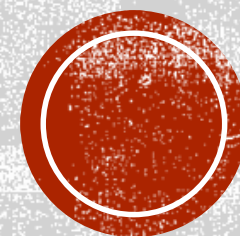


# THE PERCEPTRON

Eitan Kosman



# TOPICS

- **The Problem**
  - Linear Separability – Def. 1
  - Linear Separability – Def. 2
- **A Biological Neuron**
  - Structure
  - A Mathematical model
- **The Solution – Perceptron**
  - A brief history
  - The algorithm
  - Intuitive interpretation for weights update
  - Theorem: Mistake bound
  - The fall of perceptron
  - Rebirth



# THE PROBLEM

Given a set of  $n$  points in  $\mathbb{R}^d$  and labels:

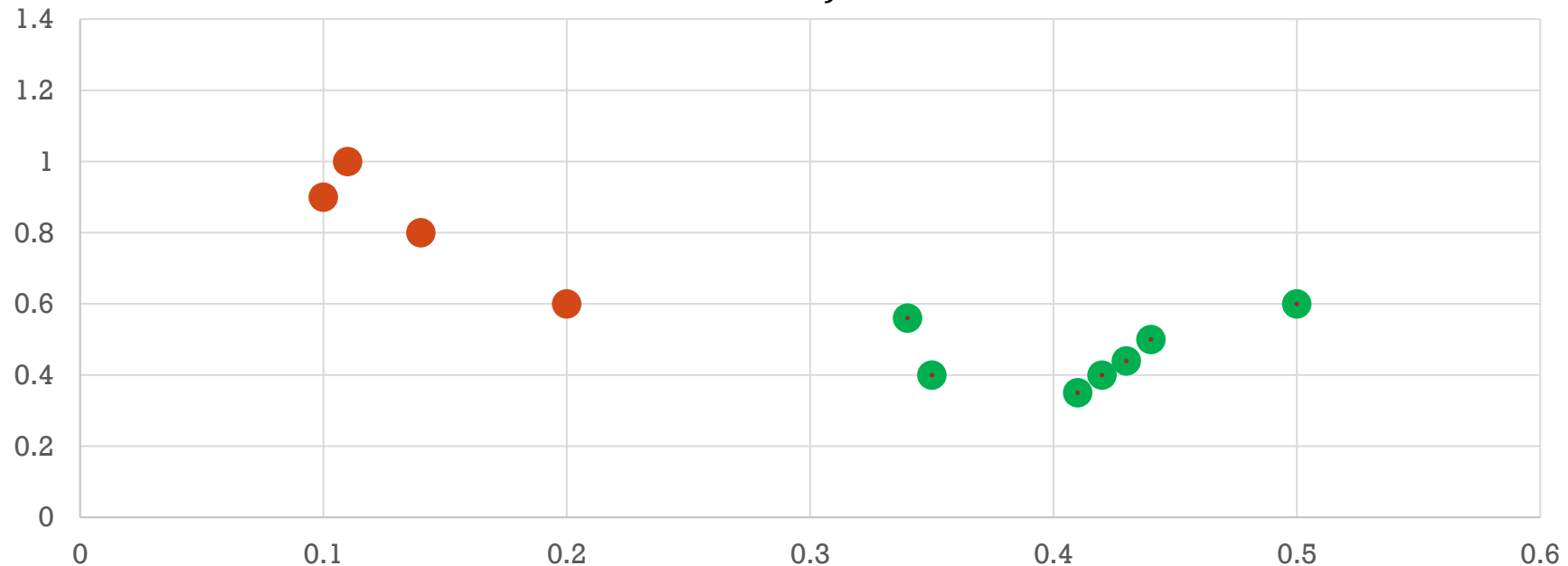
$$X = \{x_i | i \in [n], x \in \mathbb{R}^d\}, Y = \{y_i | i \in [n]\}$$

We want to find a transformation:

$$f: X \rightarrow Y$$

s.t.

$$x_i \mapsto_f y_i$$



# DEFINITION (1): LINEAR SEPARABILITY

Let  $X_0, X_1 \subseteq \mathbb{R}^d$  be 2 sets of points.  $X_0, X_1$  are linearly separable if there exist  $n + 1$  real numbers  $w_1, w_2, \dots, w_n, k$  such that:

$$\forall x \in X_0: \sum_{i=1}^n w_i x_i > k$$
$$\forall x \in X_1: \sum_{i=1}^n w_i x_i < k$$

The above terms could also be represented as inner product:  $\langle w, x \rangle$  where:  $w = (w_1, w_2, \dots, w_n)$  and  $x = (x_1, x_2, \dots, x_n)$

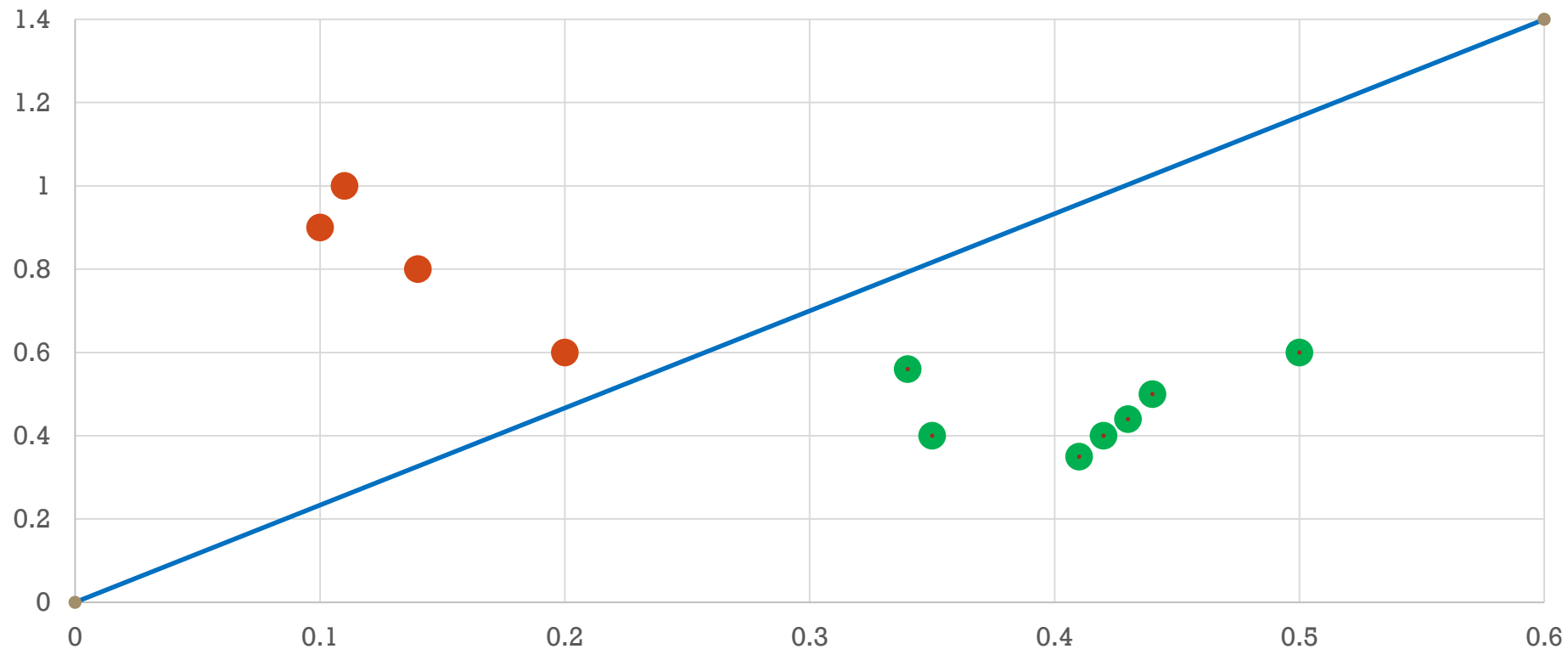


# DEFINITION 1 INTERPRETATION

Given vector  $w$ , we can define a hyper-plane by:

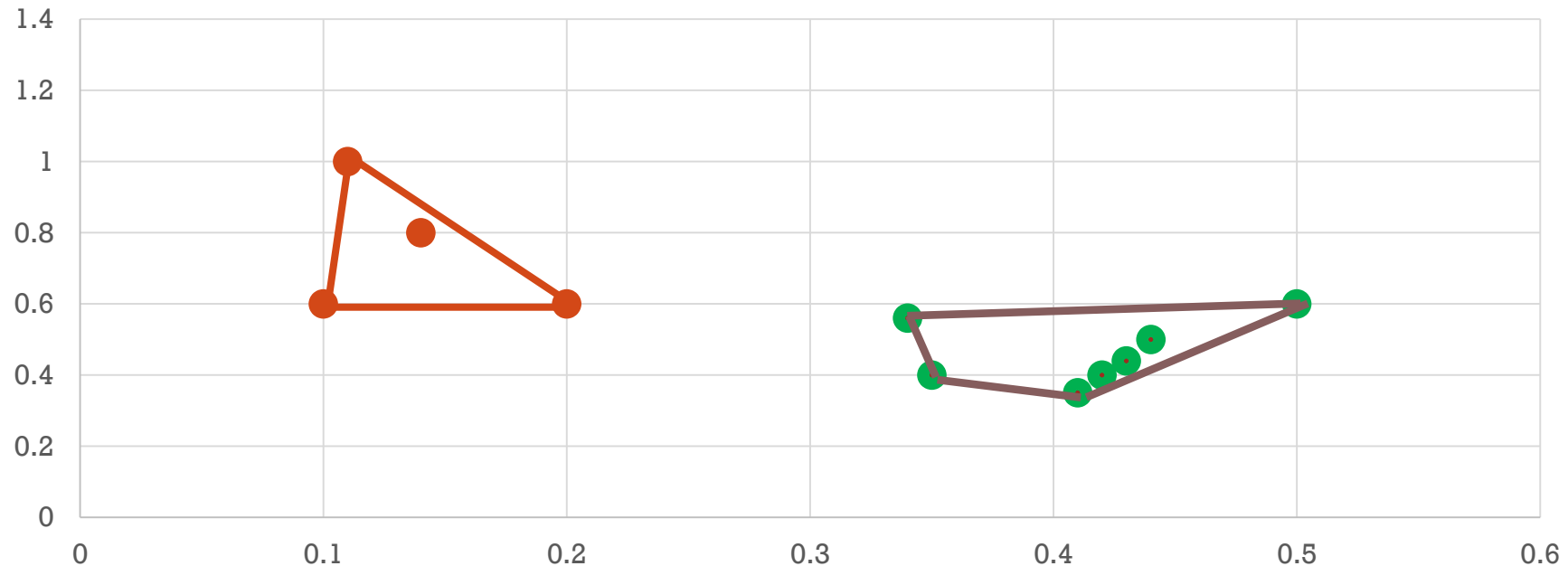
$$\langle w, x \rangle + d = 0$$

Thus, the hyper-plane separates the field into 2 regions such that all points belong to  $X_0$  are in one region and all points belong to  $X_1$  are in the other region.



