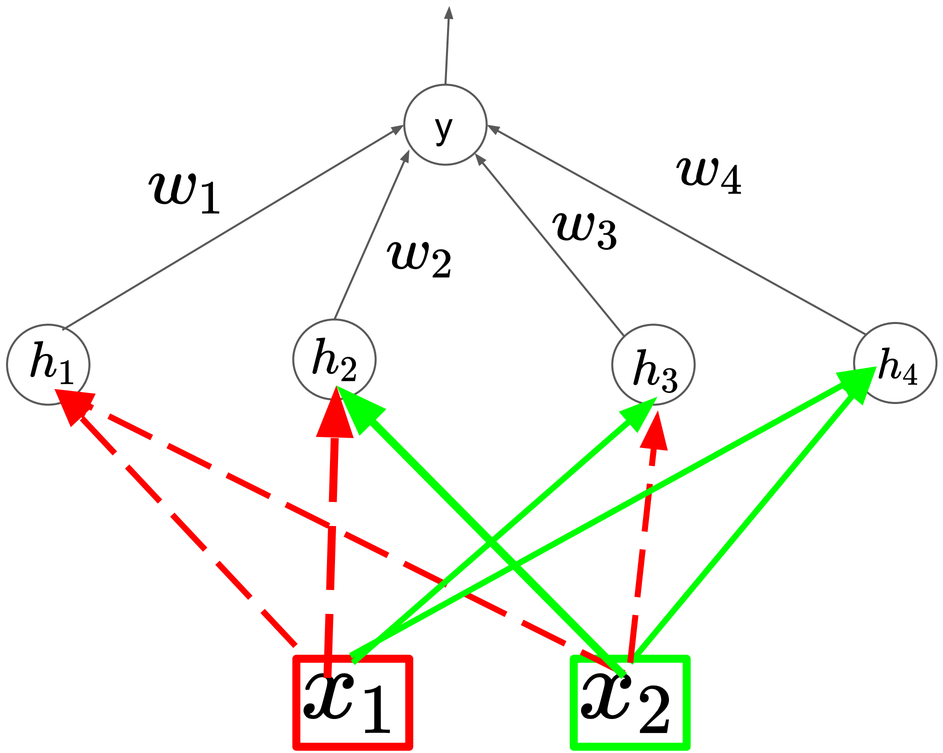


$$h_k = f(wx_1 + wx_2 - 2)$$

$$y = f(\sum_{i=1}^{4} w_i h_i)$$

x_1	x_2	h_1	h_2	h_3	h_4	y
-1	-1	$f(0)$	$f(-2)$	$f(-2)$	$f(-4)$	
-1	1	$f(-2)$	$f(0)$	$f(-4)$	$f(-2)$	
1	-1	$f(-2)$	$f(-4)$	$f(0)$	?	
1	1	$f(-2)$	$f(-2)$	$f(-2)$	$f(0)$	

