

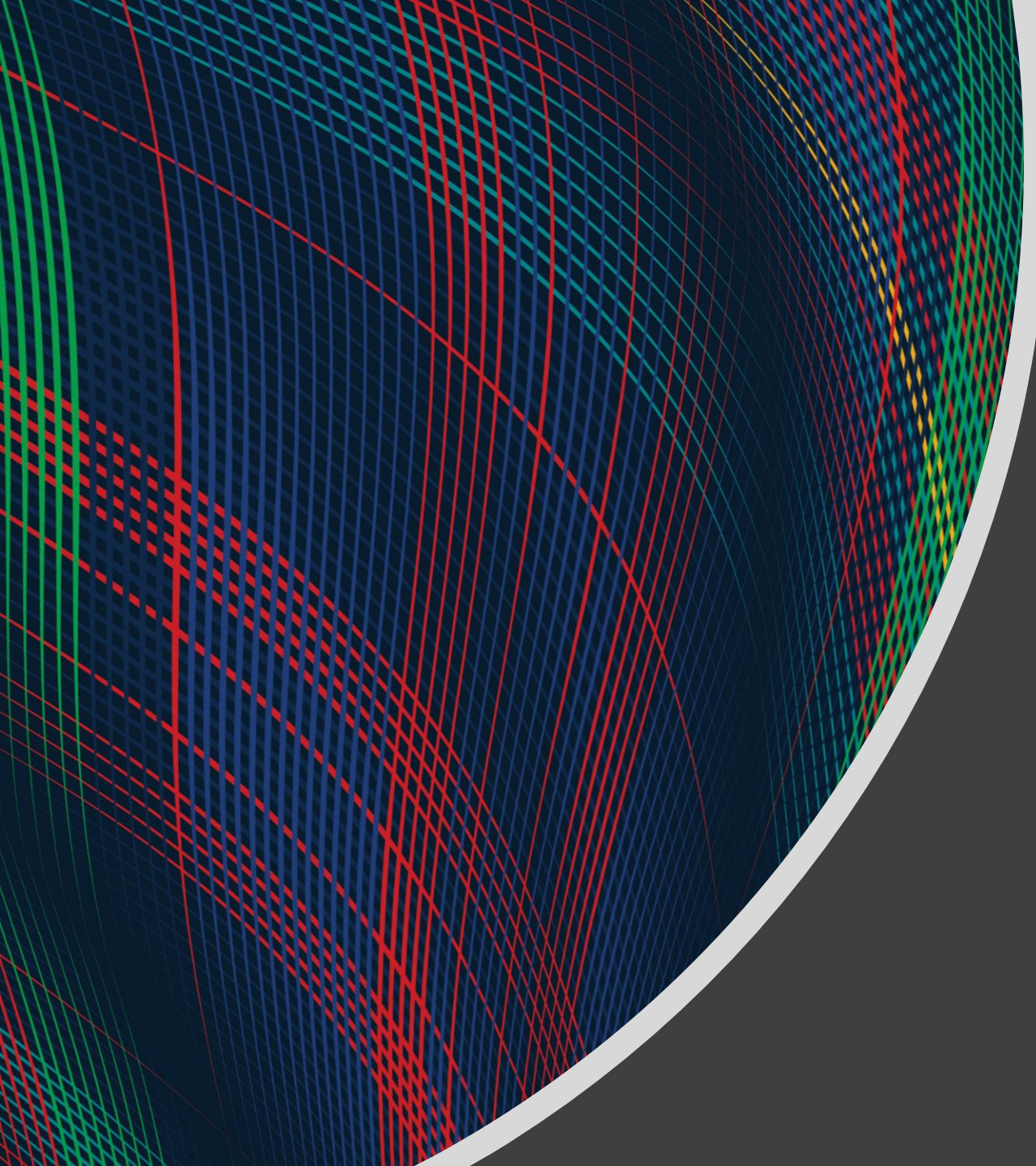
# Plan

Wrap-up Overview → Lecture 1 slides

## Proof Techniques

- ✓ Proof by cases
- ✓ Disprove by counterexample
  - Proof by contrapositive
  - Proof by contradiction

## Perceptron Algorithm

An abstract graphic on the left side of the slide, featuring a sphere-like shape composed of a dense grid of intersecting red, green, and blue lines. The lines are curved and follow the contours of the sphere, creating a complex, woven pattern. The sphere is set against a dark gray background.

