

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



Outline for March 20

- Perceptron
 - Intuition behind dot product
 - Perceptron algorithm
 - Cost function and SGD interpretation
 - Example (Handout 10)

Partnership reminders

- Make a plan for times when you will work together with your partner

Partnership reminders

- Make a plan for times when you will work together with your partner
- I don't recommend "divide and conquer", but if you must sometimes, make sure divisions are equal and deadlines are agreed upon in advance

Partnership reminders

- Make a plan for times when you will work together with your partner
- I don't recommend "divide and conquer", but if you must sometimes, make sure divisions are equal and deadlines are agreed upon in advance
- Respect your partner's thoughts and opinions about the work

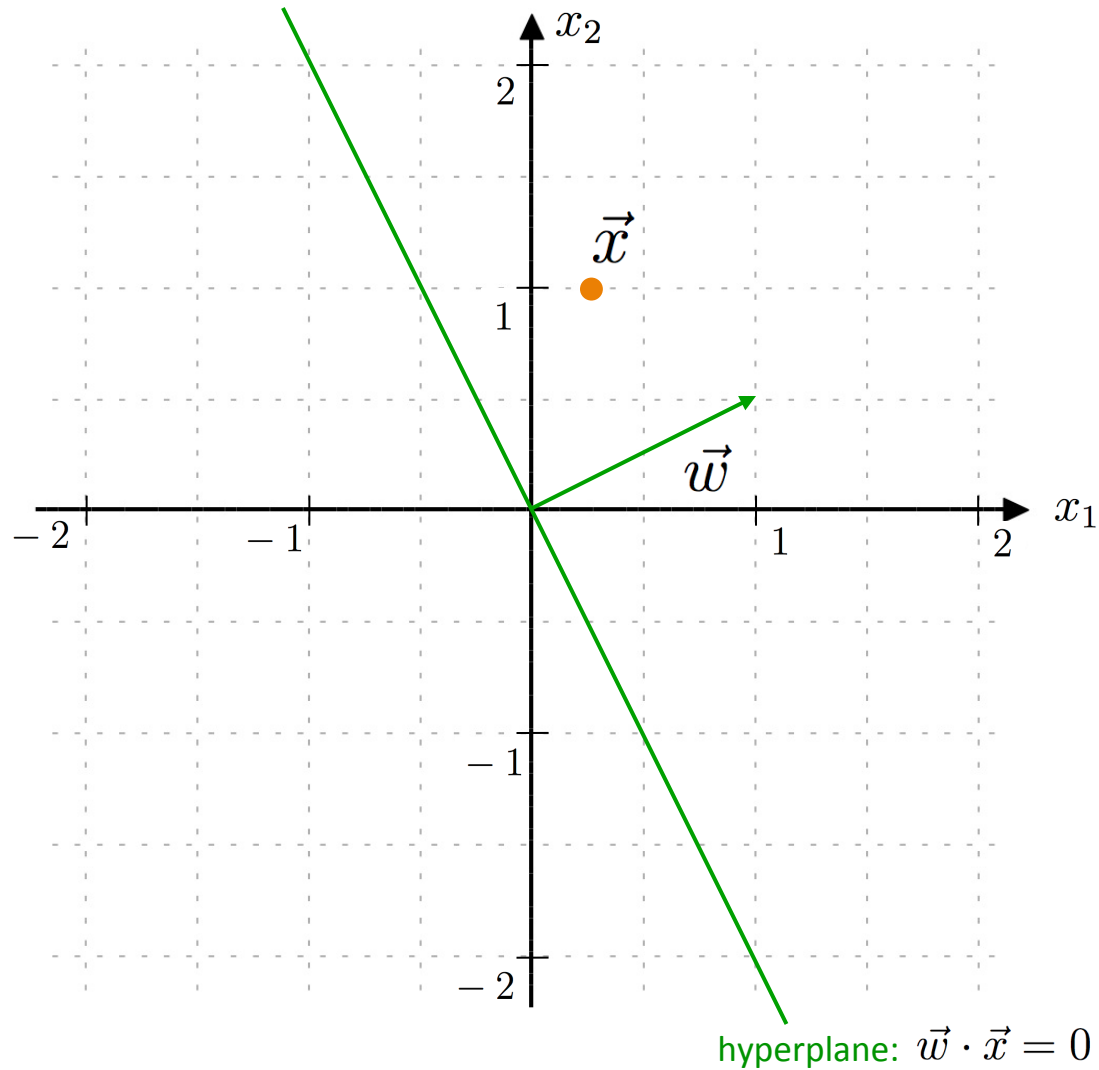
Partnership reminders

- Make a plan for times when you will work together with your partner
- I don't recommend "divide and conquer", but if you must sometimes, make sure divisions are equal and deadlines are agreed upon in advance
- Respect your partner's thoughts and opinions about the work
- Make sure you both understand everything – explaining something can be a great way to learn it better!