Network of Perceptron

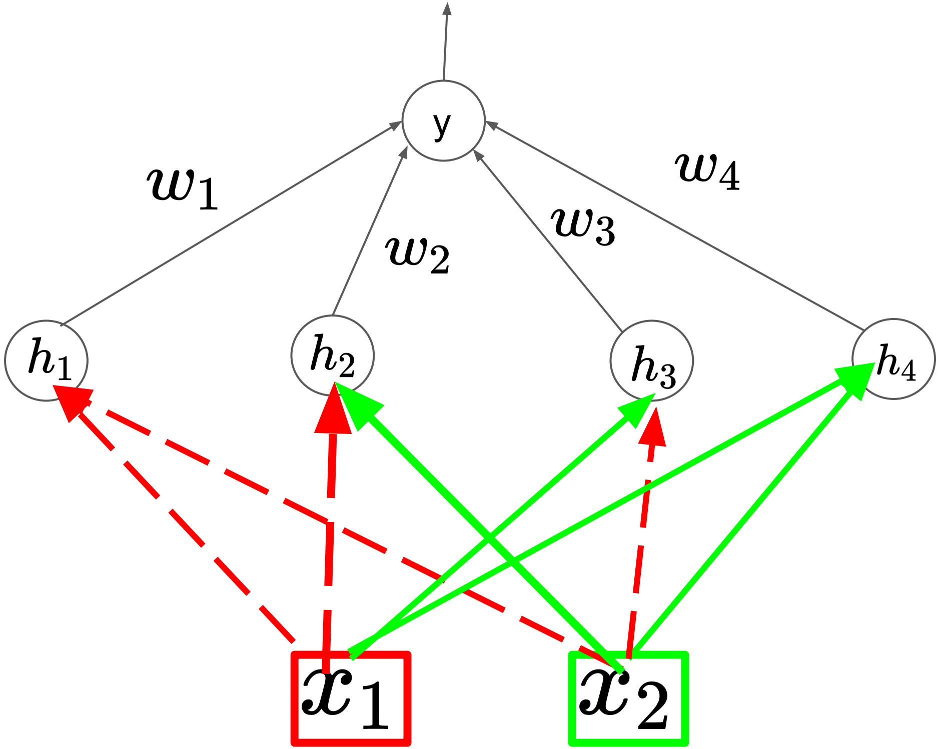


$$h_k = f(wx_1 + wx_2 - 2)$$

$$y = f(\textstyle\sum_{i=1}^{4} w_i h_i)$$

x_1	x_2	h_1	h_2	h_3	h_4	y
-1	-1	$f(0)$	$f(-2)$	$f(-2)$	$f(-4)$	
-1	1	$f(-2)$	$f(0)$	$f(-4)$	$f(-2)$	
1	-1	$f(-2)$	$f(-4)$	$f(0)$	$f(-2)$	
1	1	$f(-2)$	$f(-2)$	$f(-2)$	$f(0)$	