# DEFINITION (2): LINEAR SEPERABILITY

Let  $X_0, X_1 \subseteq \mathbb{R}^d$  be 2 sets of points.  $X_0, X_1$  are linearly separable precisely when their respective convex hulls are disjoint (do not overlap)



# TOPICS

- The Problem
  - Linear Separability – Def. 1
  - Linear Separability – Def. 2
- The Biological Neuron
  - Structure
  - A Mathematical model
- The Solution – Perceptron
  - A brief history
  - The algorithm
  - Intuitive interpretation for weights update
  - Theorem: Mistake bound
  - The fall of perceptron
  - Rebirth

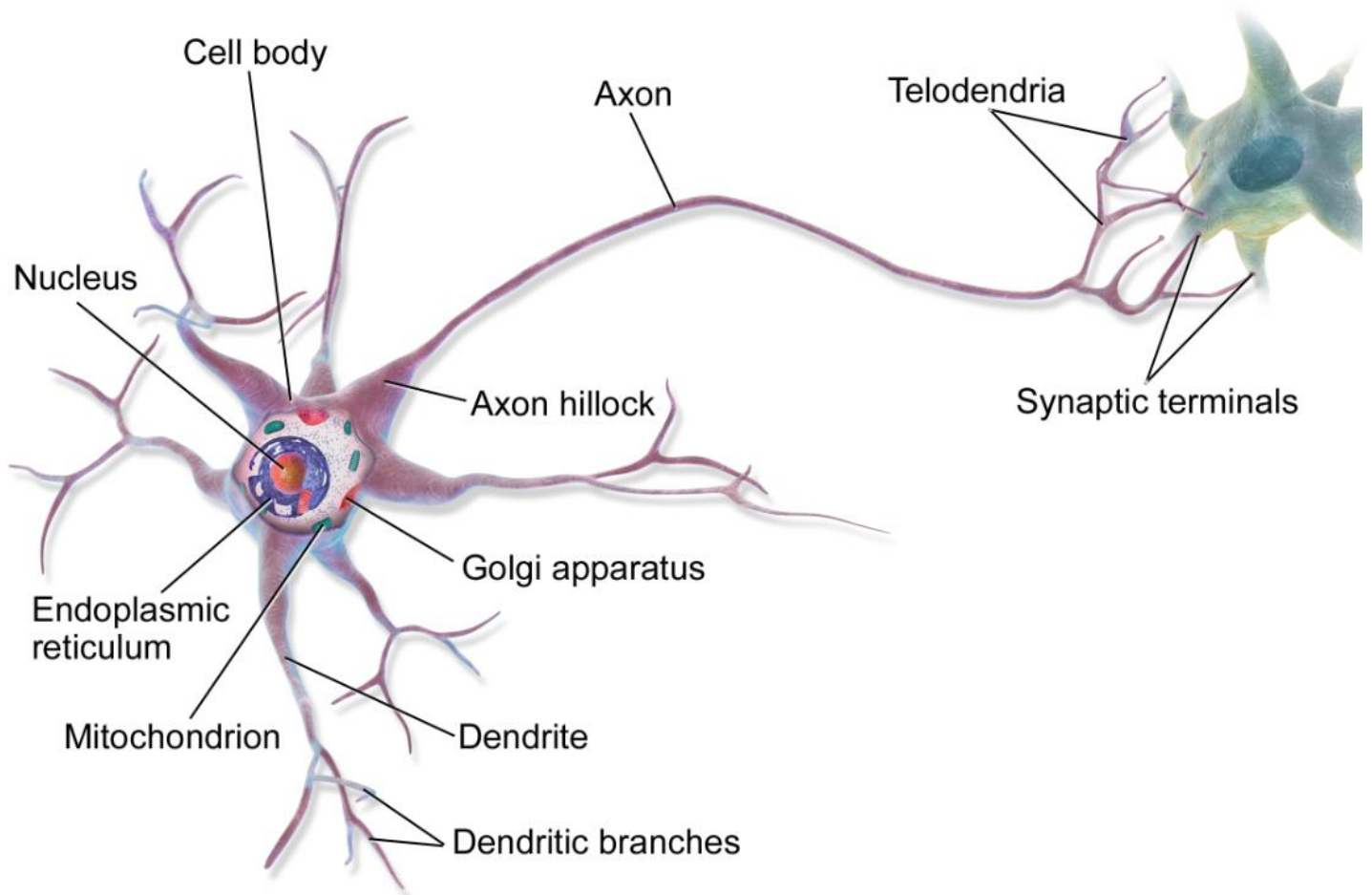


# THE BIOLOGICAL NEURON - STRUCTURE

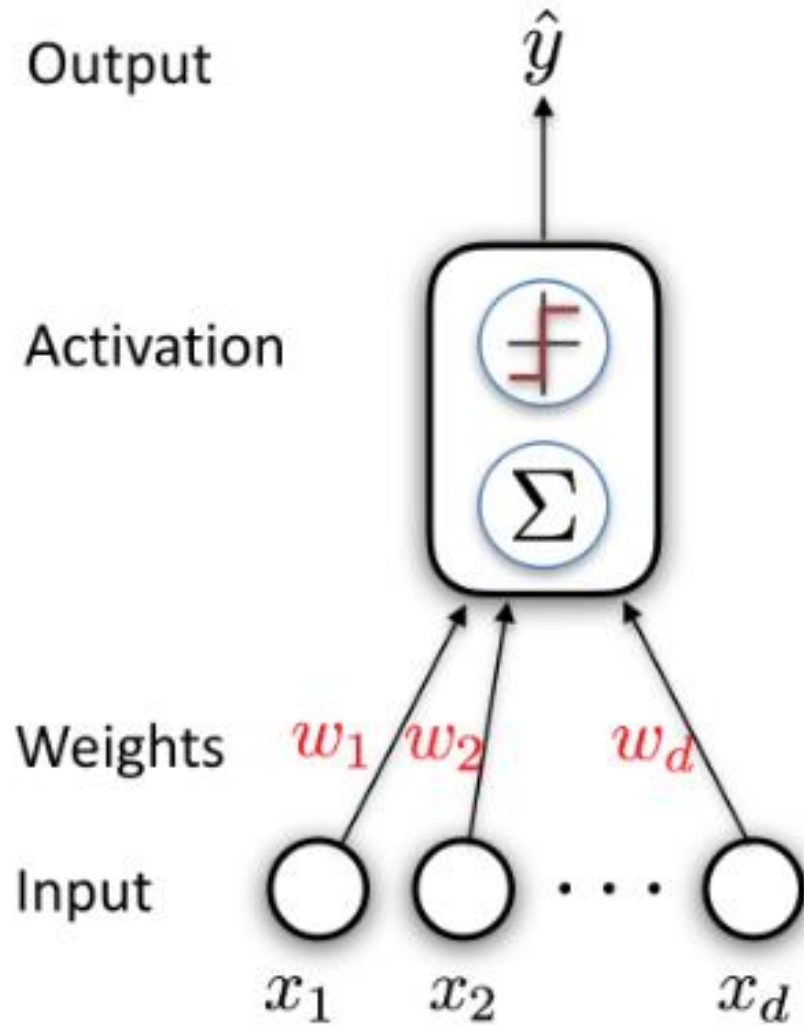
Like any other body cell, the neuron has a cell body which contains a nucleus where the DNA is stored.

From our perspective, the interesting parts are:

- Dendrites – make connections with tens of thousand of other cells; other neurons. They behave as “inputs”.
- Axon – transmits information to different neurons, muscles, and other body cells based on the signals the cell receives. Its signals are received by other cells' dendrites.







