

6.036/6.862: Introduction to Machine Learning

Lecture: starts Tuesdays 9:35am (Boston time zone)

Course website: introml.odl.mit.edu

Who's talking? Prof. Tamara Broderick

Questions? Ask on Discourse: discourse.odl.mit.edu

Materials: Will all be available at course website

Today's Plan

- I. (More) logistics
- II. Machine learning setup
- III. Linear classifiers

6.036/6.862: Introduction to Machine Learning

Lecture: starts Tuesdays 9:35am (Boston time zone)

Course website: introml.odl.mit.edu

Who's talking? Prof. Tamara Broderick

Questions? Ask on Discourse: discourse.odl.mit.edu

Materials: Will all be available at course website

Today's Plan

- I. (More) logistics
- II. Machine learning setup
- III. Linear classifiers



The image shows a screenshot of a Discourse forum interface. At the top, it says "Example Question for Lecture 1" with a pencil icon. Below that is a legend: a brown square for "Lecture Questions" and a red square for "Lecture 1". A user profile is shown with the letter "C" in a red circle and the name "cwang506" next to a shield icon. The background of the interface is white.

(set “Lecture 1” category)

Is Introduction to Machine Learning (6.036/6.862) right for you?

Is Introduction to Machine Learning
(6.036/6.862) right for you?

Computer Science Prerequisites

Is Introduction to Machine Learning
(6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

- Matrix manipulations (inverse, transpose, multiplication, etc.)

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

- Matrix manipulations (inverse, transpose, multiplication, etc.)
- Points and planes in dimension > 2

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

- Matrix manipulations (inverse, transpose, multiplication, etc.)
- Points and planes in dimension > 2
- Gradients

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

- Matrix manipulations (inverse, transpose, multiplication, etc.)
- Points and planes in dimension > 2
- Gradients
- Basic discrete probability (random variables, independence, conditioning, etc.)

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

- Matrix manipulations (inverse, transpose, multiplication, etc.)
- Points and planes in dimension > 2
- Gradients
- Basic discrete probability (random variables, independence, conditioning, etc.)

▼ Welcome to 6.036

[Announcements](#)

[Schedule Survey](#)

[Basic Information](#)

[Readiness Assessment](#)

[Grading Policies](#)

[Collaboration Policy](#)

[Teaching Staff](#)

[Software](#)

[Numpy Tutorial](#)

[Course calendar](#)

Is Introduction to Machine Learning (6.036/6.862) right for you?

Computer Science Prerequisites

- Python programming
- Algorithms (read & understand pseudocode)

Math Prerequisites

- Matrix manipulations (inverse, transpose, multiplication, etc.)
- Points and planes in dimension > 2
- Gradients
- Basic discrete probability (random variables, independence, conditioning, etc.)

▼ Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assessment

Grading Policies

Collaboration Policy

Teaching Staff

Software

Numpy Tutorial

Course calendar

6.036/6.862: Introduction to Machine Learning

6.036/6.862: Introduction to Machine Learning, Staff

6.036/6.862: Introduction to Machine Learning, Staff

Instructors:



**Jehangir
Amjad**



**Duane
Boning**



**Tamara
Broderick**



**Ike
Chuang**



**Iddo
Drori**



**Phillip
Isola**



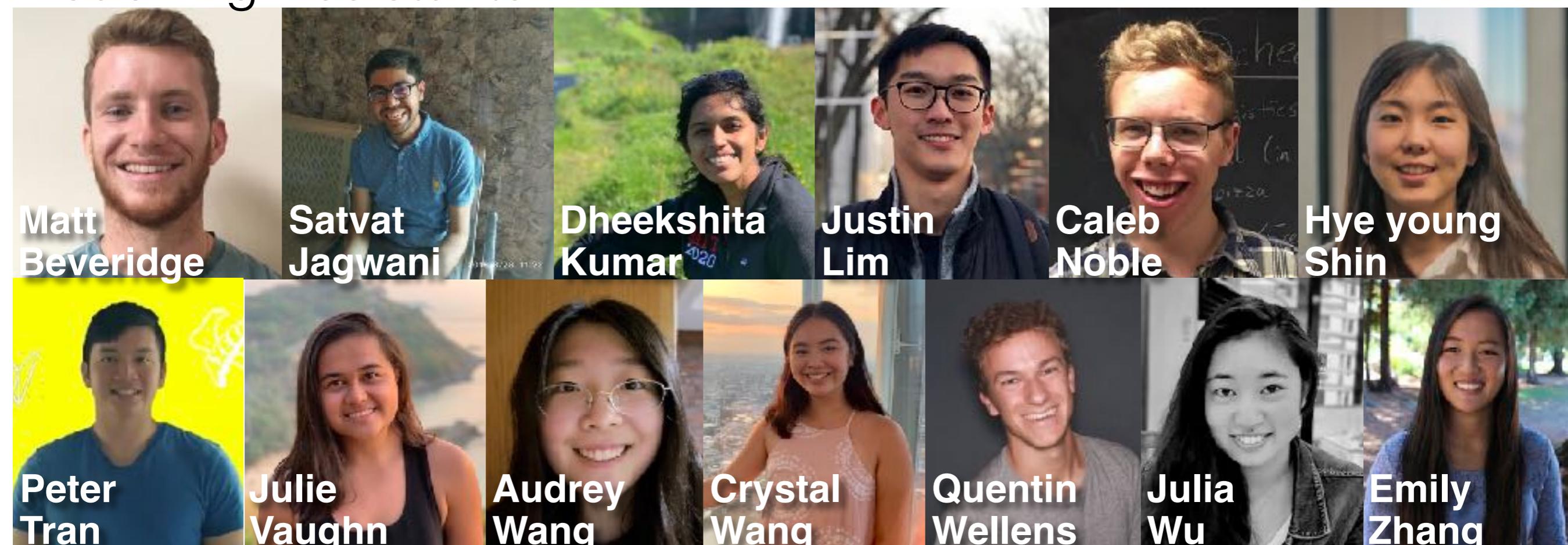
**David
Sontag**

6.036/6.862: Introduction to Machine Learning, Staff

Instructors:

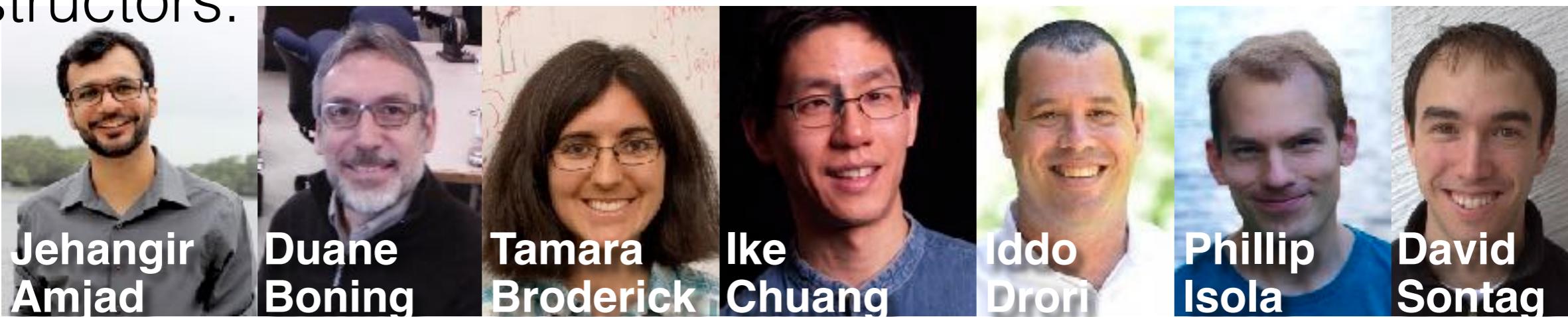


Teaching Assistants:

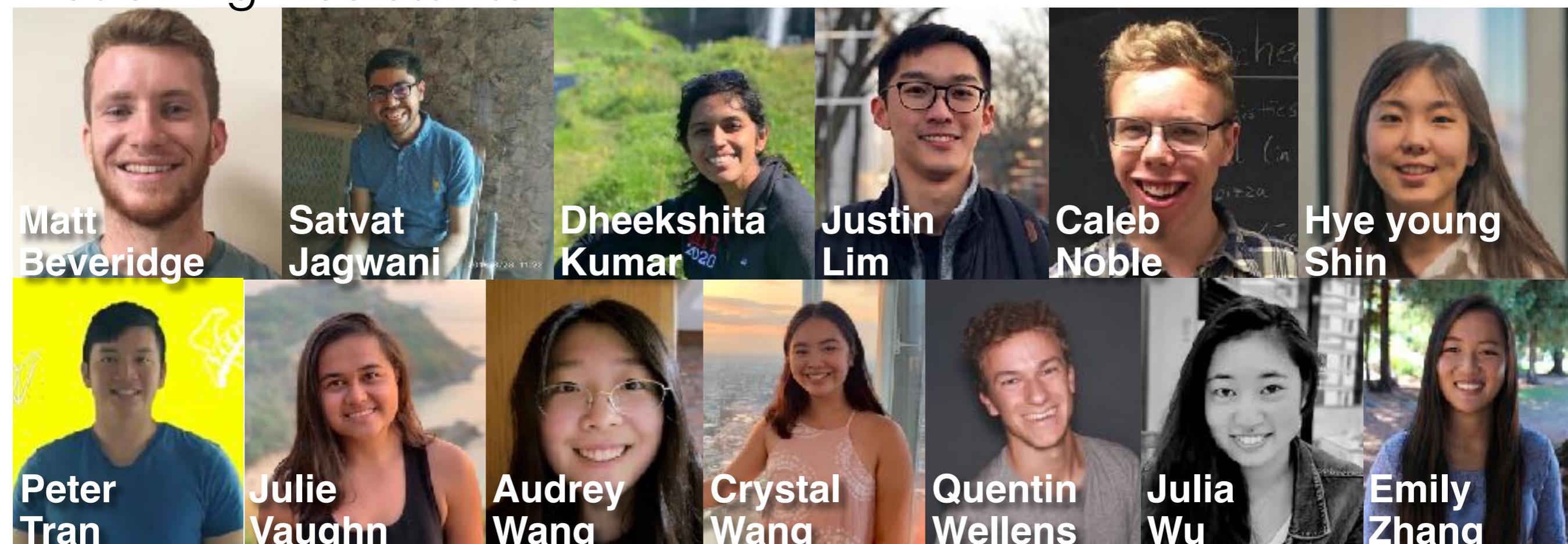


6.036/6.862: Introduction to Machine Learning, Staff

Instructors:



Teaching Assistants:



6.036/6.862: Introduction to Machine Learning, Weekly Plan

▼ Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assessment

Grading Policies

Collaboration Policy

Teaching Staff

Software

Numpy Tutorial

Course calendar

6.036/6.862: Introduction to Machine Learning, Weekly Plan

▼ Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assessment

Grading Policies

Collaboration Policy

Teaching Staff

Software

Numpy Tutorial

Course calendar

6.036/6.862: Introduction to Machine Learning, Weekly Plan

▼ Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assess

Grading Policies

Collaboration Policy

Teaching Staff

Software

Numpy Tutorial

Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes

▼ Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assess

Grading Policies

Collaboration Policy

Teaching Staff

Software

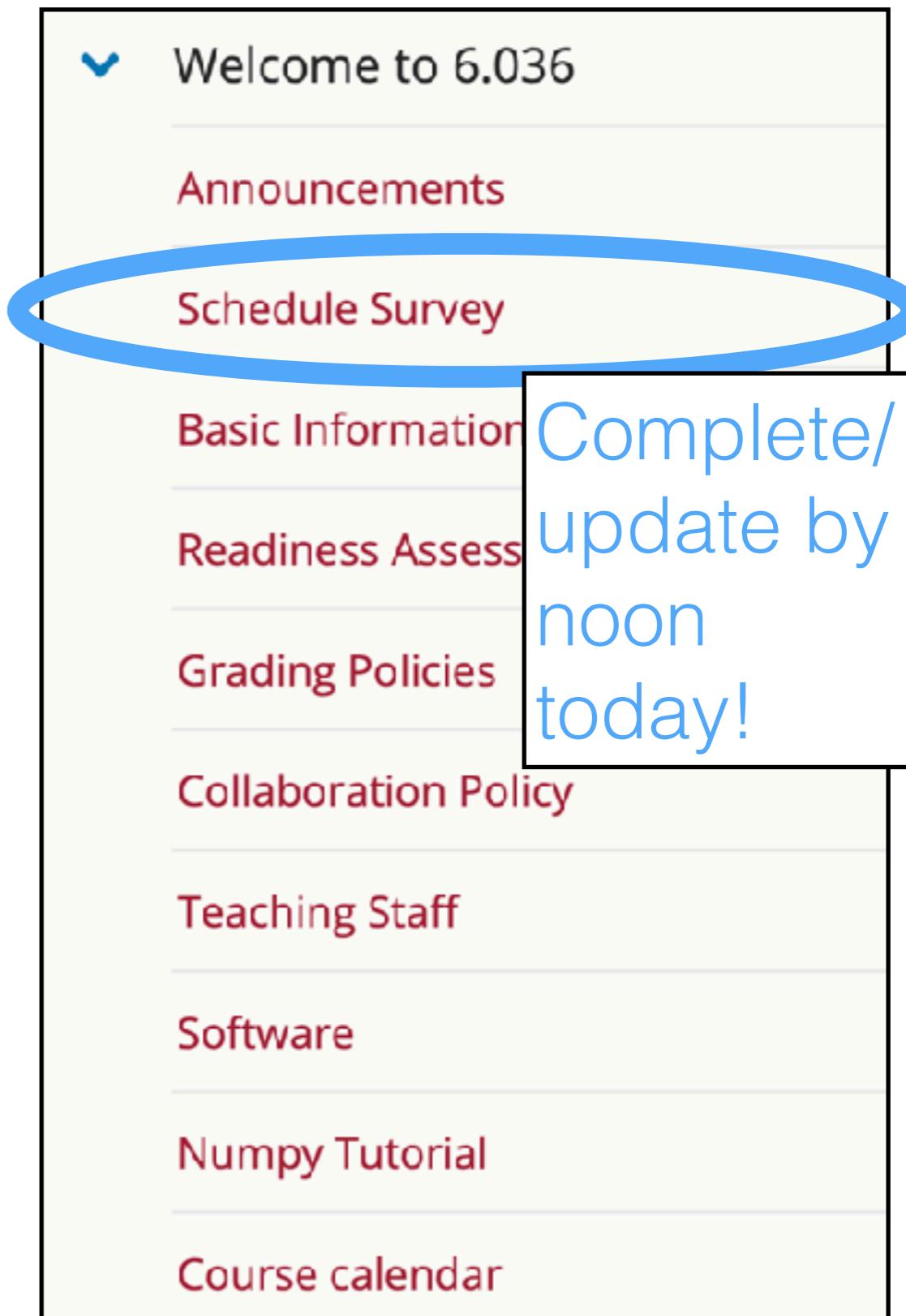
Numpy Tutorial

Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture



Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assess

Grading Policies

Collaboration Policy

Teaching Staff

Software

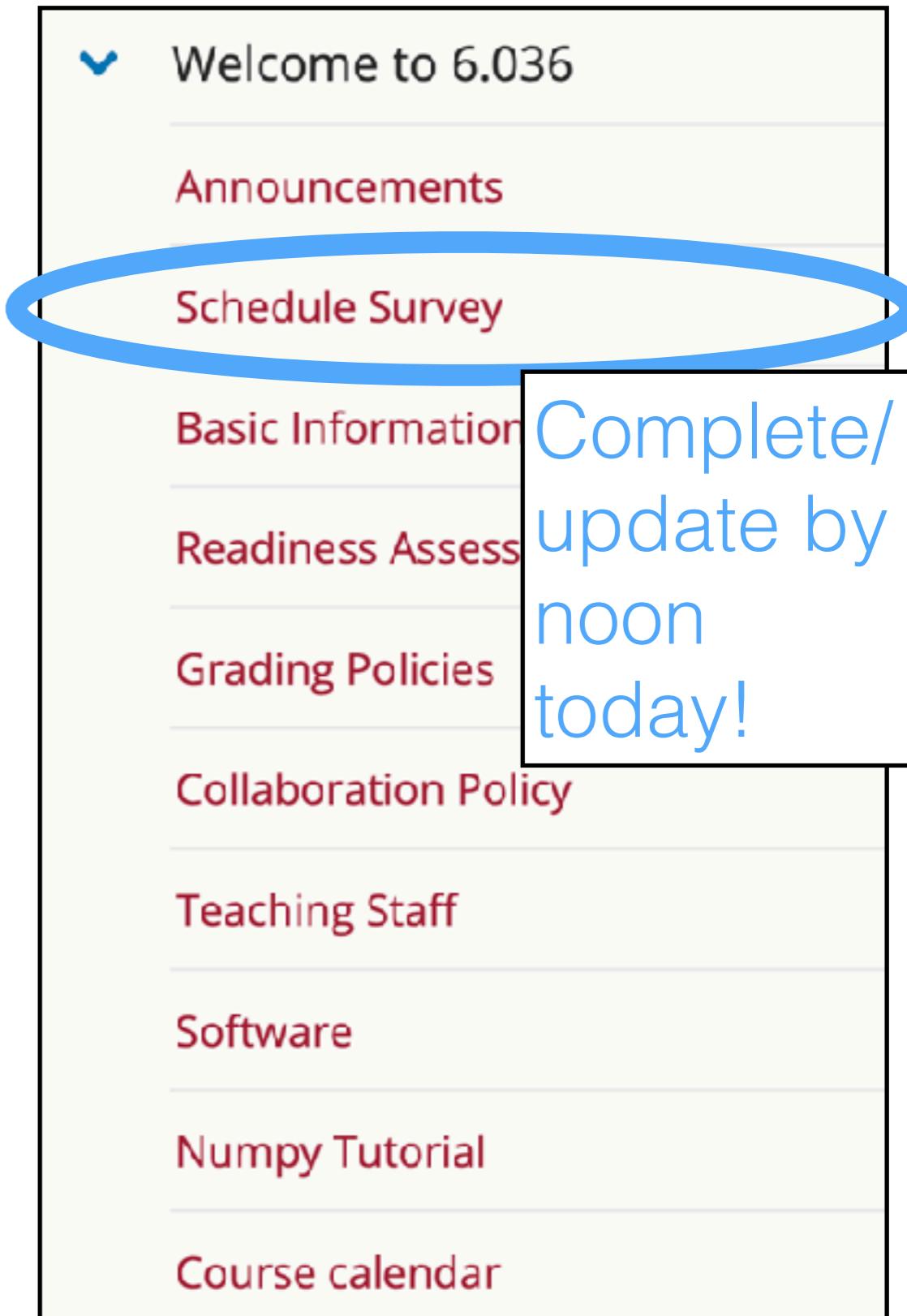
Numpy Tutorial

Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)

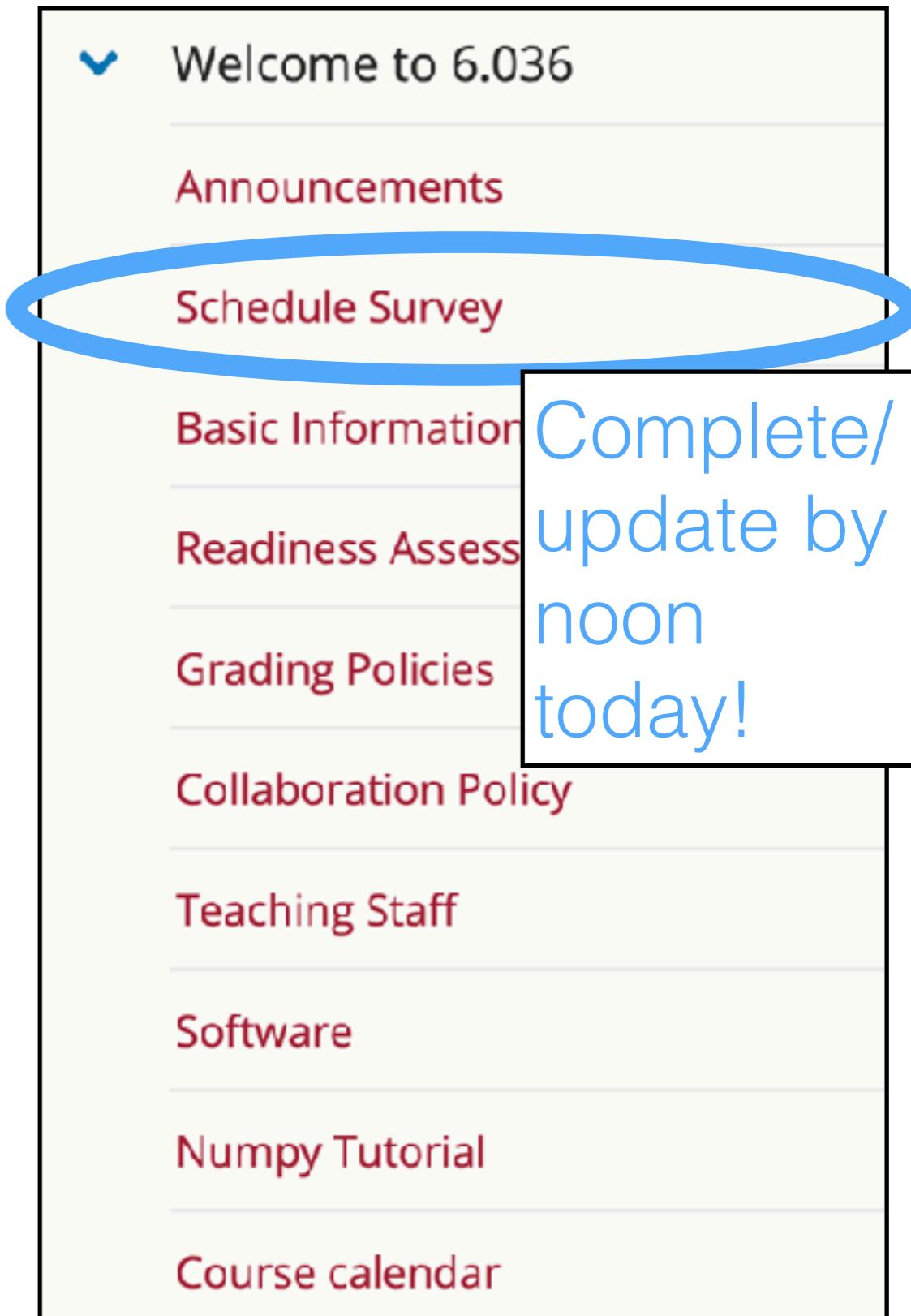


- ▼ Welcome to 6.036
- Announcements
- Schedule Survey
- Basic Information
- Readiness Assess
- Grading Policies
- Collaboration Policy
- Teaching Staff
- Software
- Numpy Tutorial
- Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students



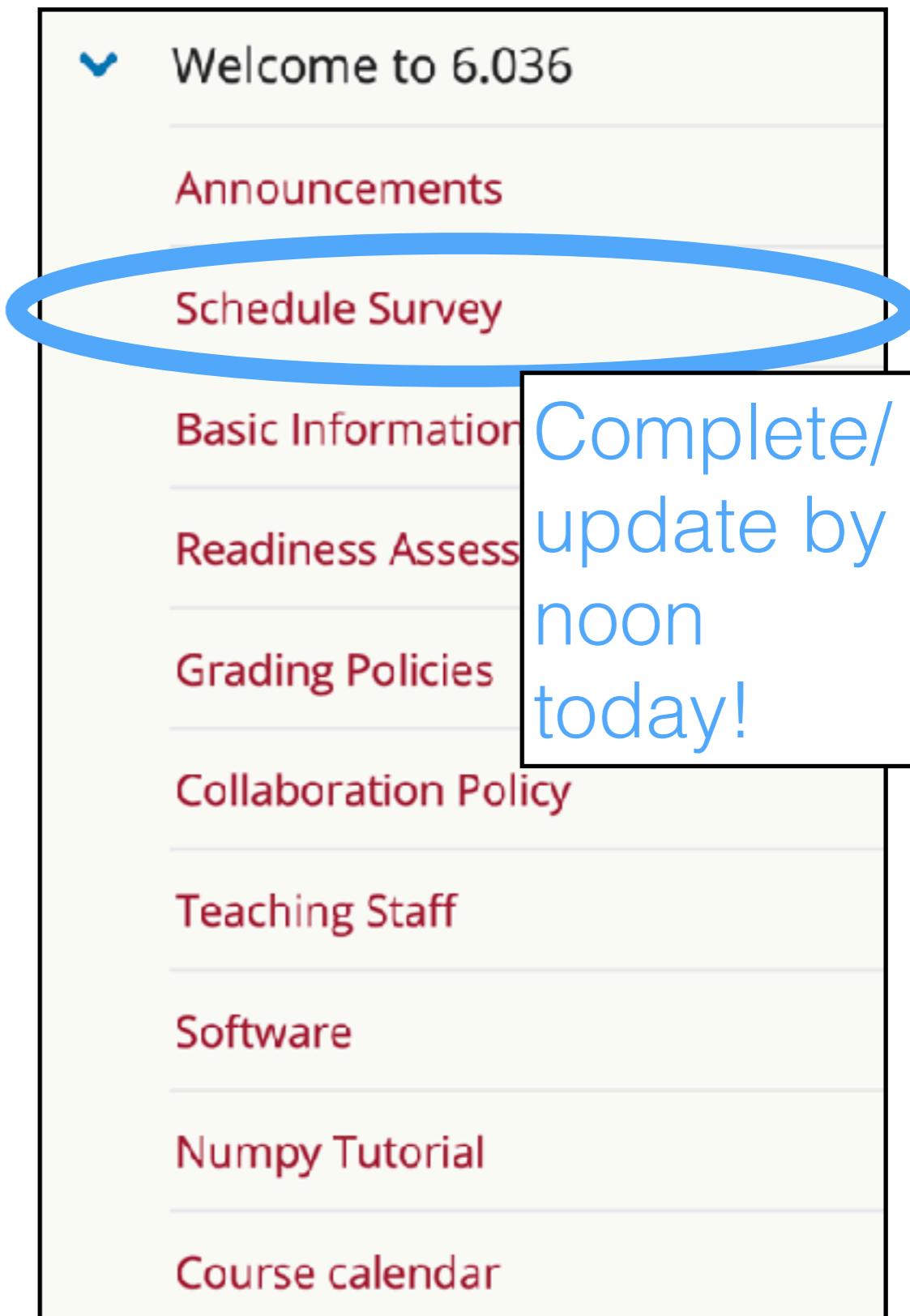
The image shows a vertical list of course links. The link 'Schedule Survey' is highlighted with a blue oval. A blue callout box is positioned to the right of this link, containing the text 'Complete/ update by noon today!'. The other links in the list are: Welcome to 6.036, Announcements, Basic Information, Readiness Assess, Grading Policies, Collaboration Policy, Teaching Staff, Software, Numpy Tutorial, and Course calendar.

