

# CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019



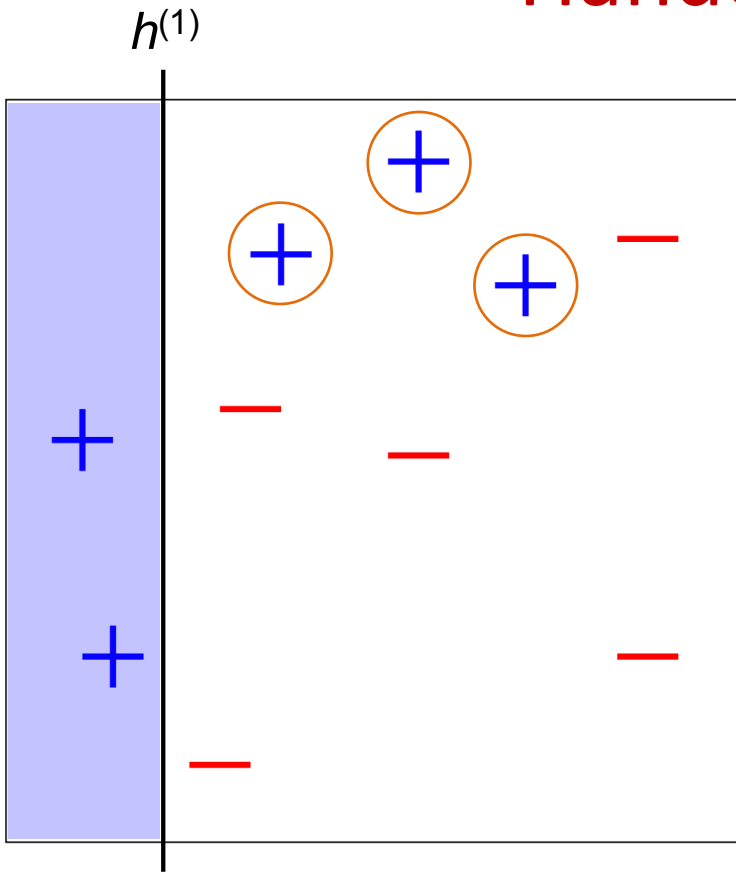
# Outline for October 29

- Recap AdaBoost and weighted entropy
  - Perceptron Algorithm
  - Mid-semester feedback
- 
- Reading quiz Thursday: Duame 4.1 (2 pages)
  - Lab check in Thursday: Parts 1&2 complete

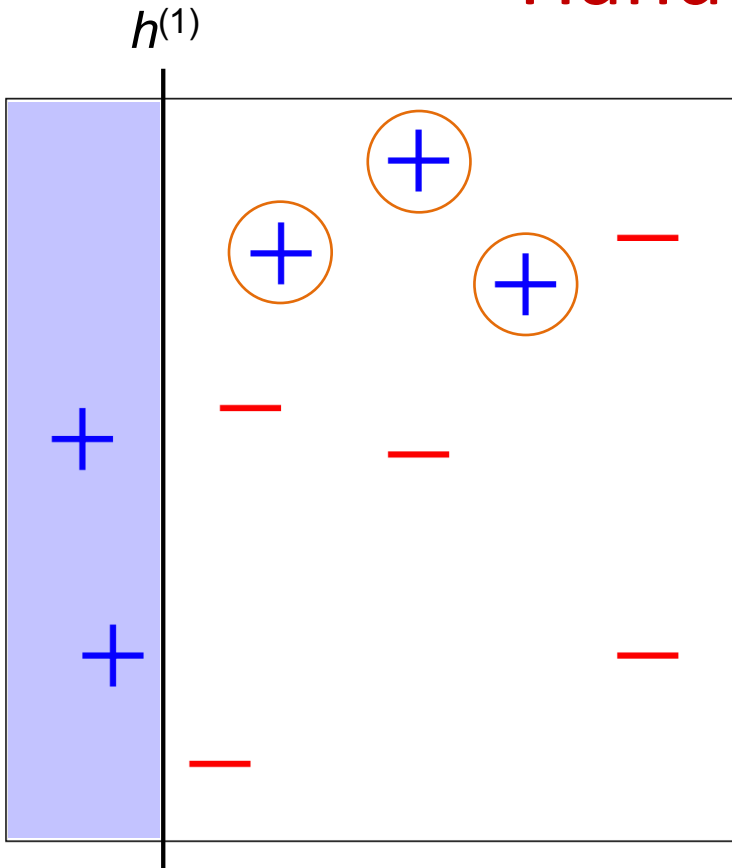
# Outline for October 29

- Recap AdaBoost and weighted entropy
- Perceptron Algorithm
- Mid-semester feedback

# Handout 14: Round 1



# Handout 14: Round 1



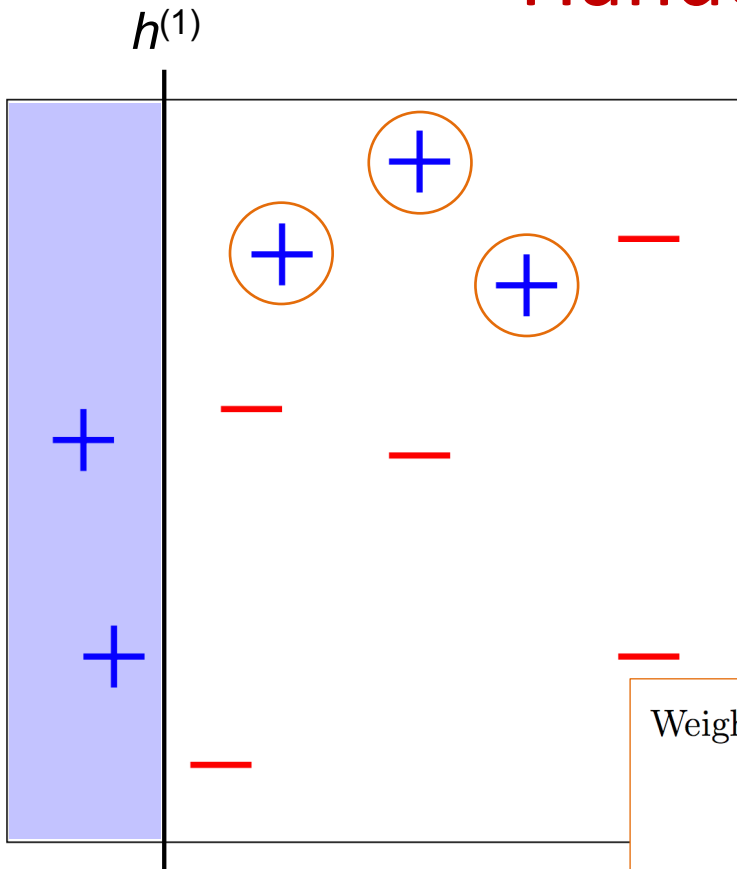
$$w_i^{(1)} = \frac{1}{10} \text{ for all } i = 1, 2, \dots, 10.$$

$$\epsilon_1 = \frac{3}{10} \text{ (three points incorrectly classified, all with weight } \frac{1}{10}\text{)}$$

$$\alpha_1 = \frac{1}{2} \ln \left( \frac{1 - \frac{3}{10}}{\frac{3}{10}} \right) = \ln \sqrt{\frac{7}{3}} \approx 0.42$$

- correctly classified:  $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left( -\ln \sqrt{\frac{7}{3}} \right)$
- incorrectly classified:  $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left( \ln \sqrt{\frac{7}{3}} \right)$

# Handout 14: Round 1



$$w_i^{(1)} = \frac{1}{10} \text{ for all } i = 1, 2, \dots, 10.$$

$$\epsilon_1 = \frac{3}{10} \text{ (three points incorrectly classified, all with weight } \frac{1}{10})$$

$$\alpha_1 = \frac{1}{2} \ln \left( \frac{1 - \frac{3}{10}}{\frac{3}{10}} \right) = \ln \sqrt{\frac{7}{3}} \approx 0.42$$

- correctly classified:  $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left( -\ln \sqrt{\frac{7}{3}} \right)$
- incorrectly classified:  $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left( \ln \sqrt{\frac{7}{3}} \right)$