# THE BIOLOGICAL NEURON – A MATHEMATICAL MODEL

- We will try to mimic the function of a neuron using mathematical tools. Given an input vector  $x$ :
  - $x$  will be the inputs of the neuron (dendrites).
  - Define a weight,  $w_i$ , for each input, and sum all the multiplications.
  - Output the result as  $\hat{y}$  (Axon)

**There's still a problem – How do we find the weights?**



# TOPICS

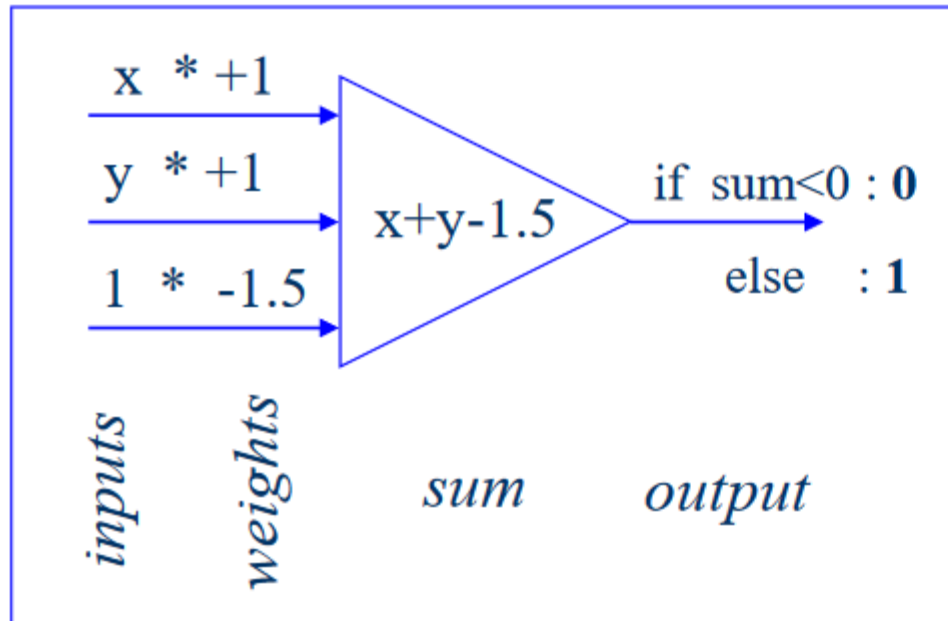
- The Problem
  - Linear Separability – Def. 1
  - Linear Separability – Def. 2
- The Biological Neuron
  - Structure
  - A Mathematical model
- The Solution – Perceptron
  - A brief history
  - The algorithm
  - Intuitive interpretation for weights update
  - Theorem: Mistake bound
  - The fall of perceptron



# PREHISTORY

Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic.

- W.S. McCulloch & W. Pitts (1943). “A logical calculus of the ideas immanent in nervous activity”, Bulletin of Mathematical Biophysics, 5, 115-137
- This seminal paper pointed out that simple artificial “neurons” could be made to perform basic logical operations such as AND, OR and NOT

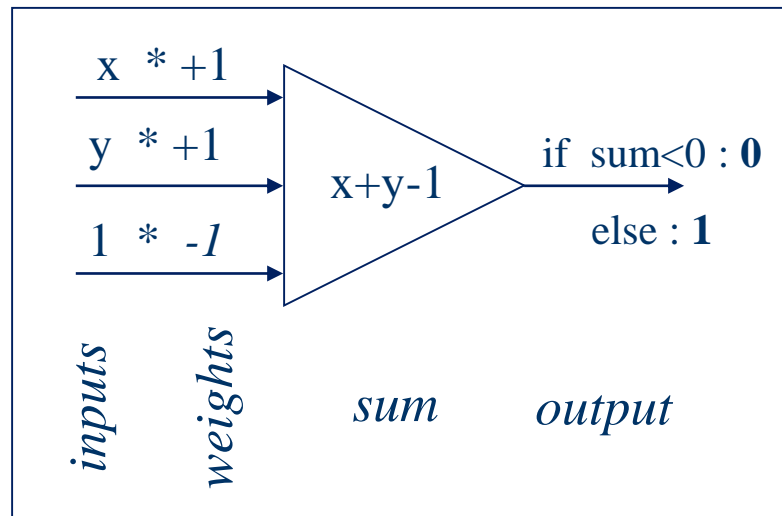


**Truth Table for Logical AND**

x	y	x & y
0	0	0
0	1	0
1	0	0
1	1	1

*inputs* *output*





**Truth Table for Logical  
OR**

$x$	$y$	$x \mid y$
0	0	0
0	1	1
1	0	1
1	1	1

*inputs* *output*



# 1958 — THE PERCEPTRON

*Psychological Review*  
Vol. 65, No. 6, 1958

## THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN<sup>1</sup>

F. ROSENBLATT

*Cornell Aeronautical Laboratory*



# NEW NAVY DEVICE LEARNS BY DOING

---

Psychologist Shows Embryo  
of Computer Designed to  
Read and Grow Wiser

---

WASHINGTON, July. 7 (UPI)  
—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.



# PERCEPTRON — THE ALGORITHM

- The goal is to find a hyper-plane separating 2 known classes.
- Consider definition (1) for linear separability:

$$\forall x \in X_0: \langle w, x \rangle > k$$

$$\forall x \in X_1: \langle w, x \rangle < k$$

$\Downarrow$

$$\forall x \in X_0: \langle w, x \rangle - k > 0$$

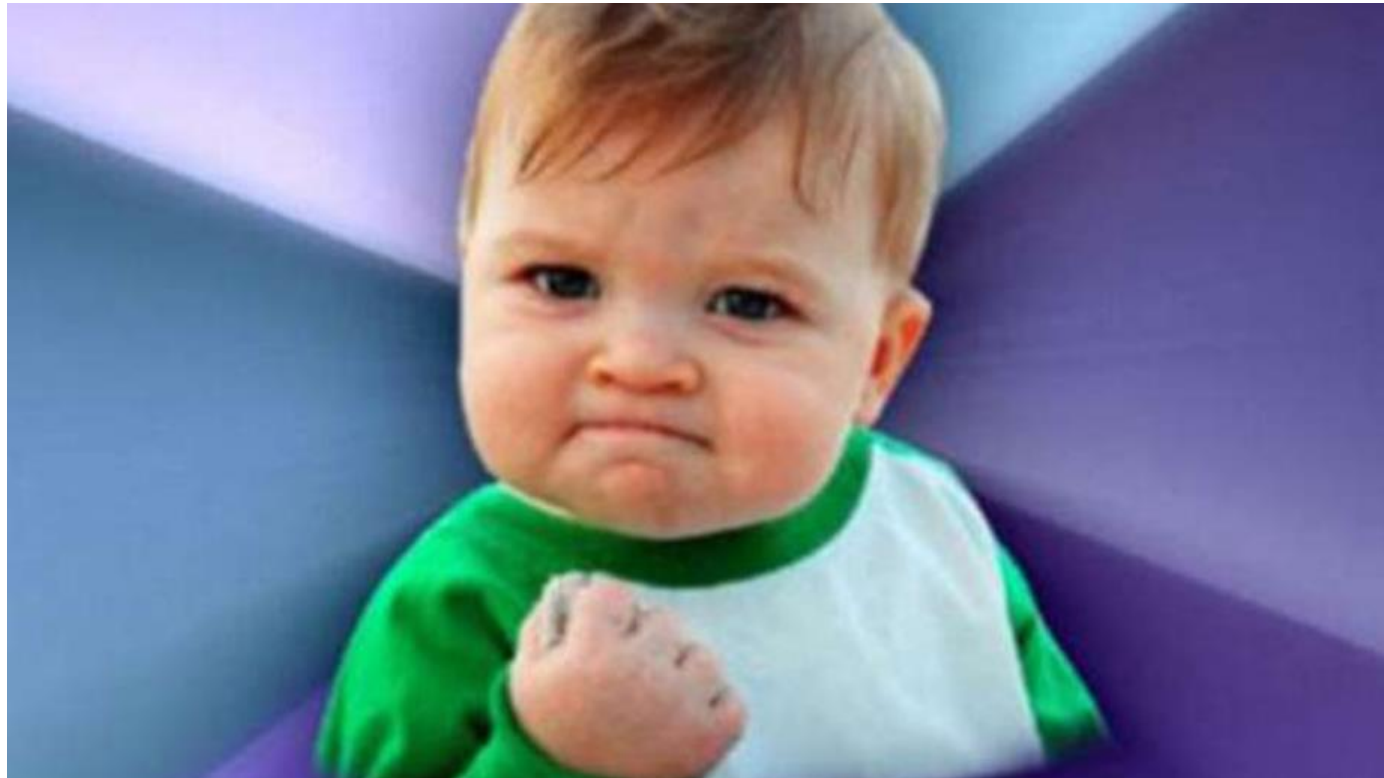
$$\forall x \in X_1: \langle w, x \rangle - k < 0$$



We can eliminate  $k$  by augmenting representation with one dimension:

$$\begin{aligned}x' &= (x, 1) \\ w' &= (w, -k)\end{aligned}$$

$$\langle w', x' \rangle = (w, -k) \begin{pmatrix} x \\ 1 \end{pmatrix} = w \cdot x - k$$





---

## Algorithm: Perceptron Learning Algorithm

---

$P \leftarrow \text{inputs with label } 1;$

$N \leftarrow \text{inputs with label } 0;$

Initialize  $\mathbf{w}$  randomly;

**while** !*convergence* **do**

    Pick random  $\mathbf{x} \in P \cup N$  ;

**if**  $\mathbf{x} \in P$  and  $\mathbf{w} \cdot \mathbf{x} < 0$  **then**

        |  $\mathbf{w} = \mathbf{w} + \mathbf{x}$  ;

**end**

**if**  $\mathbf{x} \in N$  and  $\mathbf{w} \cdot \mathbf{x} \geq 0$  **then**

        |  $\mathbf{w} = \mathbf{w} - \mathbf{x}$  ;