# 10-607 Computational Foundations for Machine Learning

## Proof Techniques & Perceptron Algorithm

Instructor: Pat Virtue

# Proof Techniques

Proof by cases

Disproof by counterexample

- One example is not sufficient to prove
- One counterexample is sufficient to disprove

Proof by contrapositive

# Previous Poll

Given model  $m$ : {A: True, B: False}

Does  $m$  satisfy the following sentence:

$$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$$

Handwritten annotations in blue:

- Under  $A$ :  $T$
- Under  $B$ :  $F$
- Under  $\neg B$ :  $T$
- Under  $\neg A$ :  $F$
- Under the implication  $\Rightarrow$  in the right-hand side:  $F$

i. Yes

ii. No

iii. Not enough information

iv. Syntax error in sentence

# Previous Poll

Truth table for  $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$

How many rows do we need (excluding a header)?

- i. 2
- ii. 4
- iii. 6
- iv. 8
- v. 16

# Previous Poll

Truth table for  $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$

How many columns should we have?

i. 2

ii. 3 ←

iii. 4

iv. 5 ←

v. 6

7 ←

		$\neg B$	$\neg A$	
A	B			
T	T			
T	F			
F	T			
F	F			

$A \Rightarrow B$	$\neg B \Rightarrow \neg A$	$\Leftrightarrow$

# Exercise

Truth table for  $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$

# Exercise

Truth table for  $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$

$A$	$B$	$\neg A$	$\neg B$	$A \Rightarrow B$	$\neg B \Rightarrow \neg A$	$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$
F	F	T	T	T	T	T
F	T	T	F	T	T	T
T	F	F	T	F	F	T
T	T	F	F	T	T	T



# Proof Techniques

## Proof by cases

## Disproof by counterexample

- One example is not sufficient to prove
- One counterexample is sufficient to disprove

## Proof by contrapositive

- Law of contrapositive

$$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$$

- Prove  $(\neg B \Rightarrow \neg A) \rightarrow$  Conclude  $(A \Rightarrow B)$

# Proof by Contrapositive

Proposition: If  $a, b \in \mathbb{Z}$  s.t.  $a+b$  is even, then  $a$  and  $b$  have the same *parity*

→ whiteboard

# Proof Techniques

## Proof by cases

## Disproof by counterexample

- One example is not sufficient to prove
- One counterexample is sufficient to disprove

## Proof by contrapositive

- Law of contrapositive

$$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$$

- Prove  $(\neg B \Rightarrow \neg A) \rightarrow$  Conclude  $(A \Rightarrow B)$

## Proof by contradiction

# Proof by Contradiction

## Template

# Example from Intro ML

Note: Just an example. Out of scope.

## PAC Learning: Theorem 1

A  $h \in \mathcal{H}$  is consistent with the training data if  $\hat{R}(h) = 0$

Theorem 1:


$$N \geq \frac{1}{\epsilon} \left[ \log |\mathcal{H}| + \log \frac{1}{\delta} \right]$$

$N$  examples is sufficient to ensure that

with prob  $1 - \delta$

all  $h \in \mathcal{H}$  with  $\hat{R}(h) = 0$  have  $R(h) \leq \epsilon$

Union bound

$$P(A \cup B) \leq P(A) + P(B)$$


$$P(\hat{R}(h_1) = 0) + P(\hat{R}(h_2) = 0) + P(\hat{R}(h_3) = 0) + \dots$$

① Assume  $k$  "bad" hypotheses  $h_1, h_2, \dots, h_k$  with  $R(h_i) > \epsilon$

② Pick bad  $h_i$ : Prob  $h_i$  is consistent w/ first train point  $\leq 1 - \epsilon$   
 Prob  $h_i$  is consistent w/ first  $N$  train points  $\leq (1 - \epsilon)^N$