- Welcome to 6.036
- Announcements
- Schedule Survey
- Basic Information
- Readiness Assess
- Grading Policies
- Collaboration Policy
- Teaching Staff
- Software
- Numpy Tutorial
- Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff



▼ Welcome to 6.036

Announcements

Schedule Survey

Basic Information

Readiness Assess

Grading Policies

Collaboration Policy

Teaching Staff

Software

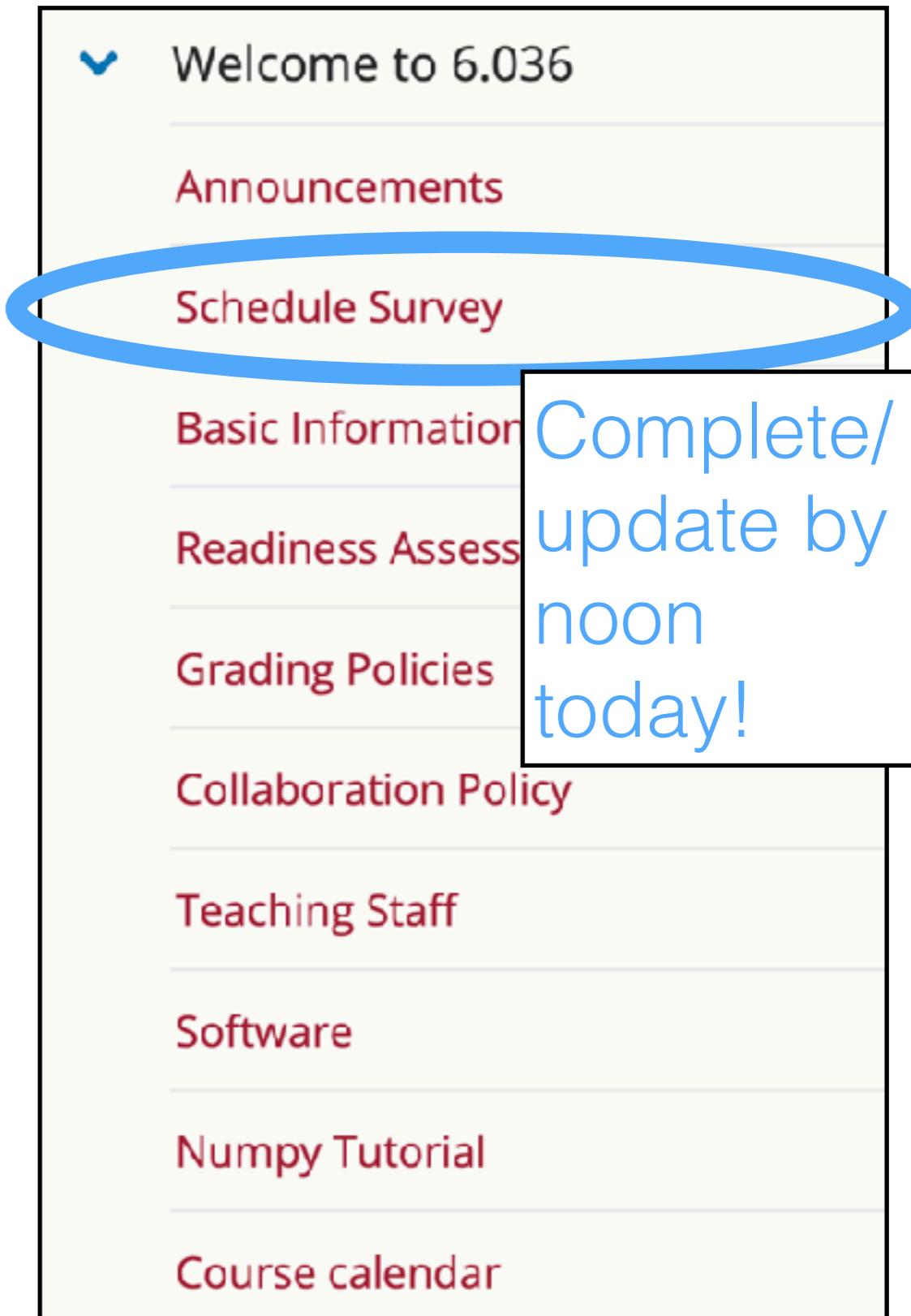
Numpy Tutorial

Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**



The image shows a vertical list of course links. The link 'Schedule Survey' is highlighted with a large blue oval. To the right of this oval, there is a black-bordered box containing the text 'Complete/ update by noon today!'. The other links in the list are: Welcome to 6.036, Announcements, Basic Information, Readiness Assess, Grading Policies, Collaboration Policy, Teaching Staff, Software, Numpy Tutorial, and Course calendar.

- Welcome to 6.036
- Announcements
- Schedule Survey
- Basic Information
- Readiness Assess
- Grading Policies
- Collaboration Policy
- Teaching Staff
- Software
- Numpy Tutorial
- Course calendar

Complete/
update by
noon
today!

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**

▼ Week 1: Basics

Week 1 Live Lecture

Introduction to ML

Linear classifiers

 **Week 1 Nanoquiz**

NQ due Sep 4, 2020 16:00 EDT

 **Week 1 Lab**

LAB due Sep 7, 2020 21:00 EDT

 **Homework 1**

HW due Sep 9, 2020 23:00 EDT

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**

▼ Week 1: Basics

Week 1 Live Lecture

Introduction to ML

Linear classifiers

Week 1 Nanoquiz

NQ due Sep 4, 2020 16:00 EDT

Week 1 Lab

LAB due Sep 7, 2020 21:00 EDT

Homework 1

HW due Sep 9, 2020 23:00 EDT

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**
- **Nanoquiz** (no midterm/final)
 - Timed

▼ Week 1: Basics

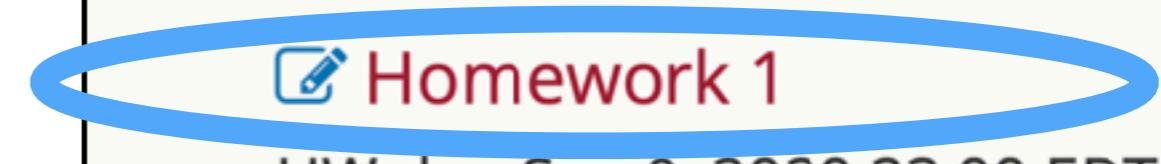
Week 1 Live Lecture

Introduction to ML

Linear classifiers

Week 1 Nanoquiz
NQ due Sep 4, 2020 16:00 EDT

Week 1 Lab
LAB due Sep 7, 2020 21:00 EDT

 **Homework 1**
HW due Sep 9, 2020 23:00 EDT

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**
- **Nanoquiz** (no midterm/final)
 - Timed
- **Office hours**

▼ Week 1: Basics

Week 1 Live Lecture

Introduction to ML

Linear classifiers

Week 1 Nanoquiz

NQ due Sep 4, 2020 16:00 EDT

Week 1 Lab

LAB due Sep 7, 2020 21:00 EDT

Homework 1

HW due Sep 9, 2020 23:00 EDT

6.036/6.862: Introduction to Machine Learning, Weekly Plan

- **Lecture** + course notes
- **Exercises**
 - Due 9am before lecture
- **Lab** (*synchronous, required!*)
 - (New!) MLyPod: 10 students
 - Work in groups of 2 to 3;
check off with staff
- **Homework**
- **Nanoquiz** (no midterm/final)
 - Timed
- **Office hours**
- **6.862:** project (canvas.mit.edu)

▼ Week 1: Basics

Week 1 Live Lecture

Introduction to ML

Linear classifiers

Week 1 Nanoquiz
NQ due Sep 4, 2020 16:00 EDT

Week 1 Lab
LAB due Sep 7, 2020 21:00 EDT

Homework 1
HW due Sep 9, 2020 23:00 EDT

Machine learning (ML): why & what

Machine learning (ML): why & what

Machine learning algorithm confirms 50 new exoplanets in historic first



by R. Dallon Adams in Innovation on August 26, 2020, 9:07 AM PST

A new machine learning technique can be used to sift through massive datasets to discern exoplanets from false positives.



Machine learning (ML): why & what

Machine learning algorithm confirms 50 new exoplanets in historic first

THE LANCET

Child & Adolescent Health

A new
dataset

ARTICLES | ONLINE FIRST

A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial

Andreea M Pavel, MD • Janet M Rennie, MD • Linda S de Vries, PhD • Mats Blennow, PhD • Adrienne Foran, MD •
Divyen K Shah, MD • et al. Show all authors

Open Access • Published: August 27, 2020 • DOI: [https://doi.org/10.1016/S2352-4642\(20\)30239-X](https://doi.org/10.1016/S2352-4642(20)30239-X)



Machine learning (ML): why & what

Machine exoplanets
THE LA Child & J

A new dataset

recognition: a

PhD • Adrienne Foran, MD •

42(20)30239-X •

At America's court of last resort, a handful of lawyers now dominates the docket

By [Joan Biskupic](#), [Janet Roberts](#) and [John Shiffman](#)

A small group of lawyers and its outsized influence at the U.S. Supreme Court



TOP TIER: In handling appeals heard by the U.S. Supreme Court, 75 lawyers have stood out — most for their success at getting cases before the high court, others for how often they argue those cases, and some for both reasons. Most of the 75 work at law firms that primarily represent businesses.

REUTERS | The Echo Chamber

Machine learning (ML): why & what

REUTERS | The Echo Chamber

A small group of lawyers and its outsized influence at the U.S. Supreme Court

THE LA Child & J

A new dataset

TOP TIER: In success at gel both reasons.

At A reso now

tion: a

Jeanne Foran, MD •

O-X •

Forbes

1,392 views | Aug 1, 2020, 01:12pm EDT

Using Machine Learning To Automate Data Coding At The Bureau Of Labor Statistics (BLS)

Kathleen Walch Contributor
COGNITIVE WORLD Contributor Group ⓘ
AI

By Joan Biskupic, Janet Roberts and John Shiffman

Machine learning (ML): why & what

The screenshot shows a news article from Forbes. At the top, there's a navigation bar with the REUTERS logo and 'The Echo Chamber'. Below that is a dark header with the word 'Forbes' in white. The main title of the article is '5 Ways Machine Learning Can Thwart Phishing Attacks', written in large, bold, black font. Below the title is a photo of a man, identified as the author, Louis Columbus. He is described as a 'Senior Contributor' in 'Enterprise & Cloud'. The article has social sharing icons for Facebook, Twitter, and LinkedIn. A large, abstract image of binary code in blue and red is used as a background for the article. On the left side of the main content area, there's a sidebar with various news snippets and a sidebar menu.

Machine exoplanets
THE LA Child & A
A new dataset
TOP TIER: In success at gel both reasons.
At A reso now
By Joan Biskupic, Janet R

REUTERS | The Echo Chamber

A small group of lawy

≡ Forbes

EDITORS' PICK | 1,248 views | Aug 12, 2020, 11:29am EDT

5 Ways Machine Learning Can Thwart Phishing Attacks

Louis Columbus Senior Contributor ⓘ
Enterprise & Cloud

f
t
in

GETTY

Machine learning (ML): why & what

The screenshot shows a news article from ABC Science. The main headline is "5 Ways Machine Learning Can Thwart". Below it, a sub-headline reads "Australian Federal Police officers trialled controversial facial recognition tool Clearview AI". The article is categorized under "SCIENCE". It includes a photo of the Australian Federal Police badge on a blue fabric background. The badge features a crown at the top, a central emblem, and the words "AUSTRALIAN FEDERAL POLICE" around the bottom.

REUTERS | The Echo Chamber

A small group of lawyers are challenging the use of Clearview AI, a controversial facial recognition technology developed by a U.S. company.

≡ Forbes

EDITORS' PICK | 1,248 views | Aug 12, 2020, 11:29am EDT

5 Ways Machine Learning Can Thwart

SCIENCE

Australian Federal Police officers trialled controversial facial recognition tool Clearview AI

ABC Science / By technology reporter Ariel Bogle

Posted Tue 14 Apr 2020 at 2:54am, updated Tue 14 Apr 2020 at 8:18pm

TOP TIER: In success at getting both reasons.

At A reso now

By Joan Biskupic, Janet R

Machine learning (ML): why & what

Machine exoplanet THE LA Child & A

A new dataset

TOP TIER: In success at get both reasons.

At A reso now

By Joan Biskupic, Janet R

REUTERS | The Echo Chamber

Forbes

A small group of lawy

EDITORS' P

5 W Lead

SCIENCE

Austra controve

ABC Science / By

Posted Tue 14 Apr

f t in

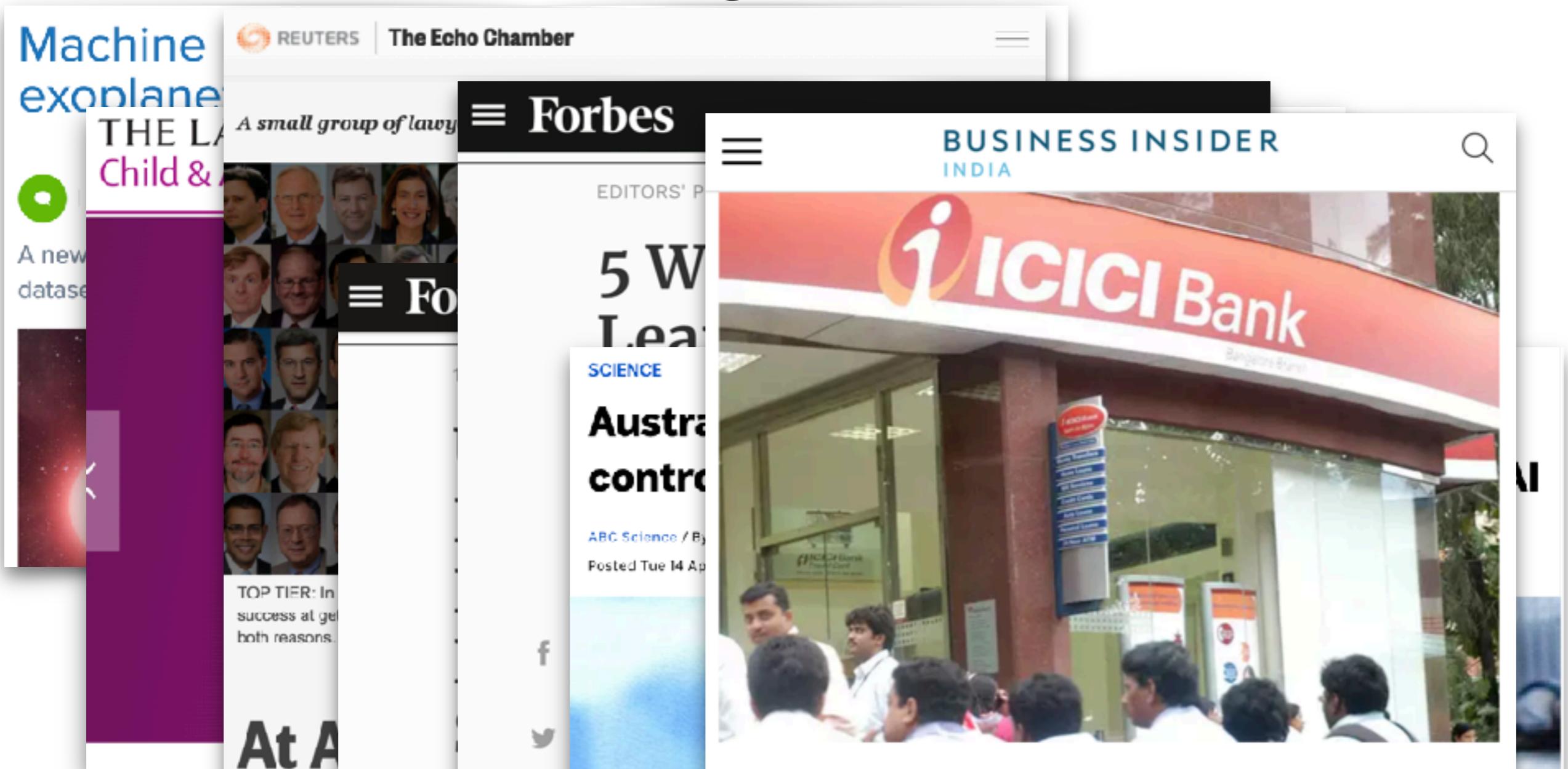
Business Insider INDIA

ICICI Bank

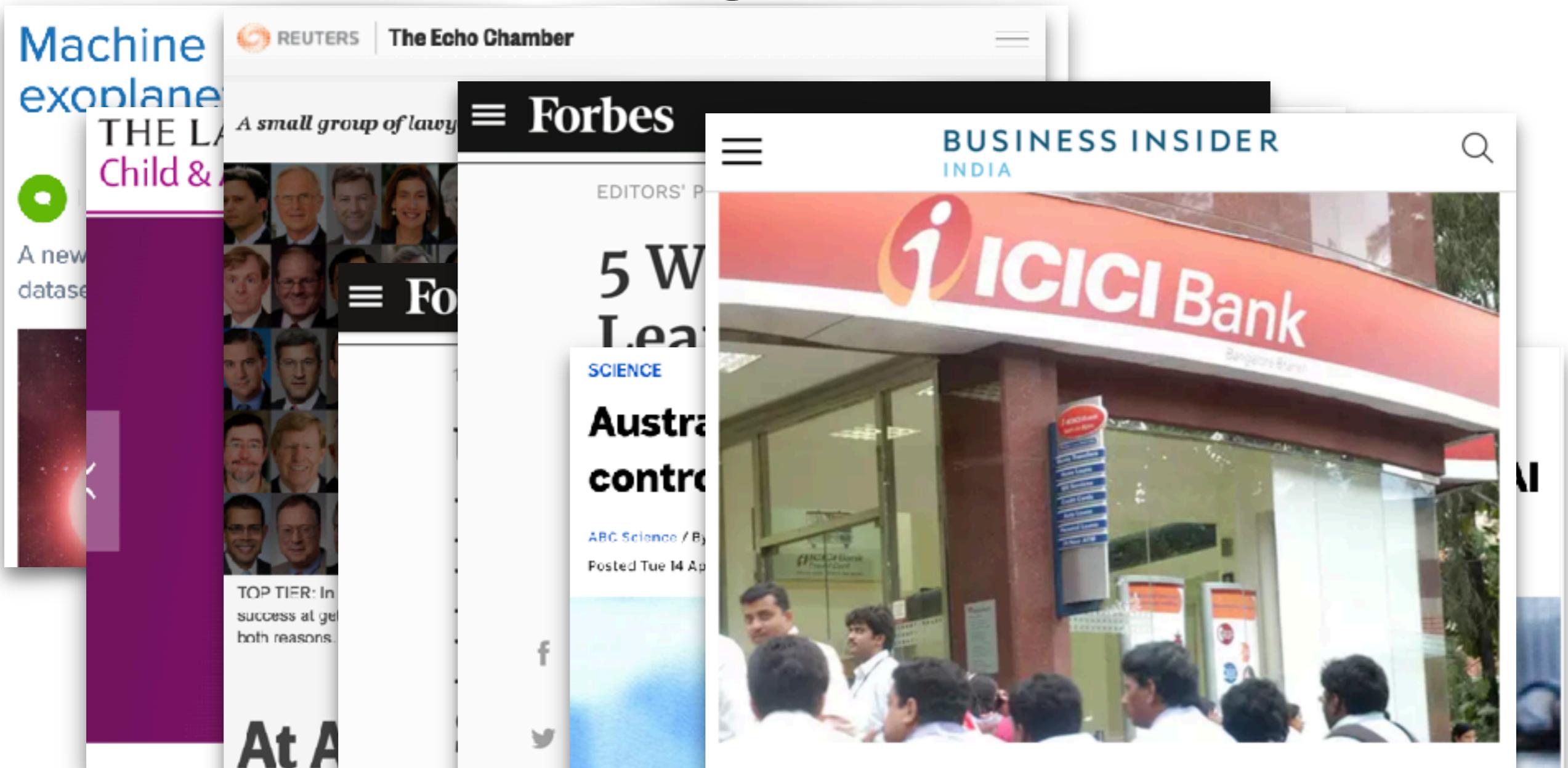
ICICI Bank will use satellite images to assess the credit worthiness of farmers BCCL

- **ICICI Bank's new machine learning (ML) algorithms use satellite data and images to determine whether a farmer is creditworthy or not.**