**end**

**end**

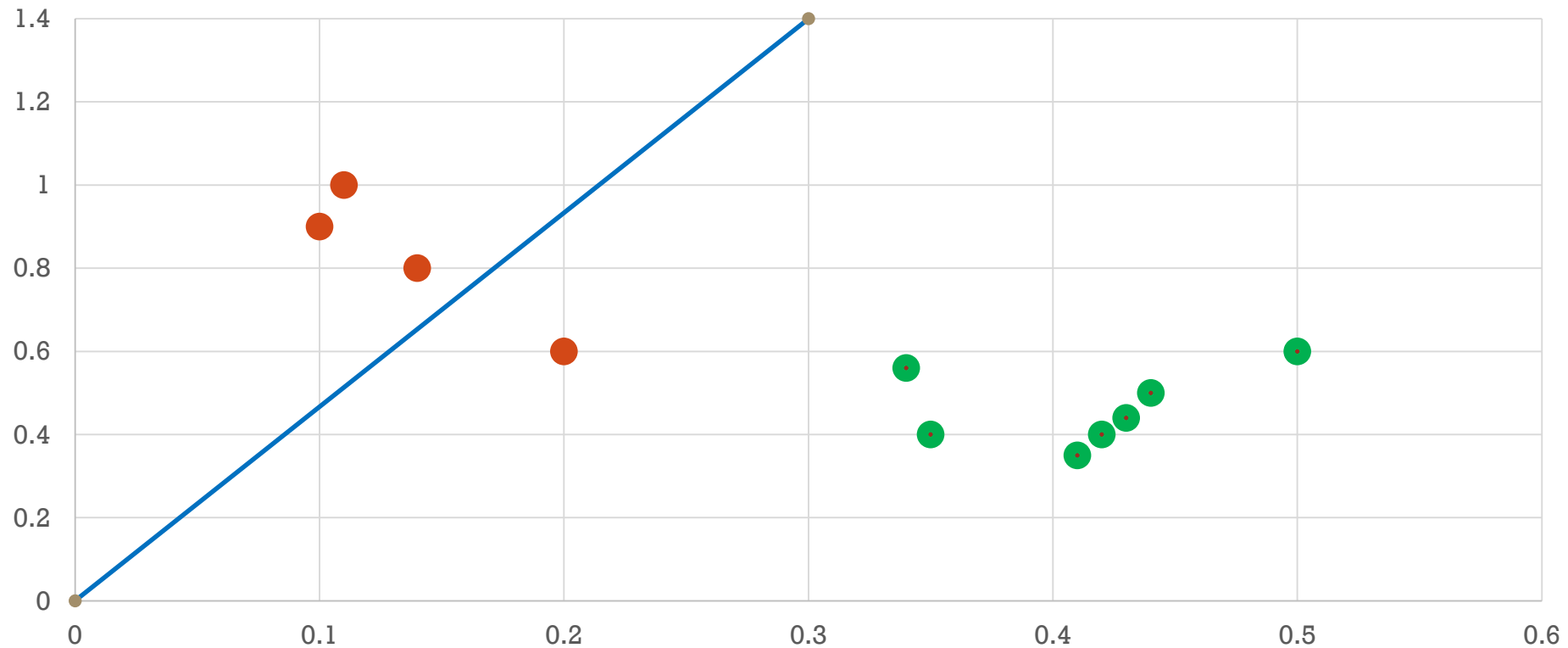
//the algorithm converges when all the  
inputs are classified correctly