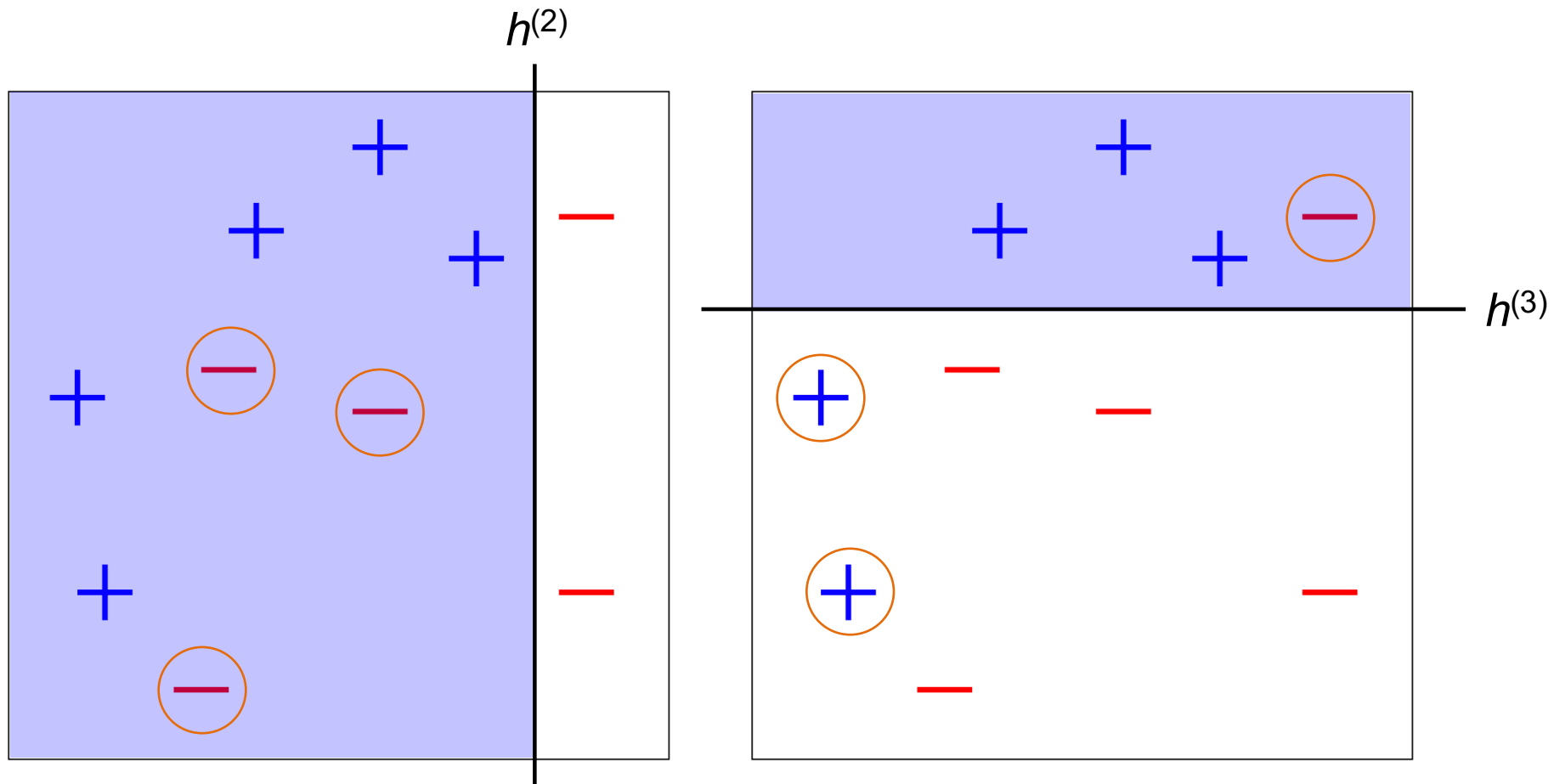
Weights must sum to 1,  $\Rightarrow$

$$7 \cdot \frac{c_1}{10} \exp \left( -\ln \sqrt{\frac{7}{3}} \right) + 3 \cdot c_1 \cdot \frac{1}{10} \exp \left( \ln \sqrt{\frac{7}{3}} \right) = 1$$

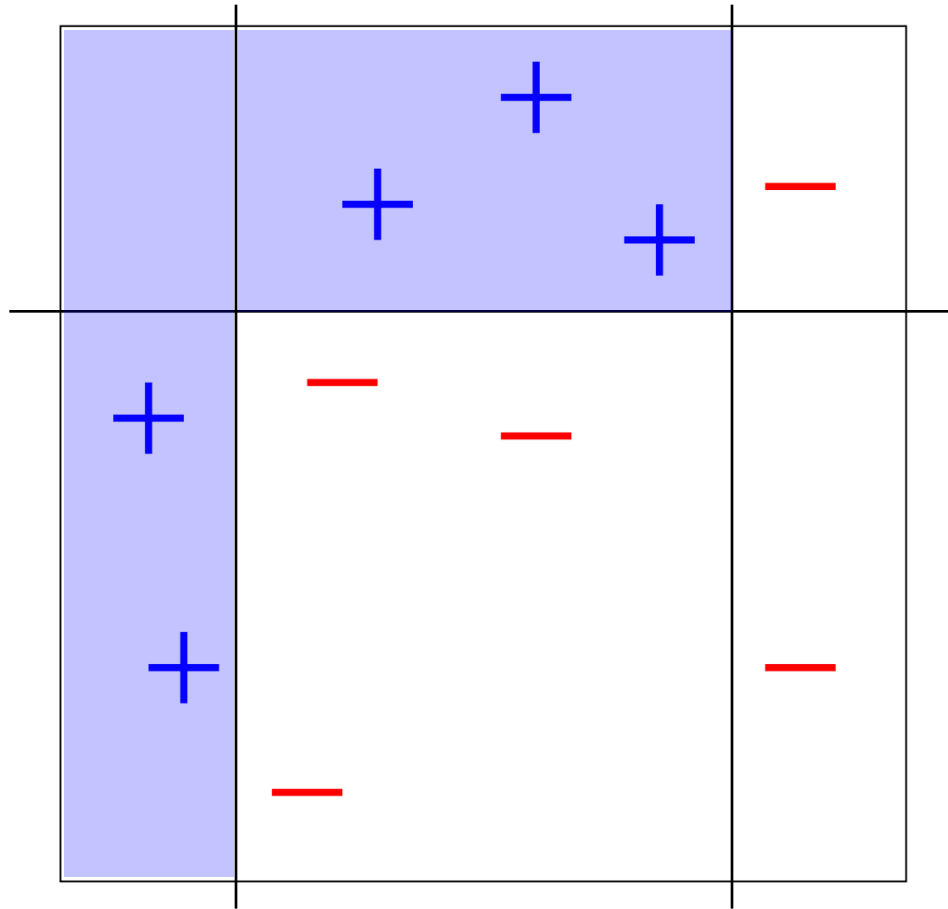
$$\Rightarrow c_1 = \frac{5}{\sqrt{21}}$$

- correctly classified:  $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{3}{7}} = \frac{1}{14}$  decrease!
- incorrectly classified:  $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{7}{3}} = \frac{1}{6}$  increase!

# Handout 14: Round 2 & 3 (exercise!)



# Handout 14: final classifier



$$h(\mathbf{x}) = \text{sign}\left(0.42 \cdot h^{(1)}(\mathbf{x}) + 0.65 \cdot h^{(2)}(\mathbf{x}) + 0.92 \cdot h^{(3)}(\mathbf{x})\right)$$



# Decision Trees with Weighted

## Entropy

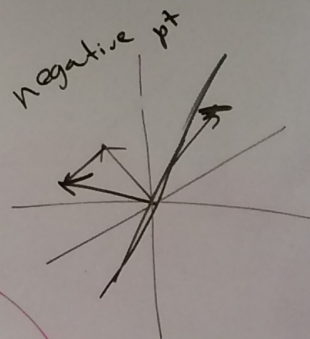
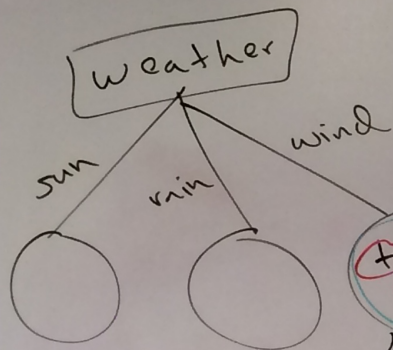
$$H(Y|X)$$

$$H(Y|X_j=v) = - \sum_{c \in \text{vals}(Y)} \underbrace{P(Y=c|X_j=v)}_{\substack{1 \text{ if criterion} \\ 0 \text{ o.w.}}} \log_2 \underbrace{P(Y=c|X_j=v)}$$

$$P(Y=c|X_j=v) = \frac{\sum_{i=1}^n \underbrace{w_i^{(t)}}_{\text{weight}} \mathbb{1}(X_{ij}=v, Y_i=c)}{\sum_{i=1}^n \underbrace{w_i^{(t)}}_{\text{weight}} \mathbb{1}(X_{ij}=v)}$$

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

+	-
-	+

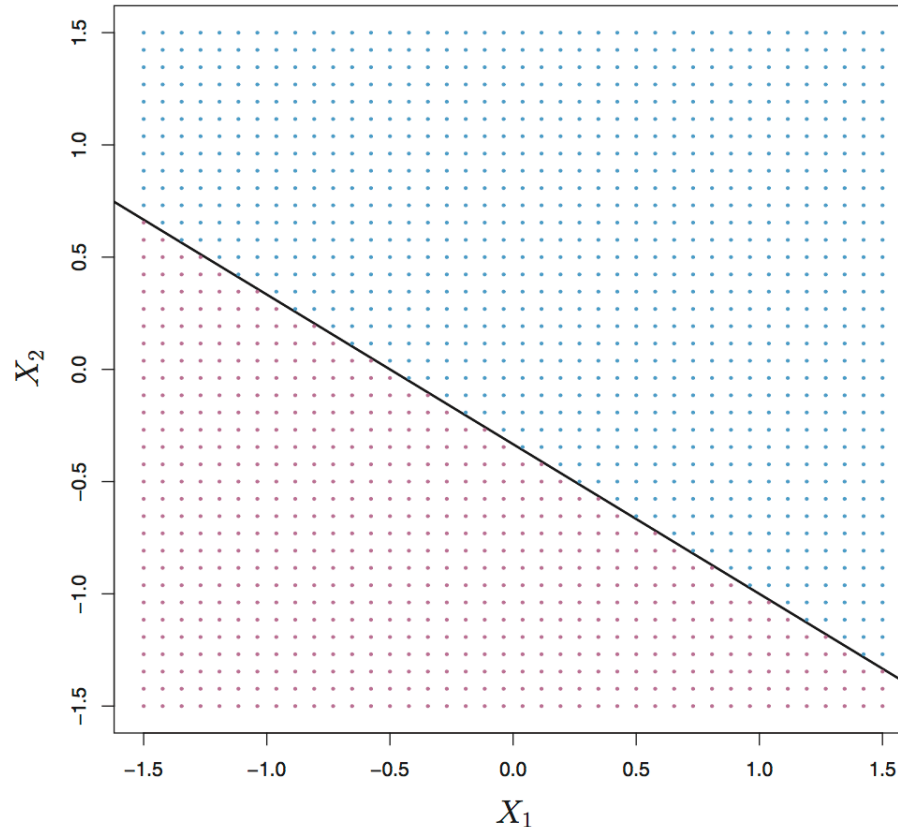


$$\rightarrow \underbrace{P(\text{label} = 1)}_{\text{positive}} = \frac{\sum_{\text{data } i \text{ in leaf}} w_i^{(+)} \mathbb{1}(y_i = 1)}{\sum_{\text{data } i \text{ in leaf}} w_i^{(+)}}$$

# Outline for October 29

- Recap AdaBoost and weighted entropy
- **Perceptron Algorithm**
- Mid-semester feedback