Network of Perceptron

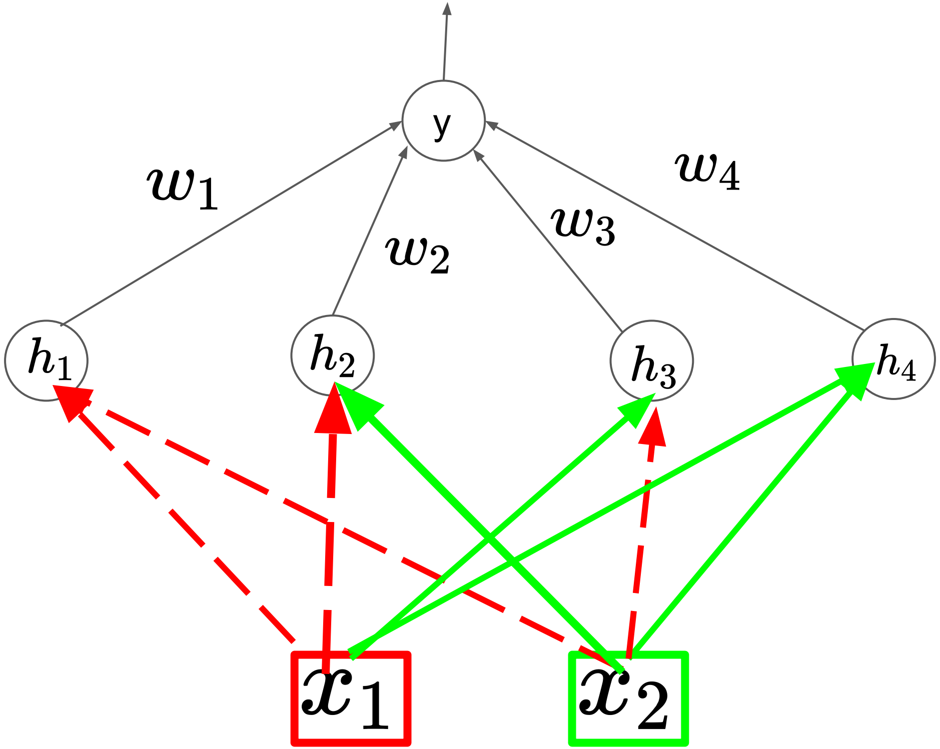


$$h_k = f(wx_1 + wx_2 - 2)$$

$$y = f(\textstyle\sum_{i=1}^{4} w_i h_i)$$

x_1	x_2	h_1	h_2	h_3	h_4	y
-1	-1	1	0	0	0	
-1	1	0	1	0	0	
1	-1	0	0	1	0	
1	1	0	0	0	1	

Network of Perceptron



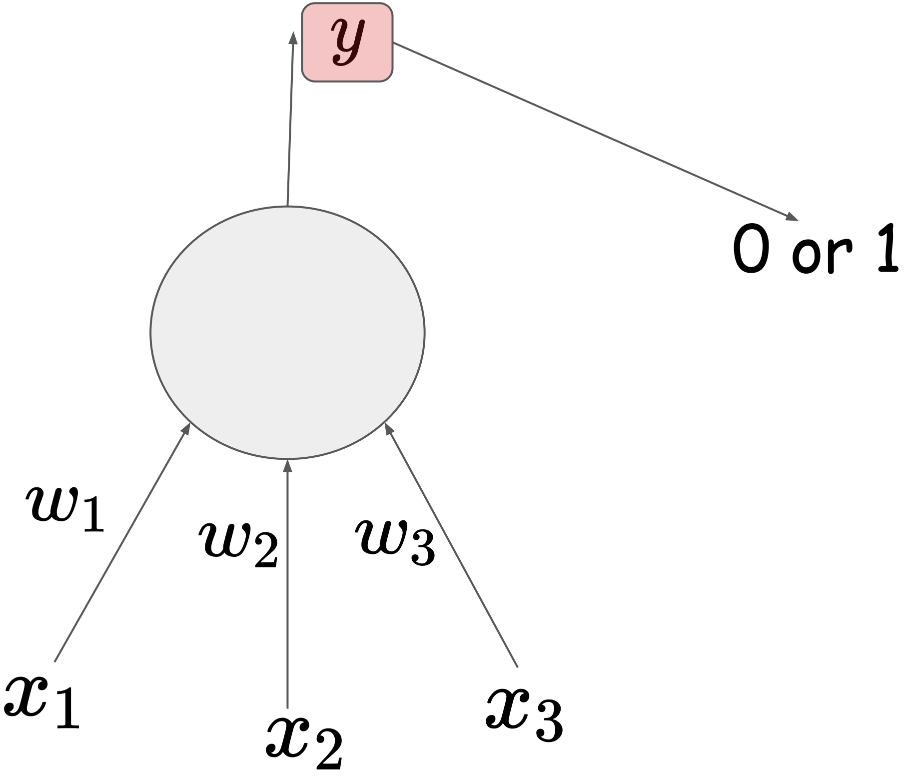
$$h_k = f(wx_1 + wx_2 - 2)$$

$$y = f(\textstyle\sum_{i=1}^{4} w_i h_i)$$

x_1	x_2	h_1	h_2	h_3	h_4	y
-1	-1	1	0	0	0	$f(w_1)$
-1	1	0	1	0	0	$f(w_2)$
1	-1	0	0	1	0	$f(w_3)$
1	1	0	0	0	1	$f(w_4)$