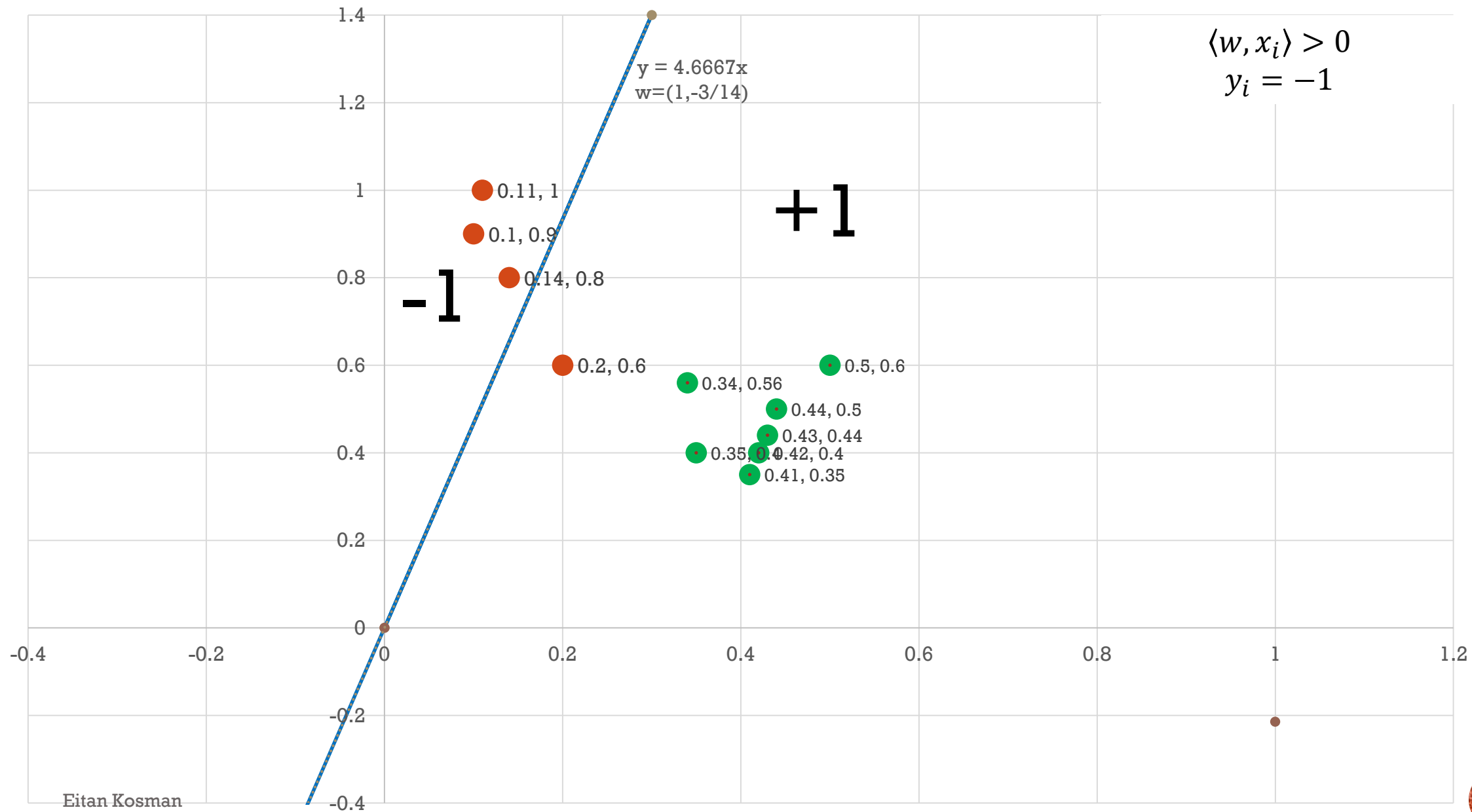
---

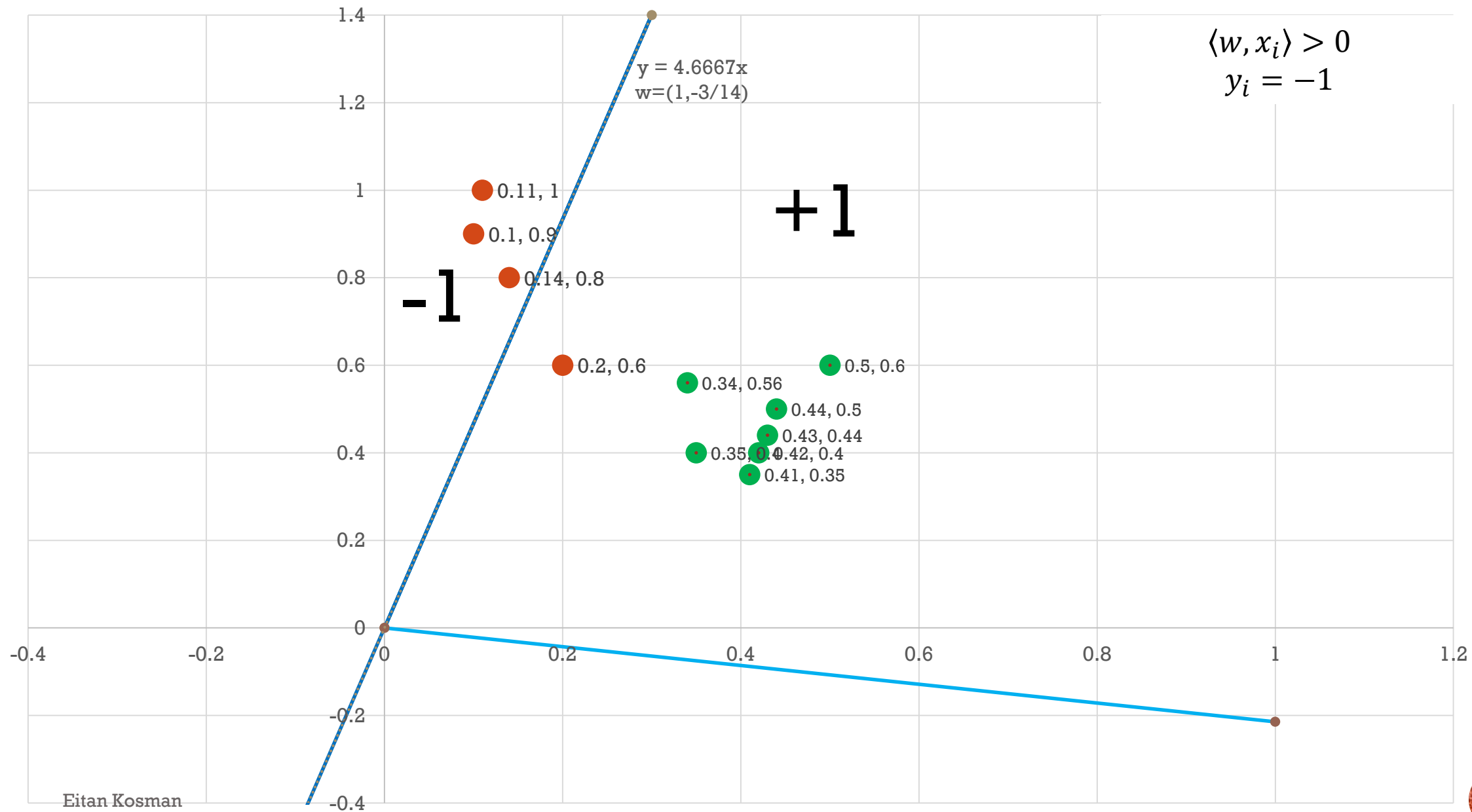


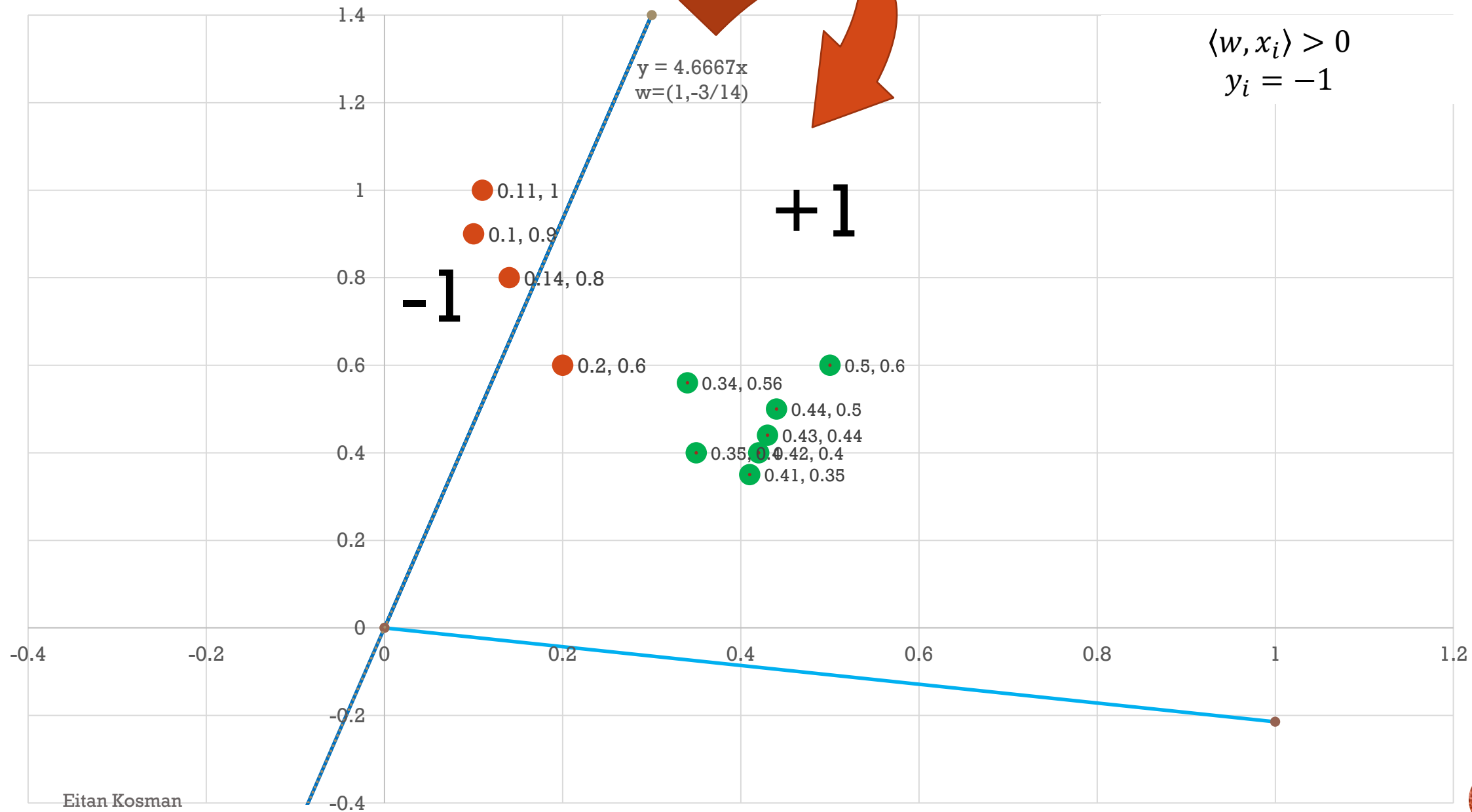
# WEIGHTS UPDATE : INTUITION

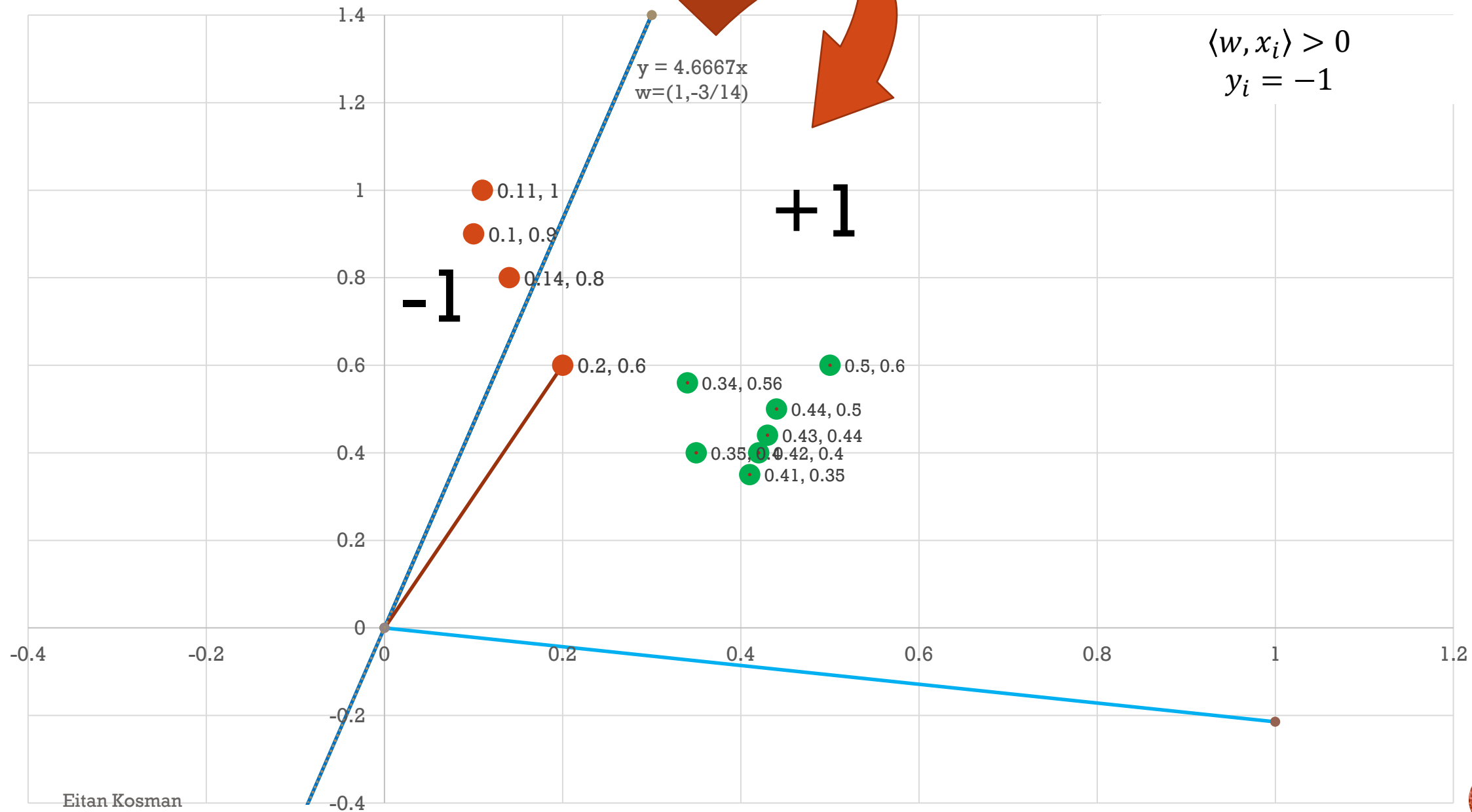
The orange points are from class -1 and the green points are from class +1.  
How would we update the decision line so that it classifies all the points correctly?