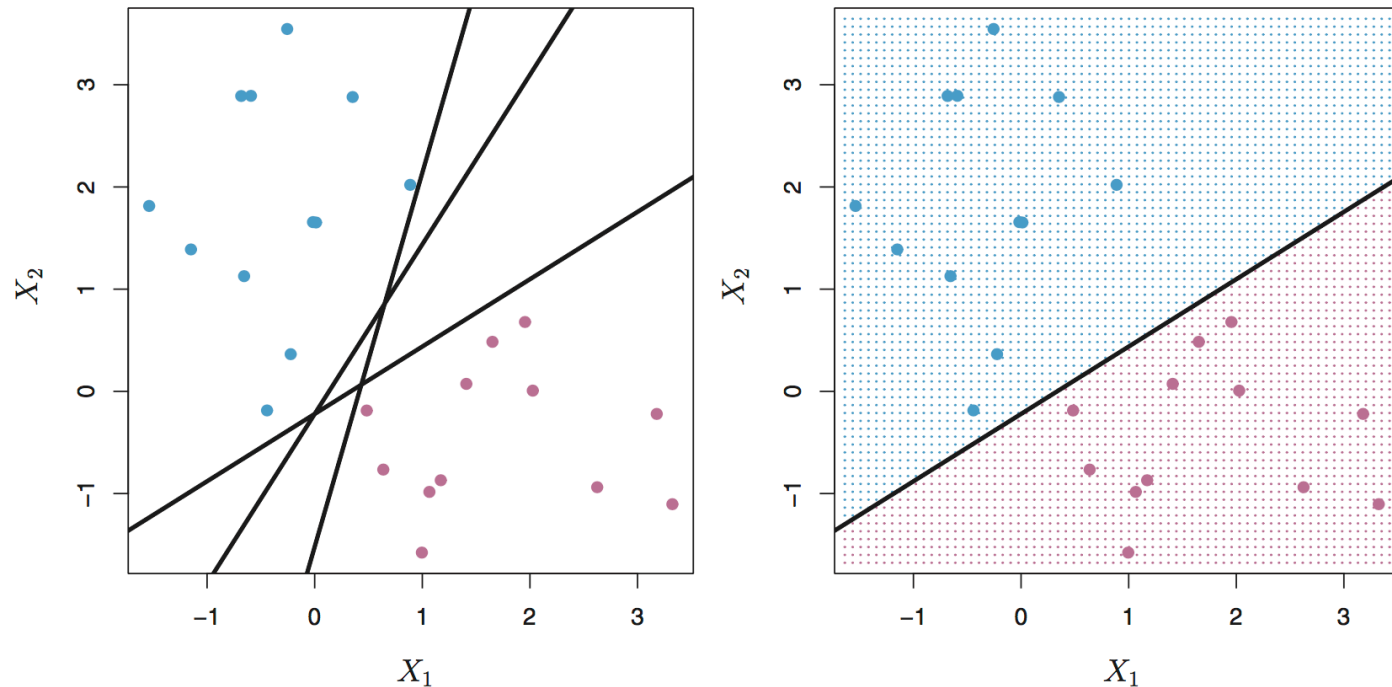
# Hyperplane divides space into positive (+1) and negative (-1)



**FIGURE 9.1.** *The hyperplane  $1 + 2X_1 + 3X_2 = 0$  is shown. The blue region is the set of points for which  $1 + 2X_1 + 3X_2 > 0$ , and the purple region is the set of points for which  $1 + 2X_1 + 3X_2 < 0$ .*

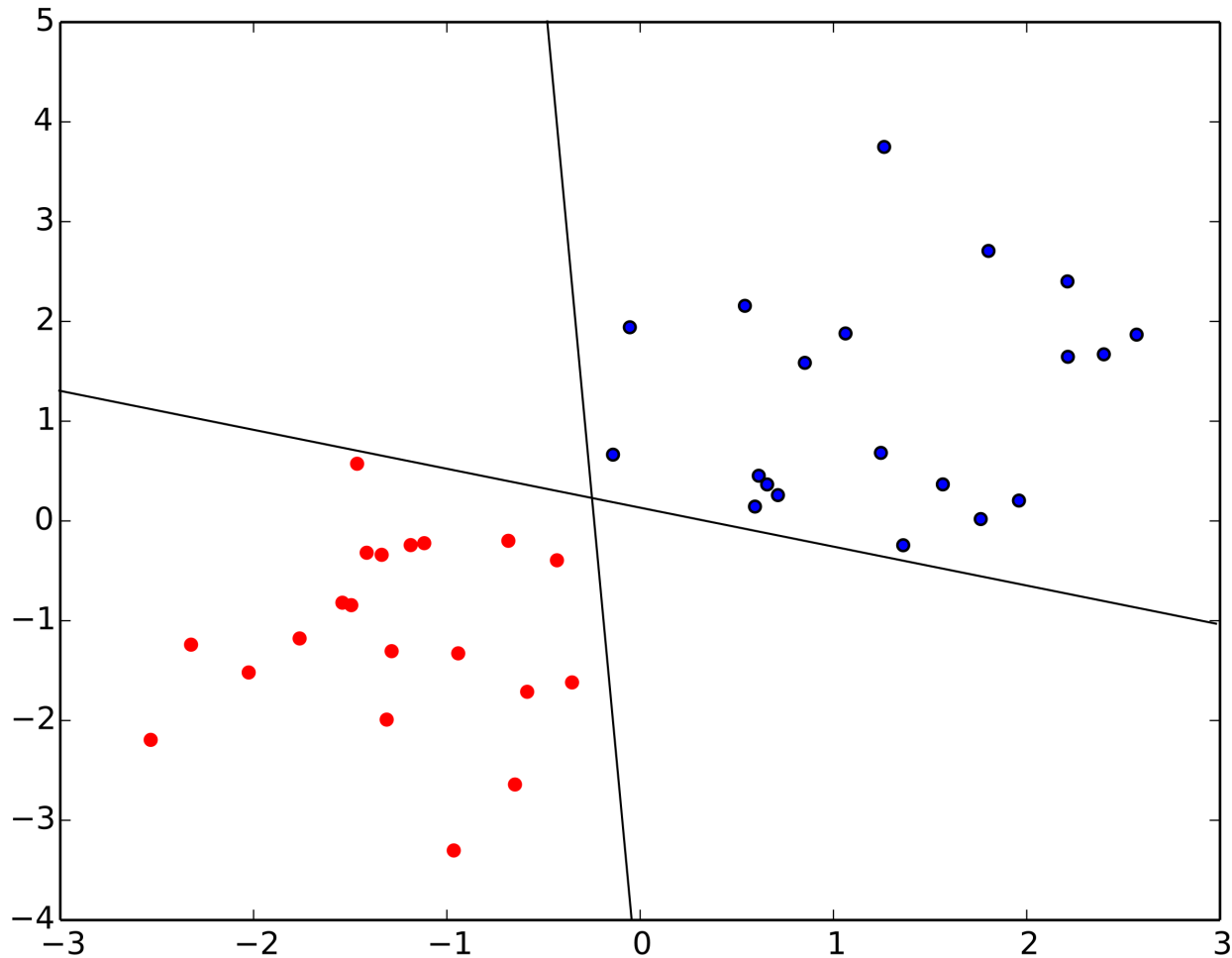


# Goal: use training data to create a *separating hyperplane*



**FIGURE 9.2.** Left: There are two classes of observations, shown in blue and in purple, each of which has measurements on two variables. Three separating hyperplanes, out of many possible, are shown in black. Right: A separating hyperplane is shown in black. The blue and purple grid indicates the decision rule made by a classifier based on this separating hyperplane: a test observation that falls in the blue portion of the grid will be assigned to the blue class, and a test observation that falls into the purple portion of the grid will be assigned to the purple class.

These two hyperplanes would likely perform very differently on test data, but they both separate the training data



# Perceptron Algorithm

(similar to logistic regression)

$$y \in \{-1, +1\}$$

goal:

model  $h(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x})$

$$\text{if } \vec{w} \cdot \vec{x} > 0 \Rightarrow \hat{y} = 1$$

$$\text{if } \vec{w} \cdot \vec{x} \leq 0 \Rightarrow \hat{y} = -1$$

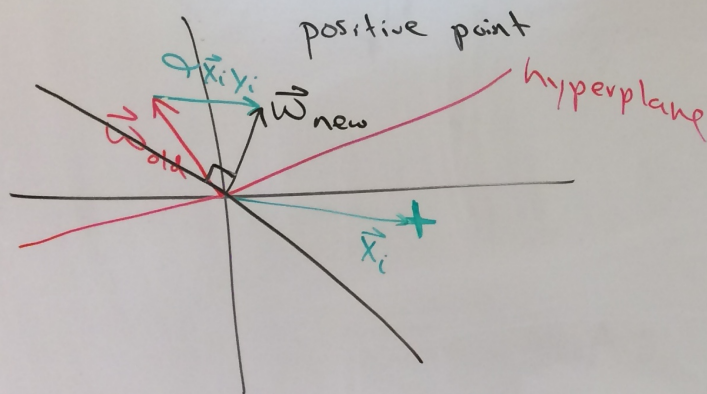
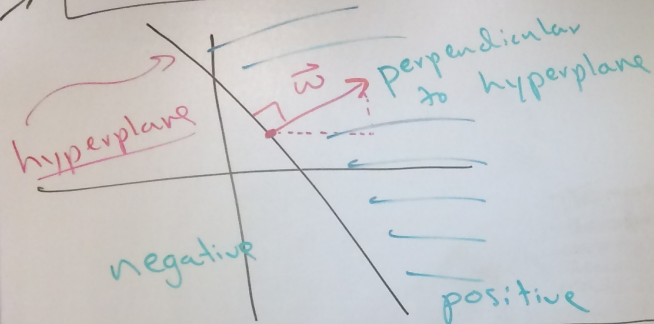
ex:  $\vec{w} = \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$

direction of  $\vec{w}$

$$-5 + 2x_1 + x_2 = 0$$

equation of the hyperplane.

$$\rightarrow x_2 = 5 - 2x_1$$



### Algorithm:

• set  $\vec{w}$  = zero vector  
 repeat until entire  
 training set classified  
 correctly

- ① choose example  $(\vec{x}_i, y_i)$
- ② if correctly classified  $\Rightarrow$   
 do nothing