③ Prob at least one bad  $h_i$  is consistent w/ first  $N$  train points  $\leq k(1 - \epsilon)^N$   
 $\leq |\mathcal{H}|(1 - \epsilon)^N$

④  $1 - x \leq e^{-x} \Rightarrow |\mathcal{H}|(1 - \epsilon)^N \leq |\mathcal{H}|e^{-\epsilon N}$

⑤ Calc  $N$  and  $\delta$  s.t.  $|\mathcal{H}|e^{-\epsilon N} \leq \delta$

⑥ Solve  $N$   $|\mathcal{H}|/\delta \leq \frac{1}{e^{-\epsilon N}}$

$$\log |\mathcal{H}| + \log 1/\delta \leq \epsilon N$$

Assume  $N \geq \frac{1}{\epsilon} [\log |\mathcal{H}| + \log 1/\delta]$

with prob  $\delta$

$\exists h$  s.t.  $R(h) > \epsilon$  and  $\hat{R}(h) = 0$  "bad"

with prob  $1 - \delta$   
 all  $h \in \mathcal{H}$  with  $R(h) > \epsilon$  have  $\hat{R}(h) > 0$  "good"

$\Rightarrow$  all  $h \in \mathcal{H}$  with  $\hat{R}(h) = 0$  have  $R(h) \leq \epsilon$

Contrapositive

$$A \Rightarrow B$$

$$\neg B \Rightarrow \neg A$$

# Plan

Wrap-up Overview → Lecture 1 slides

## Proof Techniques

- ✓ Proof by cases
- ✓ Disprove by counterexample
- ✓ Proof by contrapositive
- ✓ Proof by contradiction

## Perceptron Algorithm

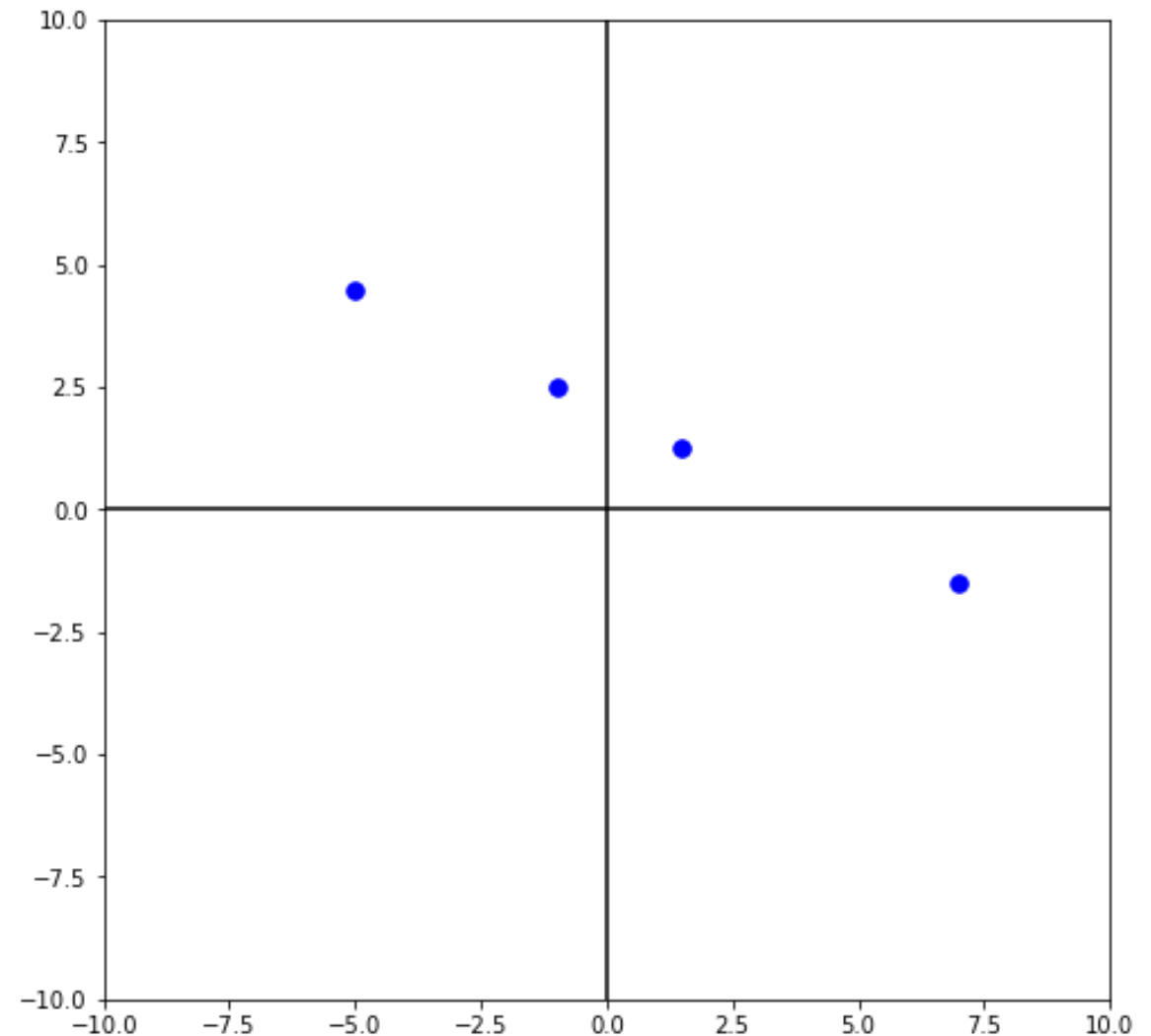
- Prep: ML tasks, data, notation
- Prep: Geometry of linear models