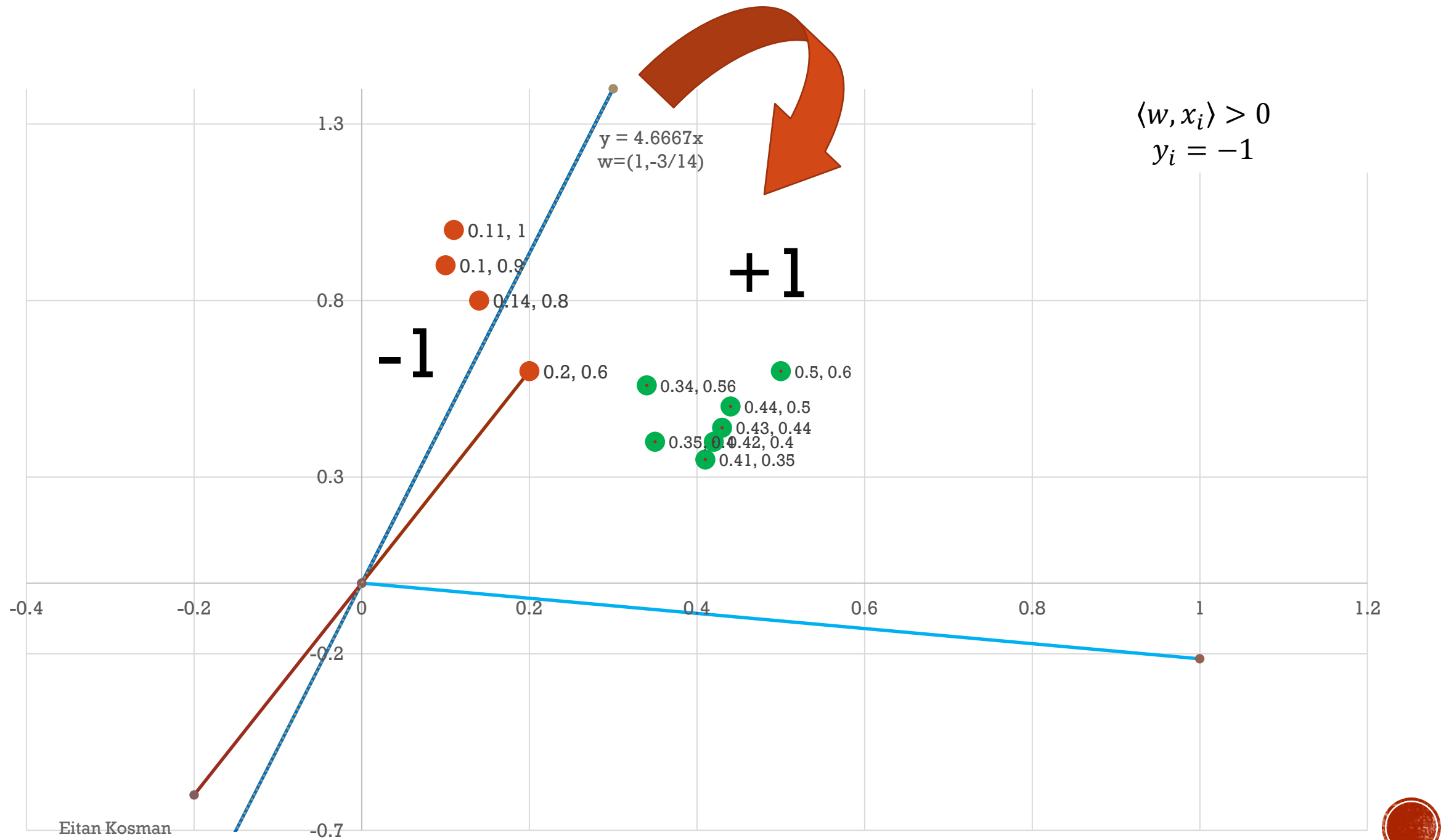


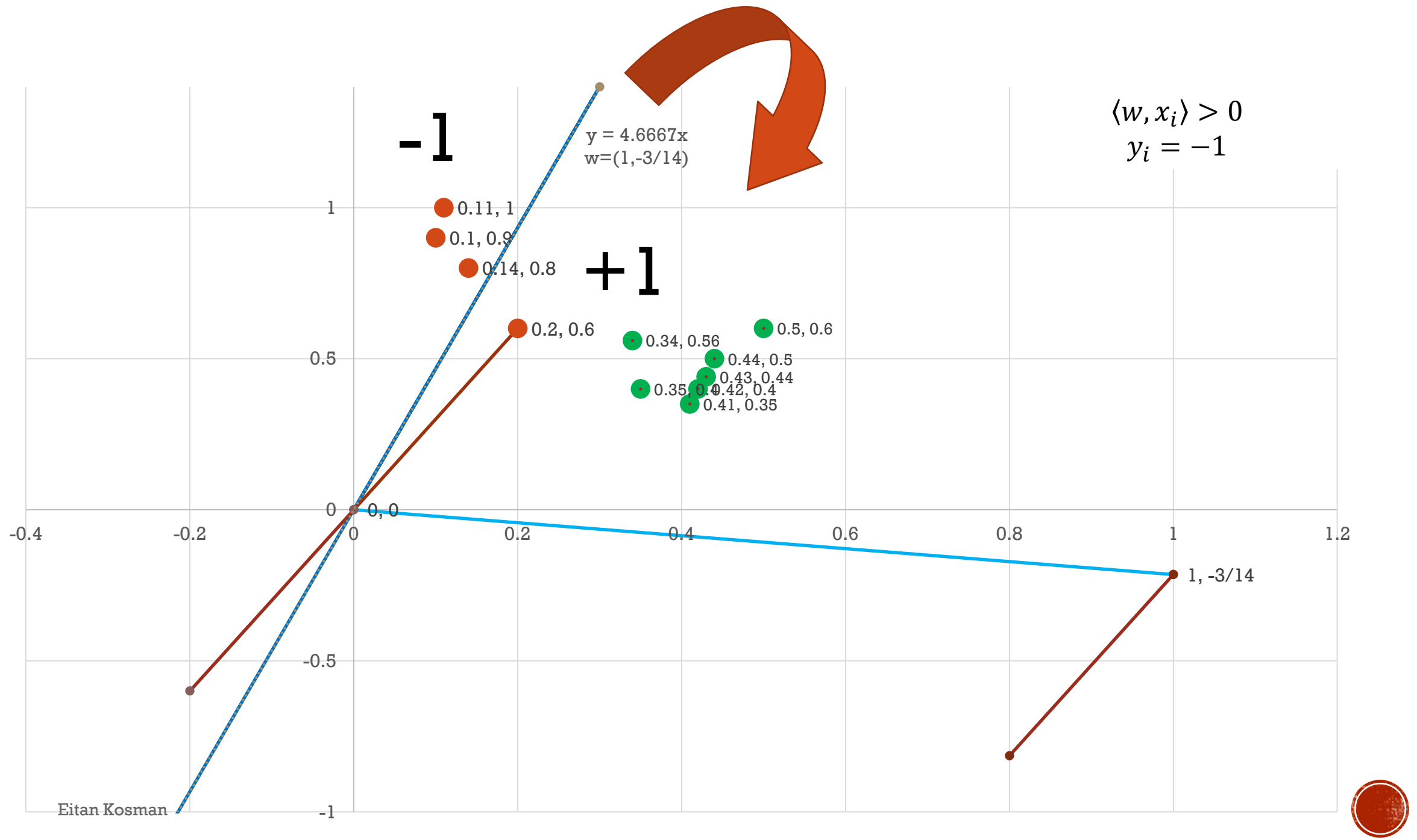




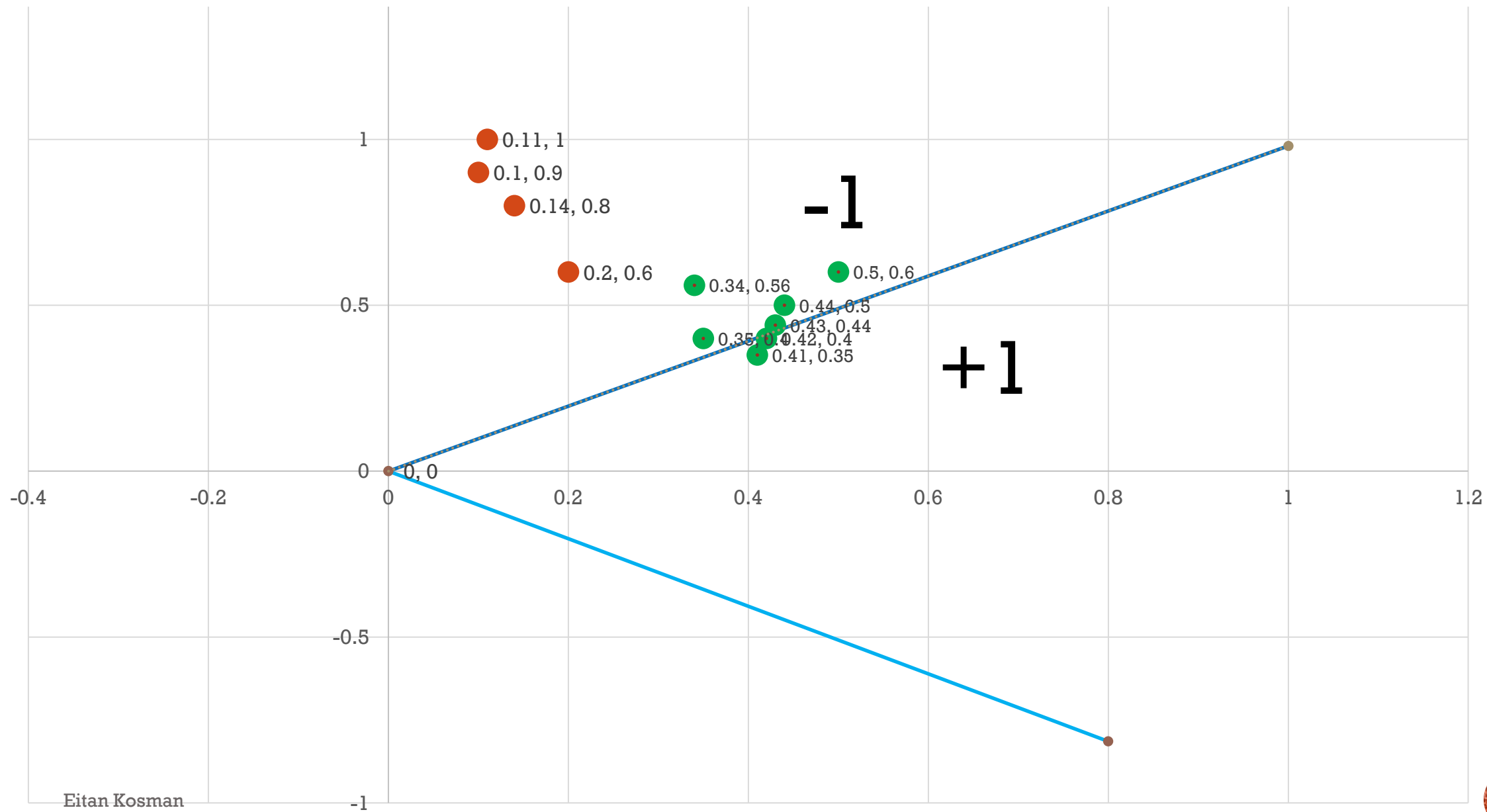


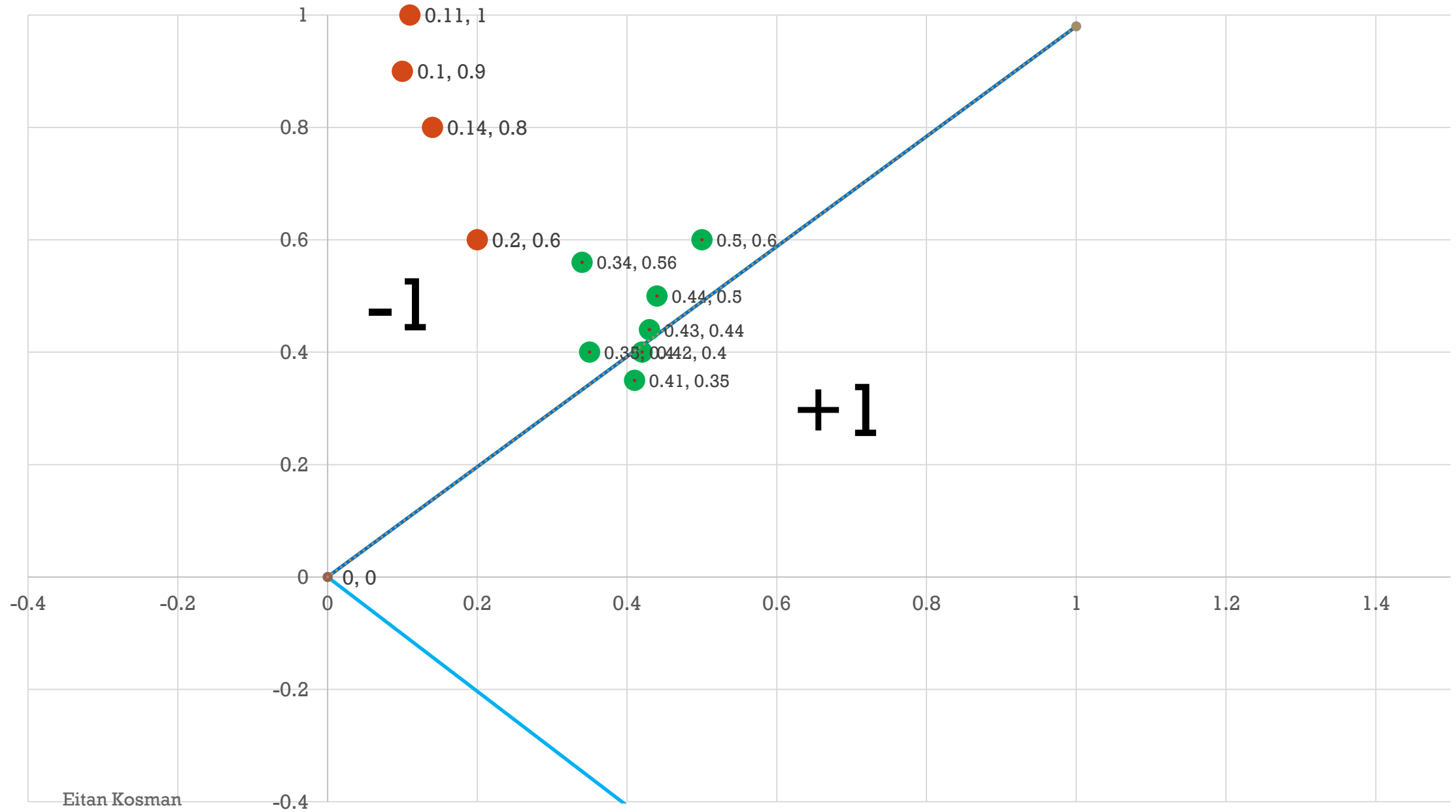












# MISTAKE BOUND

Theorem:

Let  $(x_1, y_1), \dots, (x_n, y_n)$ , where  $x_i \in \mathbb{R}^N$  and  $y_i \in \{-1, 1\}$  be a sequence of labeled examples and assume it is linearly separable.

Denote:

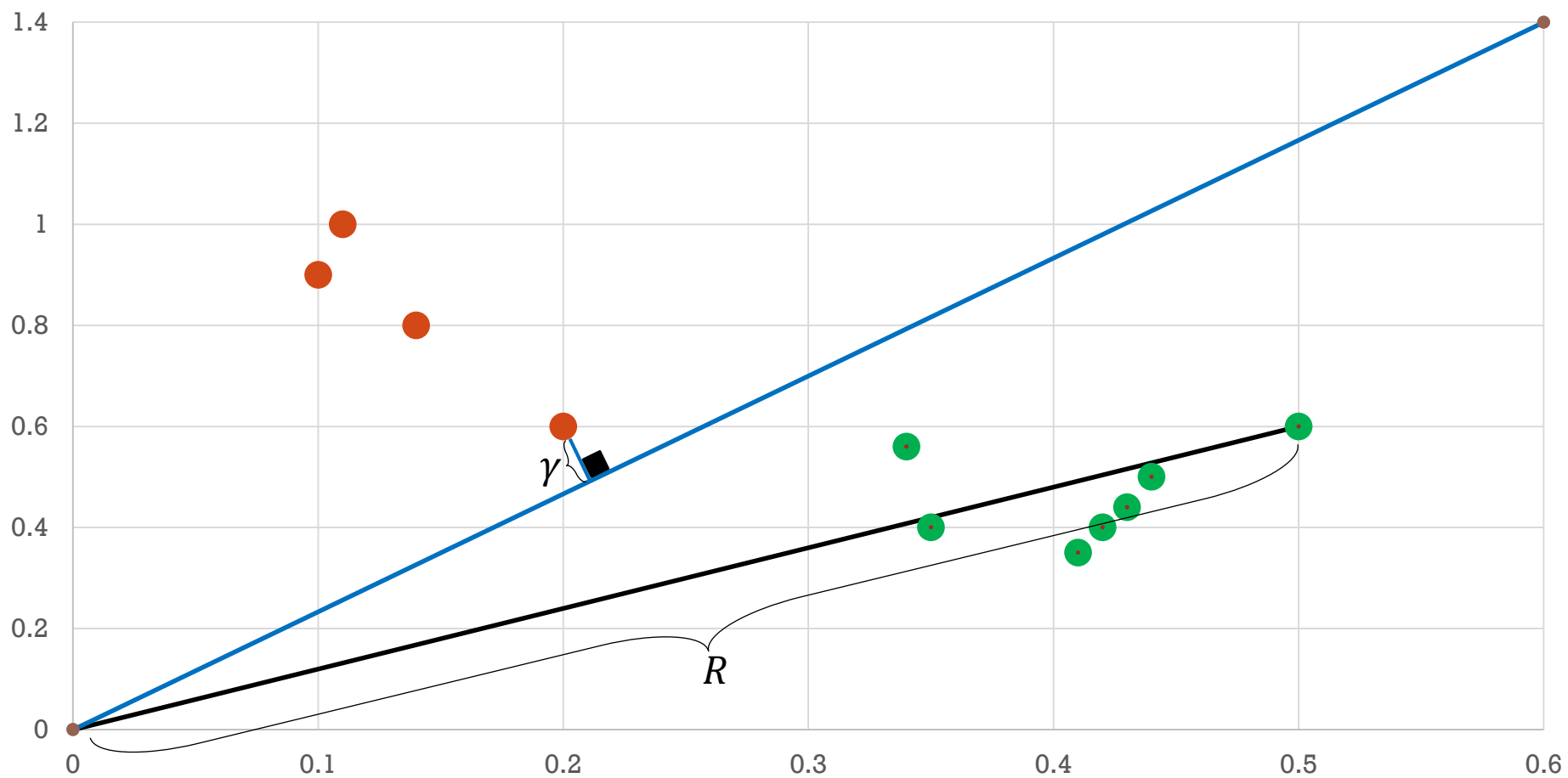
$$R = \max_i ||x_i||$$

Suppose there exists a vector  $w^*$ ,  $\gamma > 0$  such that  $||w^*|| = 1$  and  $\forall i, y_i(w^{*T} x_i) \geq \gamma$ , then the number of mistakes made by the Perceptron algorithm of this sequence of example is  $O\left(\left(\frac{R}{\gamma}\right)^2\right)$



$$R = \max_i \|x_i\|$$

$$\forall i, y_i(w^{*T} x_i) \geq \gamma$$



# MISTAKE BOUND

Let  $w_1 = 0$  (initial weight vector) and denote  $w_k$  the weight vector after the  $k'$ th mistake.