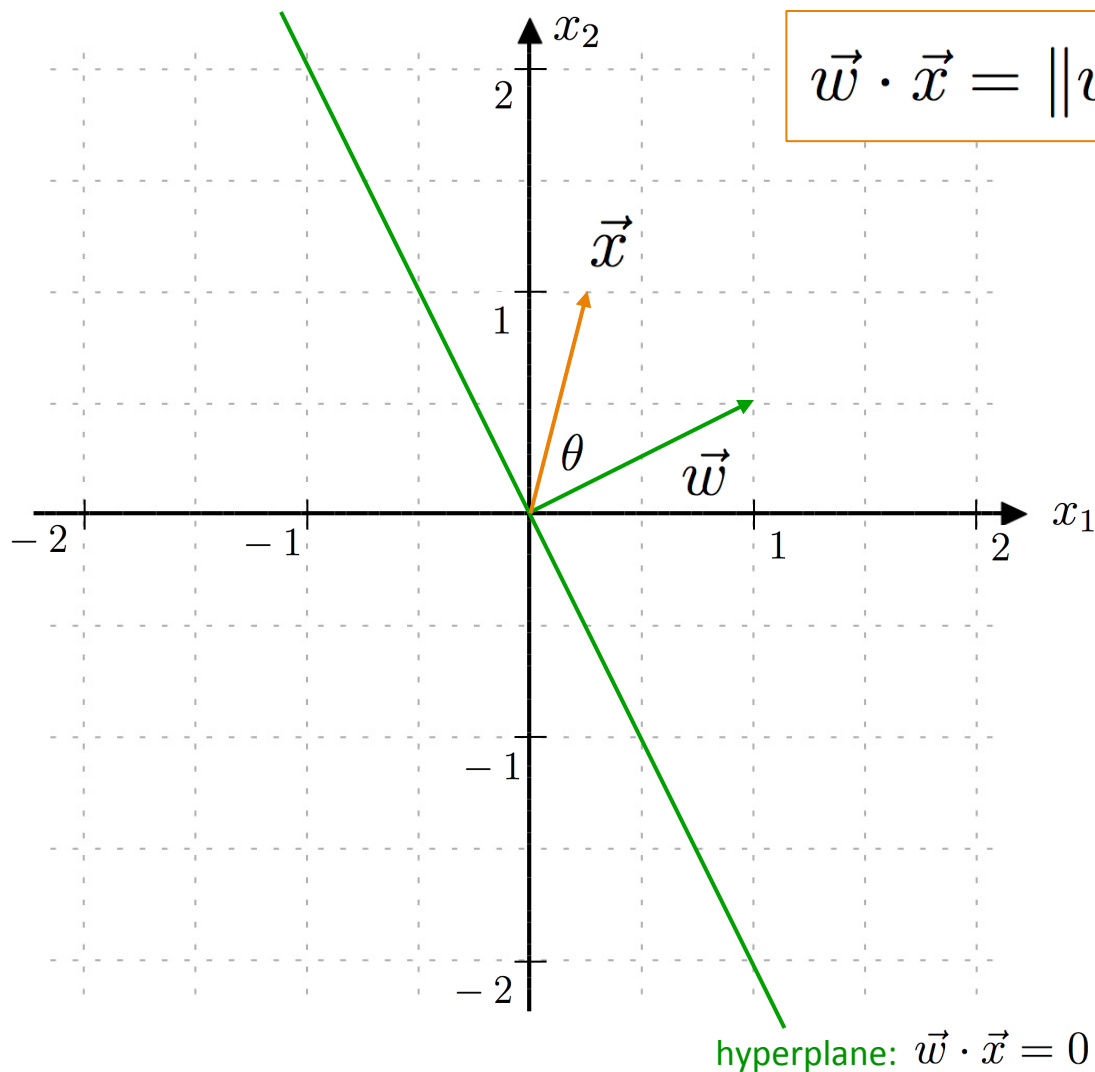
Outline for March 20

- Perceptron
 - Intuition behind dot product
 - Perceptron algorithm
 - Cost function and SGD interpretation
 - Example (Handout 10)

Intuition behind the dot product



Intuition behind the dot product

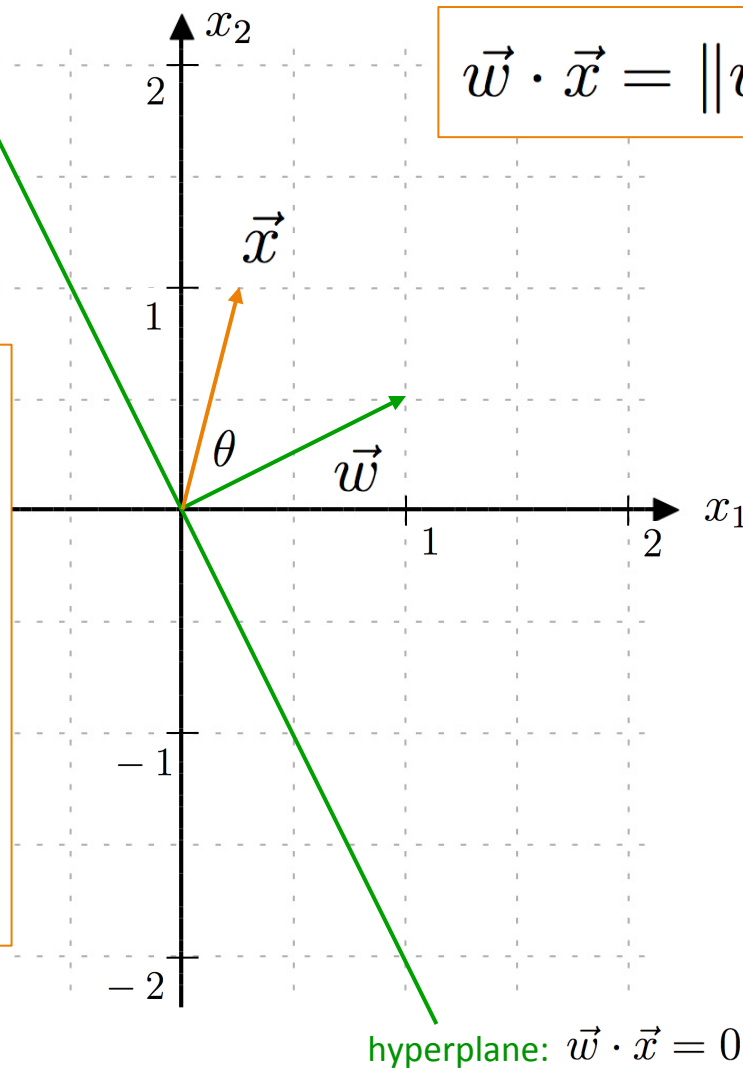


Intuition behind the dot product

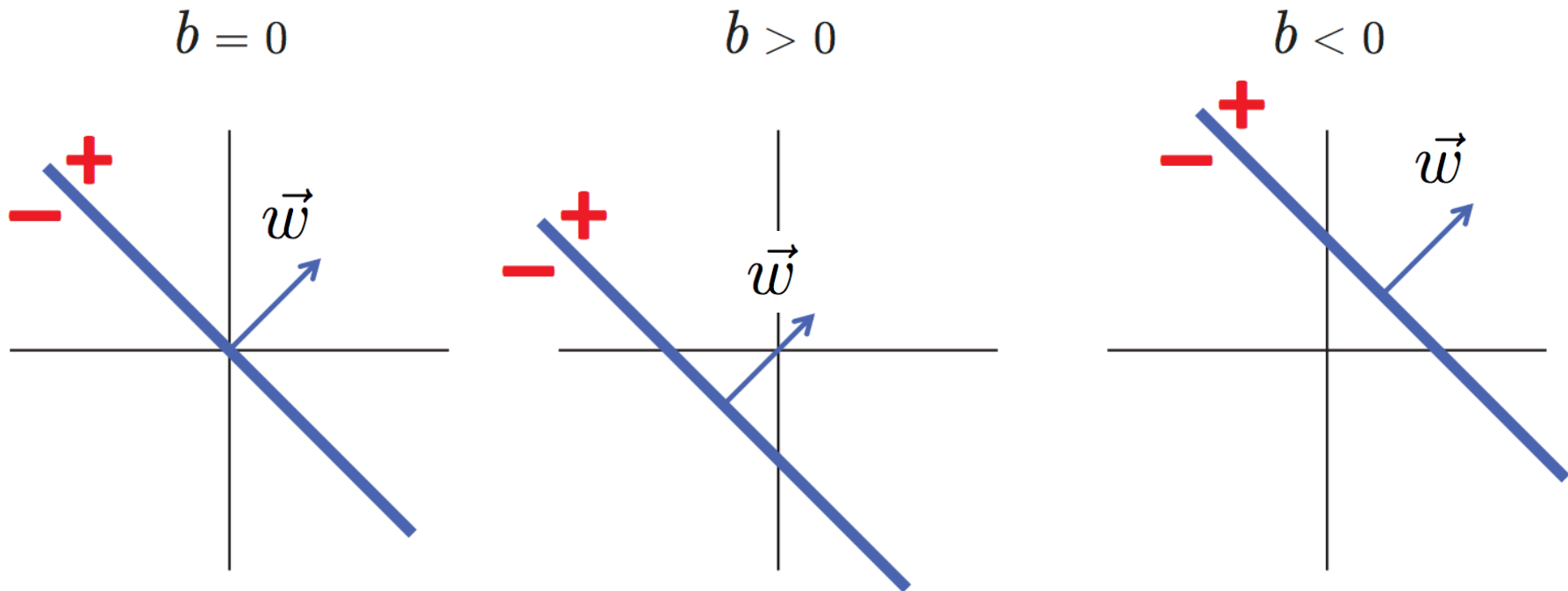
$$\vec{w} \cdot \vec{x} = \|\vec{w}\| \|\vec{x}\| \cos \theta$$