# ML Data, Tasks, Notation

Notation alert!

## Regression

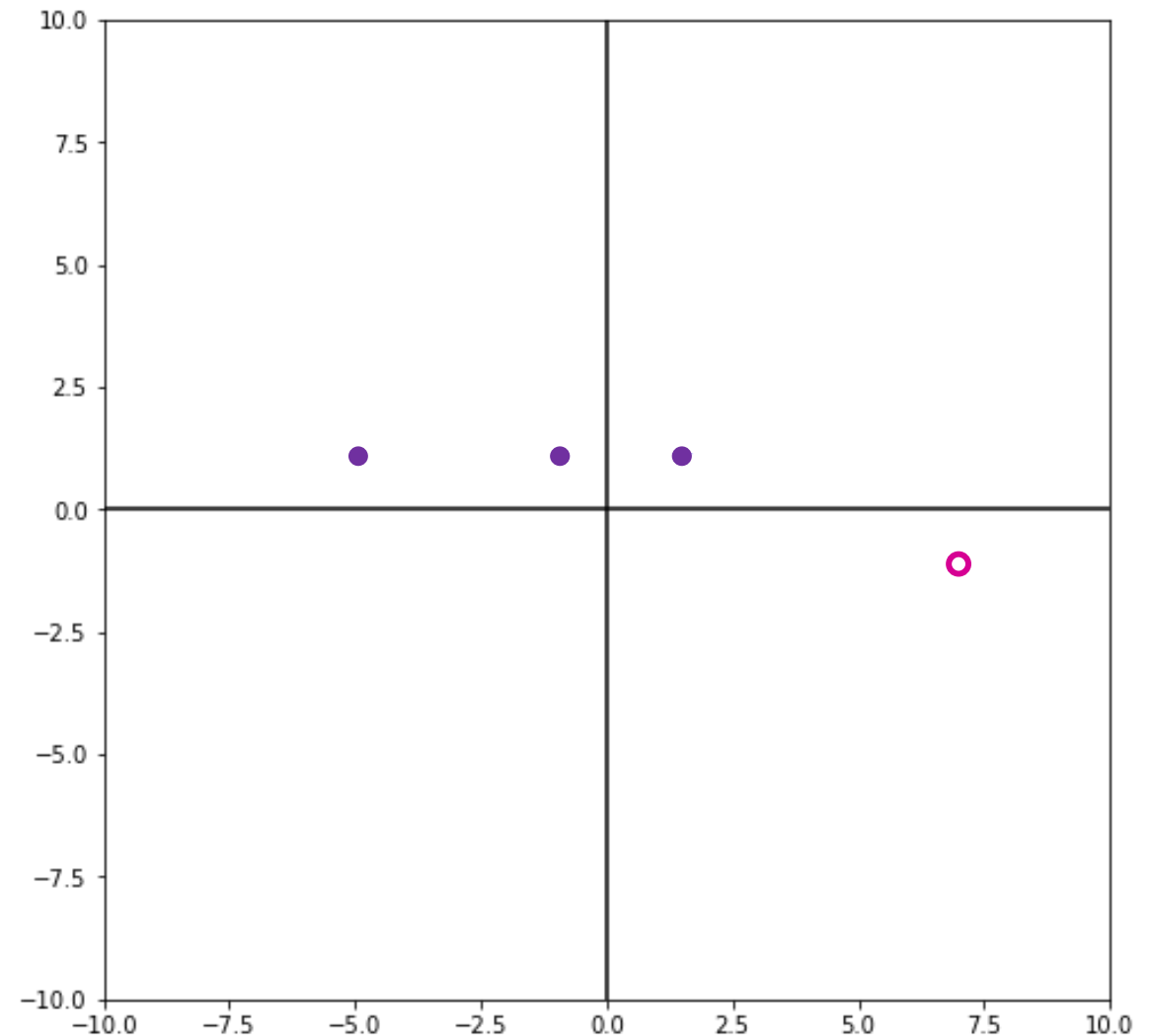
$$\begin{aligned}\mathcal{D} &= \{(x^{(i)}, y^{(i)})\}_{i=1}^4 \\ &= \{(-1, 2.5), \\ &\quad (-7, -1.5), \\ &\quad (-5, 4.5), \\ &\quad (1.5, 1.25)\}\end{aligned}$$