Machine learning (ML): why & what

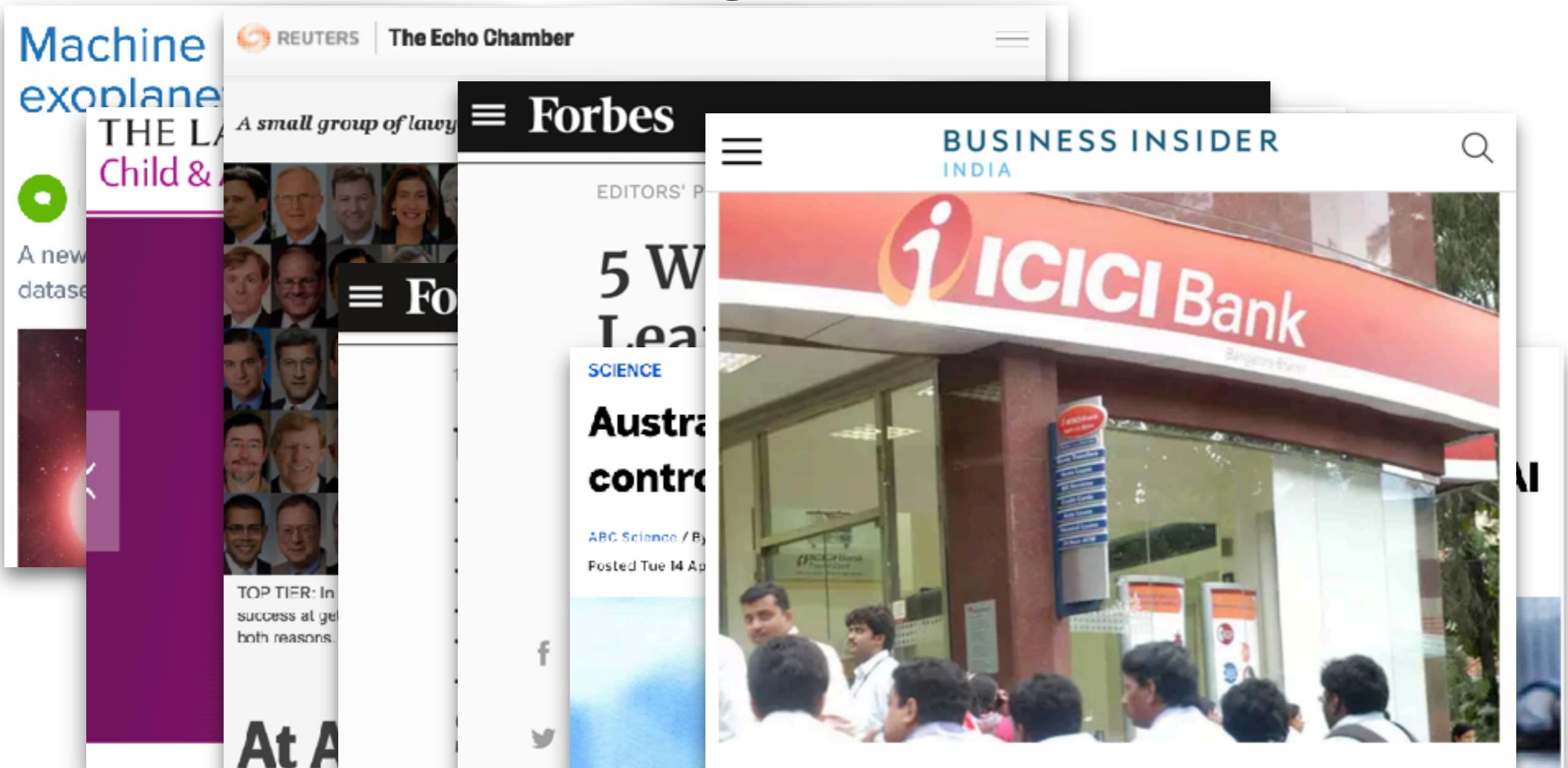


Machine learning (ML): why & what



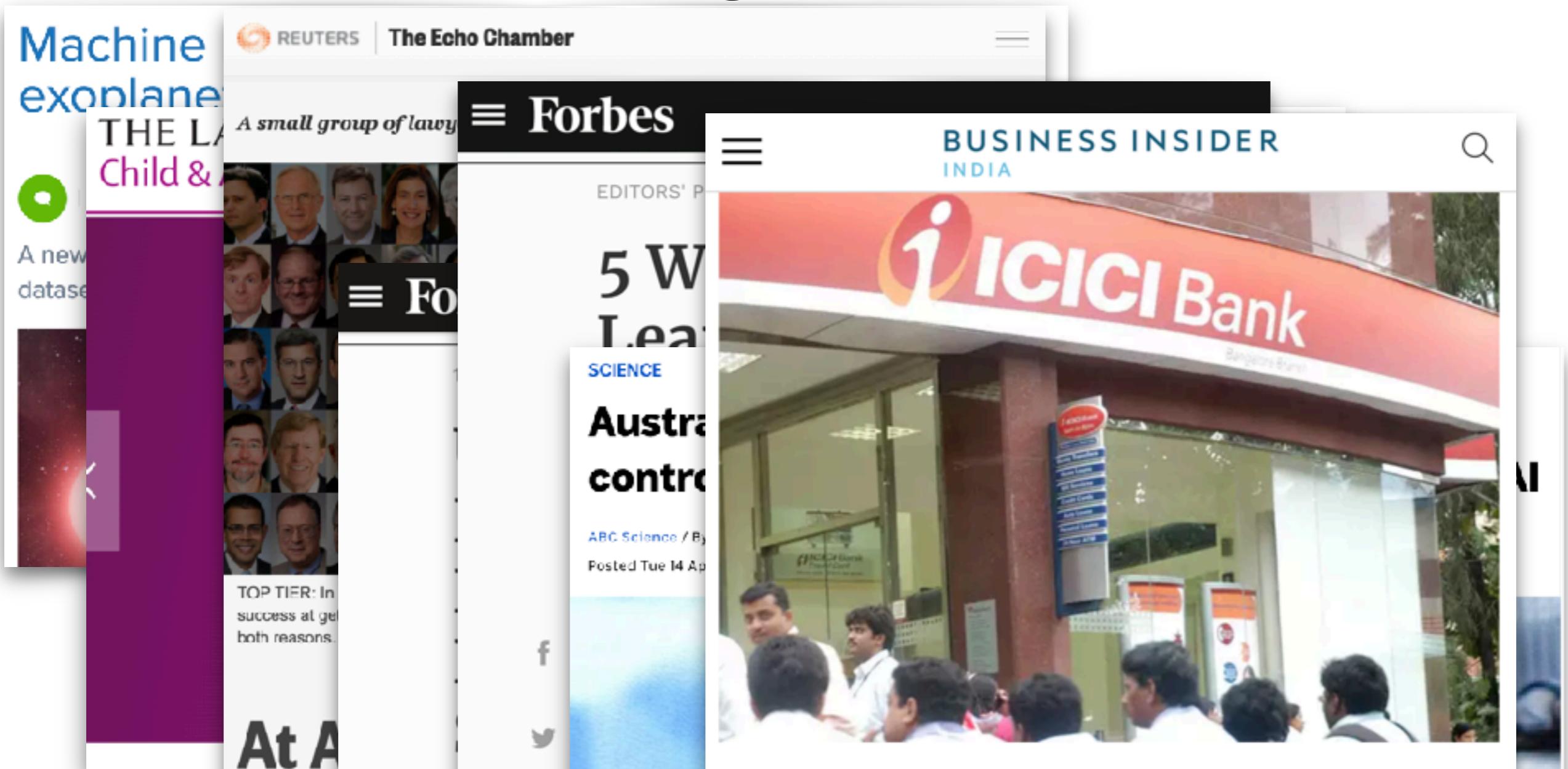
- What is ML?

Machine learning (ML): why & what



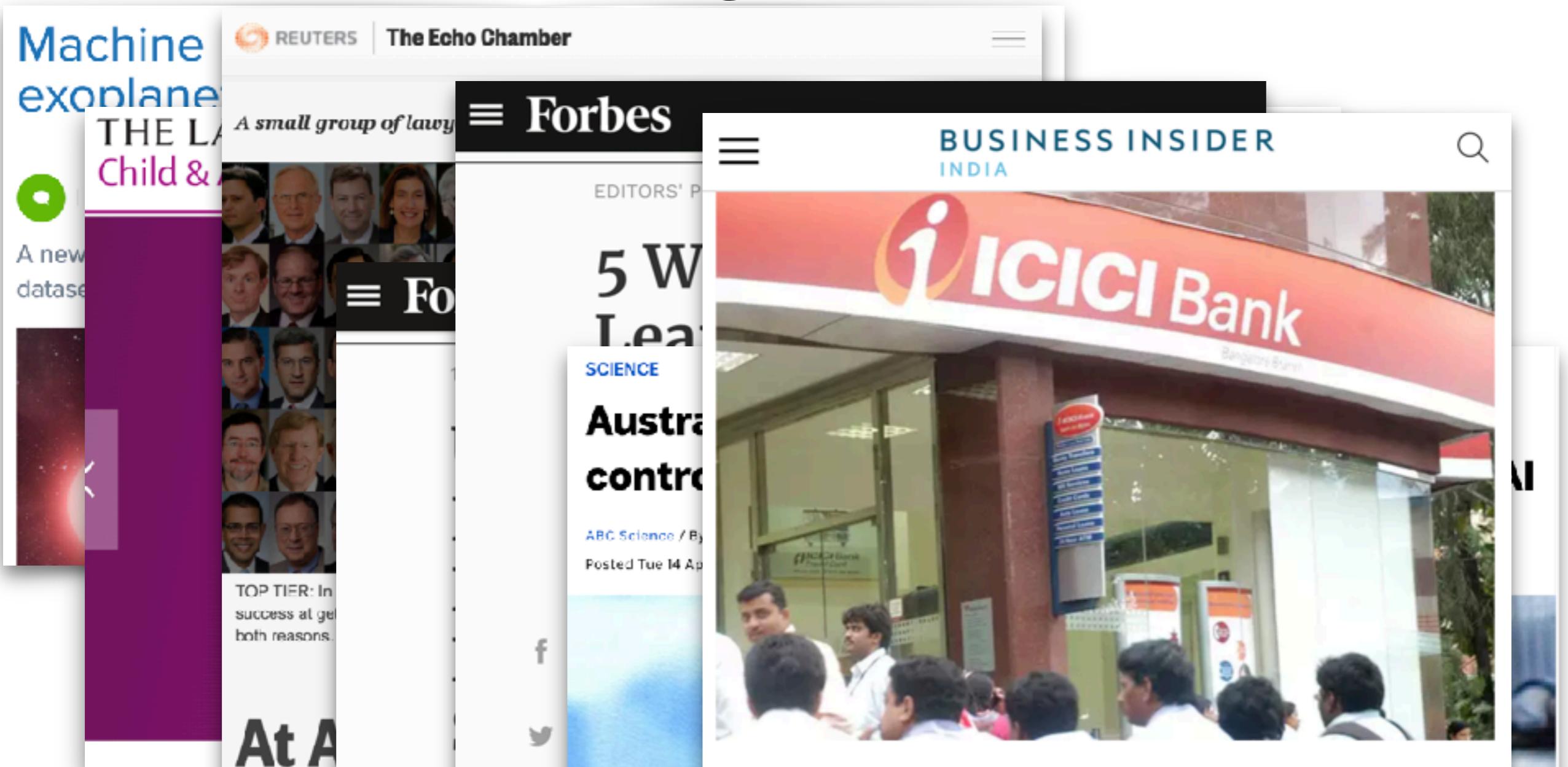
- **What is ML?** A set of methods for making decisions from data. (See the rest of the course!)

Machine learning (ML): why & what



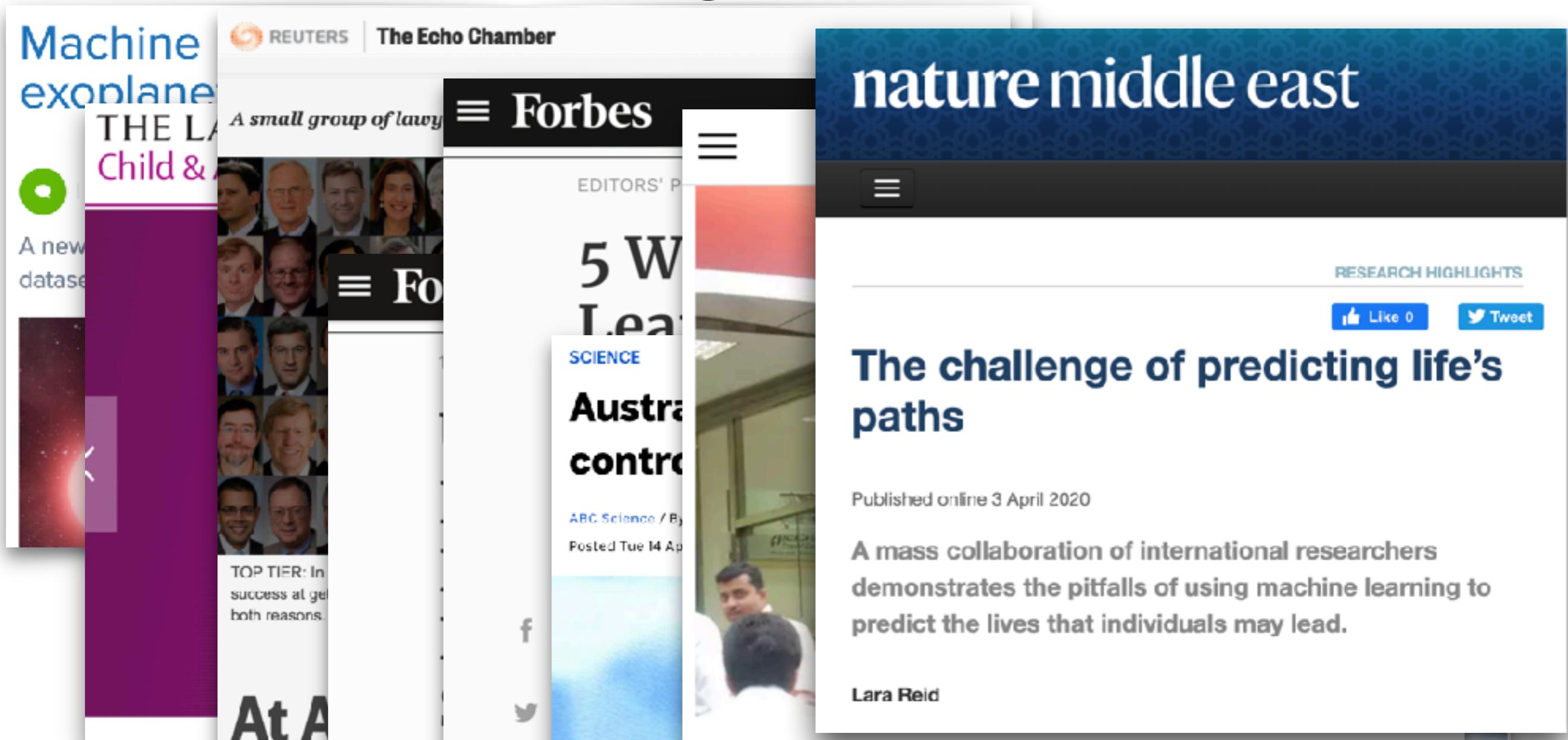
- **What is ML?** A set of methods for making decisions from data. (See the rest of the course!)
- **Why study ML?** To apply; to understand; to evaluate

Machine learning (ML): why & what



- **What is ML?** A set of methods for making decisions from data. (See the rest of the course!)
- **Why study ML?** To apply; to understand; to evaluate
- **Notes:** ML is not magic. ML is built on math.

Machine learning (ML): why & what



- **What is ML?** A set of methods for making decisions from data. (See the rest of the course!)
- **Why study ML?** To apply; to understand; to evaluate
- **Notes:** ML is not magic. ML is built on math.

Getting started

Getting started

What do we have?

Getting started

What do we have? (Training) data

Getting started

What do we have? (Training) data

- n training data points

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector

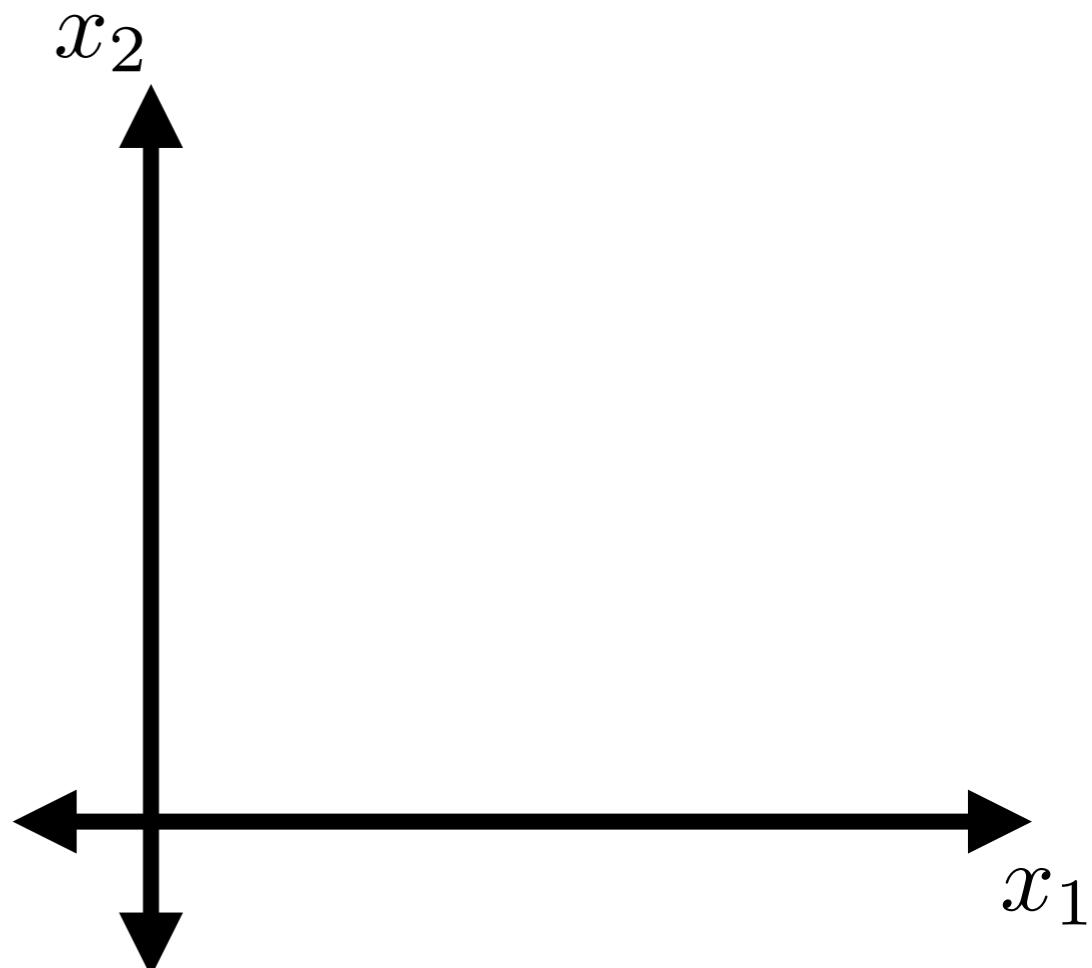
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector

$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$

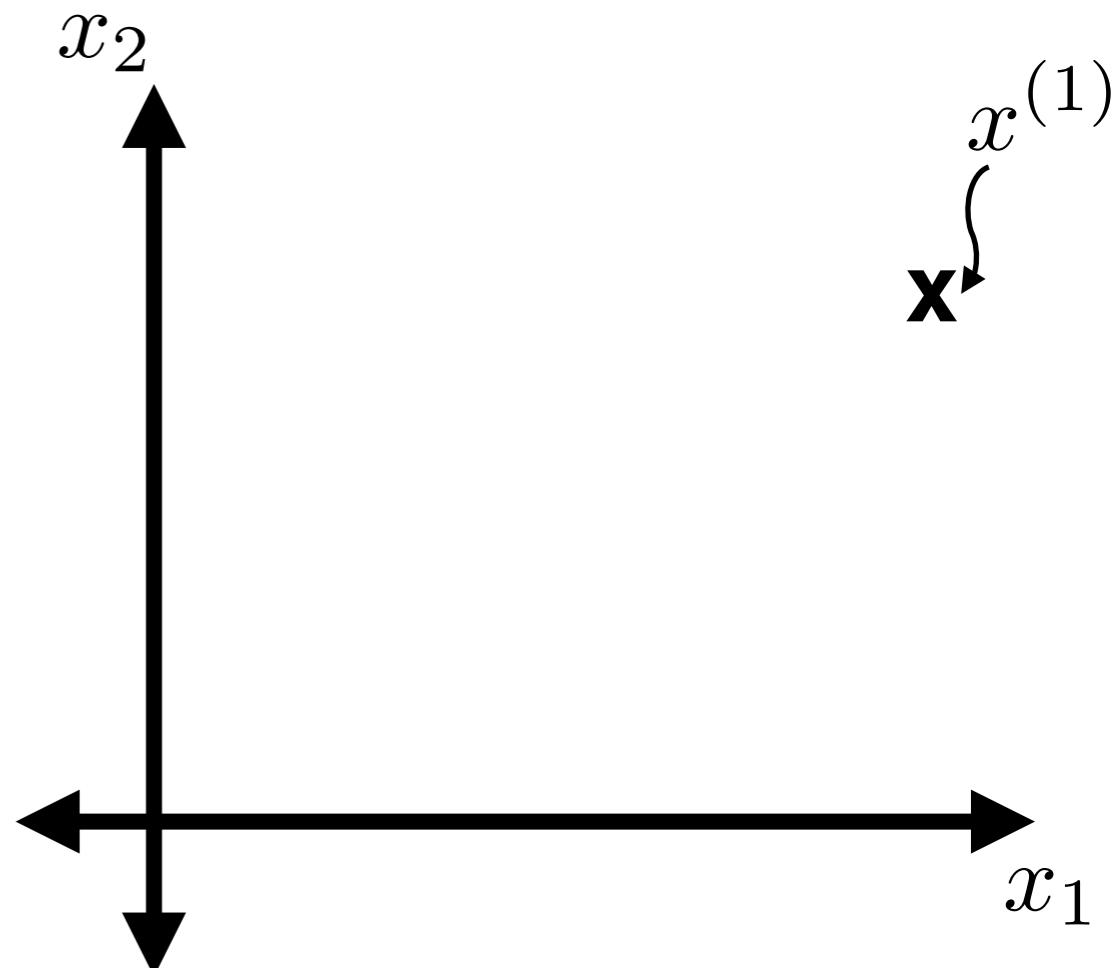


Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector

$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$

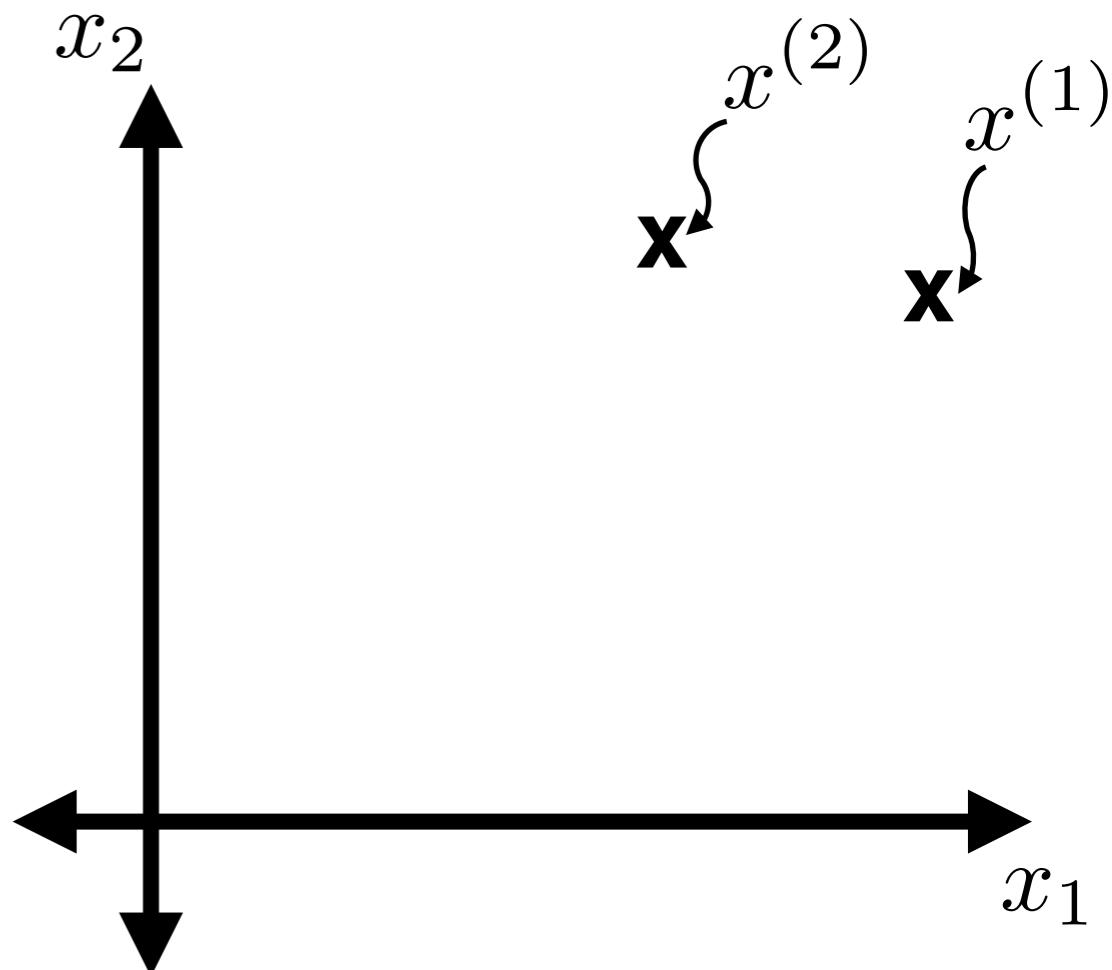


Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector

$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$

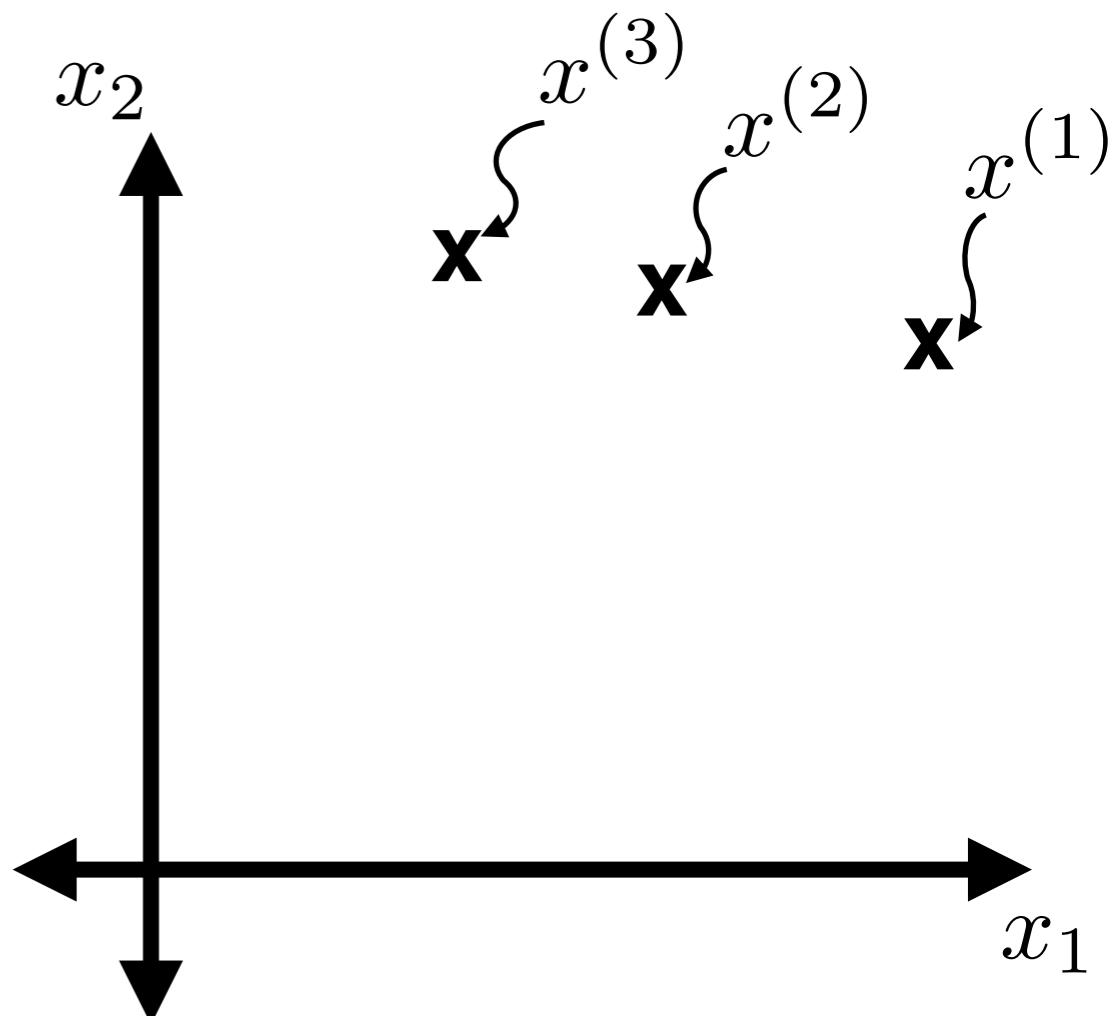


Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector

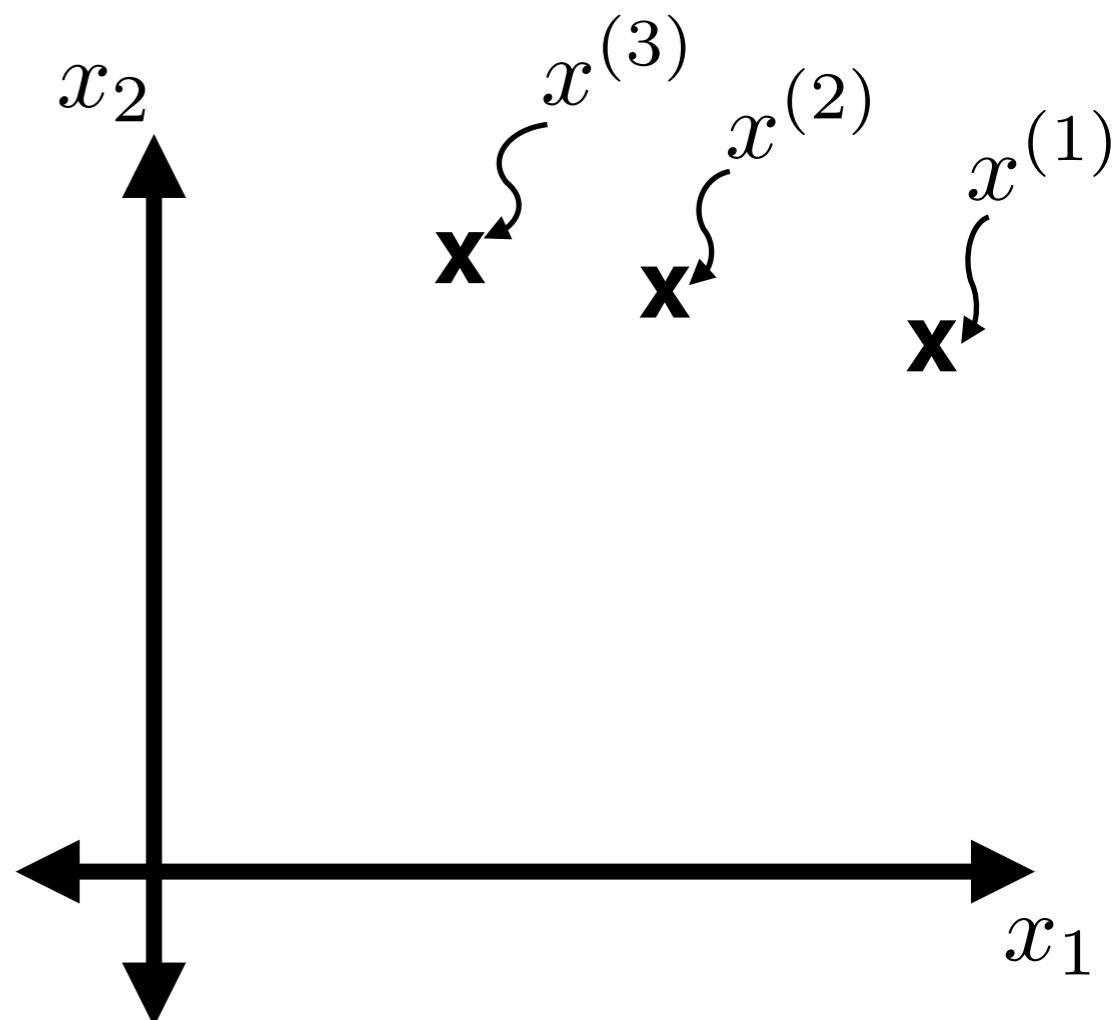
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$



Getting started

What do we have? (Training) data

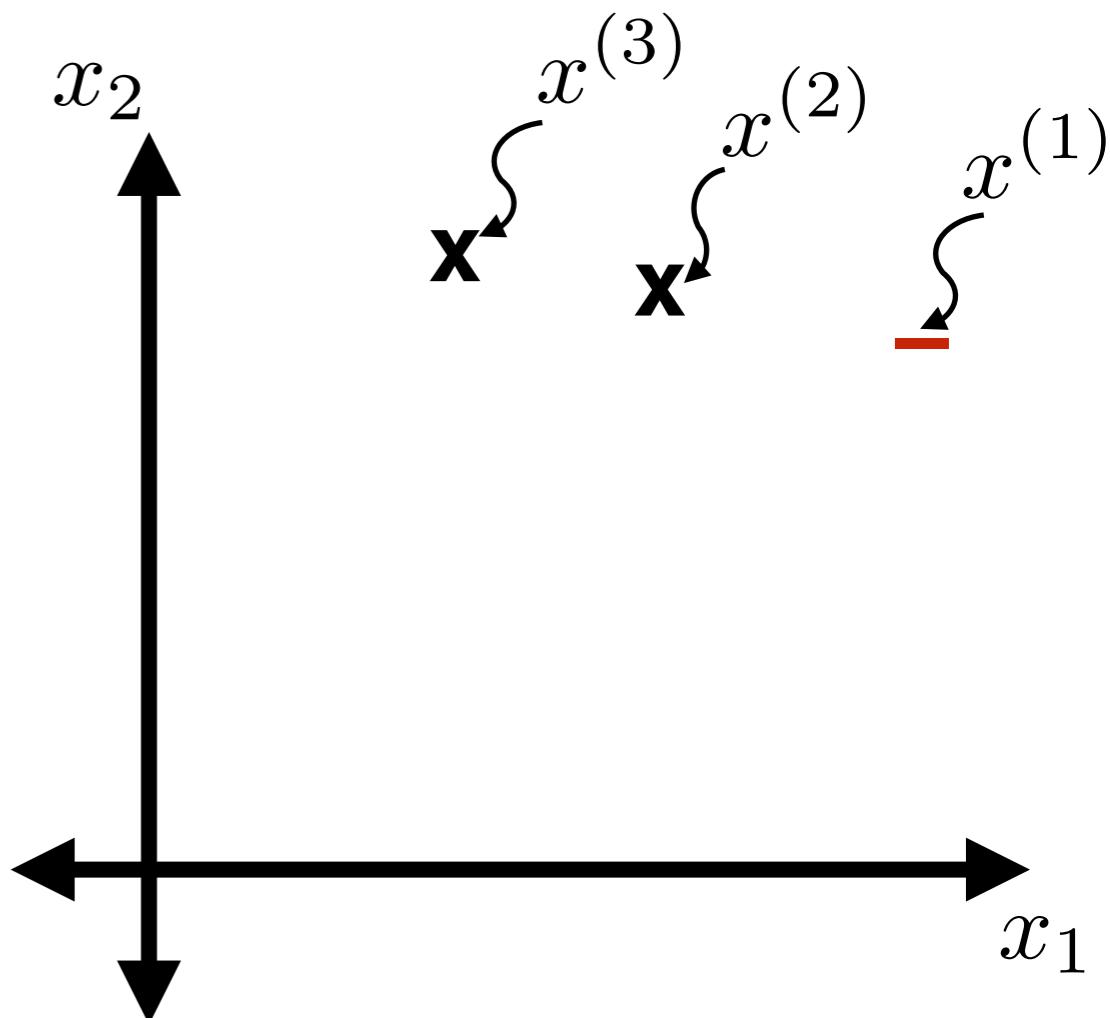
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$



Getting started

What do we have? (Training) data

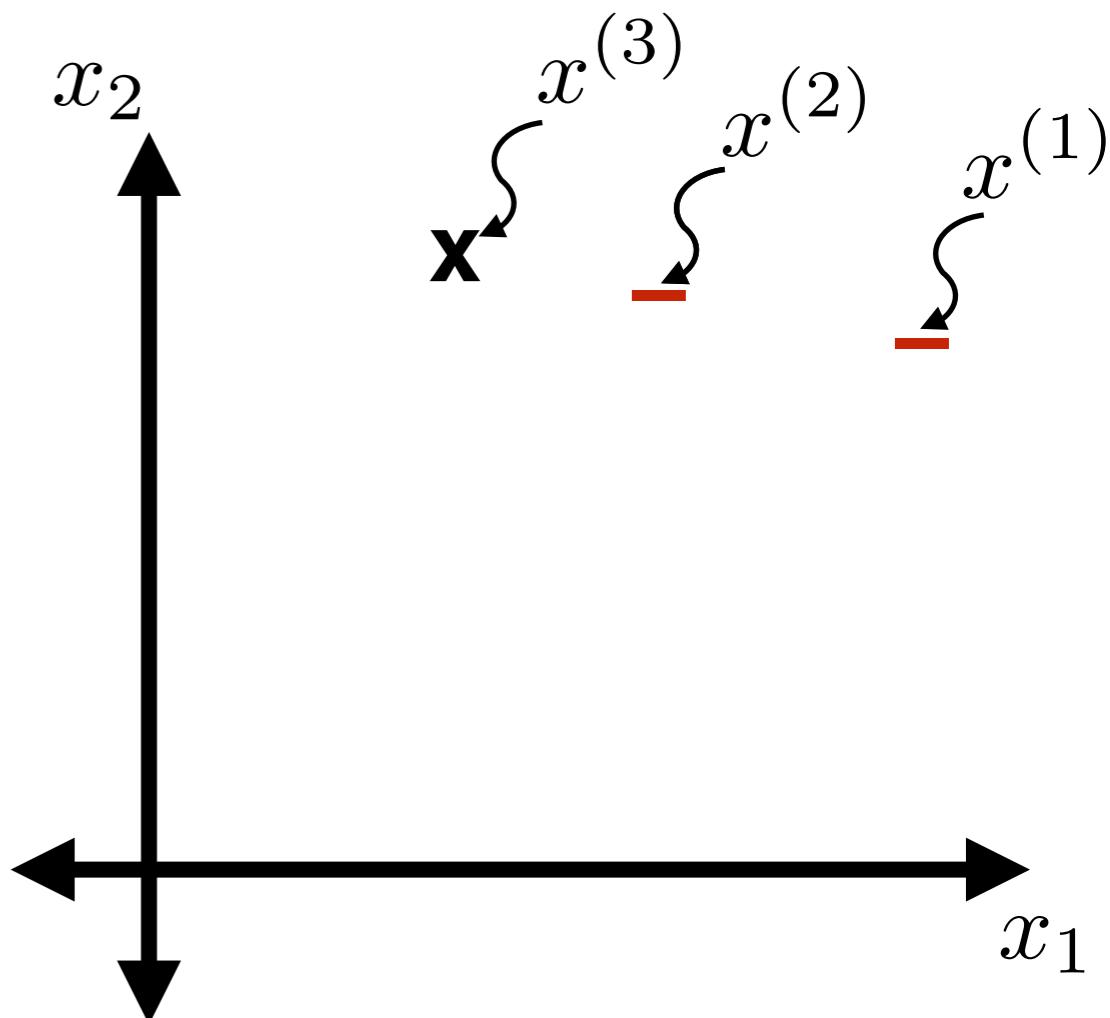
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$



Getting started

What do we have? (Training) data

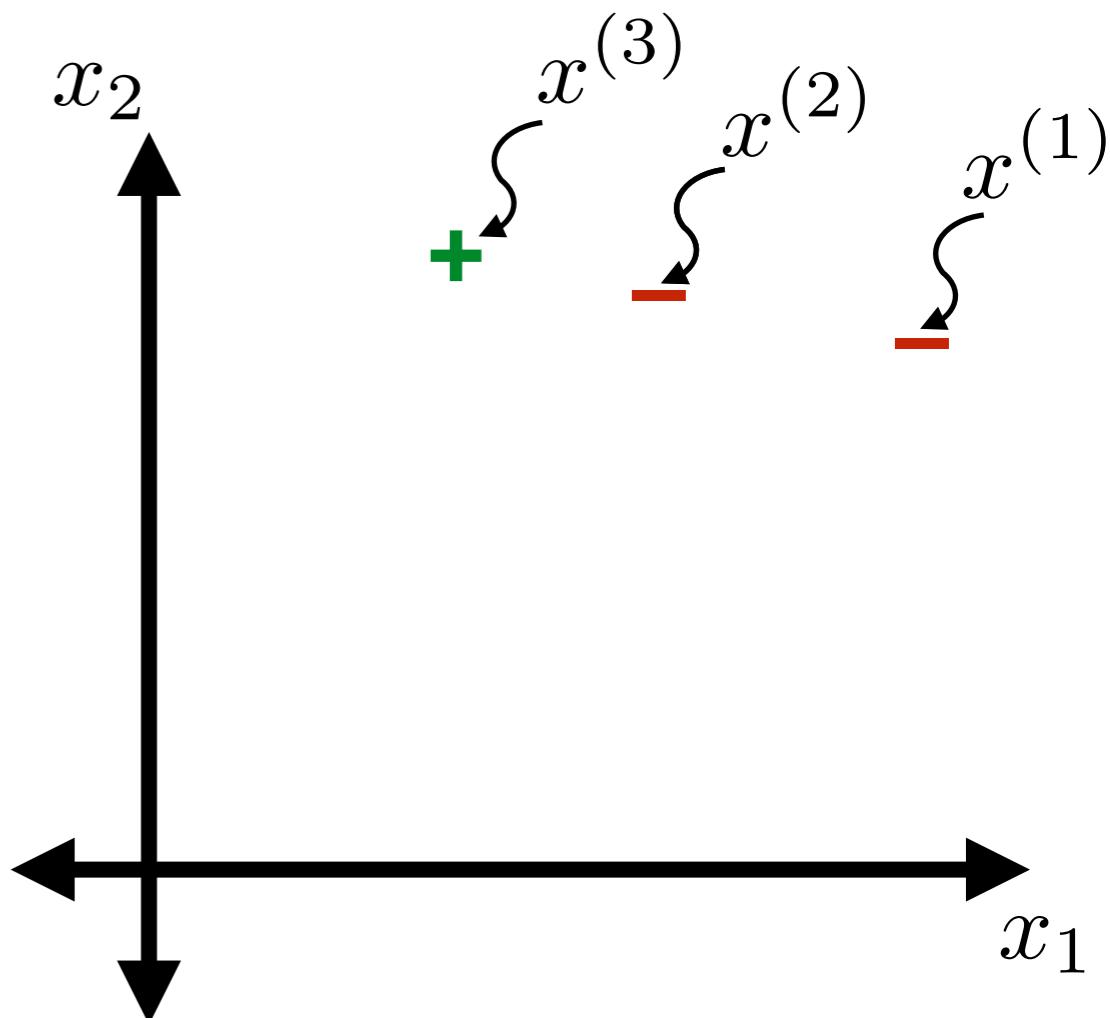
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$



Getting started

What do we have? (Training) data

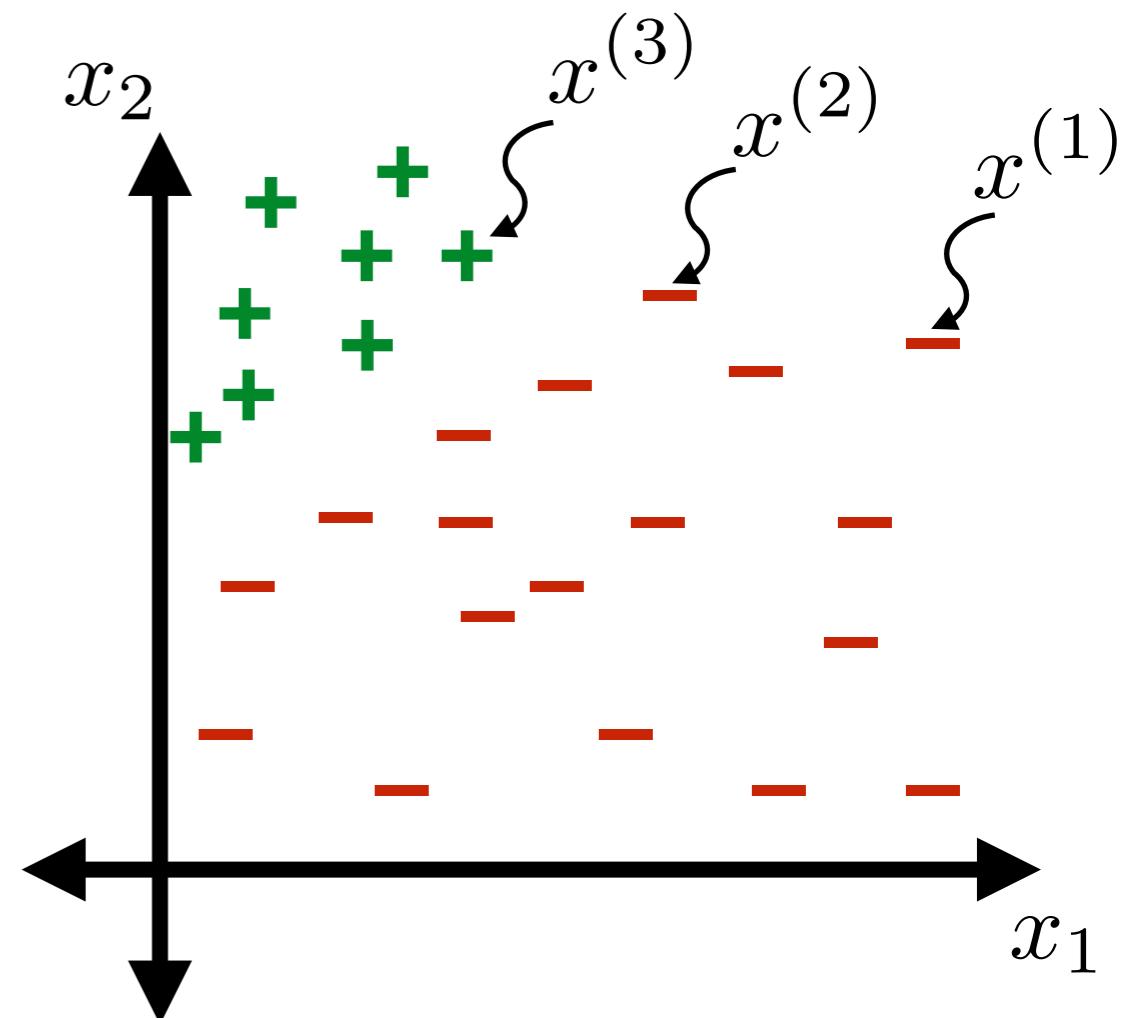
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$