③ else:

$$\vec{w} \leftarrow \vec{w} + y_i \vec{x}_i$$



consider  $\vec{x}_i$

$$\vec{w} \leftarrow \vec{w} - \frac{\alpha}{2} (h(\vec{x}_i) - y_i) \vec{x}_i$$

$$\vec{w} \leftarrow \vec{w} - \frac{\alpha}{2} (-2) y_i \vec{x}_i$$

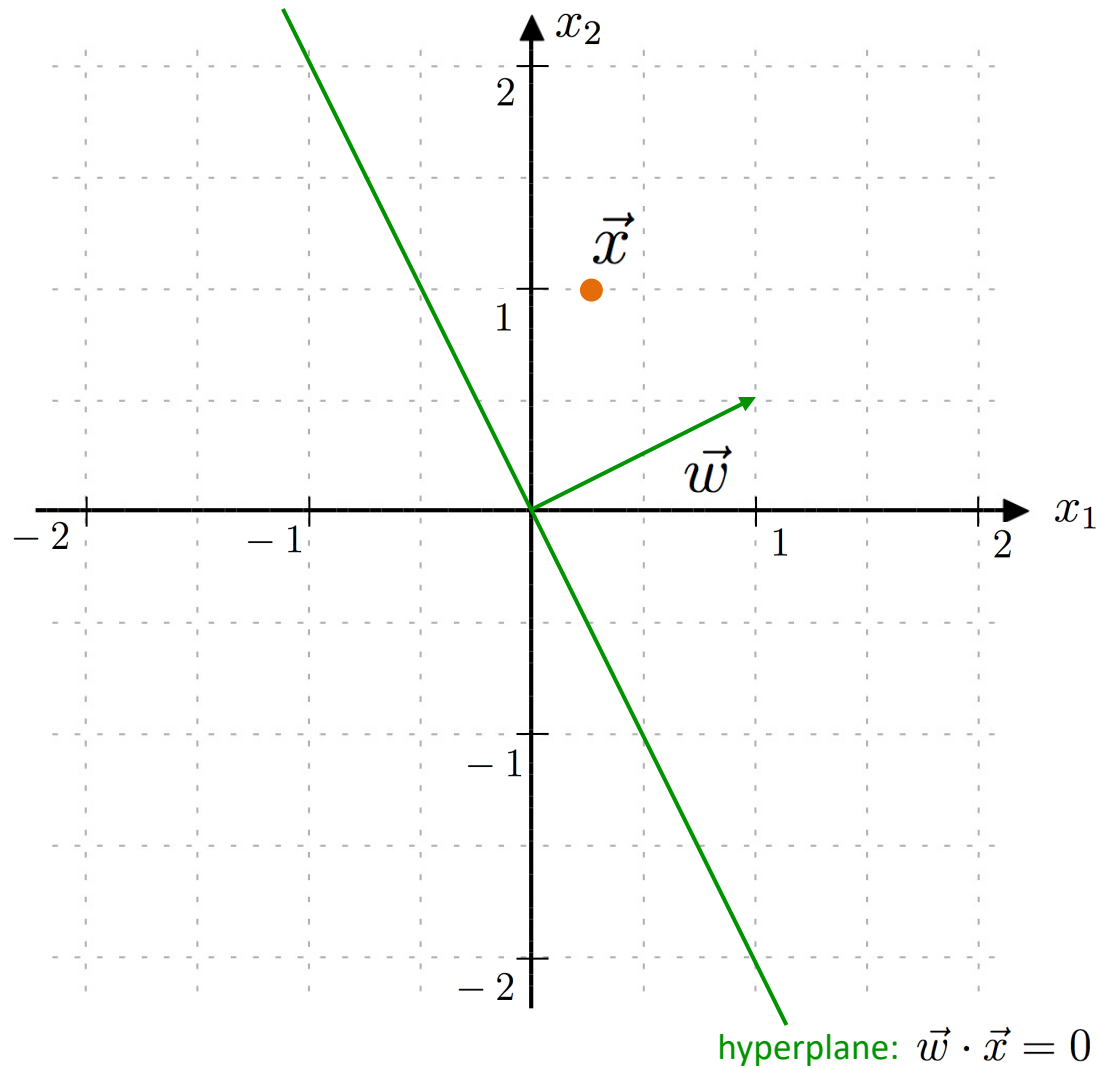
if incorrect

$h(\vec{x}_i)$	$y_i$	$h(\vec{x}_i) - y_i$	$\cdot y_i$
1	-1	2	-2
-1	1	-2	-2

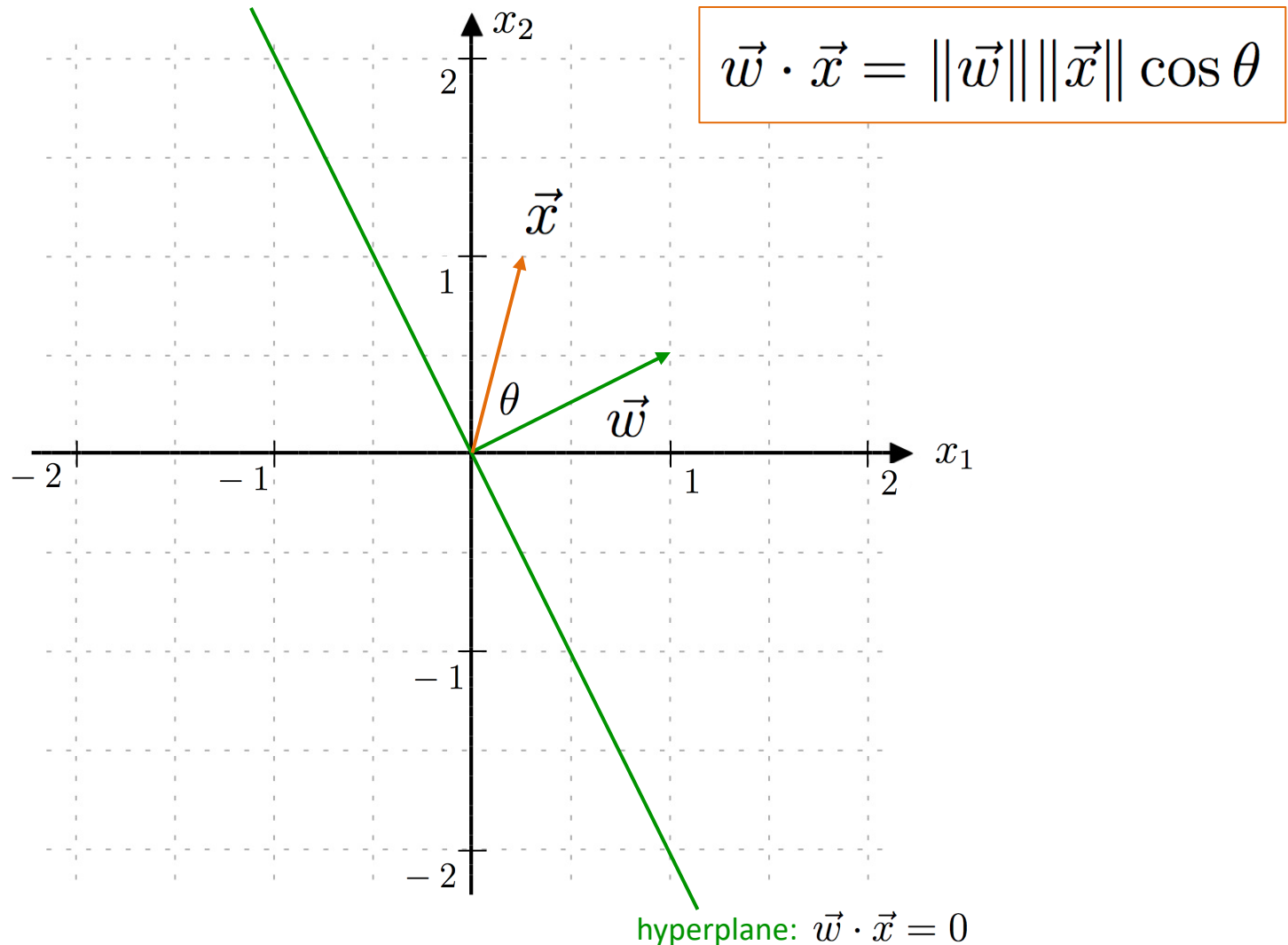
$$\vec{w} \leftarrow \vec{w} + \alpha y_i \vec{x}_i$$