# ML Data, Tasks, Notation

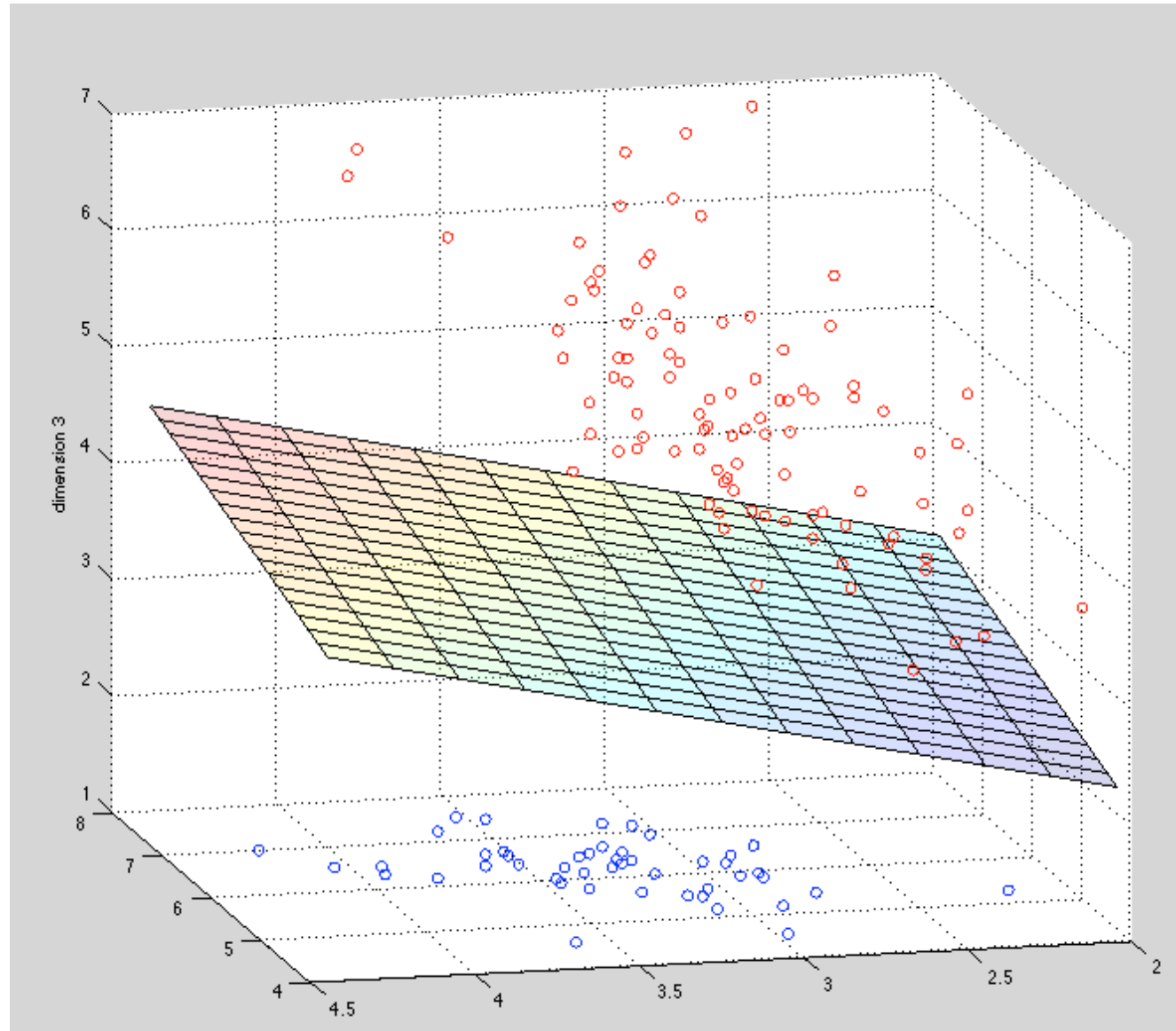
## Classification

$$\begin{aligned}\mathcal{D} &= \{(x^{(i)}, y^{(i)})\}_{i=1}^4 \\ &= \{(-1, 1), \\ &\quad (-7, -1), \\ &\quad (-5, 1), \\ &\quad (1.5, 1)\}\end{aligned}$$





# Perceptron



# Exercise

## Geometry

Draw a picture of the region corresponding to:

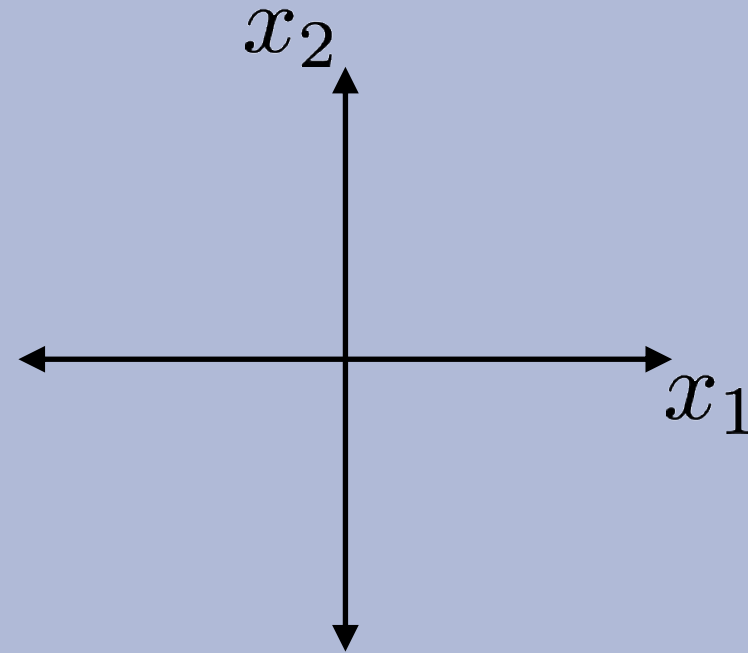
$$w_1x_1 + w_2x_2 + b > 0$$

$$\text{where } w_1 = 2, w_2 = 3, b = 6$$

Draw the vector

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

Answer Here:



# Poll 1

Which is the correct vector  $\mathbf{w}$ ?

- A.
- B.
- C.
- D.
- E. I don't know