Getting started

What do we have? (Training) data

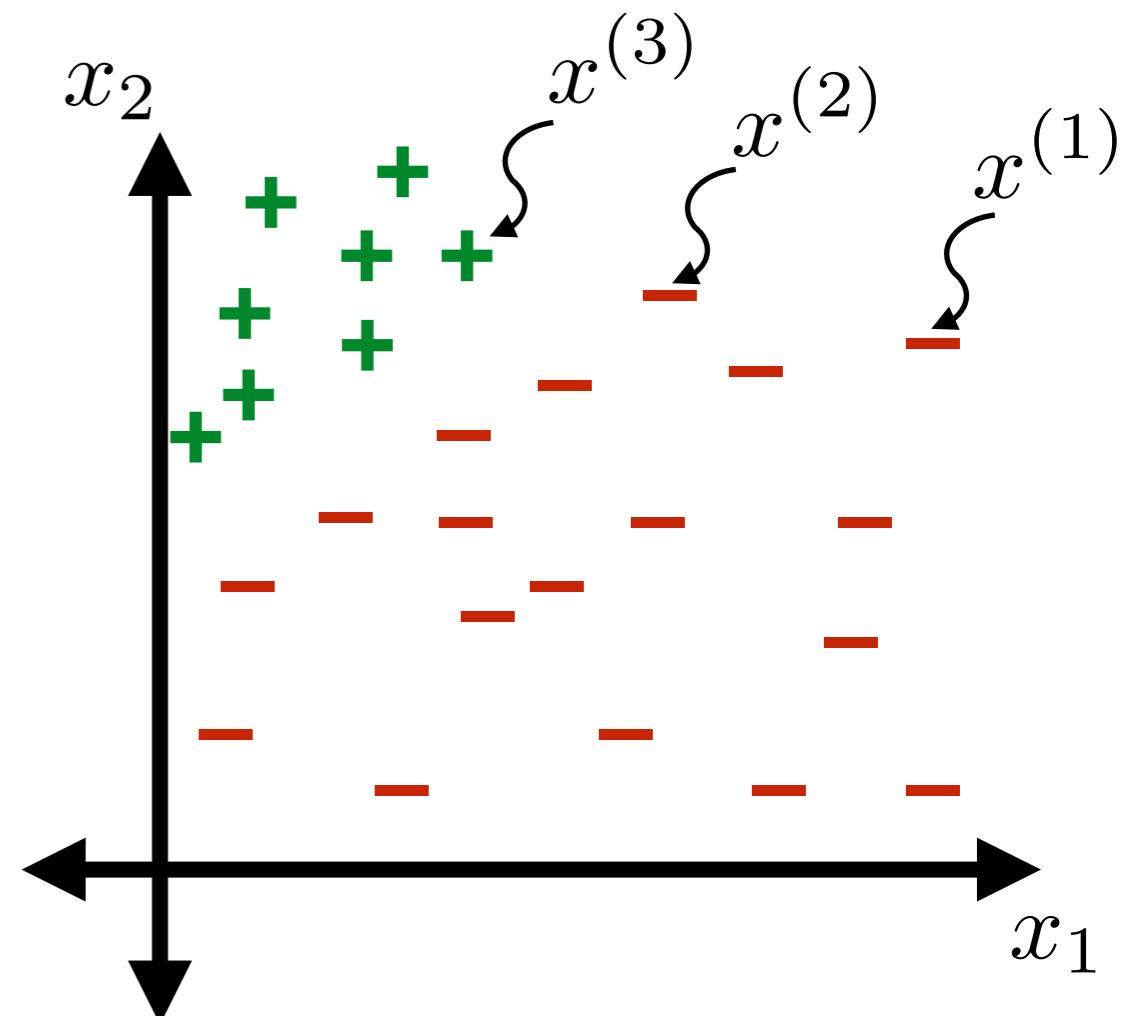
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$



Getting started

What do we have? (Training) data

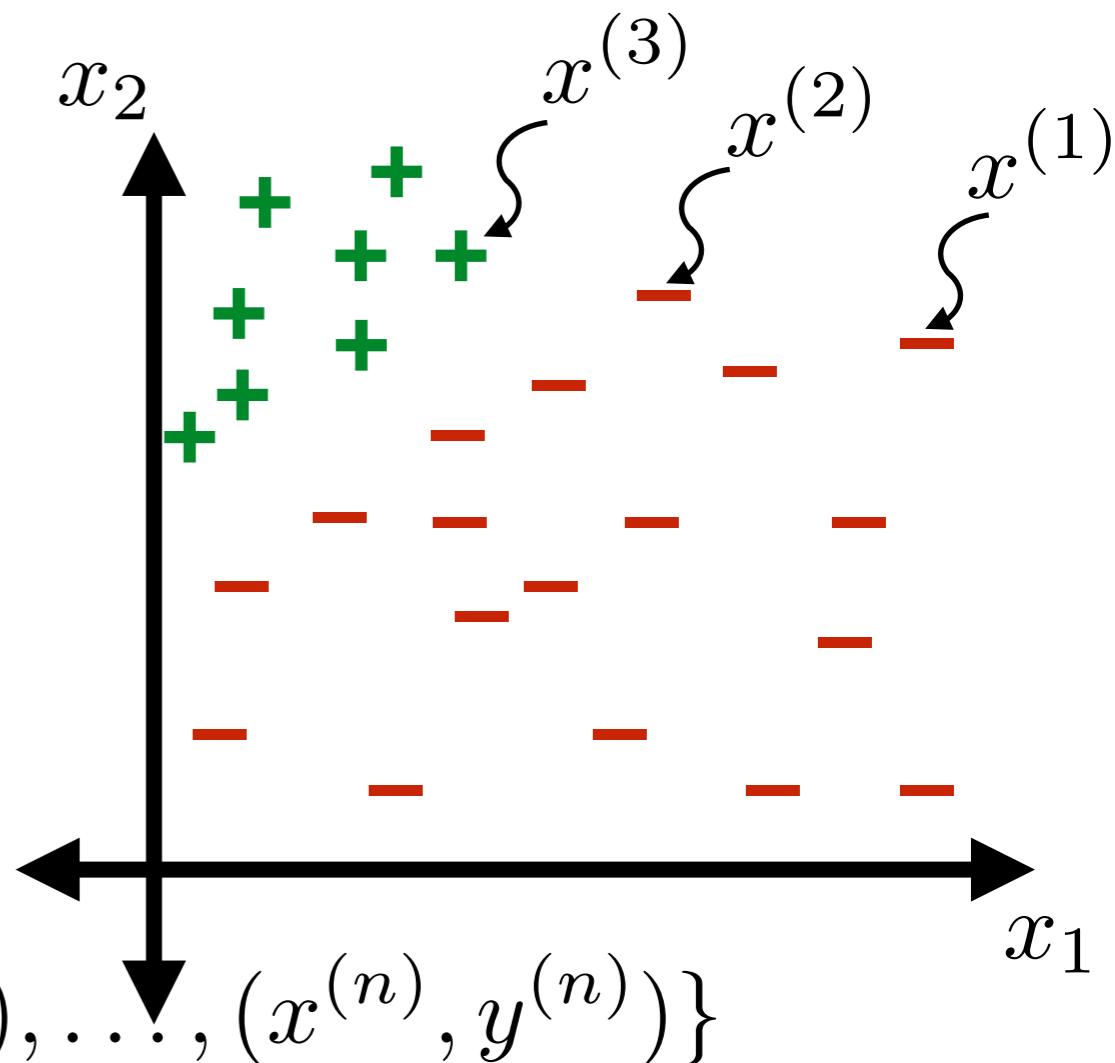
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data



Getting started

What do we have? (Training) data

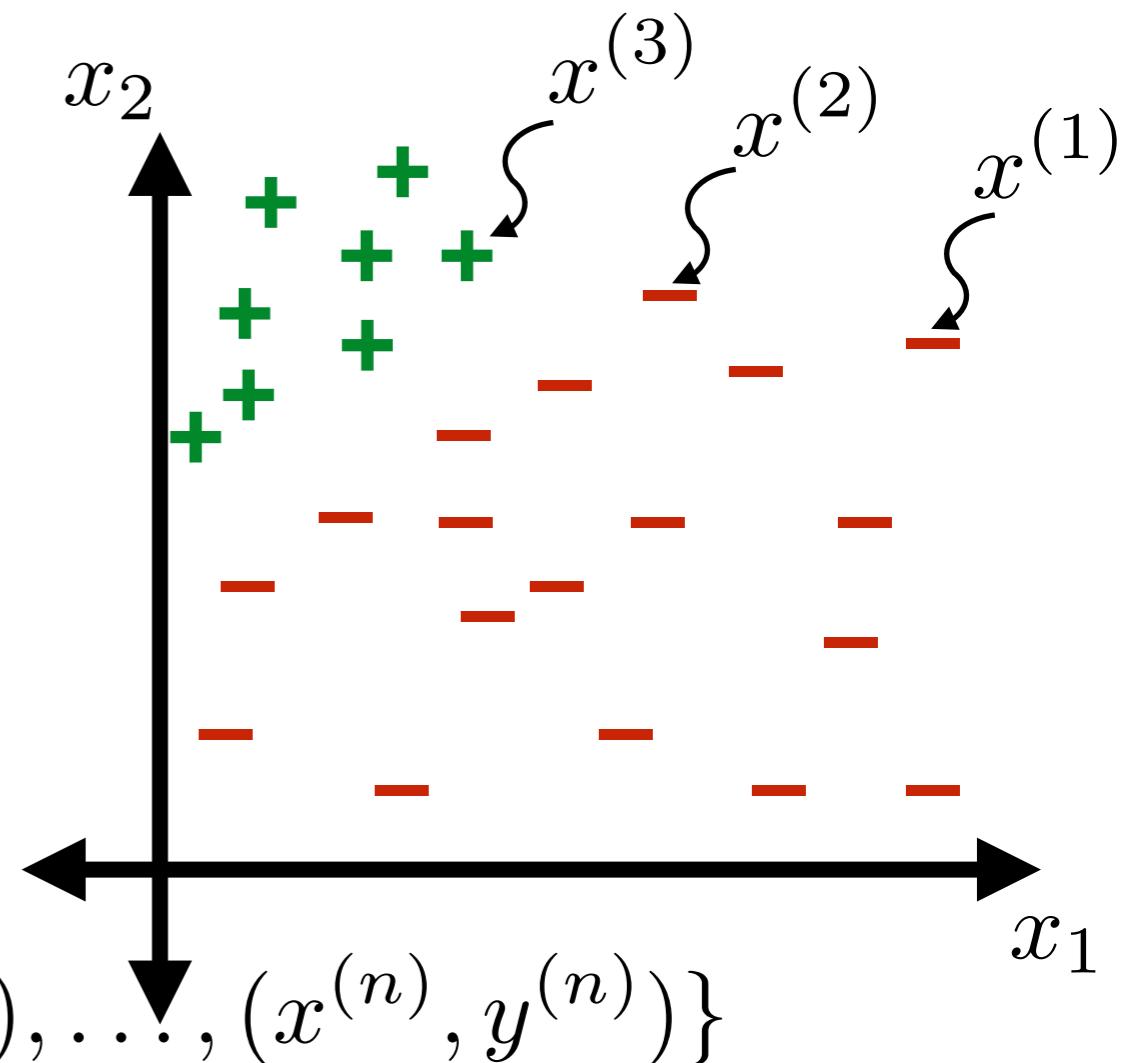
- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

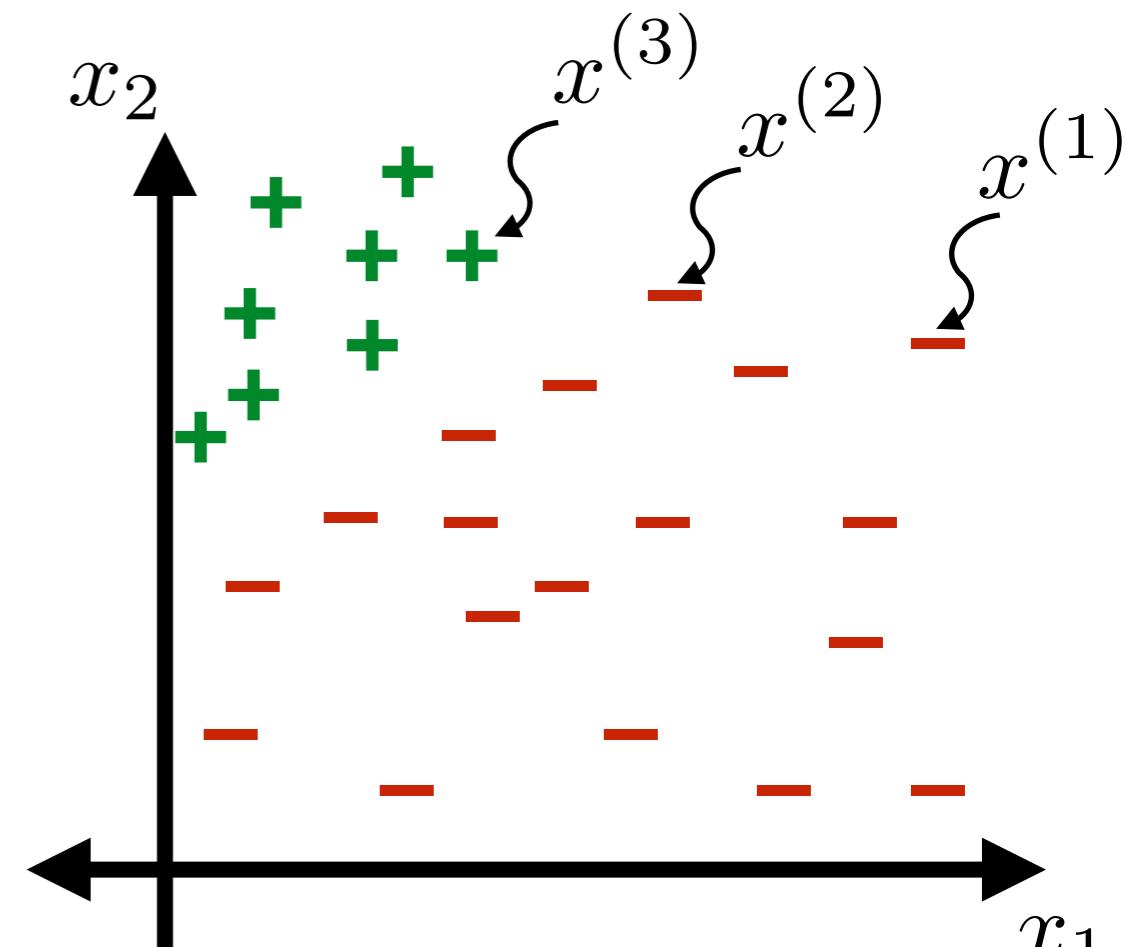


What do we want?

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

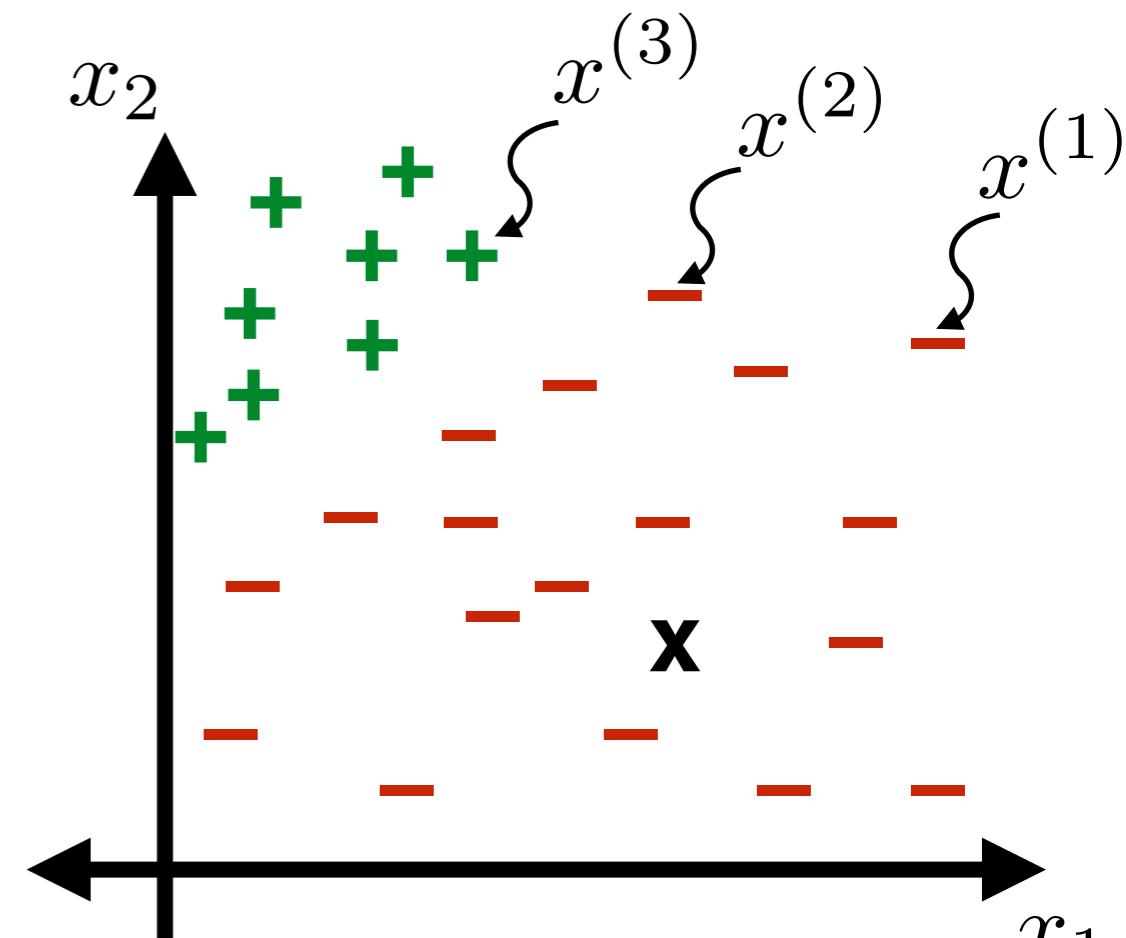


What do we want? A good way to label new points

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

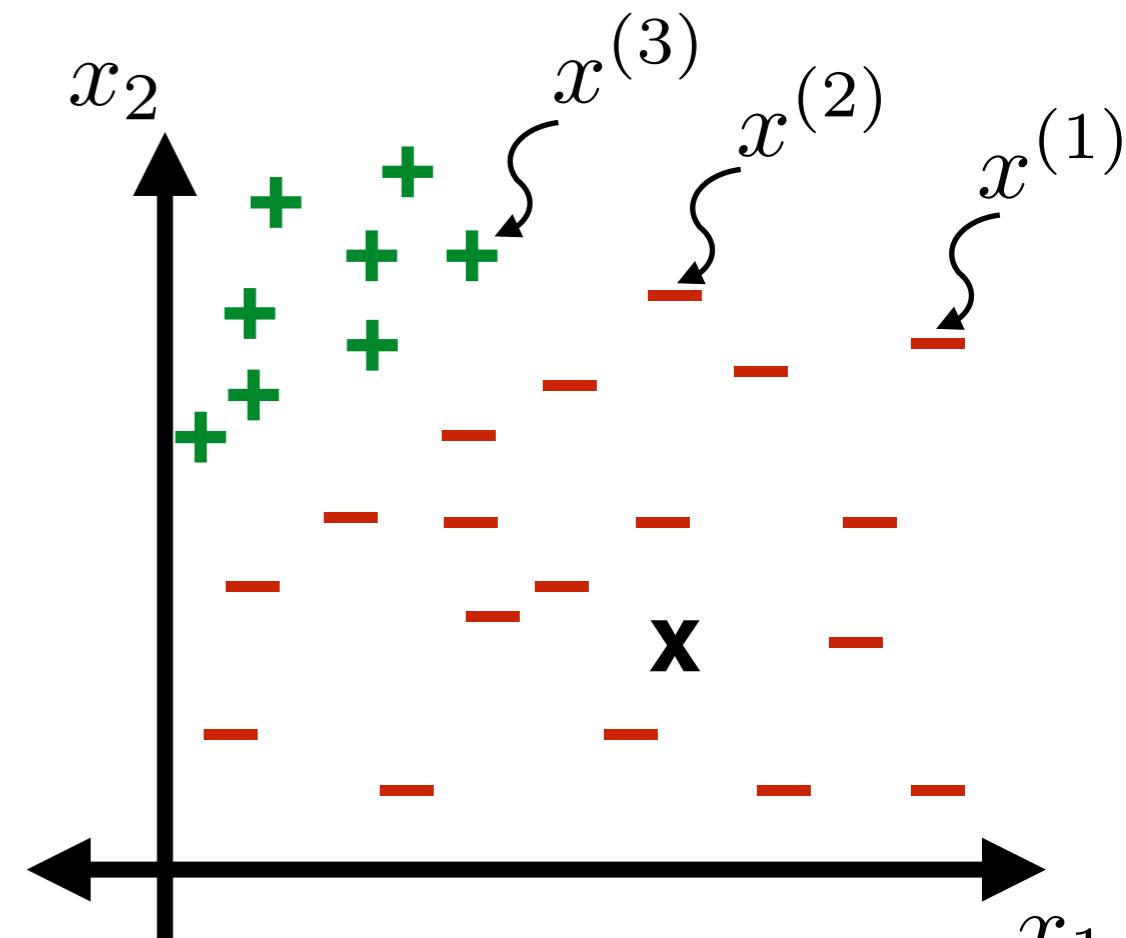


What do we want? A good way to label new points

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



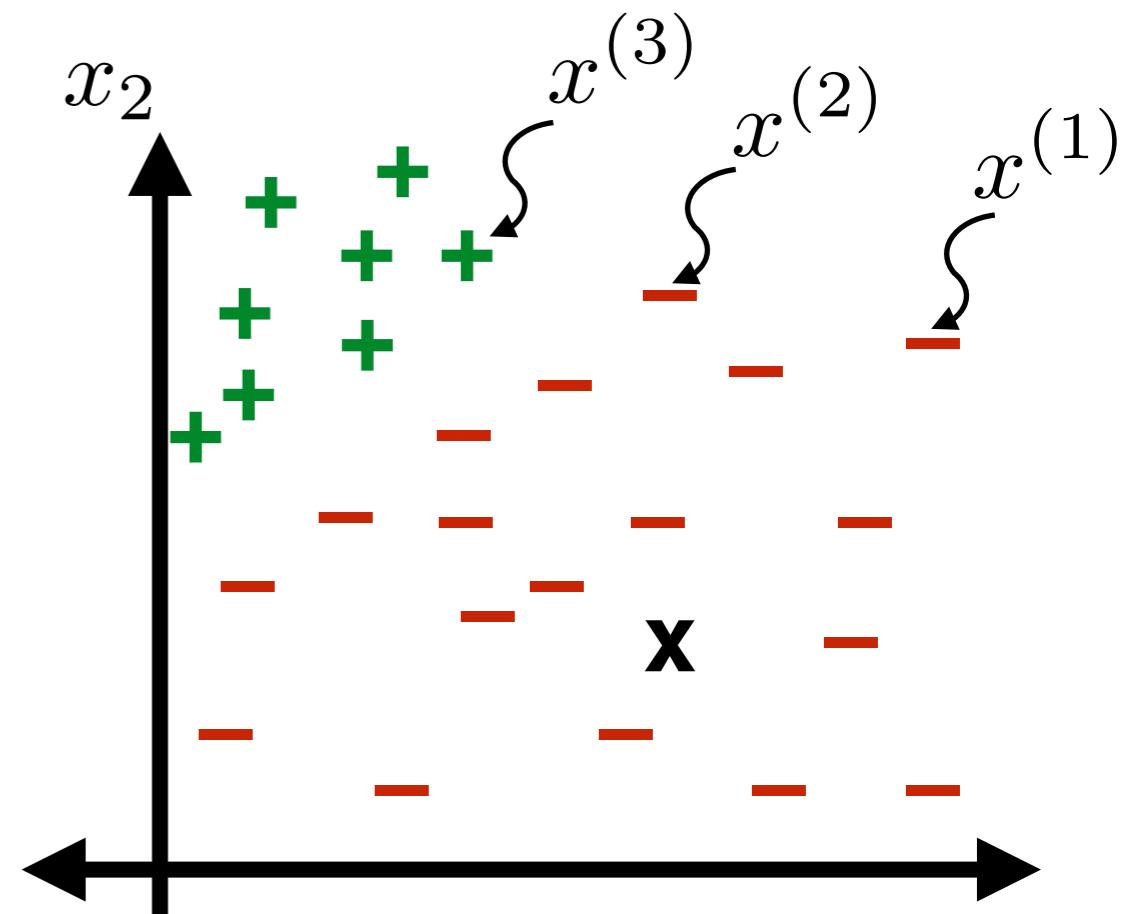
What do we want? A good way to label new points