Module 3: Sigmoid Neuron

So far only Boolean functions

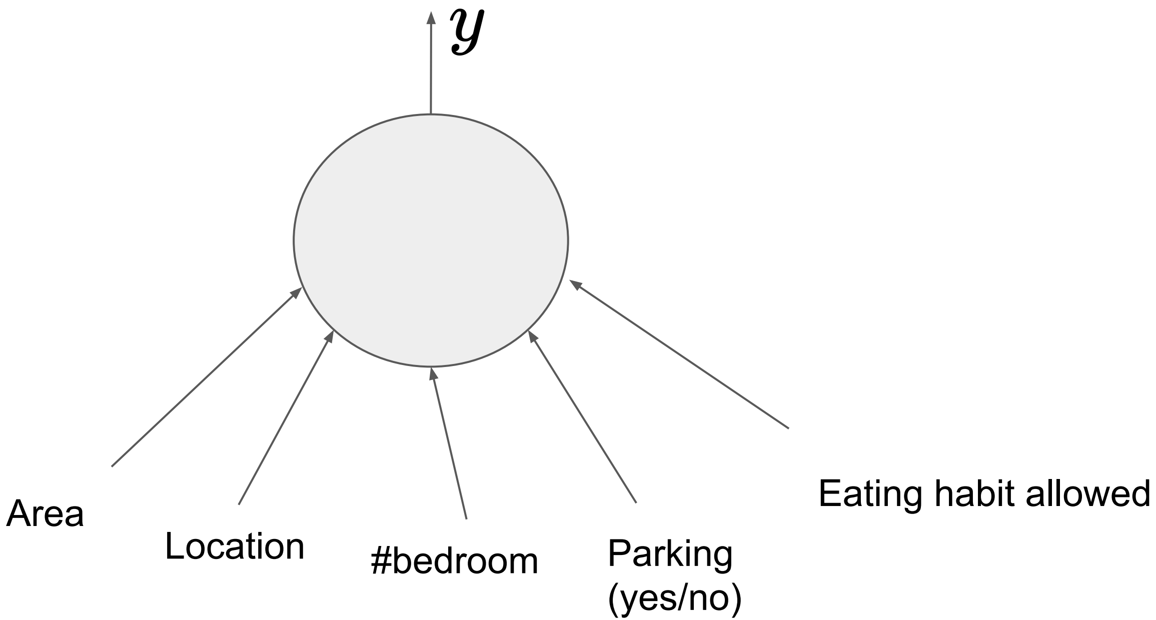


What if y is a real number?

Problems with perceptron

Problem 1: So far only models a Boolean Function.

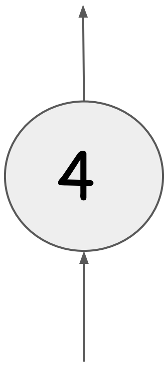
Estimate the rent cost of a real-estate property?



Problems with perceptron

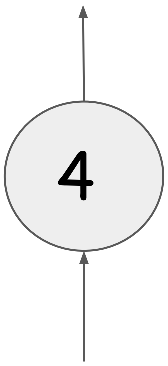
Problem 2: Hard Thresholding

You should watch movie



Average Movie Rating (4.1)