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



Edit: 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

Outline for March 20

- Perceptron
 - Intuition behind dot product
 - Perceptron algorithm
 - Cost function and SGD interpretation
 - Example (Handout 10)

Perceptron

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

$$h(\vec{x}) = \text{sign}(\vec{w}^T \vec{x})$$

linear model

$$\vec{w}^T \vec{x} \geq 0, y \rightarrow +1$$

$$\vec{w}^T \vec{x} < 0, y \rightarrow -1$$

Algorithm

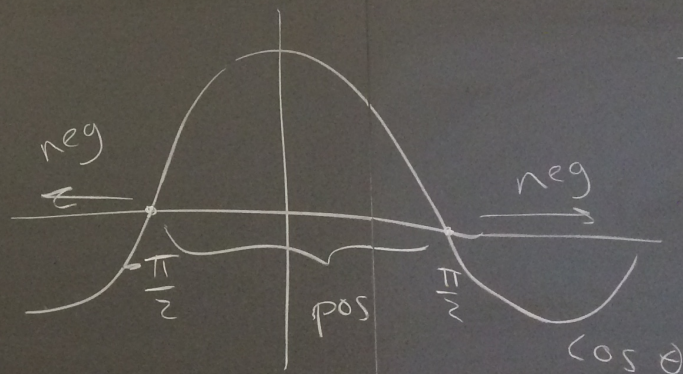
$$0 + 1 \cdot x_1 + 0.5 \cdot x_2 = 0$$

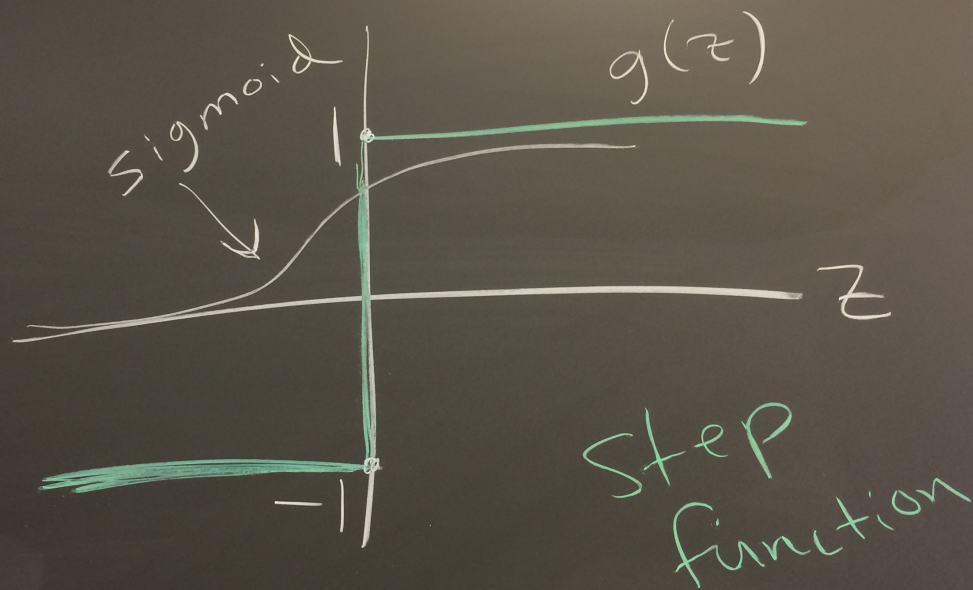
$$x_2 = -2x_1$$

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

$$\text{Sigh}(\cos(\theta))$$

"b", "bias"





$$g(z) = \text{sign}(z)$$

Algorithm

* initialize w 's to zeros or small values
 * repeat until convergence:

① choose random example \vec{x}_i

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

if correctly classified \Rightarrow do nothing

until all points classified correctly.

derivative

$\text{sign}(\vec{w} \cdot \vec{x}_i)$

Incorrect either

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

$\Rightarrow -2$ or 2

updates

$$w_j \leftarrow w_j + \alpha y_i x_{ij}$$

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

Perceptron id if incor up

Incorrect
either

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

$\Rightarrow -2$ or 2

updates

$$w_j \leftarrow w_j + \alpha y_i x_{ij}$$

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

perceptron idea
if incorrect,
update
weights

Handout 10, part (a)

$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix} - 0.2 \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

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

Cost Function

$$J(\vec{w}) = \sum_{i=1}^n \max(0, -y_i(\vec{w}^T \vec{x}_i))$$

if correct

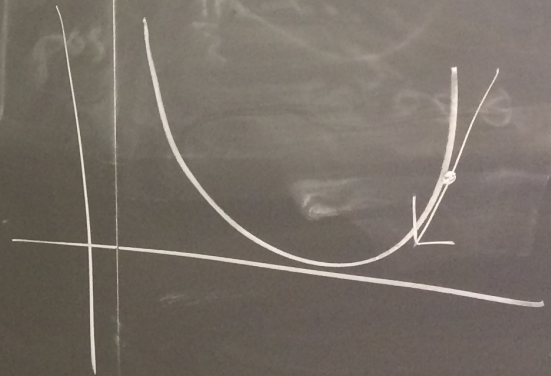
$$-y_i(\vec{w}^T \vec{x}_i) < 0 \Rightarrow \max \text{ is } \underline{0}$$

Same
Sign

$$\frac{\partial J(\vec{w})}{\partial \vec{w}} = -y_i \vec{x}_i$$

if not

max is > 0



Go in the direction of the *negative* gradient,
that's why our updates are "plus" $y_i \vec{x}_i$