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

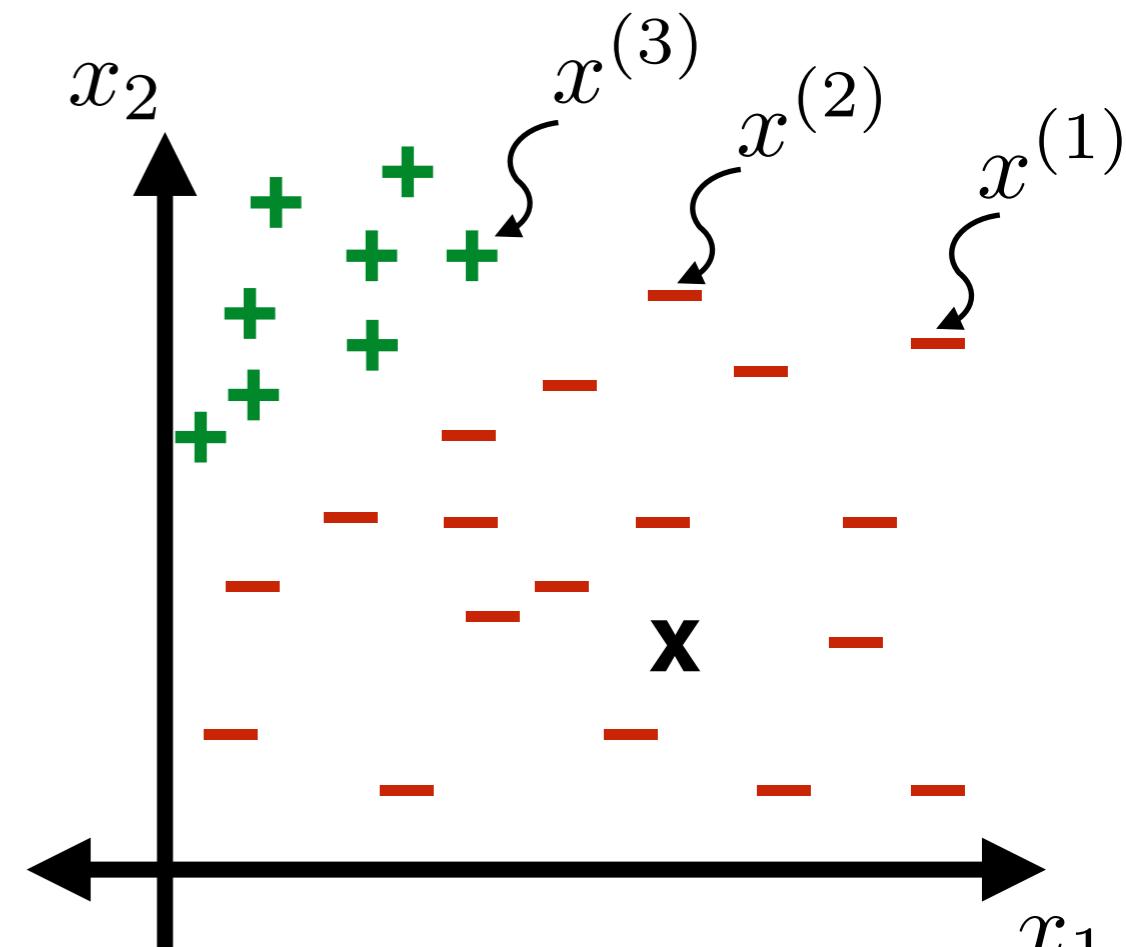
What do we want? A **good** way to label new points



Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

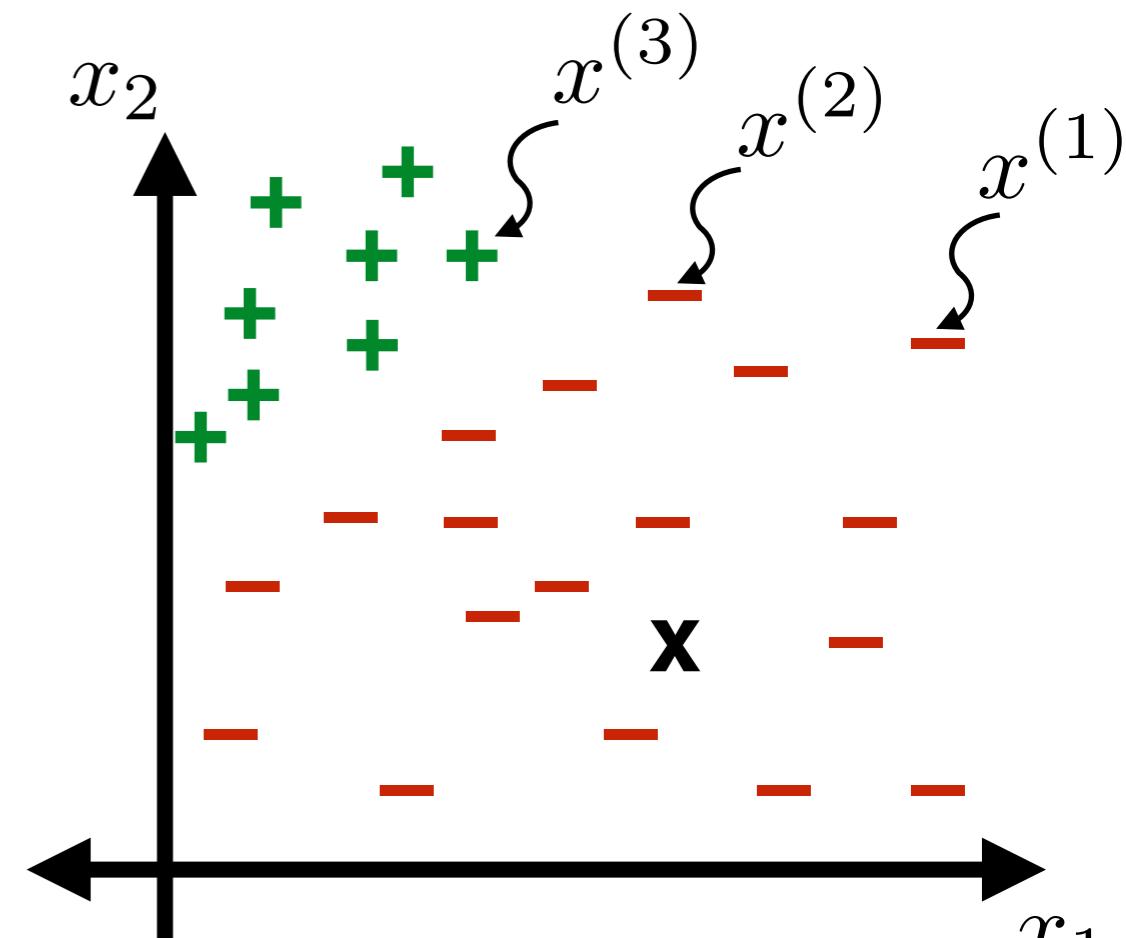


What do we want? A good way to label new points

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



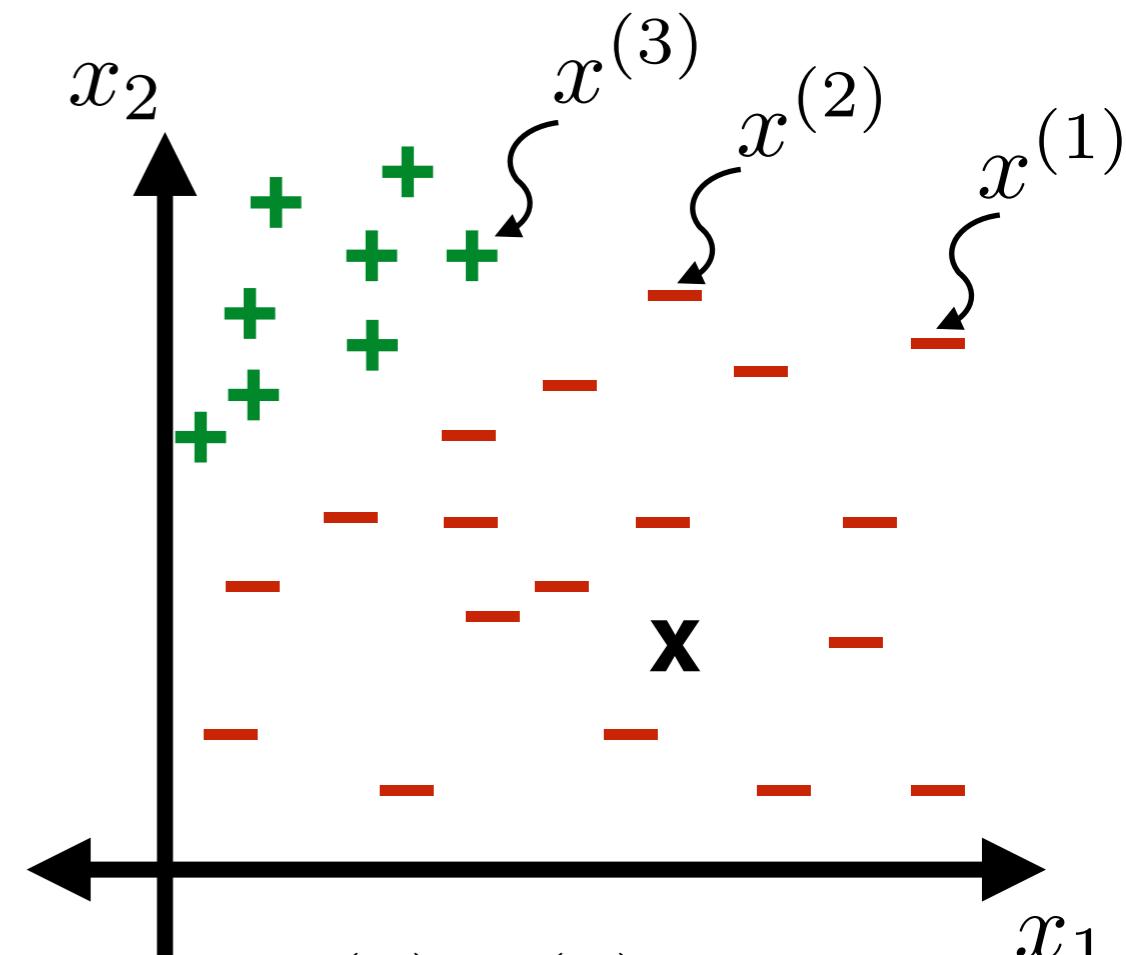
What do we want? A good way to label new points

- How to label?

Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



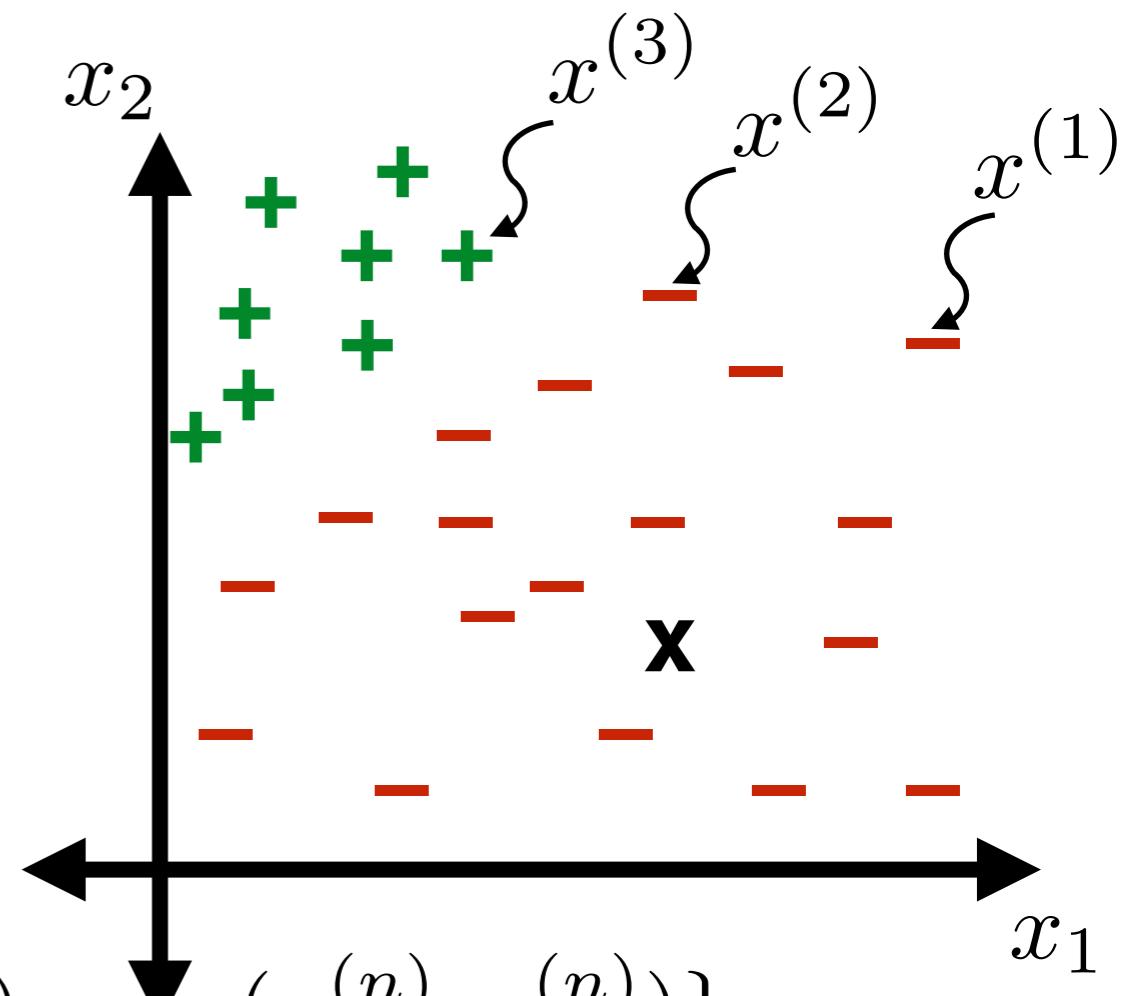
What do we want? A good way to label new points

- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$

Getting started

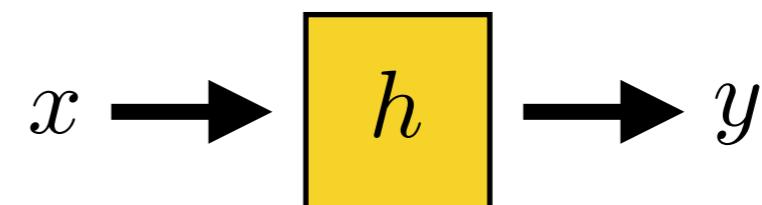
What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



What do we want? A good way to label new points

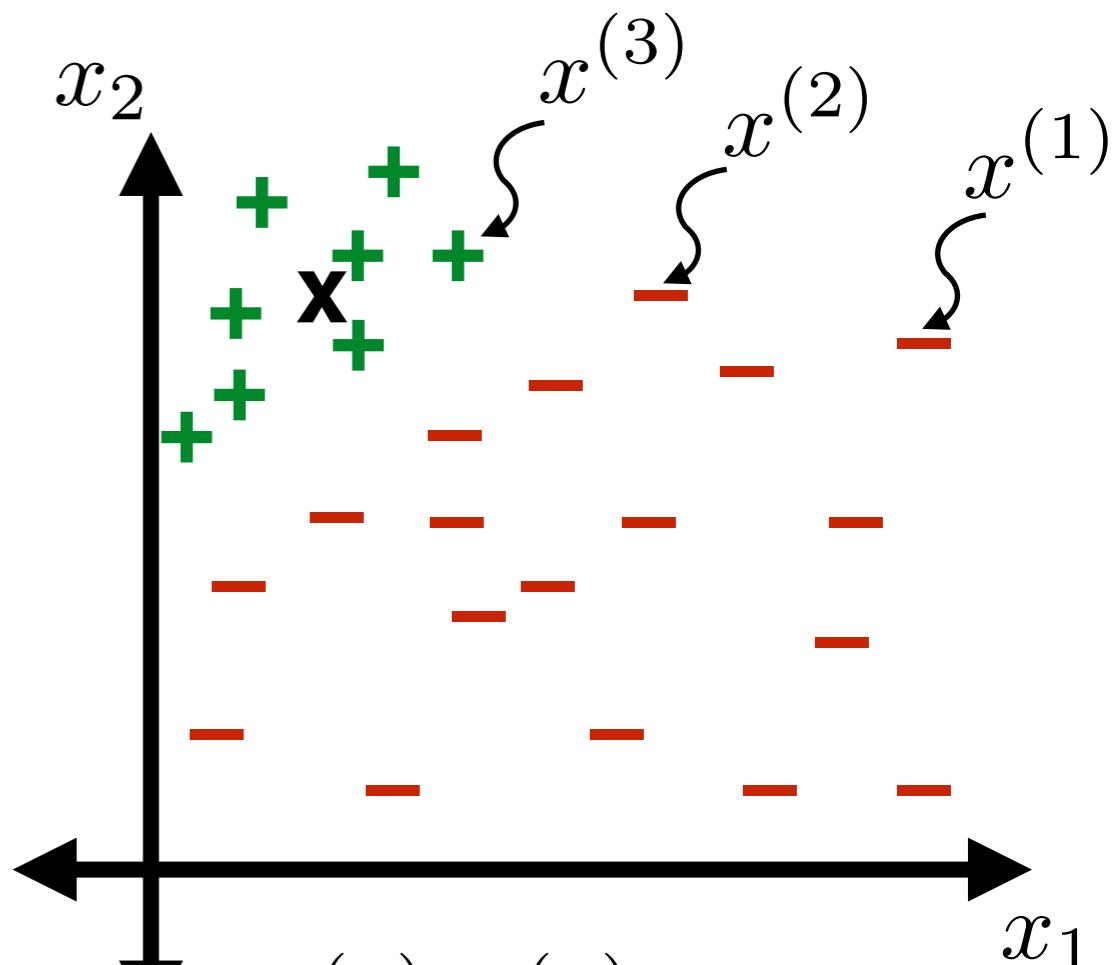
- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$



Getting started

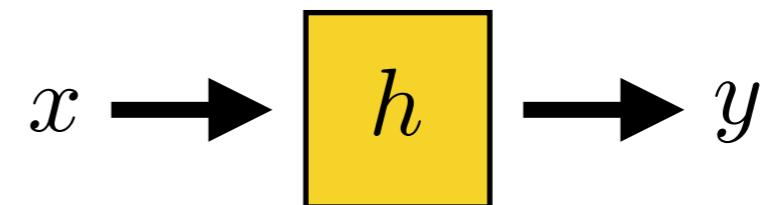
What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



What do we want? A good way to label new points

- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$



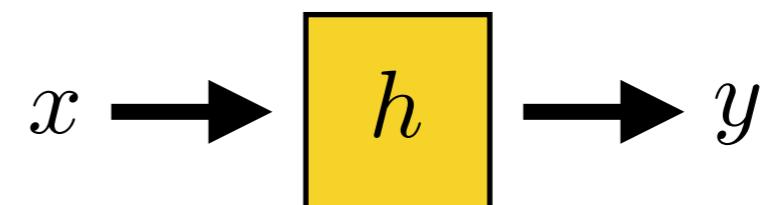
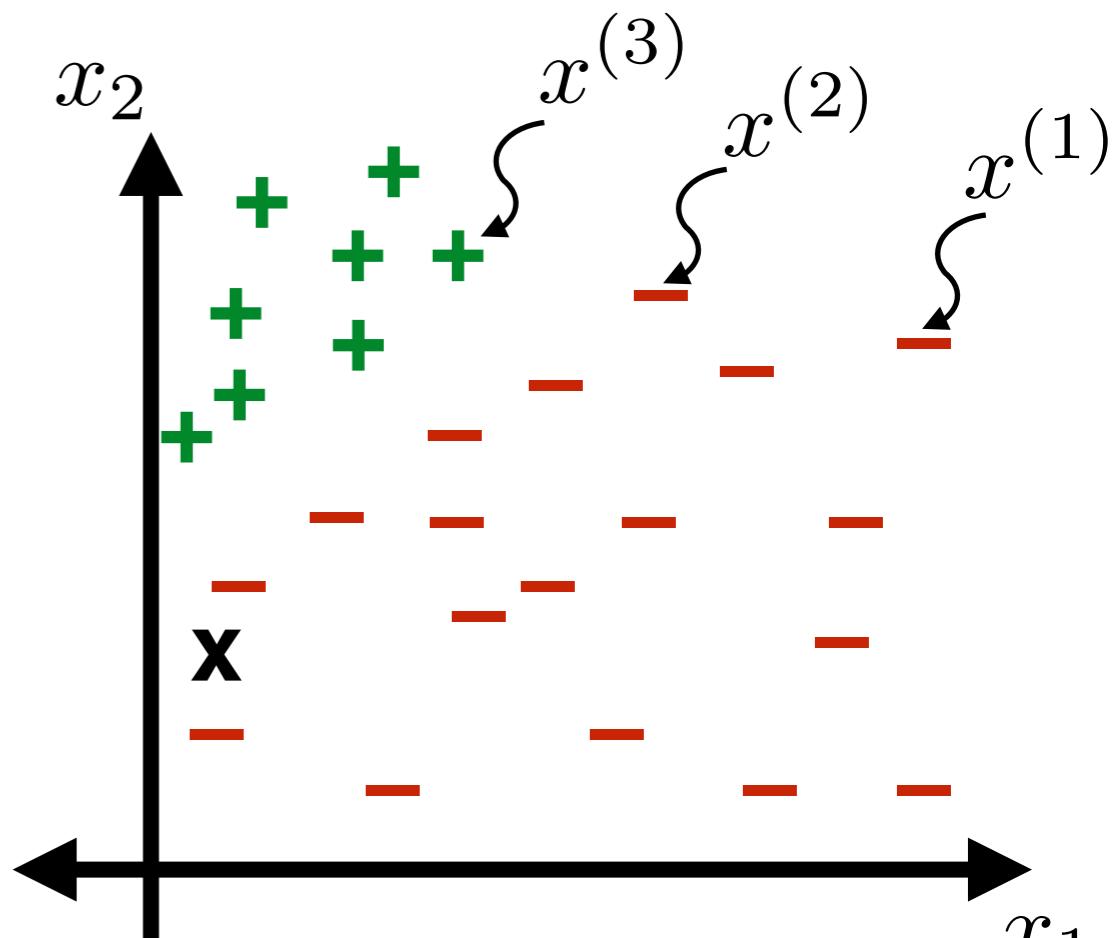
Getting started