You should not watch movie

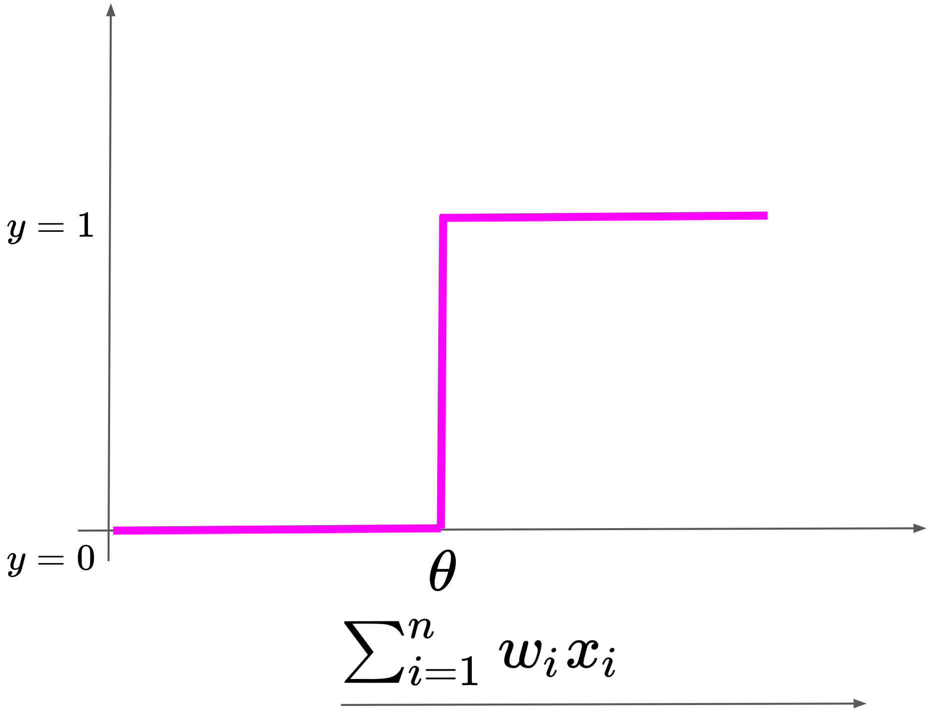


Average Movie Rating (3.9)

Decision function

$$y = 0 \;\; if \;\; \sum_{i=1}^{n} w_i x_i < \theta$$

$$y = 1 \;\; if \;\; \sum_{i=1}^{n} w_i x_i >= \theta$$

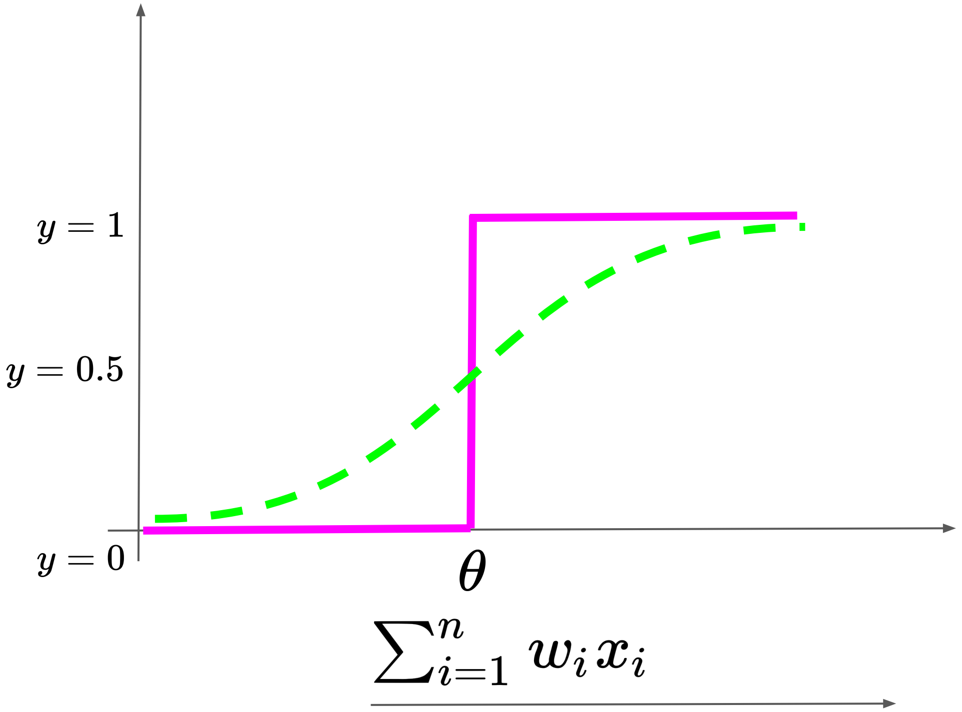


Decision function

$$y = 0 \;\; if \;\; \sum_{i=1}^{n} w_i x_i < \theta$$

$$y = 1 \;\; if \;\; \sum_{i=1}^{n} w_i x_i >= \theta$$

$$y = \frac{1}{1+e^{\sum_{i=0}^{n} w_i x_i}}$$



Decision function

Perceptron decision function

Non smooth, non-differentiable, non-continuous

Sigmoid decision function

Smooth, differentiable, continuous

Summary

- Perceptrons are powerful enough to do a good job for classifying linearly-separable samples.
- There exists a network of perceptron which can classify samples correctly for a given problem. (Network of Perceptron is extremely powerful)