Convergence Guarantee

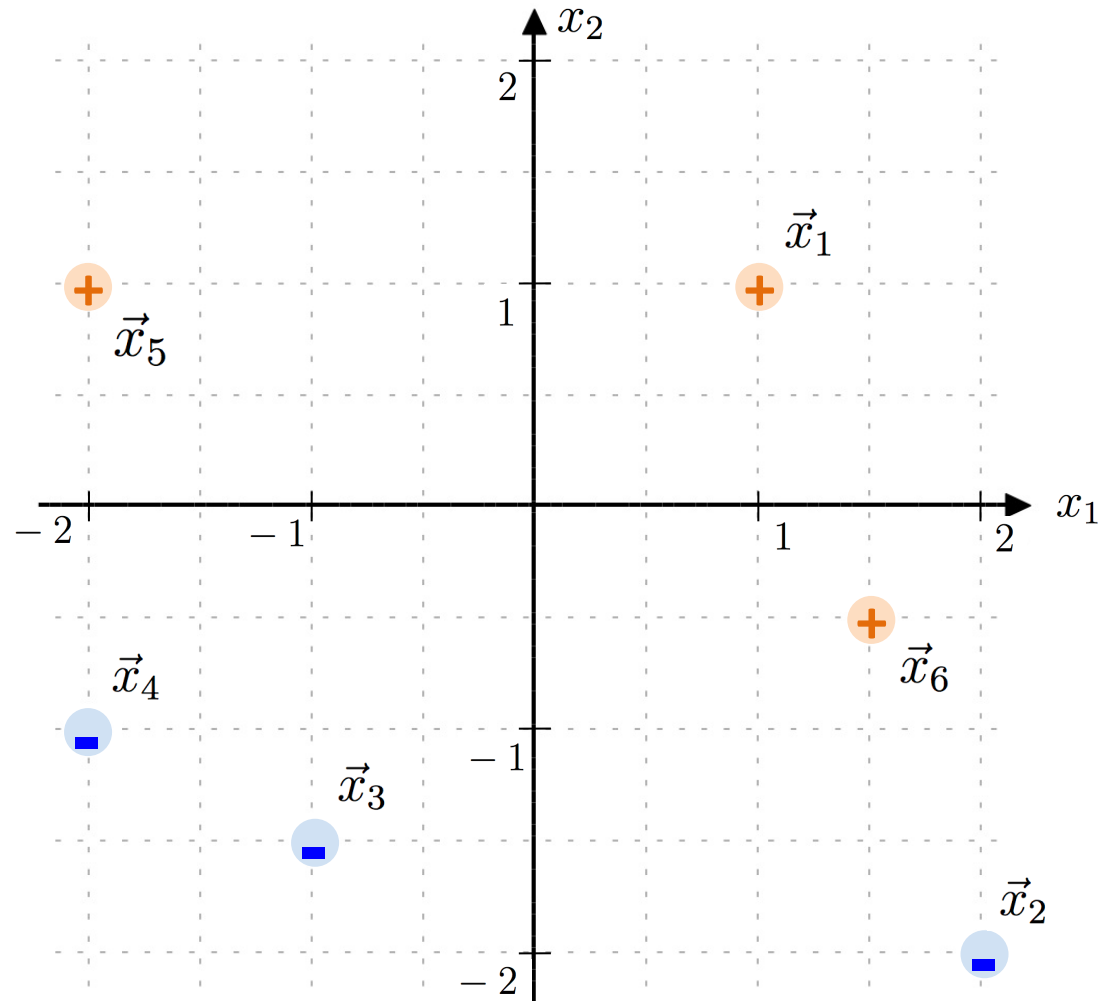
- Perceptron is guaranteed to converge to a solution if a separating hyperplane exists
- Not guaranteed to converge to a “good” solution
- No guarantees about behavior if a separating hyperplane does not exist!

Handout 10 example

Initial values:

$$\alpha = 0.2$$

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

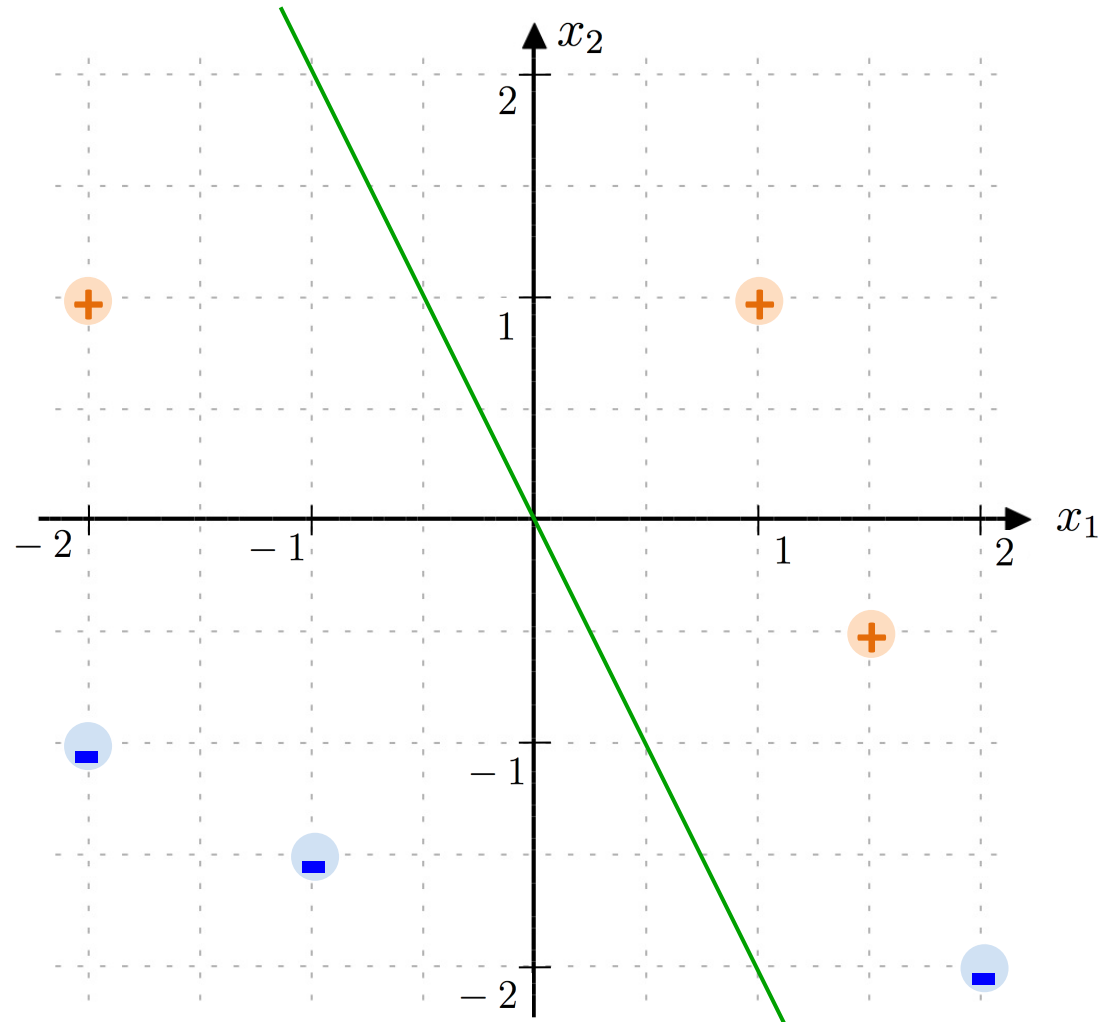


Handout 10 example

Initial values:

$$\alpha = 0.2$$

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



Handout 10 example

$$\alpha = 0.2$$

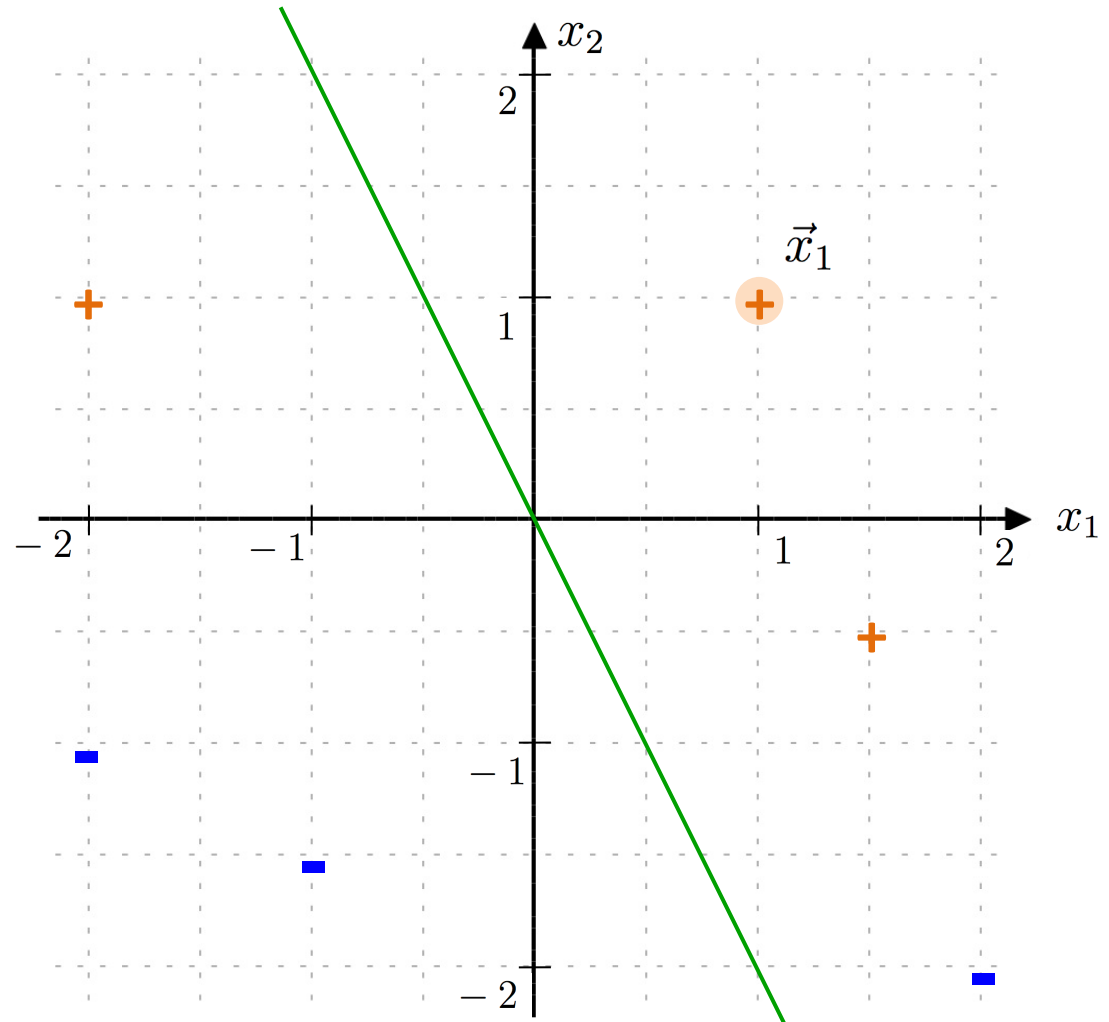
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 1:

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

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

Correct classification, no action



Handout 10 example

$$\alpha = 0.2$$

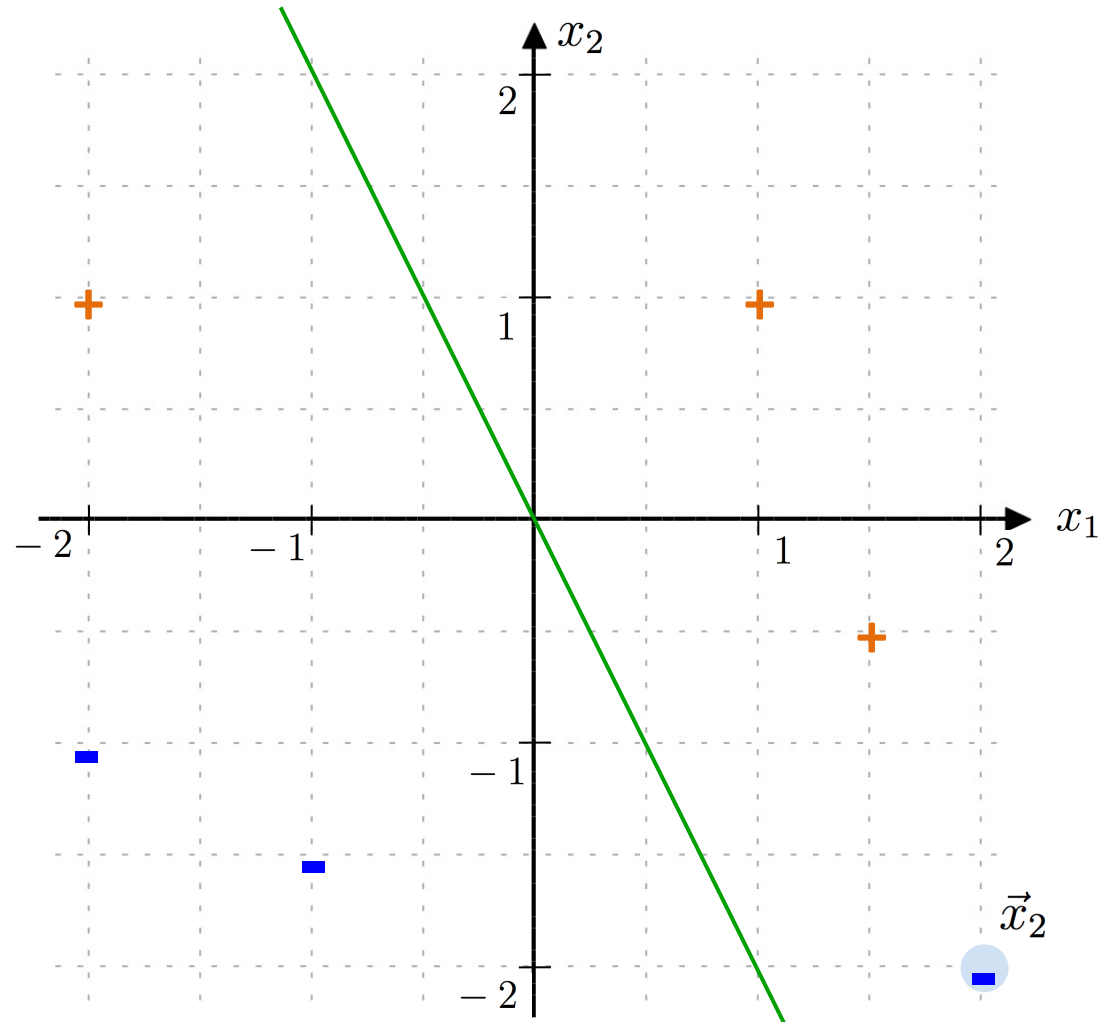
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 2:

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

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

Incorrect classification



Handout 10 example

$$\alpha = 0.2$$

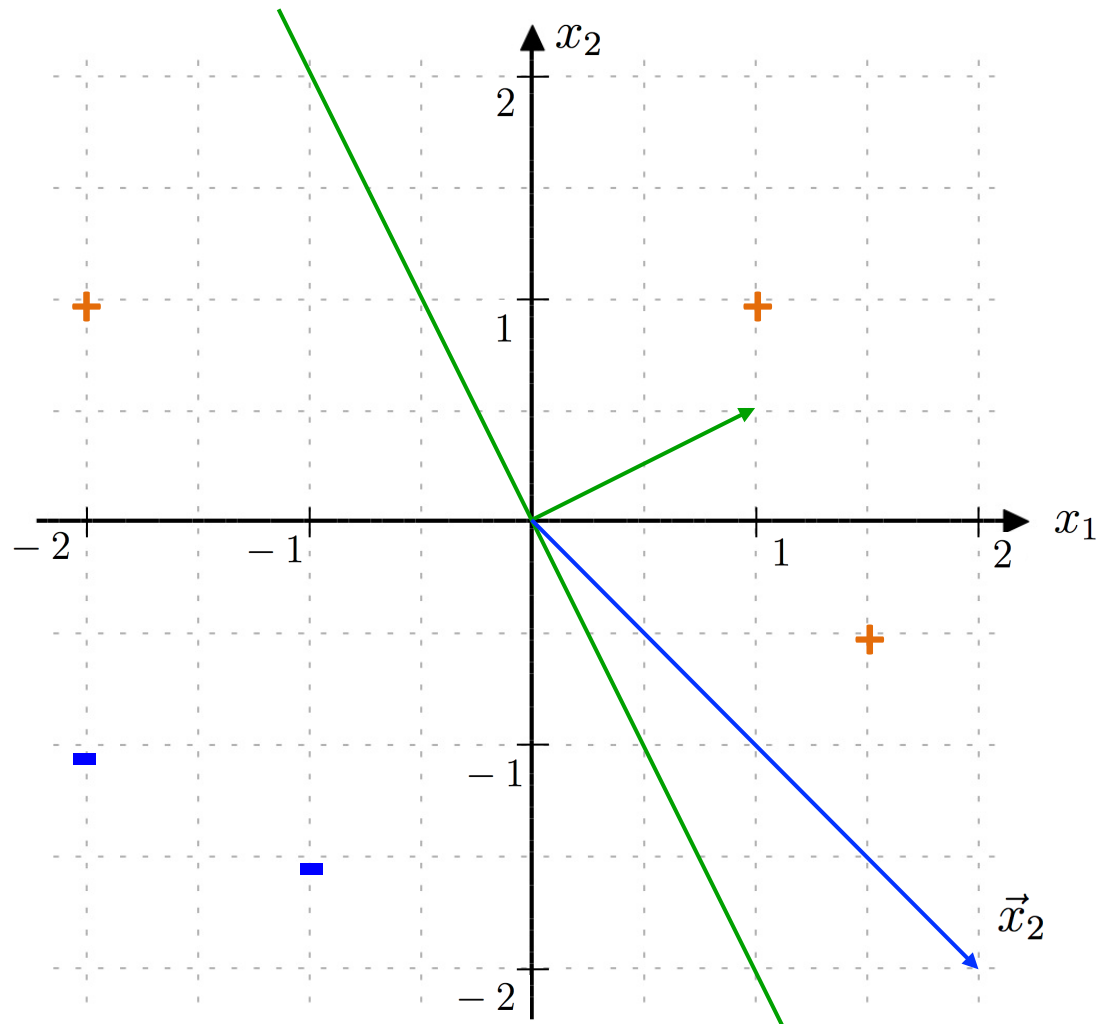
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

Round 2:

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

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

Incorrect classification



Handout 10 example

$$\alpha = 0.2$$

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

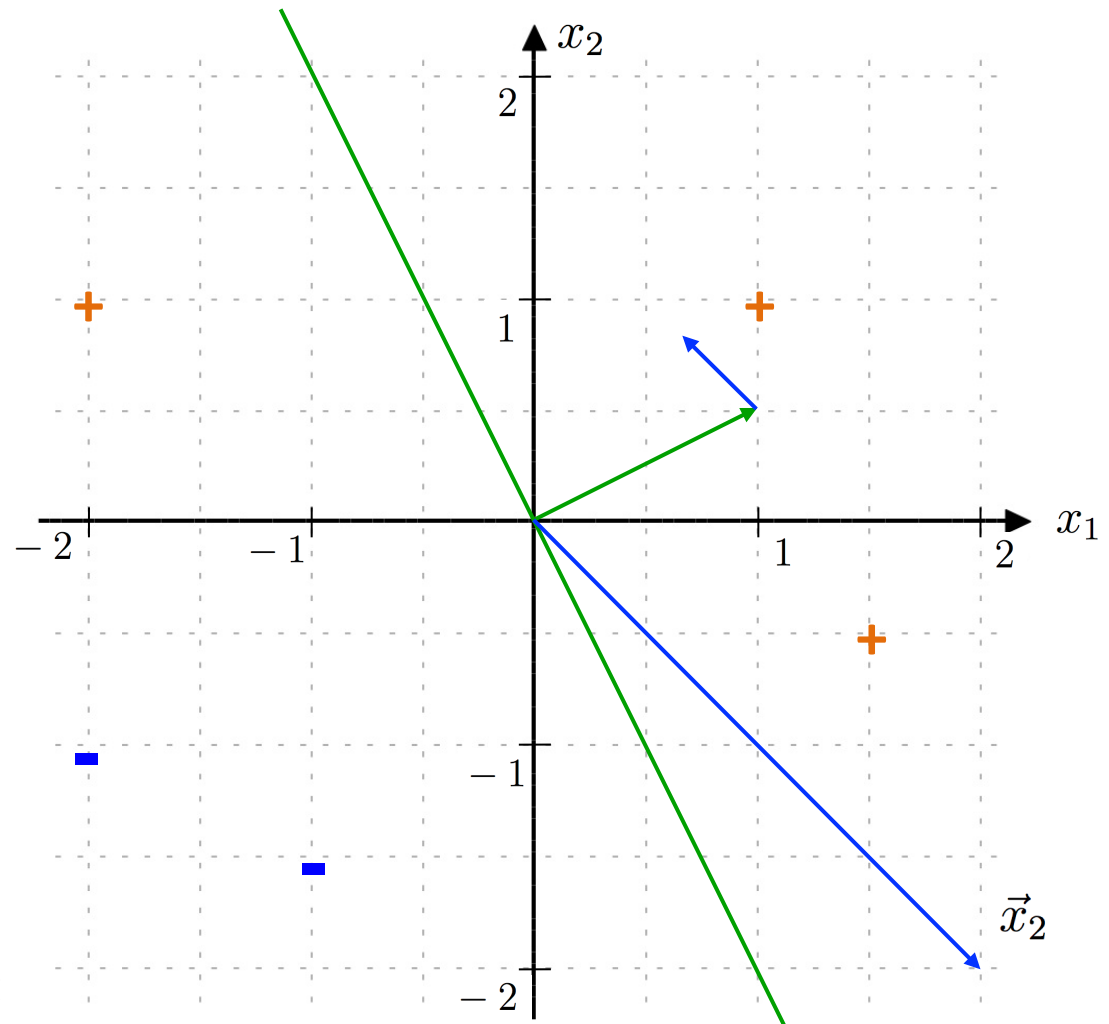
Round 2:

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

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

Incorrect classification

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



Handout 10 example

$$\alpha = 0.2$$

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

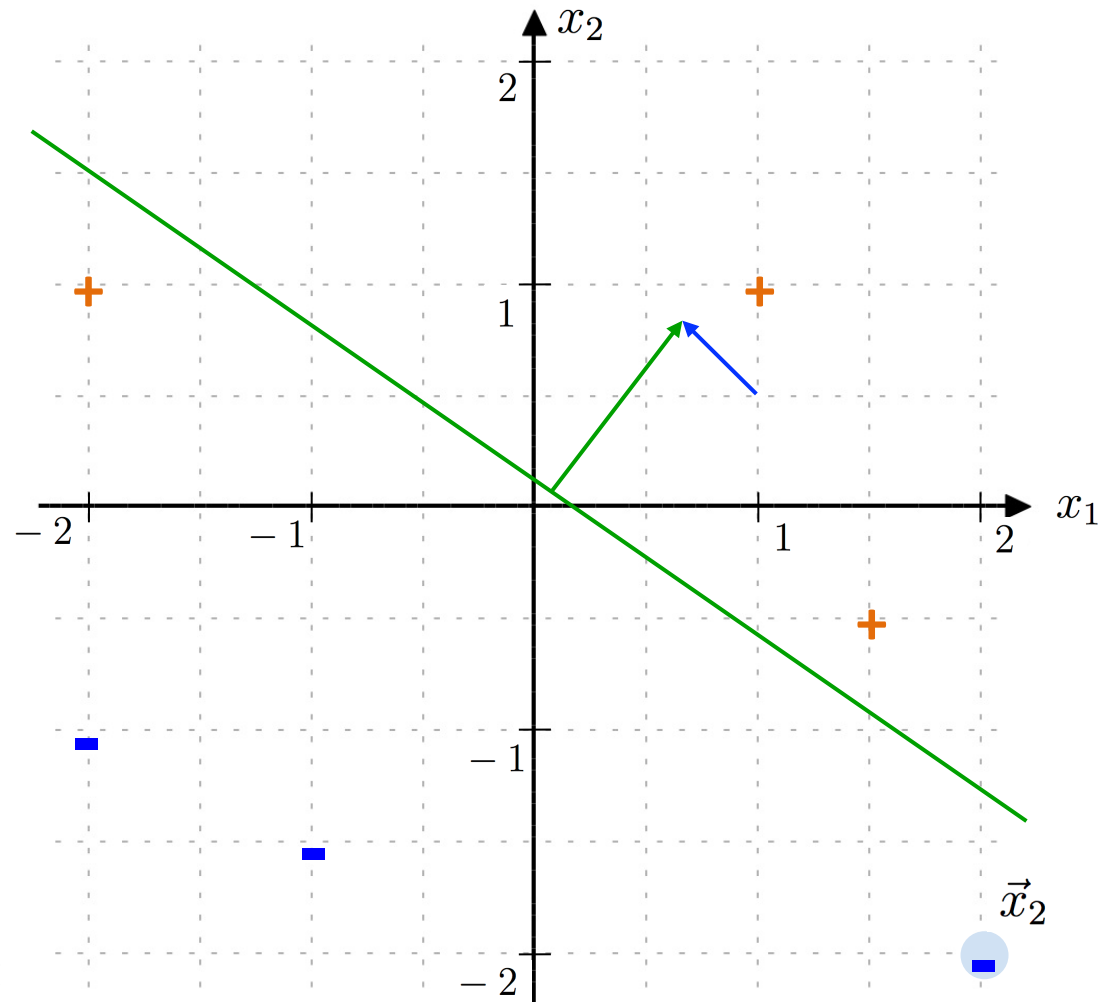
Round 2:

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

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

Incorrect classification

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



Handout 10 example

$$\alpha = 0.2$$

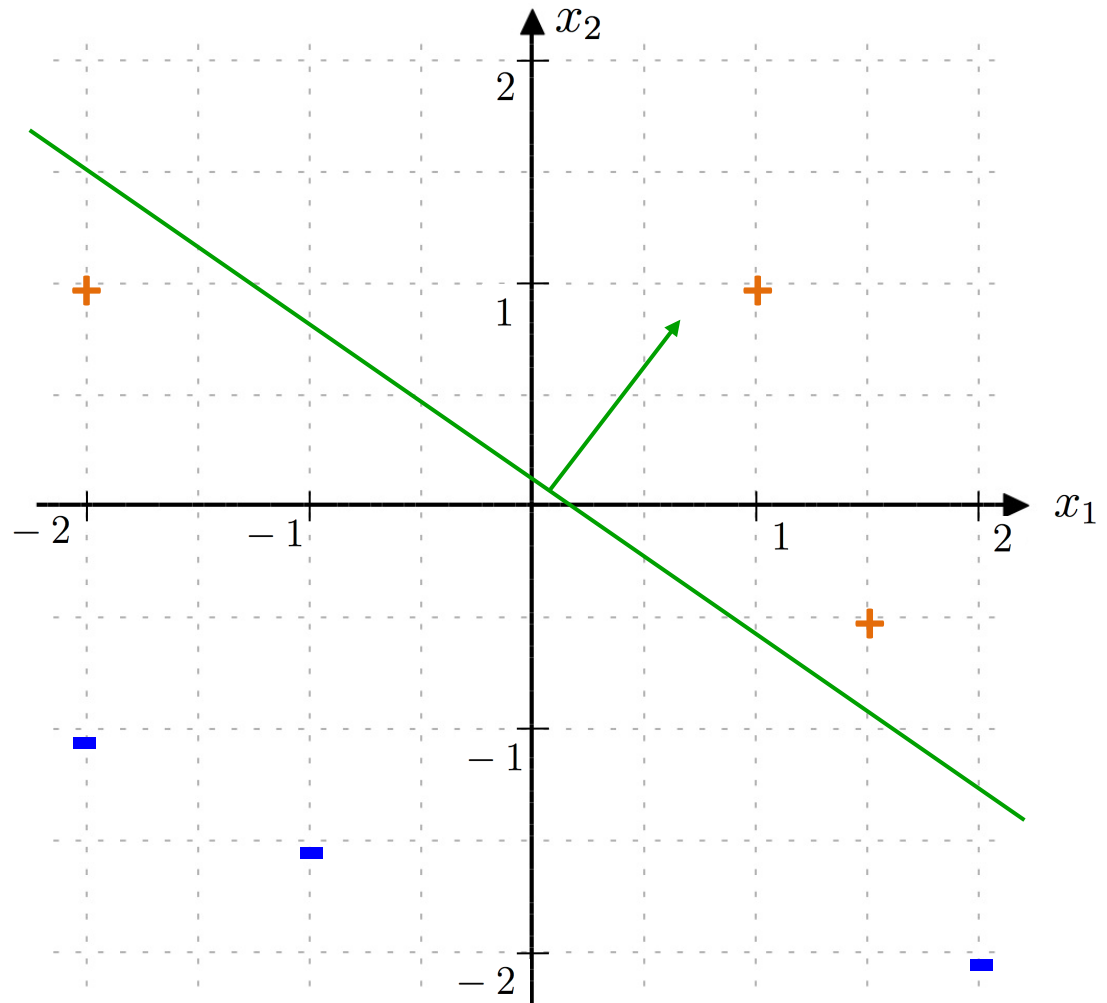
$$\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 0.5 \end{bmatrix}$$