What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

What do we want? A good way to label new points

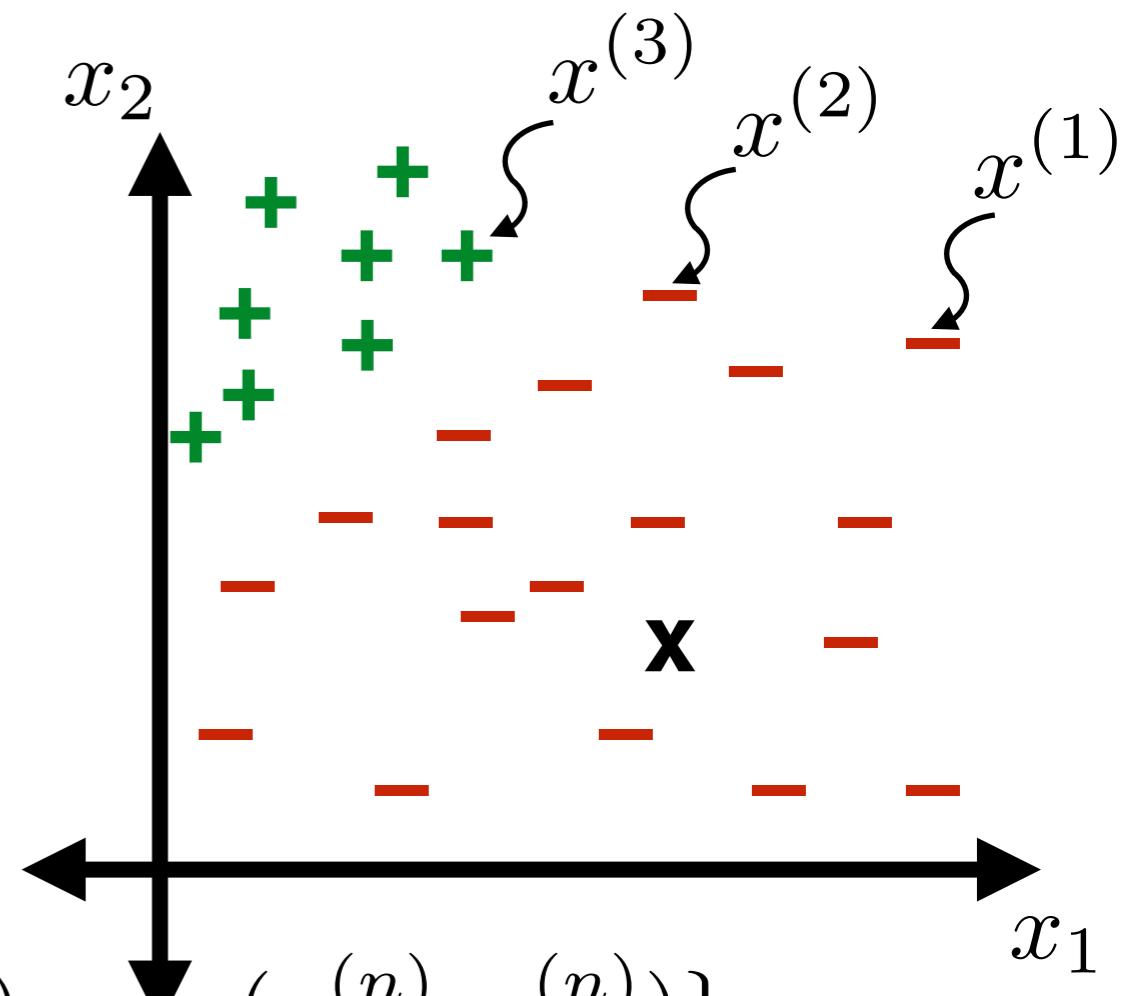
- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$



Getting started

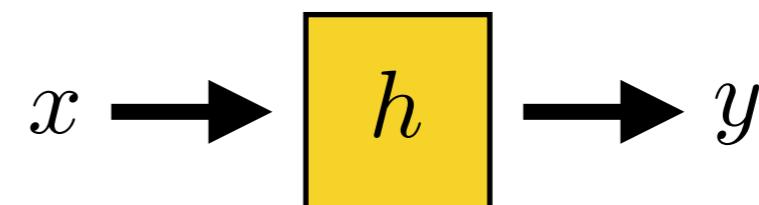
What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



What do we want? A good way to label new points

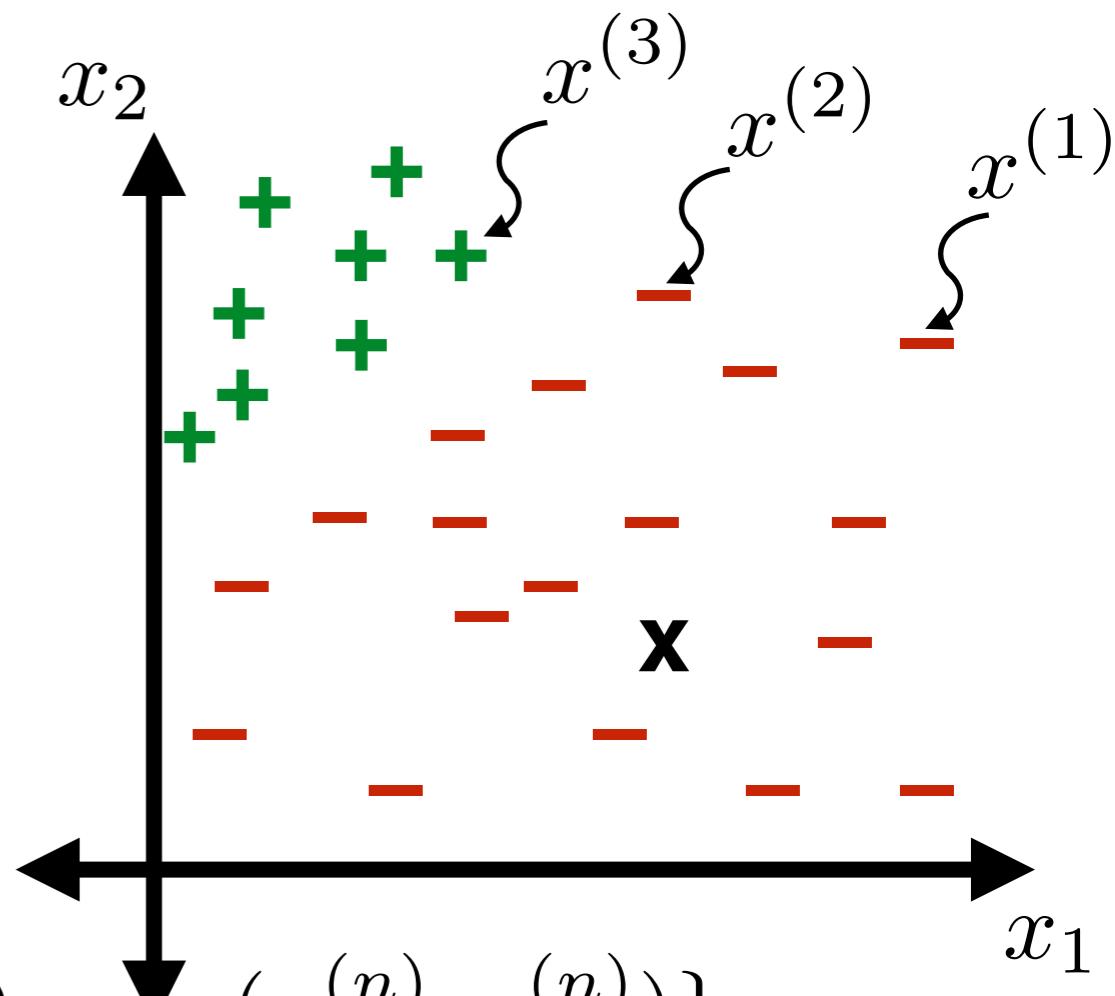
- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$



Getting started

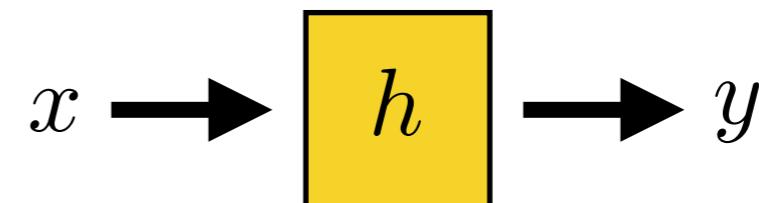
What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



What do we want? A good way to label new points

- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$

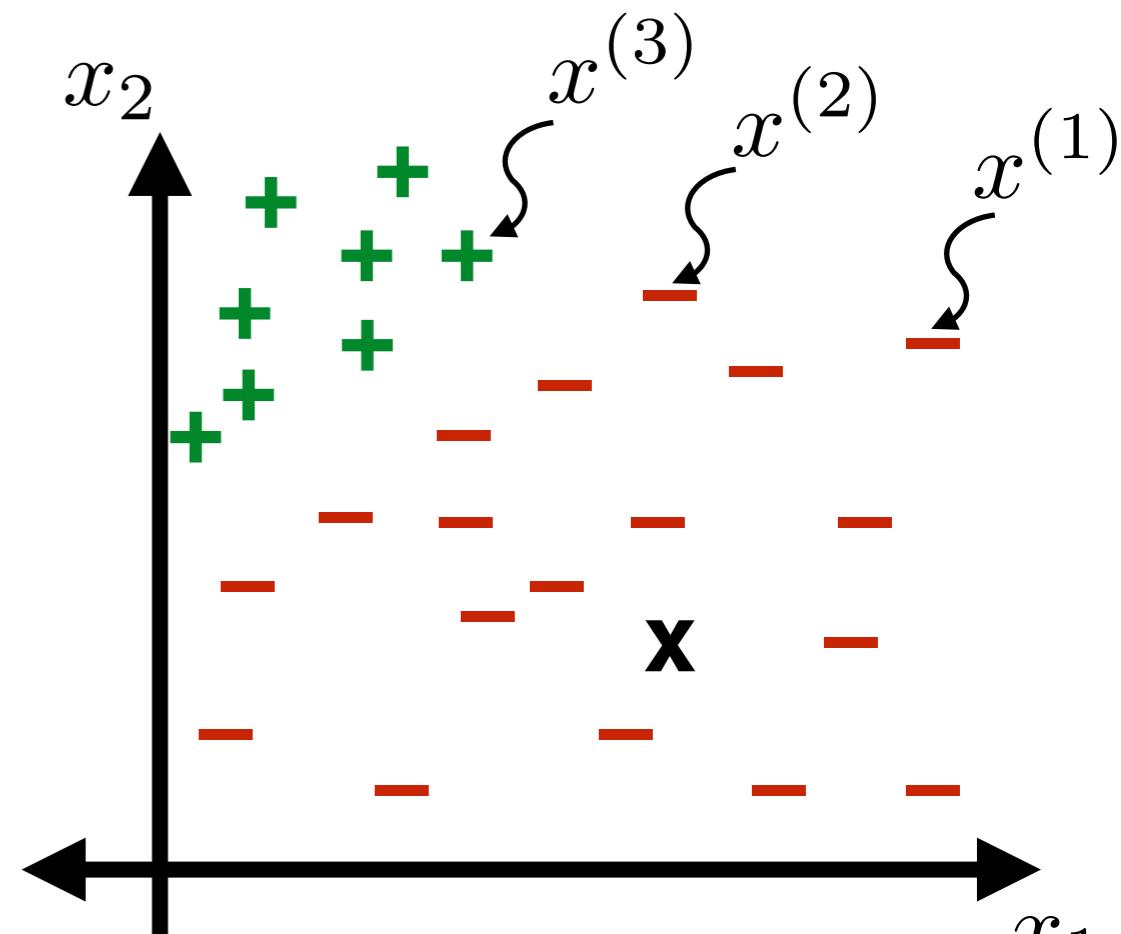


- Example h : For any x , $h(x) = +1$

Getting started

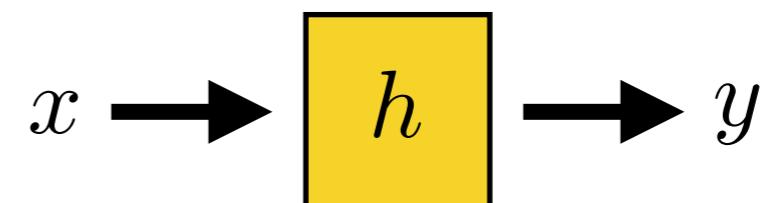
What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



What do we want? A good way to label new points

- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$

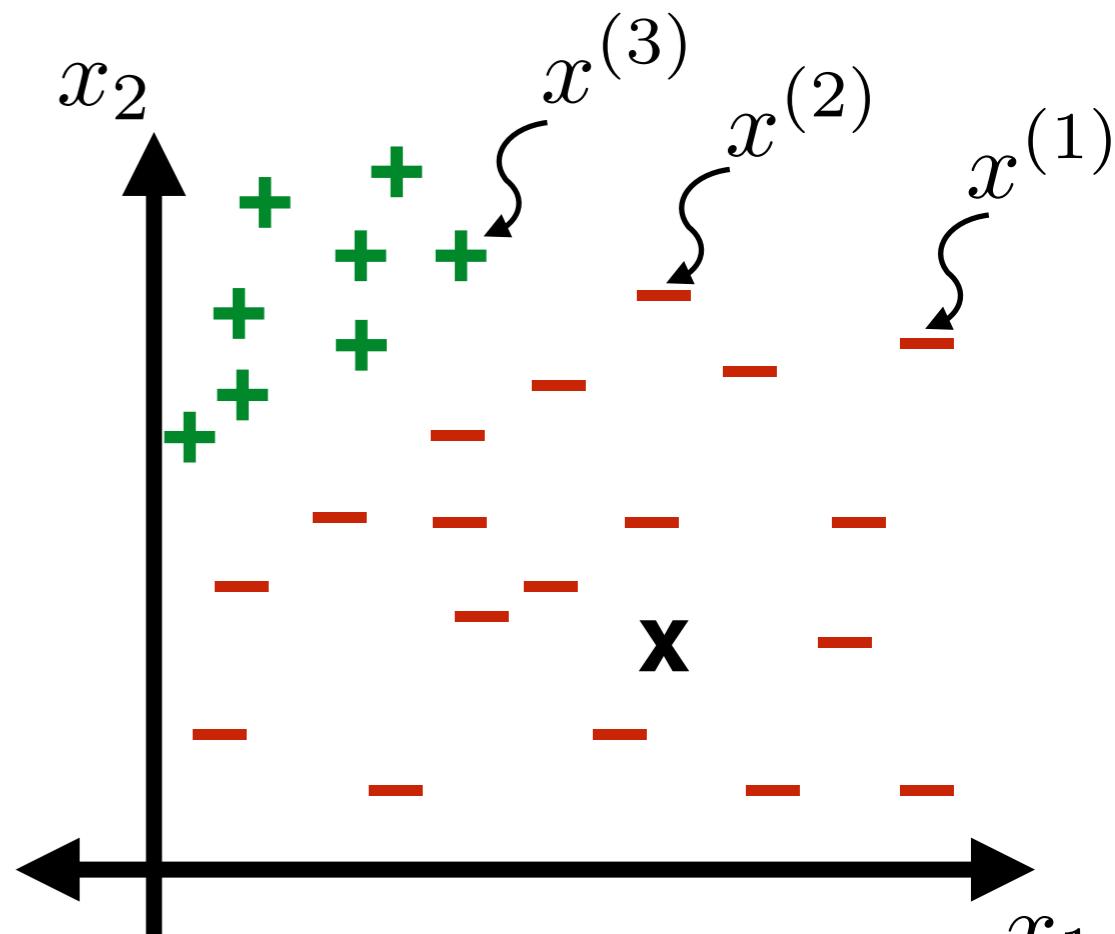


- Example h : For any x , $h(x) = +1$
- Is this a hypothesis?

Getting started

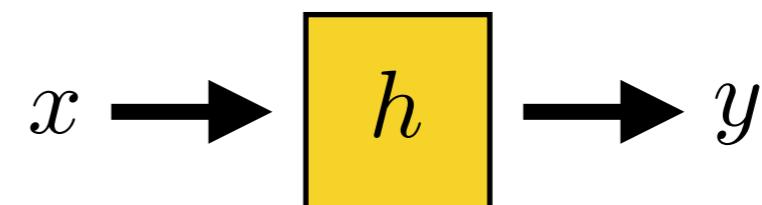
What do we have? (Training) data

- n training data points
- For data point $i \in \{1, \dots, n\}$
 - Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})^\top \in \mathbb{R}^d$$
 - Label $y^{(i)} \in \{-1, +1\}$
- Training data $\mathcal{D}_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$



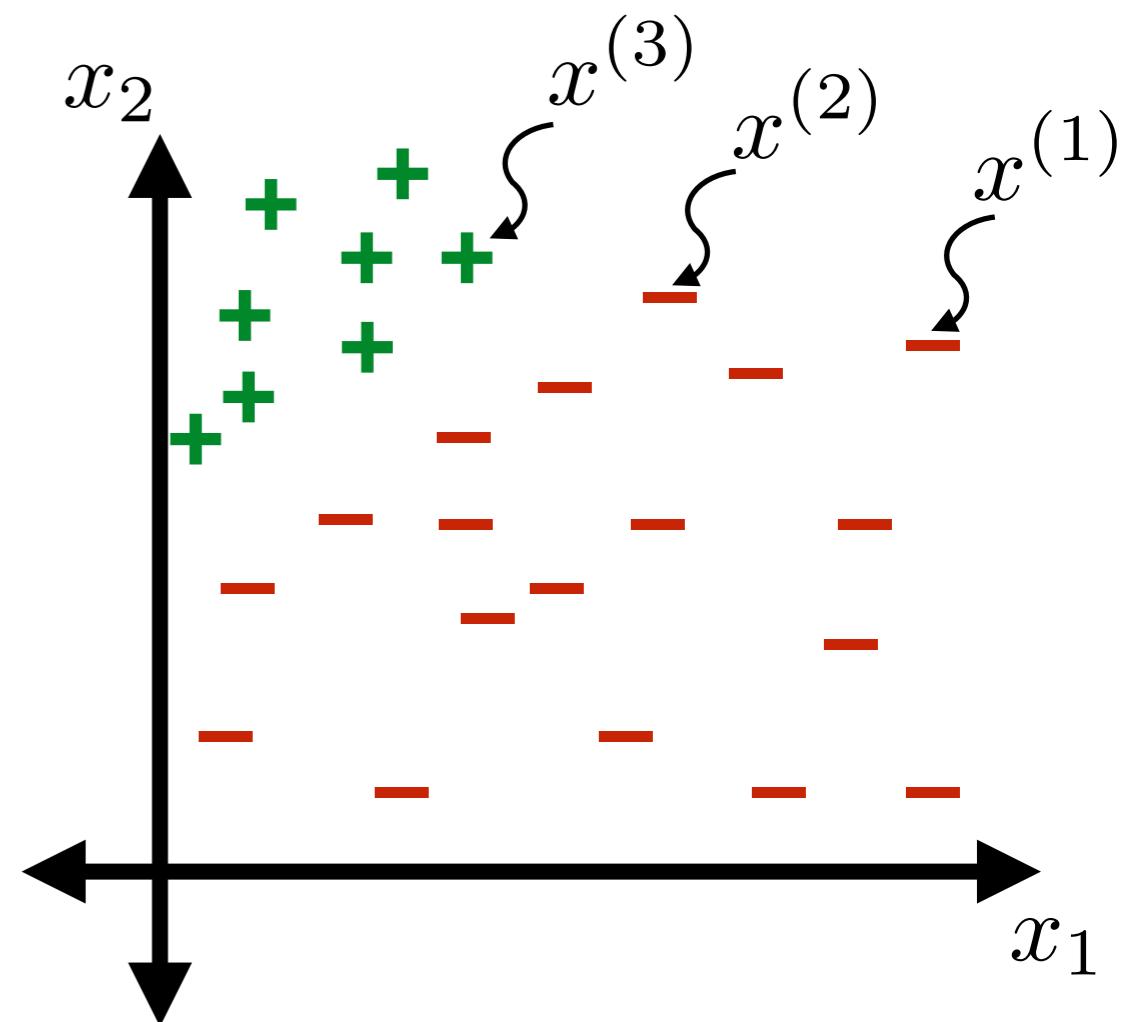
What do we want? A good way to label new points

- How to label? Hypothesis $h : \mathbb{R}^d \rightarrow \{-1, +1\}$



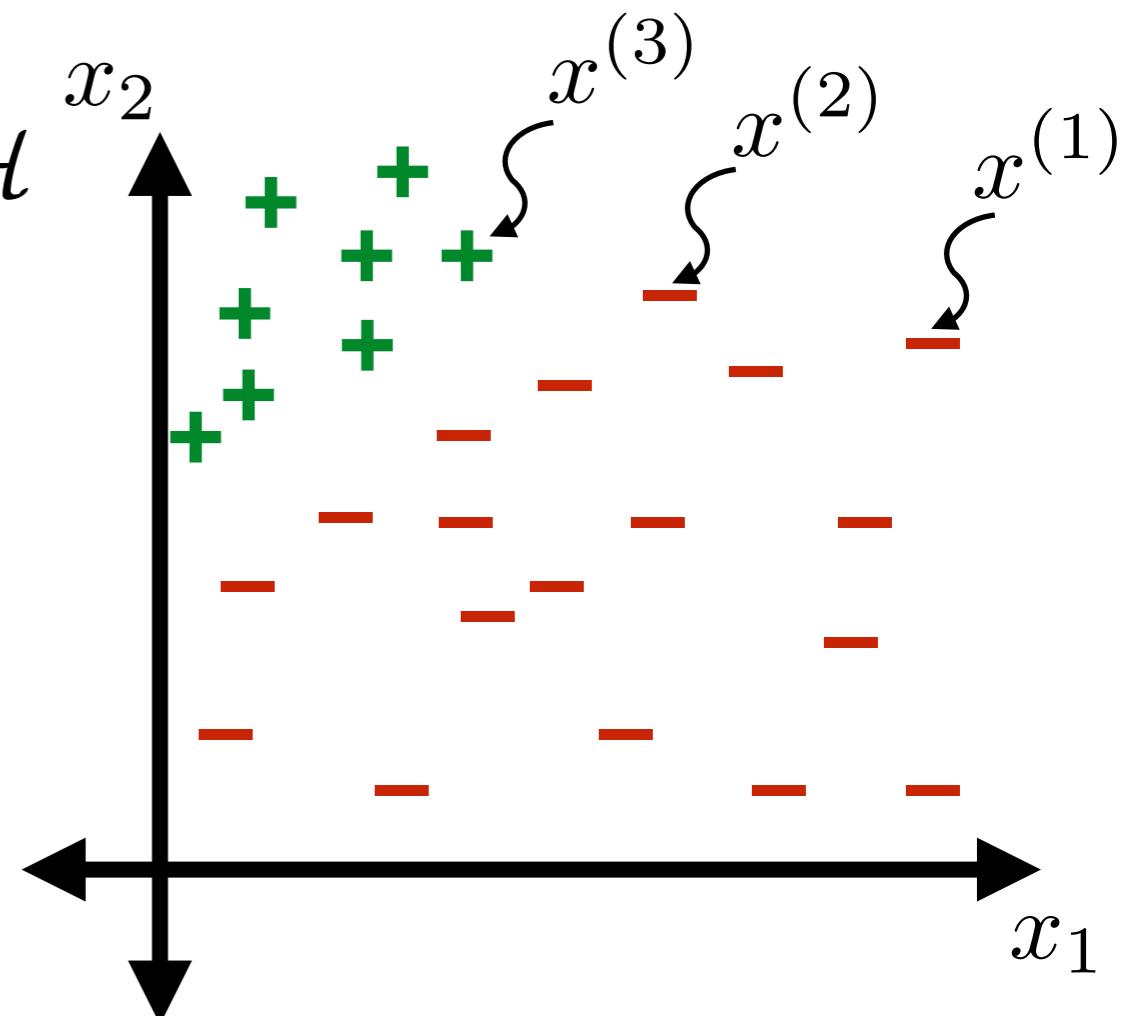
- Example h : For any x , $h(x) = +1$
- Is this a good hypothesis?