Lemma 1:  $w_{t+1} \cdot w^* \geq w_t \cdot w^* + \gamma$

Lemma 2:  $\|w_{t+1}\|^2 \leq \|w_t\|^2 + R^2$



# MISTAKE BOUND

Lemma 1:  $w_{t+1} \cdot w^* \geq w_t \cdot w^* + \gamma$

The  $t$ 's update occurred when the perceptron did a mistake on sample  $(x_i, y_i)$ .

If  $y_i = 1$ :

$$w_{t+1} \cdot w^* = (w_t + x_i) \cdot w^* = w_t \cdot w^* + \underbrace{x_i \cdot w^*}_{\geq \gamma} = w_t \cdot w^* + \gamma$$

If  $y_i = -1$ :

$$w_{t+1} \cdot w^* = (w_t - x_i) \cdot w^* = w_t \cdot w^* - \underbrace{x_i \cdot w^*}_{\geq \gamma} = w_t \cdot w^* + \gamma$$



# MISTAKE BOUND

Lemma 2:  $||w_{t+1}||^2 \leq ||w_t||^2 + R^2$

The  $t$ 's update occurred when the perceptron did a mistake on sample  $(x_i, y_i)$ .

If  $y_i = 1$ :

$$||w_{t+1}||^2 = ||w_t + x_i||^2 = ||w_t||^2 + 2 \underbrace{w_t \cdot x_i}_{\substack{< 0, \text{ since} \\ \text{a mistake} \\ \text{has occurred}}} + \underbrace{||x_i||^2}_{\leq R^2} \leq ||w_t||^2 + R^2$$

If  $y_i = -1$ :

$$||w_{t+1}||^2 = ||w_t - x_i||^2 = ||w_t||^2 - 2 \underbrace{w_t \cdot x_i}_{\substack{> 0, \text{ since} \\ \text{a mistake} \\ \text{has occurred}}} + \underbrace{||x_i||^2}_{\leq R^2} \leq ||w_t||^2 + R^2$$



# MISTAKE BOUND

Now, equipped with the two lemmas, we know that from Lemma 1:

$$\begin{aligned}w_1 &= \bar{0} \\w_2 \cdot w^* &\geq w_1 \cdot w^* + \gamma = \gamma \\w_3 \cdot w^* &\geq w_2 \cdot w^* + \gamma \geq \gamma + \gamma = 2\gamma\end{aligned}$$

Assume:  $w_t \cdot w^* \geq (t - 1) \cdot \gamma$

Thus –

$$w_{t+1} \cdot w^* \geq w_t \cdot w^* + \gamma \geq (t - 1) \cdot \gamma + \gamma = t \cdot \gamma$$

Moreover, from lemma 2:

$$\begin{aligned}|w_1| &= 0 \\|w_2|^2 &\leq |w_1|^2 + R^2 = R^2 \\|w_3|^2 &\leq |w_2|^2 + R^2 \leq R^2 + R^2 = 2R^2\end{aligned}$$

Assume:  $|w_t|^2 \leq (t - 1)R^2$

Thus –

$$|w_{t+1}|^2 \leq |w_t|^2 + R^2 \leq (t - 1)R^2 + R^2 = tR^2$$





# MISTAKE BOUND

Recap:

After  $T$  mistakes:

$$w_{T+1} \cdot w^* \geq T \cdot \gamma$$

$$|w_{T+1}|^2 \leq TR^2$$

$\Downarrow$

$$\gamma T \leq \underbrace{w_{T+1} \cdot w^*}_{\text{scalar}} = |w_{T+1} \cdot w^*| \stackrel{\substack{\text{Cauchy} \\ \text{Schwarz}}}{\leq} |w_{T+1}| \cdot \underbrace{|w^*|}_{=1} = |w_{T+1}|$$

$\Downarrow$

$$\gamma^2 T^2 \leq |w_{T+1}|^2 \leq TR^2$$

$\Downarrow$

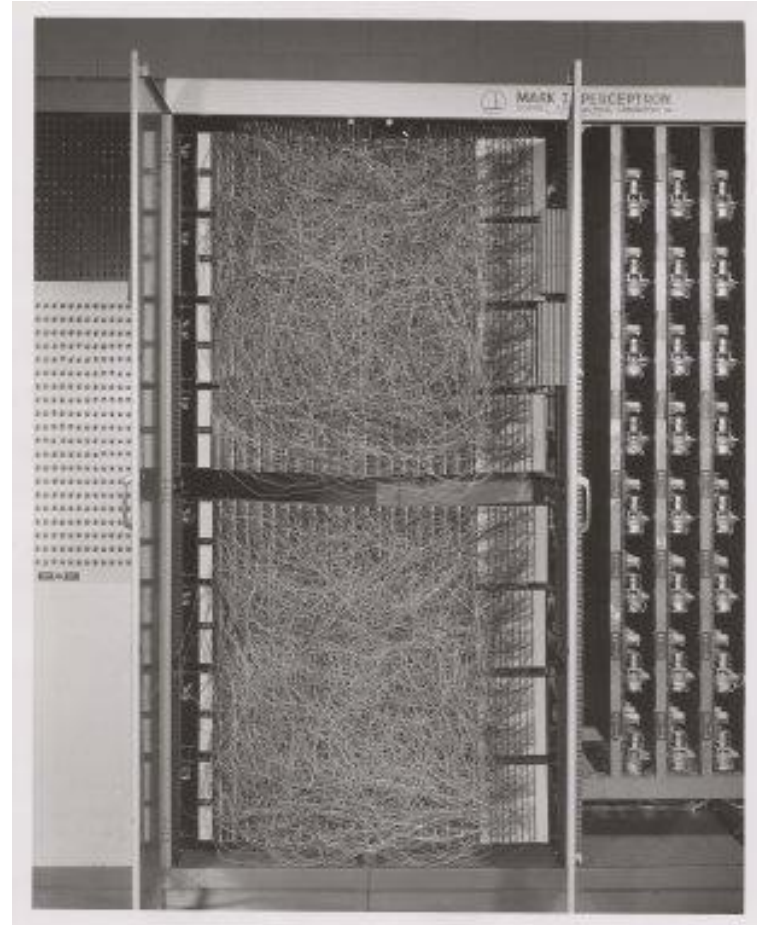
$$T \leq \frac{R^2}{\gamma^2}$$



# THE FALL OF THE PERCEPTRON

The first computer built around the concept of perceptron looked like this.

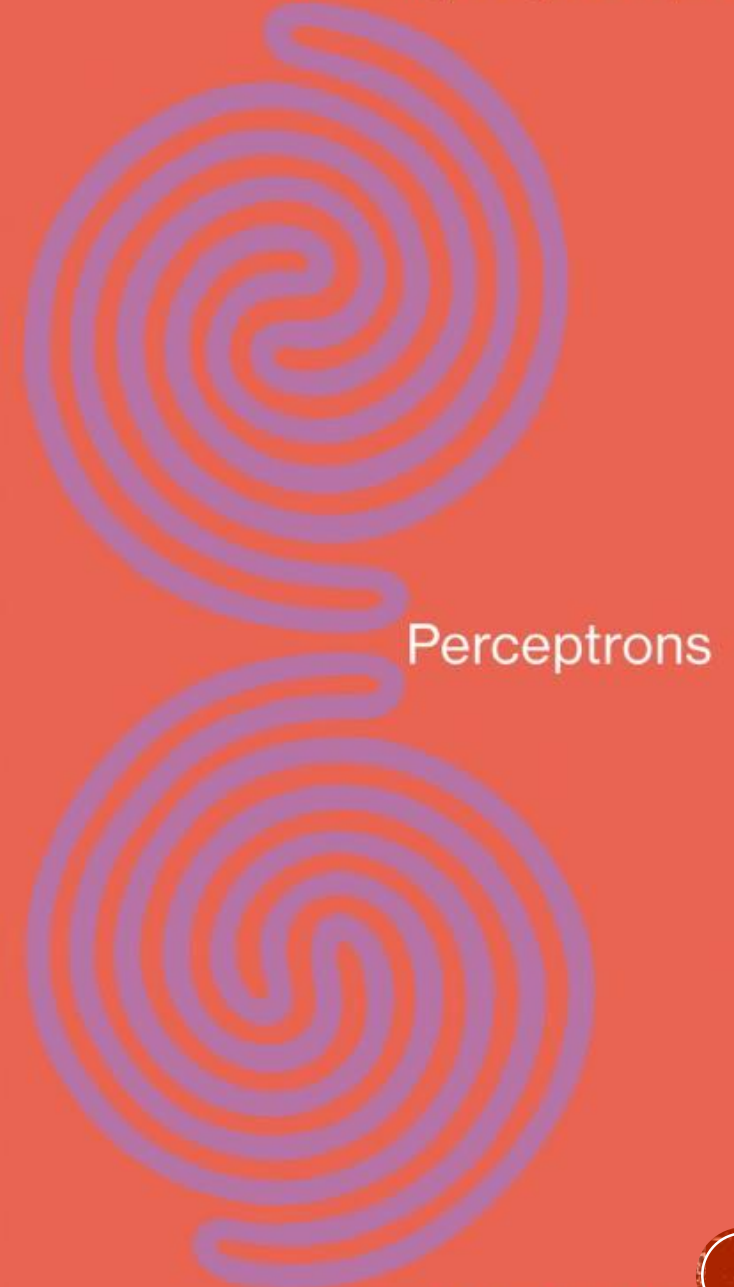
Even the wiring was supposed to simulate the connections of neurons.



# THE FALL OF THE PERCEPTRON

However, a paper describing the perceptron's shortcomings, particularly that it was effective only at solving simple problems, led to a drastic drop in interest in artificial neural networks in the 1960's.

Unless input categories were “linearly separable”, a perceptron could not learn to discriminate between them. **Example:**



# XOR