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?



$$\vec{x}_3 \quad \checkmark$$

$$\vec{x}_4 \quad \checkmark$$

$$\vec{x}_5 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad \left. \begin{array}{l} \vec{x}_3 \\ \vec{x}_4 \\ \vec{x}_5 \end{array} \right\}$$

$$\vec{w} = \begin{bmatrix} -0.2 \\ 0.6 \\ 0.9 \end{bmatrix} + 0.2 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.2 \\ 1.1 \end{bmatrix}$$

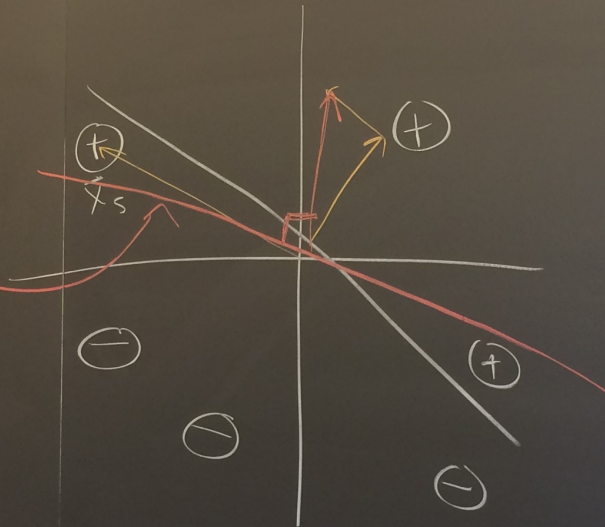
$$0 + 0.2x_1 + 1.1x_2 = 0$$

$$x_2 = -\frac{0.2}{1.1} x_1$$

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.2 \\ 1.1 \end{bmatrix}$$

$$0 + 0.2x_1 + 1.1x_2 = 0$$

$$x_2 = -\frac{0.2}{1.1}x_1$$



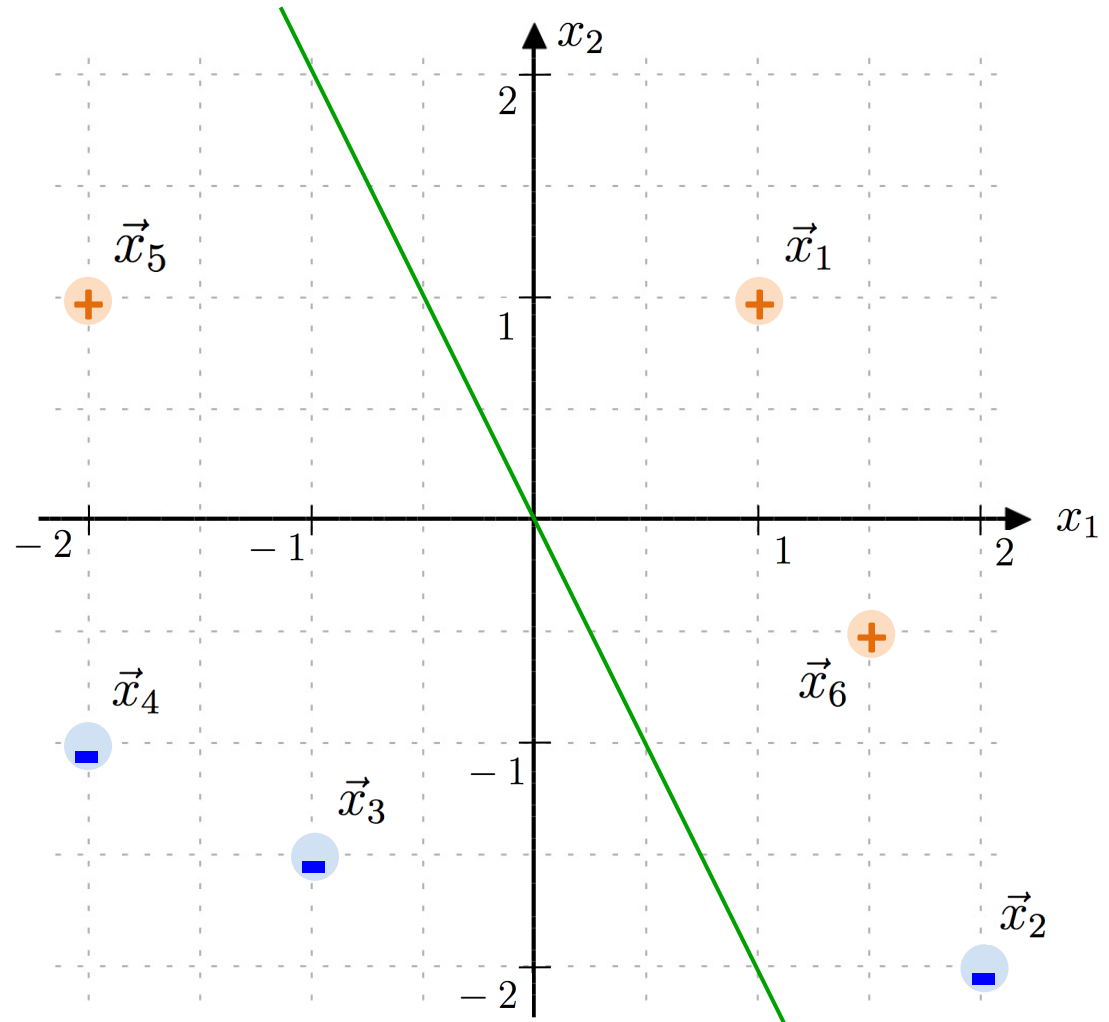
Handout 10, part (b)

Handout 10 example

Initial values:

$$\alpha = 0.2$$

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



Handout 10 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

$$x_2 = -0.2 - 0.5x_1$$