$\alpha = 1$  since  
we only care  
about direction  
of  $\vec{w}$

# Intuition behind the dot product



# Intuition behind the dot product

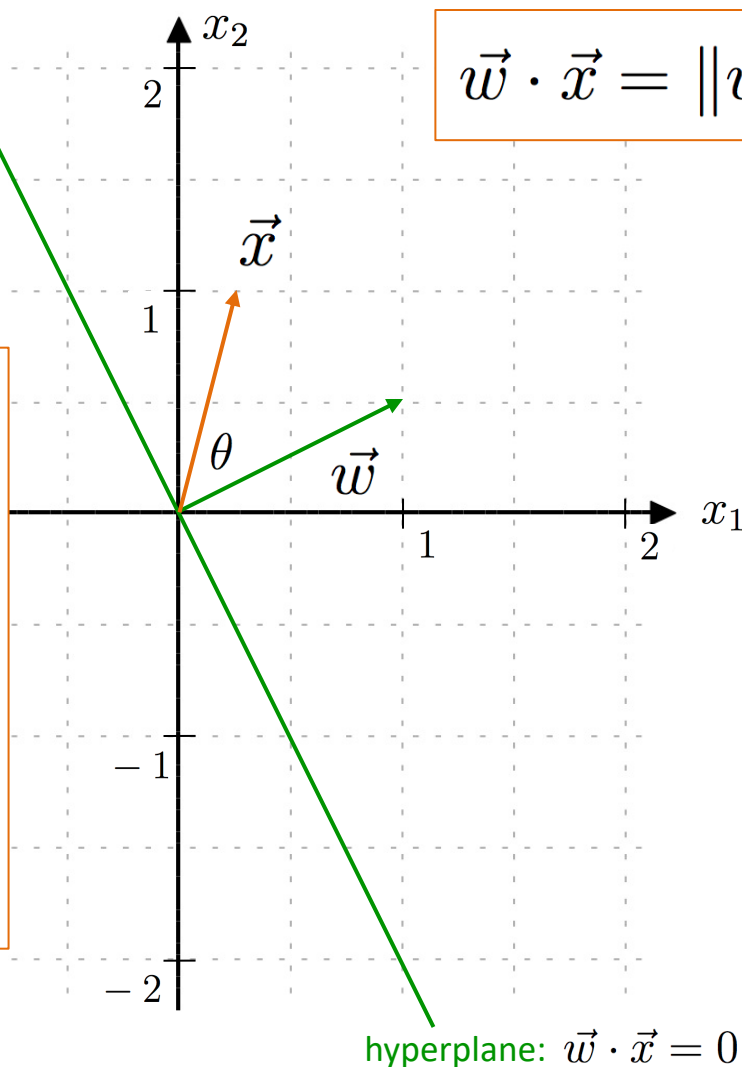


# Intuition behind the dot product

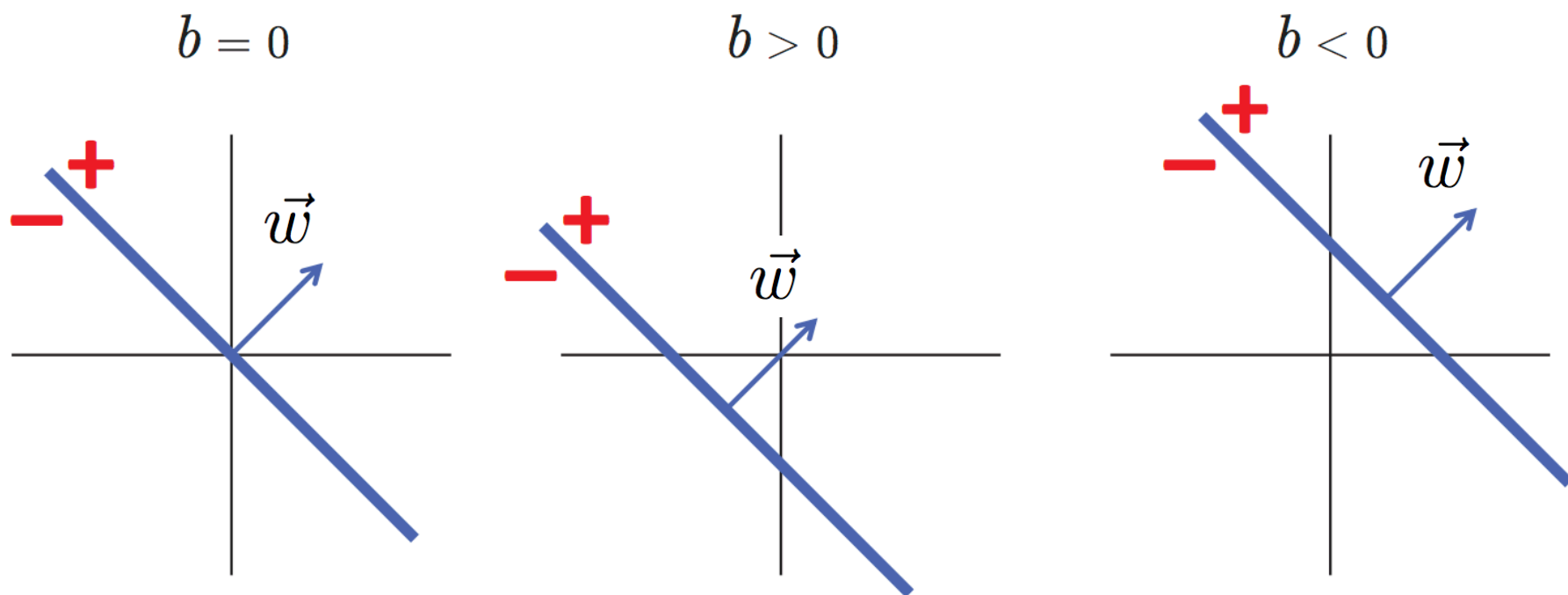
$$\vec{w} \cdot \vec{x} = \|\vec{w}\| \|\vec{x}\| \cos \theta$$

Takeaway: we only care about the sign of the angle between  $\vec{x}$  and  $\vec{w}$

- If  $\cos \theta > 0$ ,  $\vec{x}$  is on the same side of the hyperplane as  $\vec{w}$ , so we classify it as positive
- If  $\cos \theta < 0$ ,  $\vec{x}$  is on the opposite side from  $\vec{w}$ , so we classify it as negative



The **bias** ( $b$ ) and the  $y$ -intercept are different, but they both capture a “shift” away from the origin.



With  $p=2$ , if  $w_2$  is positive, then the above example holds

# Convergence Guarantee

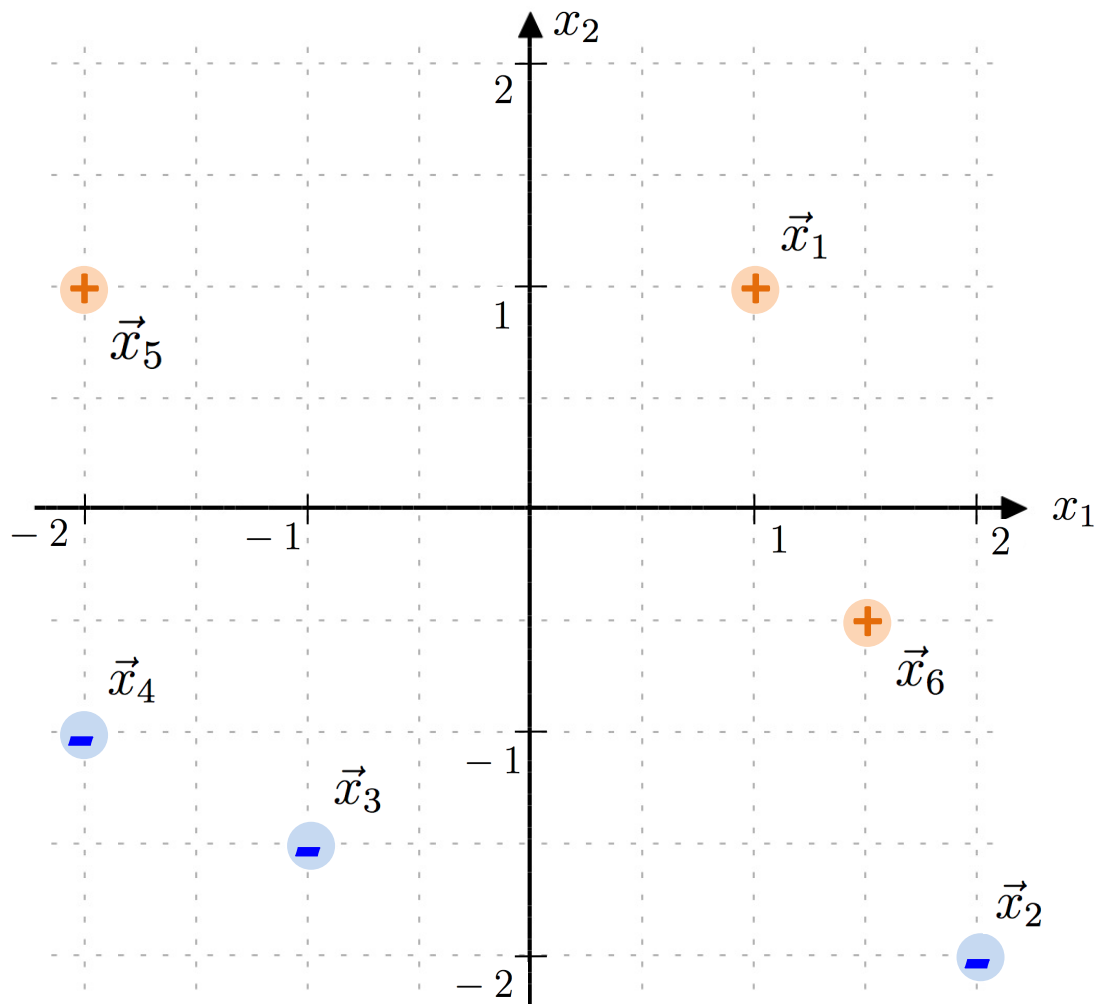
- Perceptron is guaranteed to converge to a solution if a separating hyperplane exists
- Not guaranteed to converge to a “good” solution
- No guarantees about behavior if a separating hyperplane does not exist!

# Handout 15 example

Initial values:

$$\alpha = 0.2$$

$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

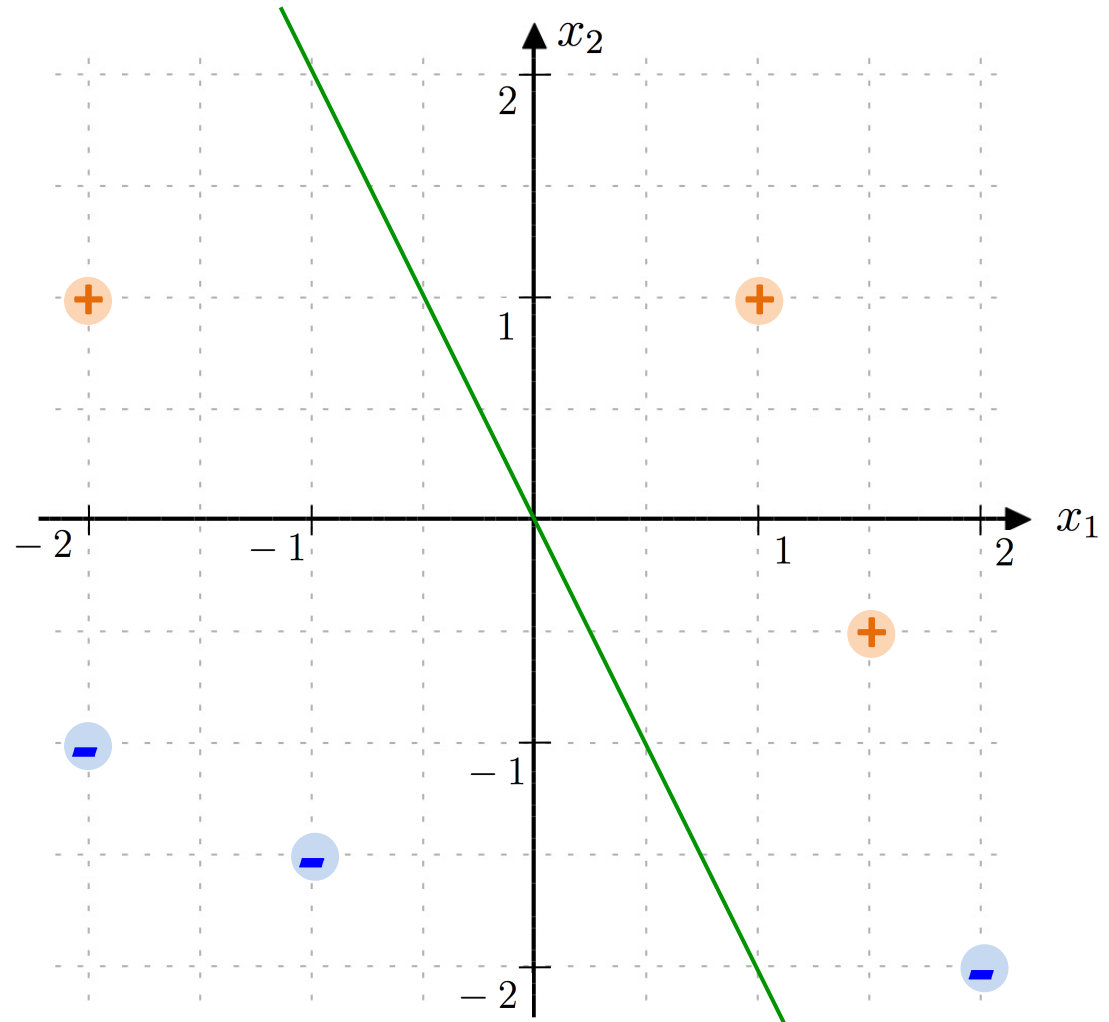


# Handout 15 example

Initial values:

$$\alpha = 0.2$$

$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$





# Handout 15 example

$$\alpha = 0.2$$

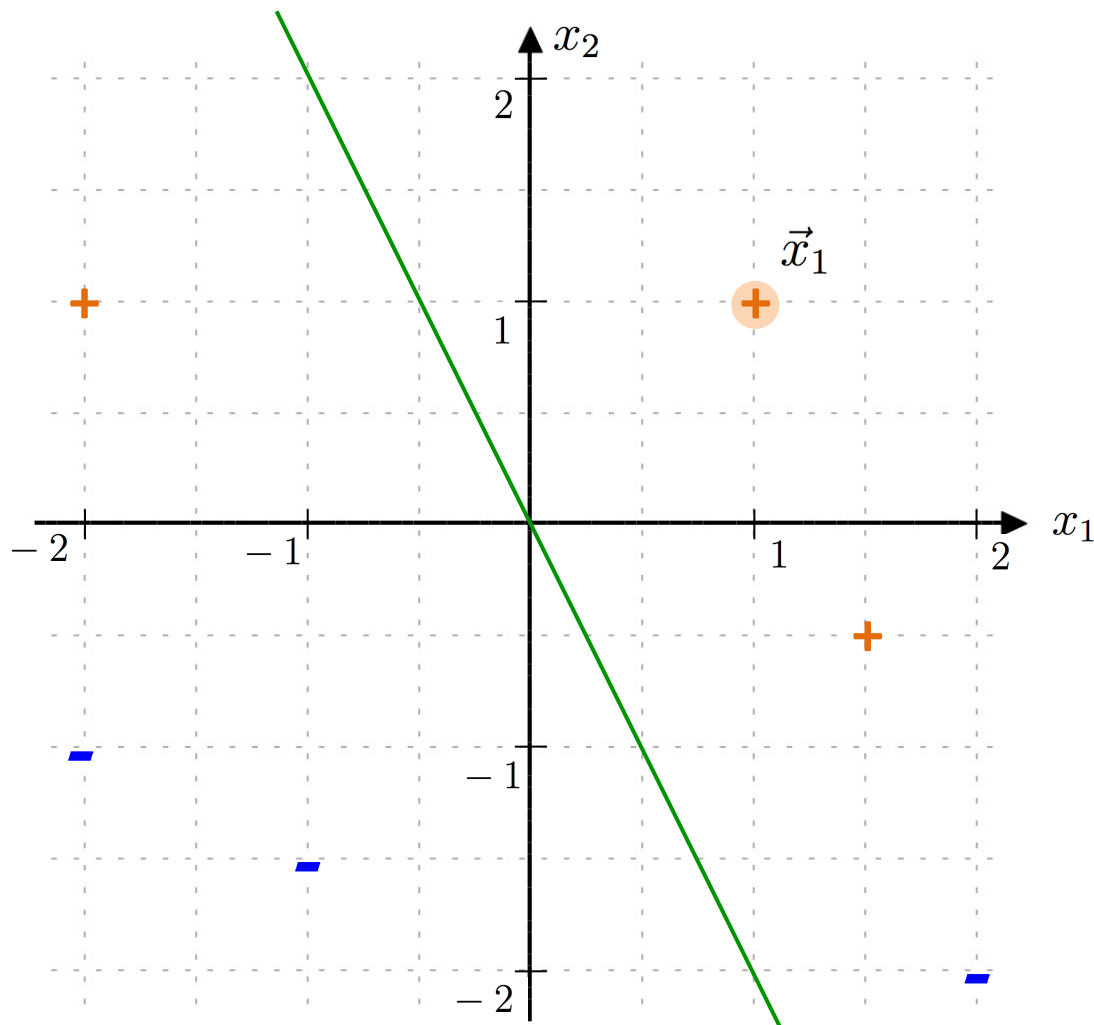
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 1:

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{w} \cdot \vec{x}_1 > 0$$

Correct classification, no action



# Handout 15 example

$$\alpha = 0.2$$

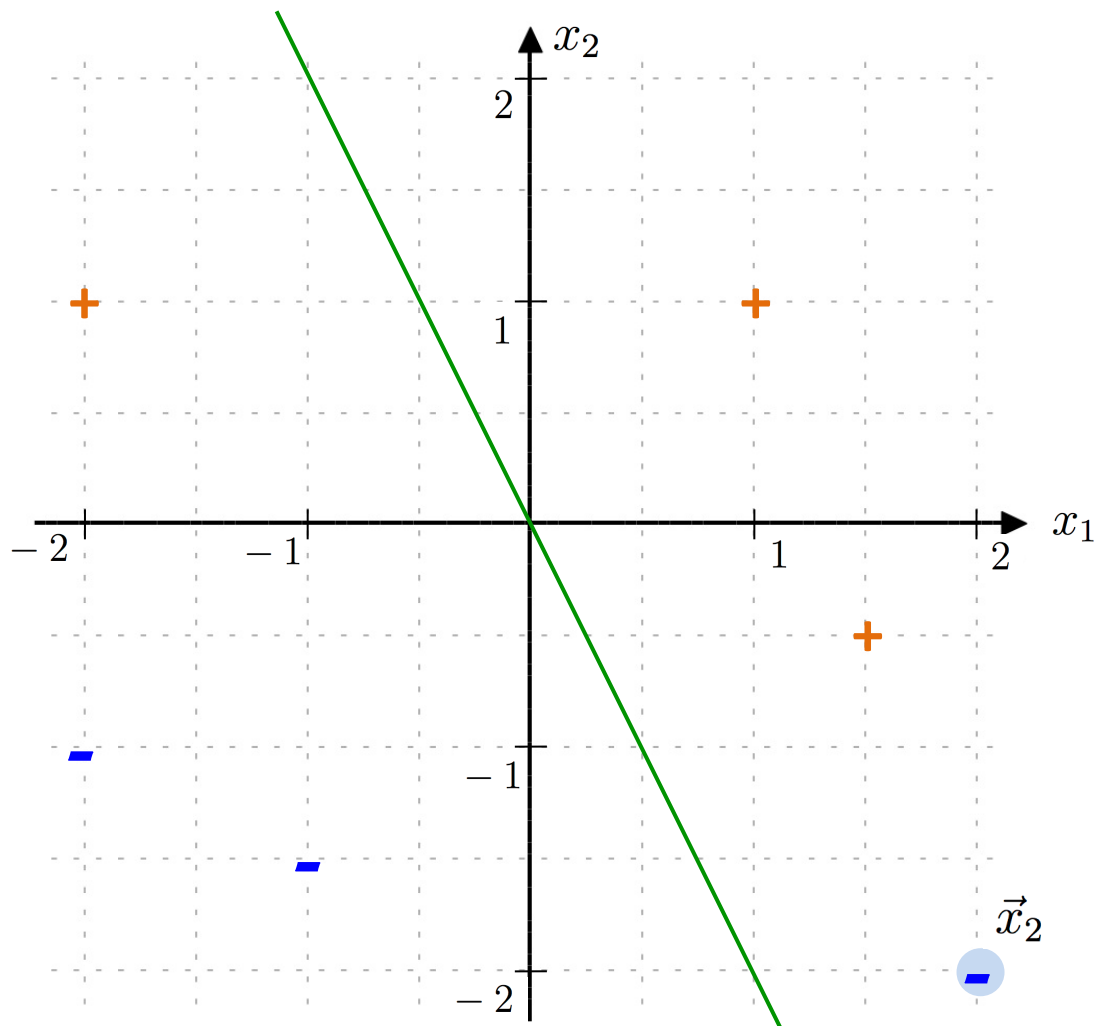
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 2:

$$\vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

$$\vec{w} \cdot \vec{x}_2 > 0$$

Incorrect classification



# Handout 15 example

$$\alpha = 0.2$$

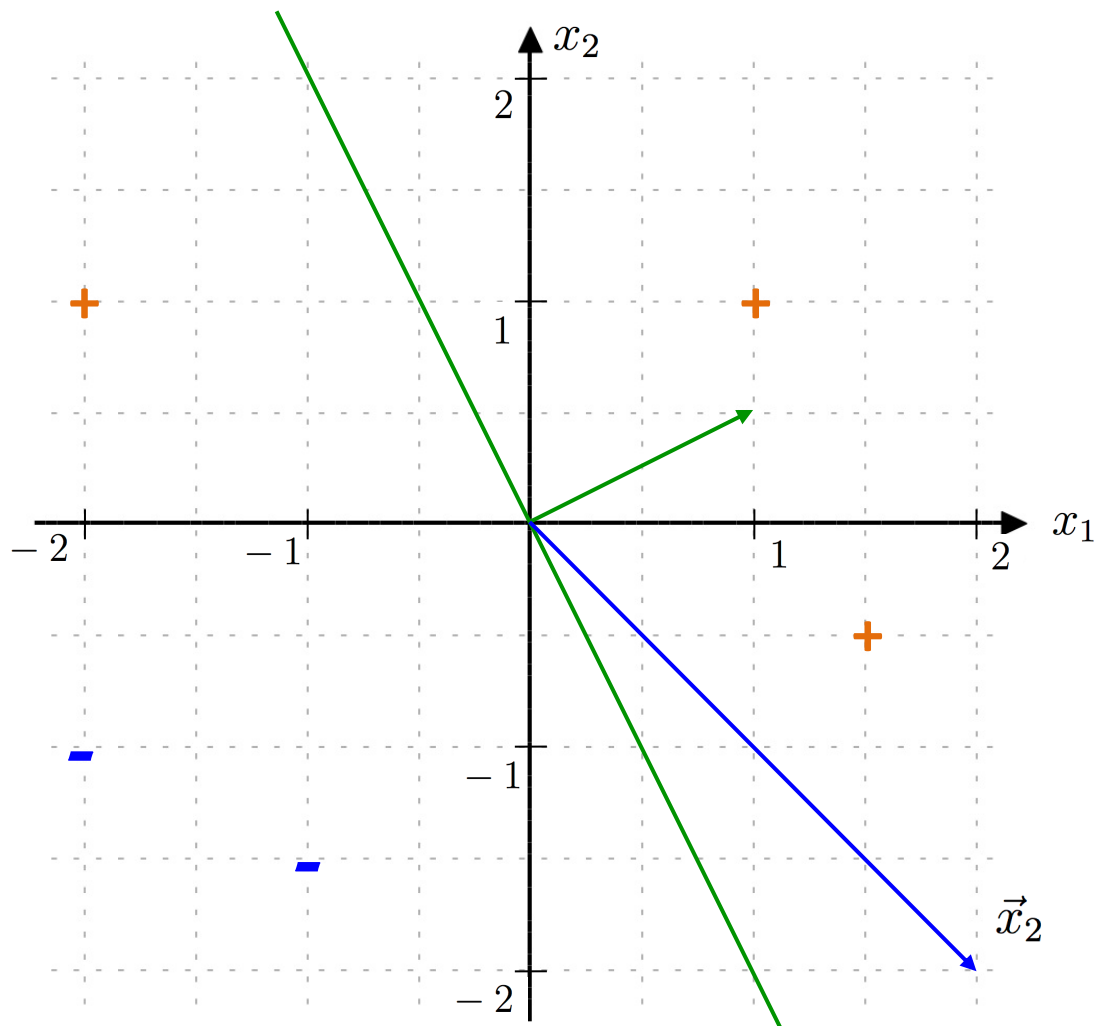
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 2:

$$\vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

$$\vec{w} \cdot \vec{x}_2 > 0$$

Incorrect classification



# Handout 15 example

$$\alpha = 0.2$$

$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

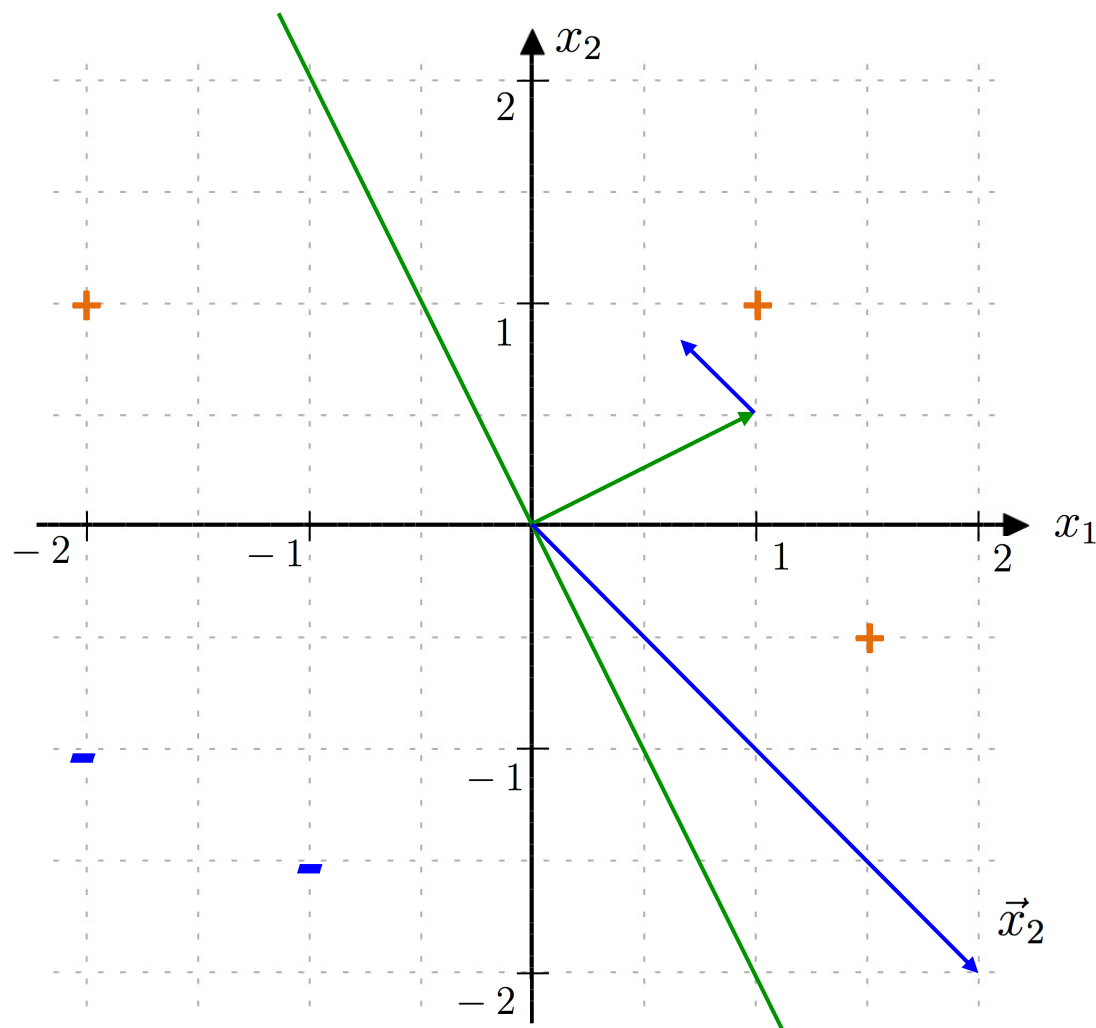
Round 2:

$$\vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

$$\vec{w} \cdot \vec{x}_2 > 0$$

Incorrect classification

“Push”  $\vec{w}$  away from negative point



# Handout 15 example

$$\alpha = 0.2$$

$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

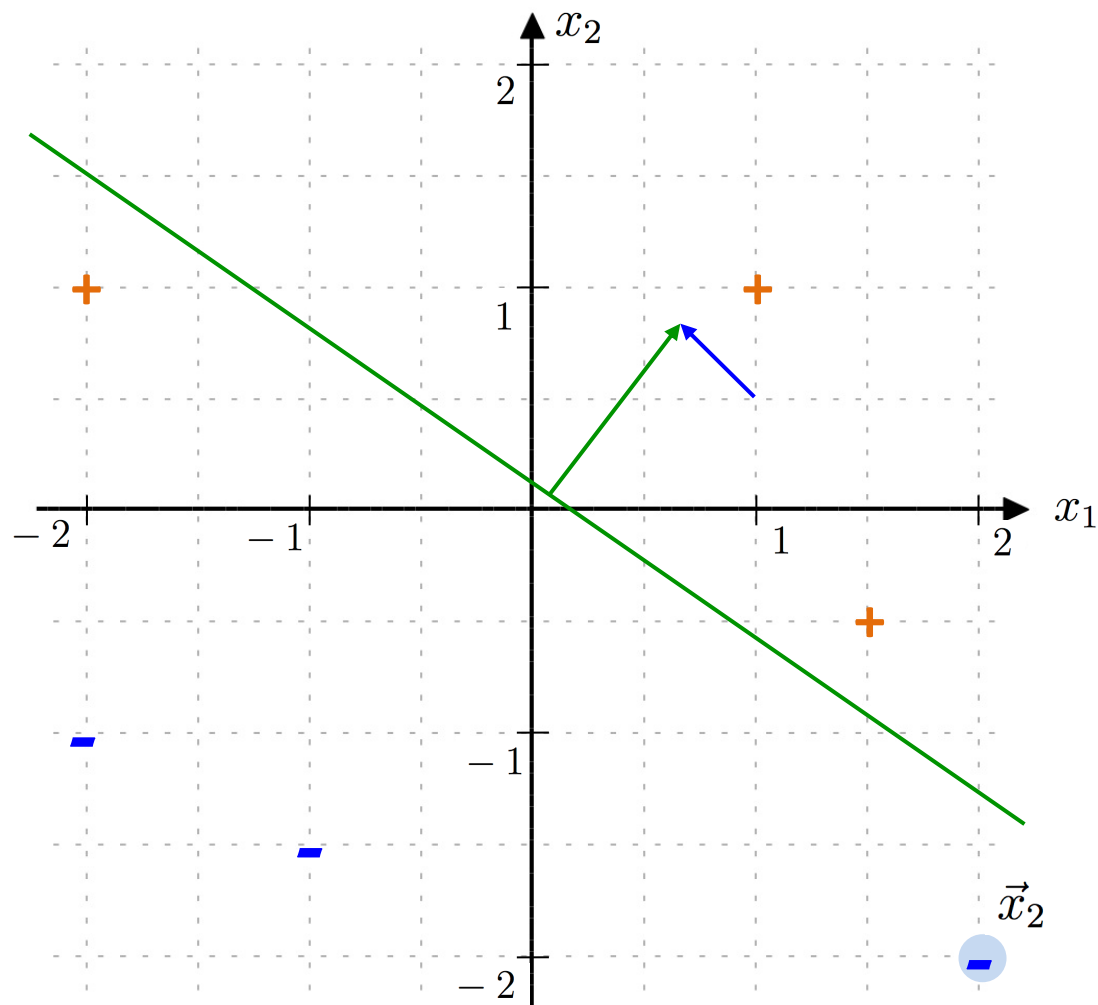
Round 2:

$$\vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

$$\vec{w} \cdot \vec{x}_2 > 0$$

Incorrect classification

“Push”  $\vec{w}$  away from negative point



# Handout 15 example

$$\alpha = 0.2$$

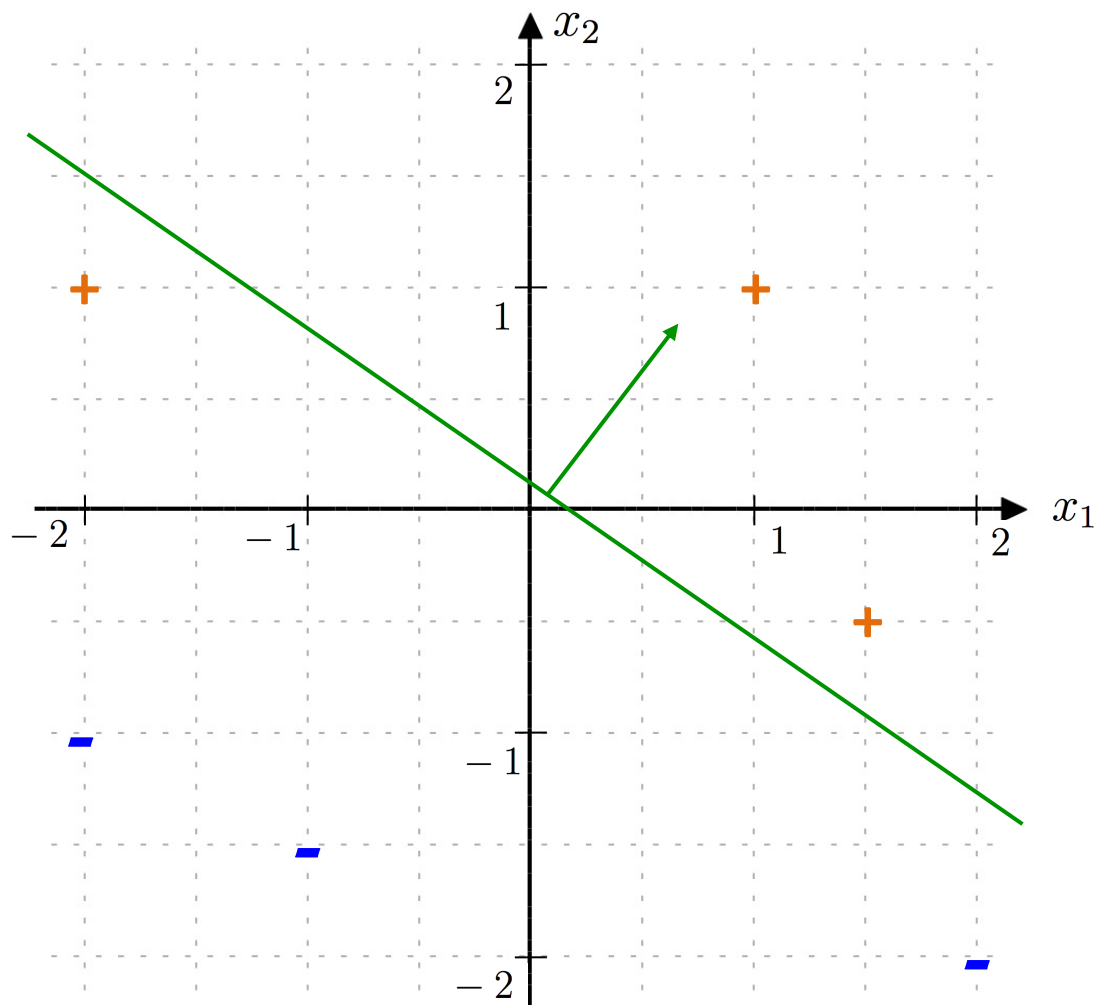
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 2:

$$\vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

$$\vec{w} \cdot \vec{x}_2 > 0$$

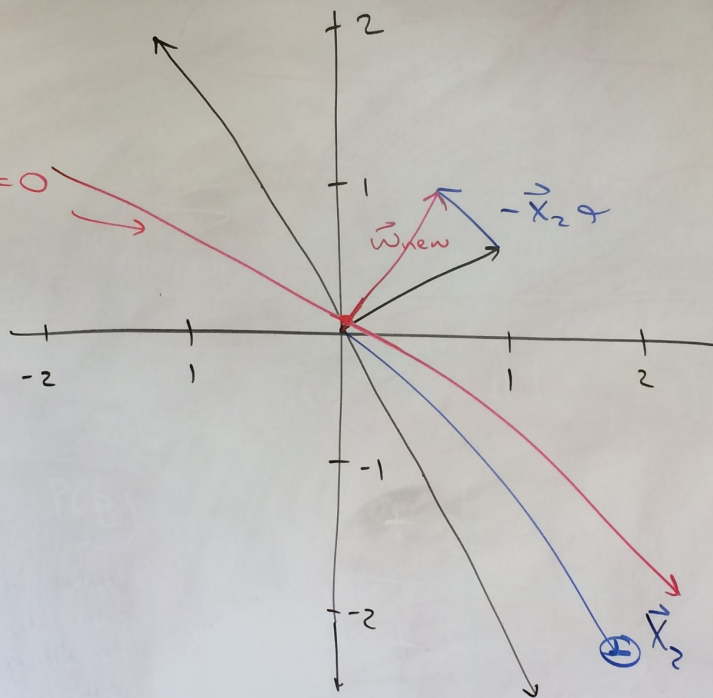
What is the new weight vector?



$$-0.2 + 0.6x_1 + 0.9x_2 = 0$$

$$x_2 = \frac{0.2 - 0.6x_1}{0.9}$$

$$x_2 = 0.22 - 0.66x_1$$



①  $\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$   $\bar{x}_1$  correct! ✓

$\bar{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$  fake one for the bias  $y_h$

$\vec{b} \leftarrow \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix} + 0.2(-1) \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$

$\vec{b} = \begin{bmatrix} -0.2 \\ 0.6 \\ 0.9 \end{bmatrix}$

# Handout 15 example

Final solution (so you can check your work):

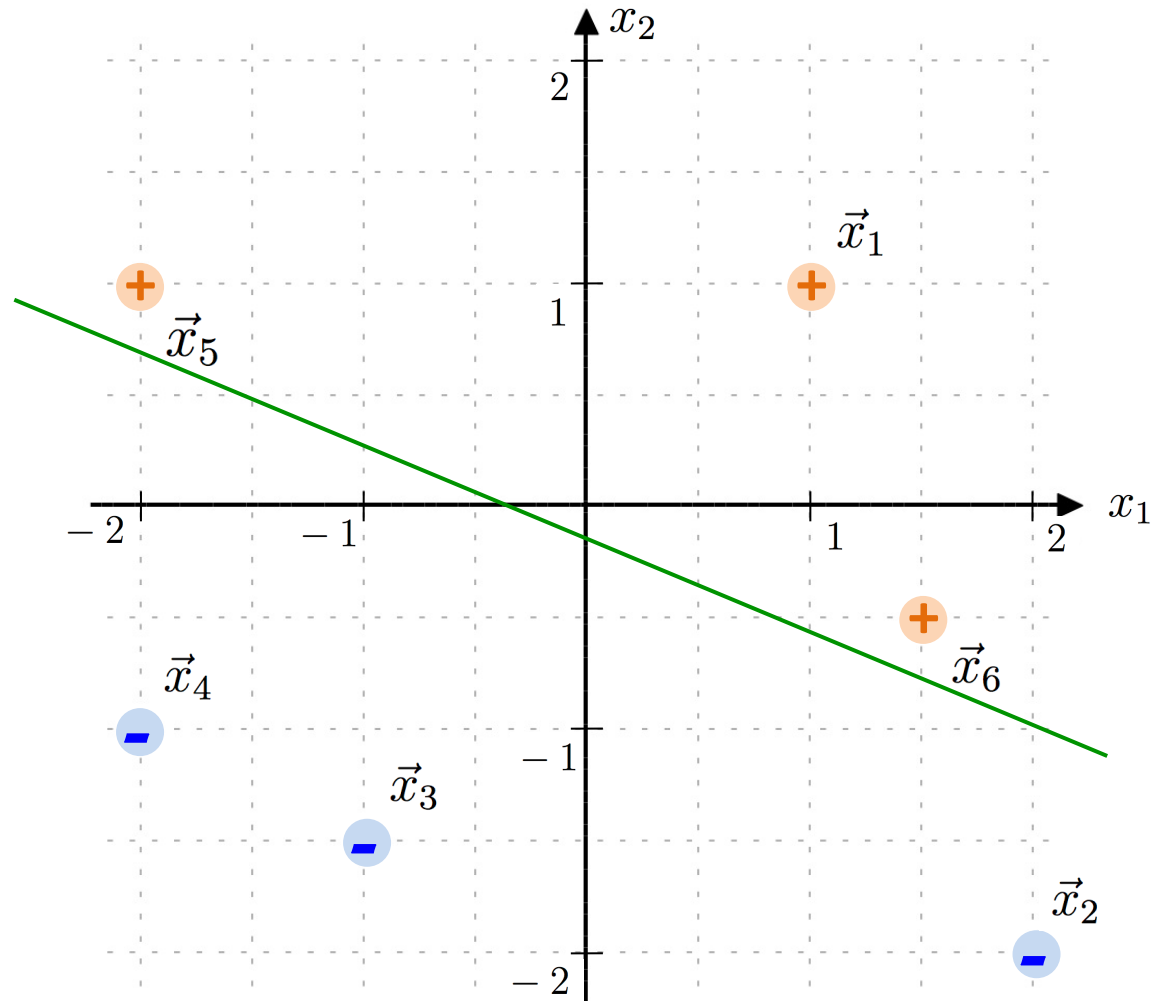
$$\vec{w}^* = \begin{bmatrix} 0.2 \\ 0.5 \\ 1 \end{bmatrix}$$

Final hyperplane:

$$0.2 + 0.5x_1 + x_2 = 0$$

$\Rightarrow$

$$x_2 = -0.2 - 0.5x_1$$





# Outline for October 29

- Recap AdaBoost and weighted entropy
- Perceptron Algorithm
- Mid-semester feedback