Linear classifiers



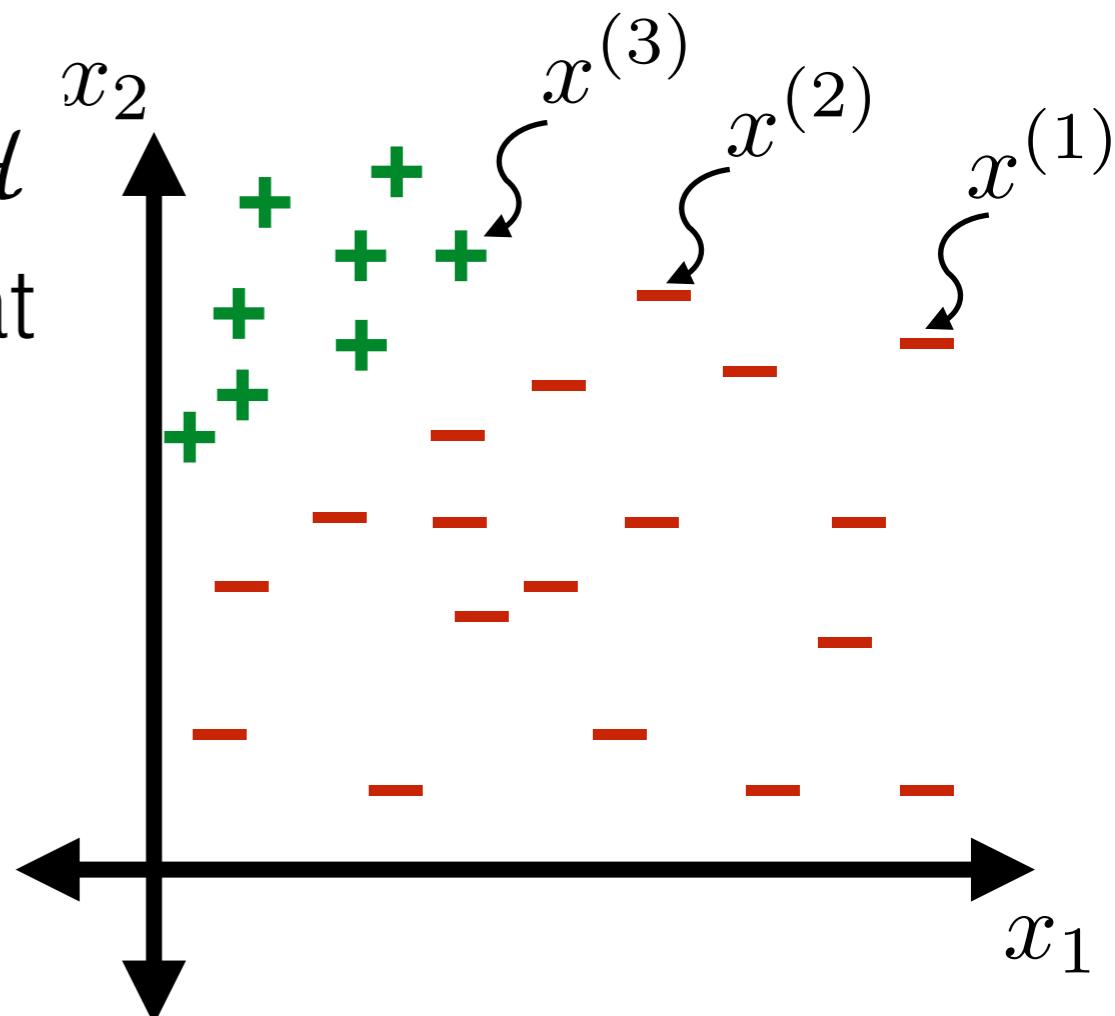
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$



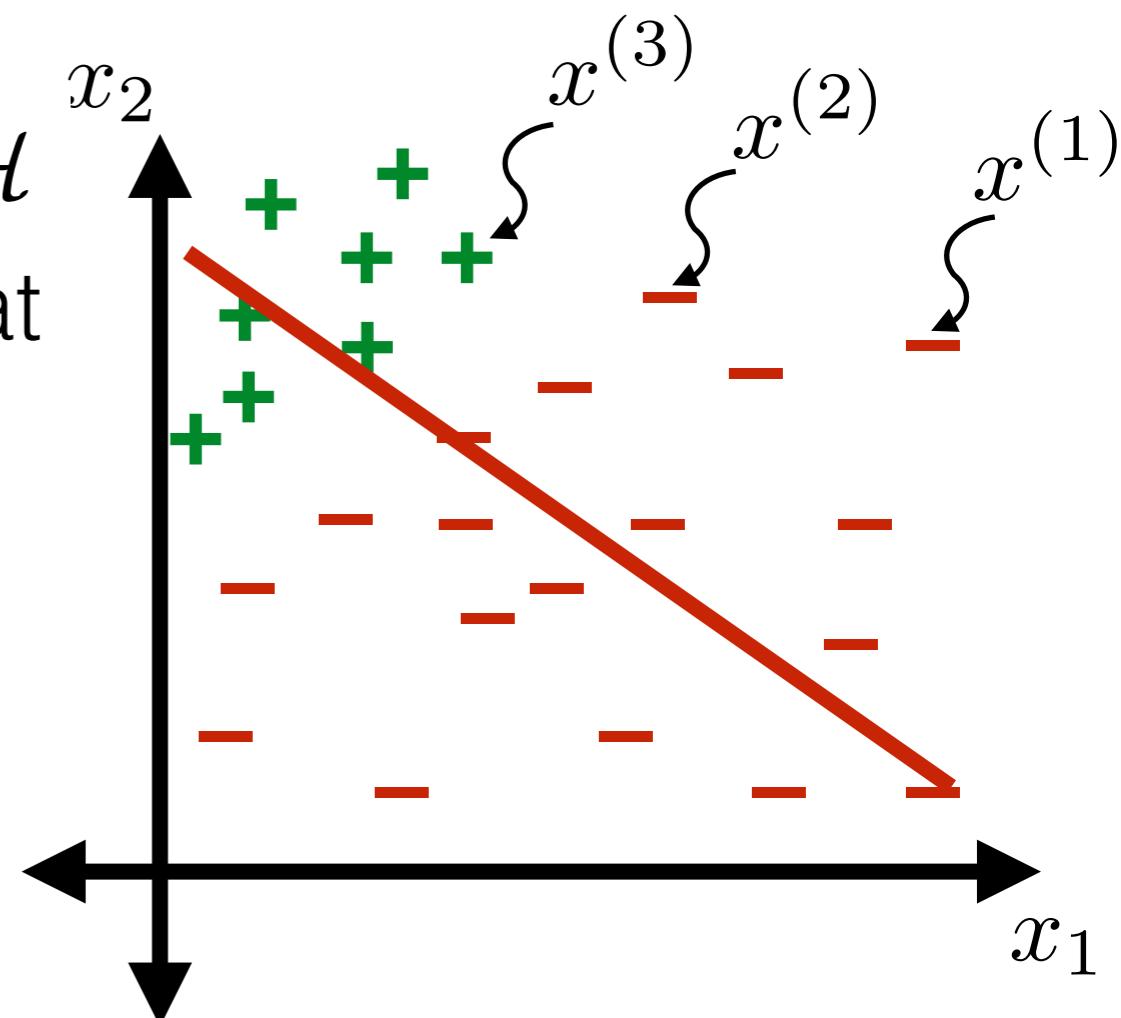
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



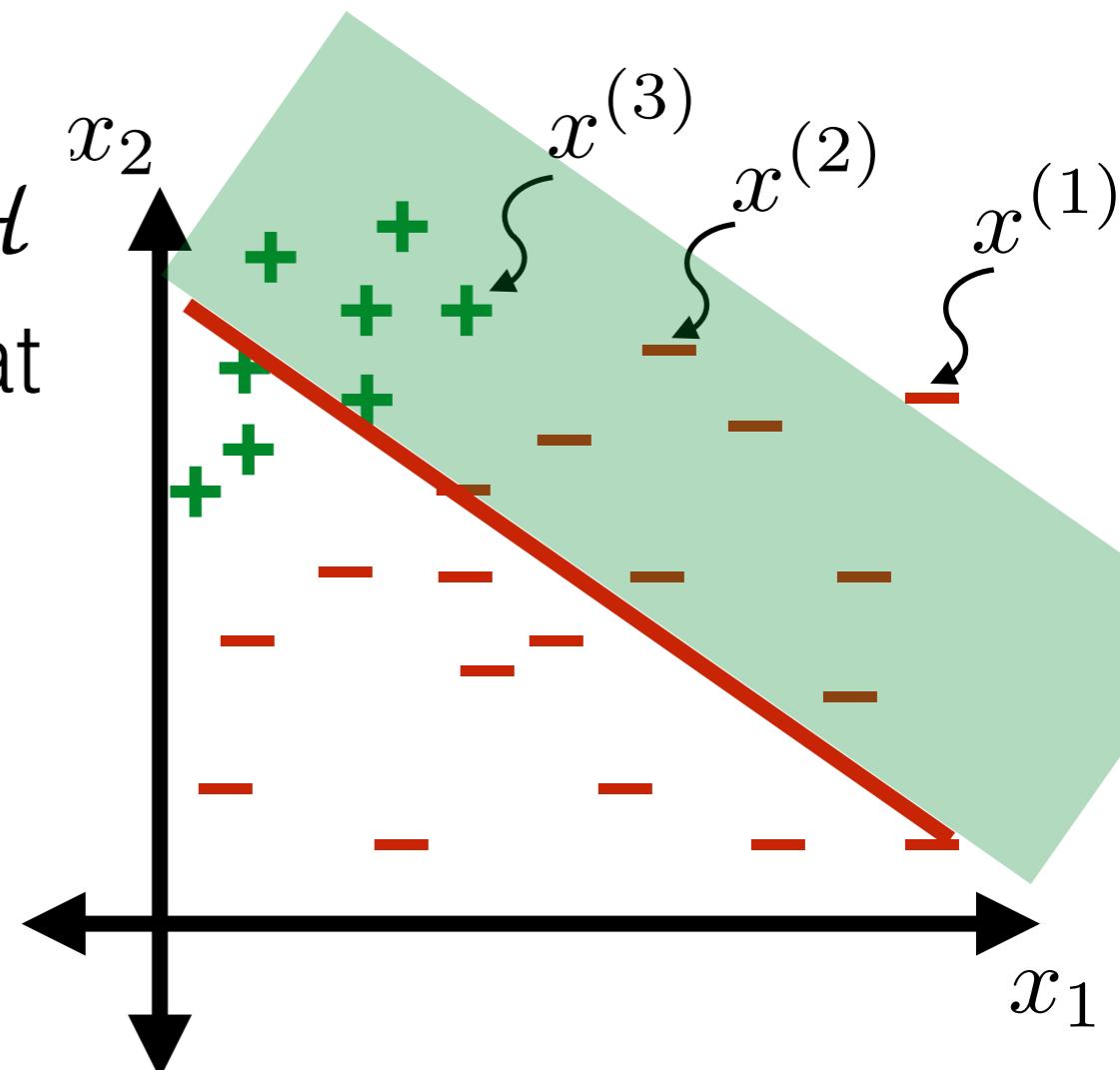
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



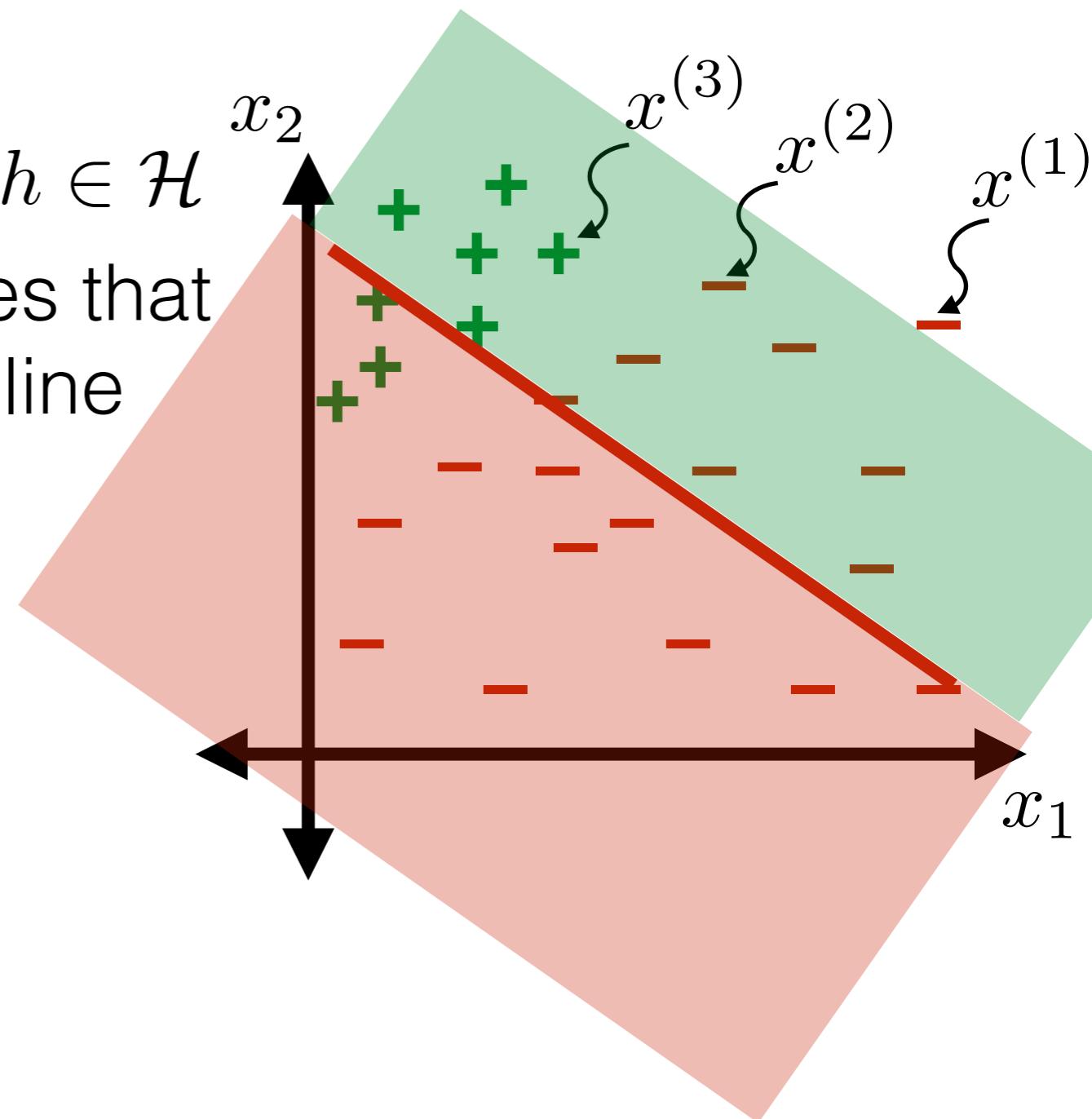
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



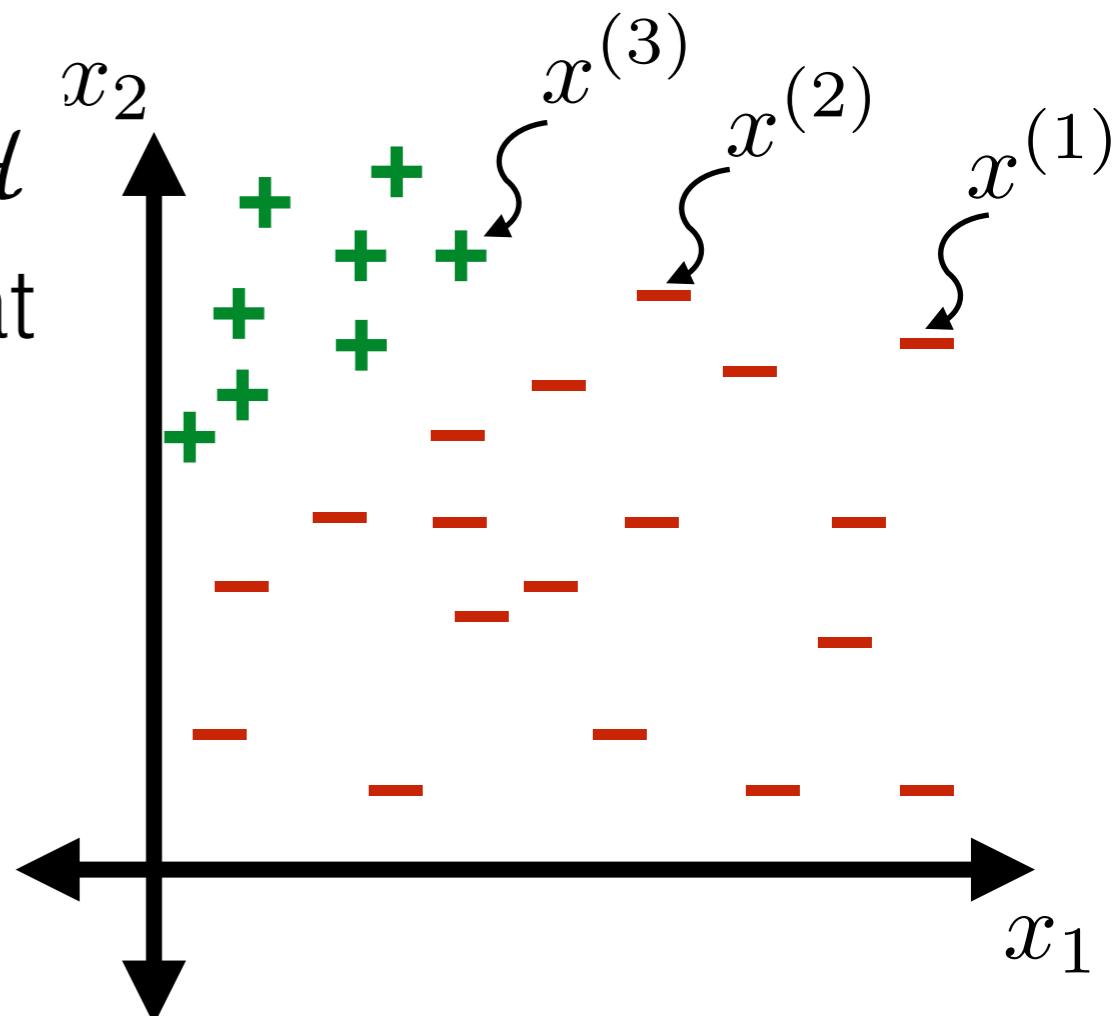
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



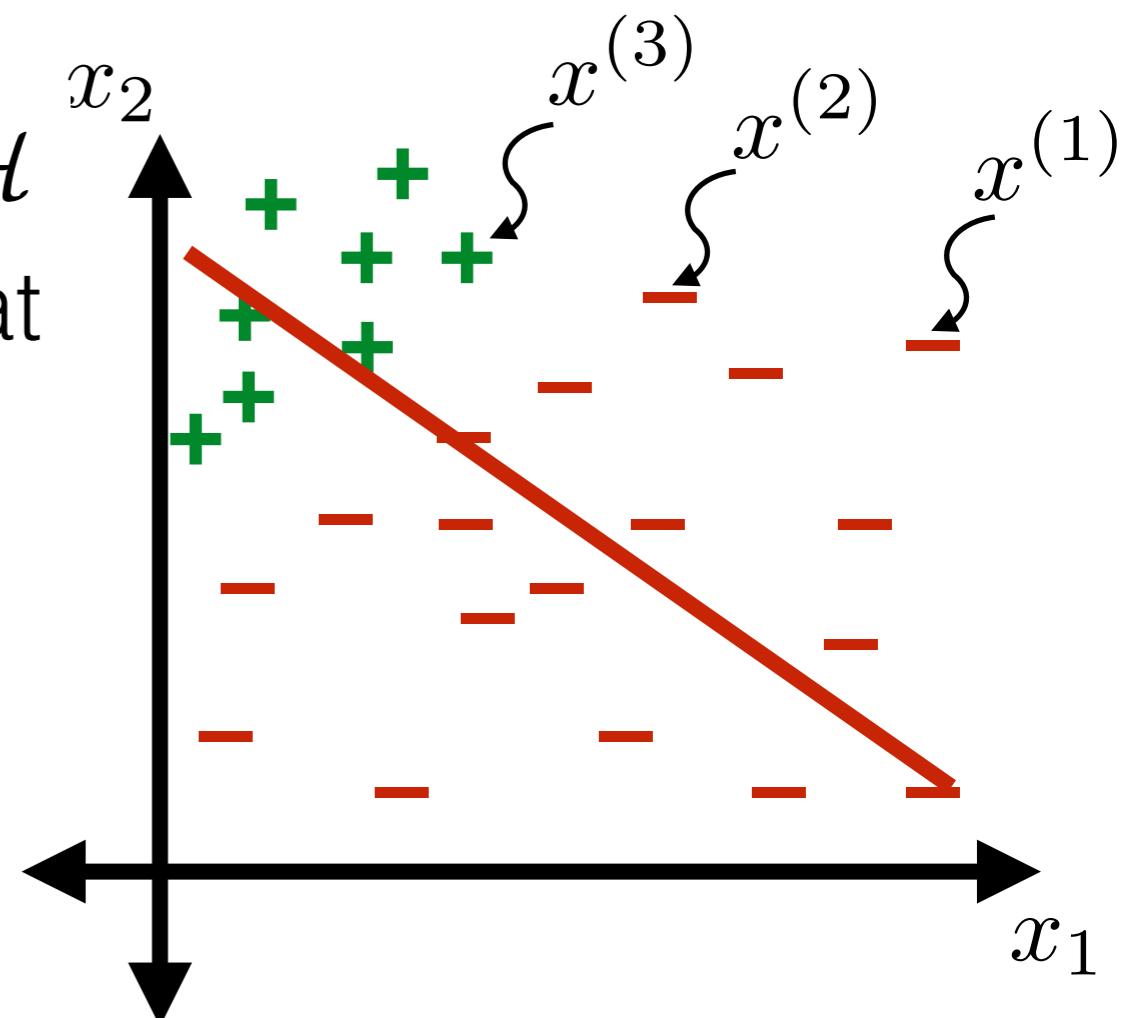
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



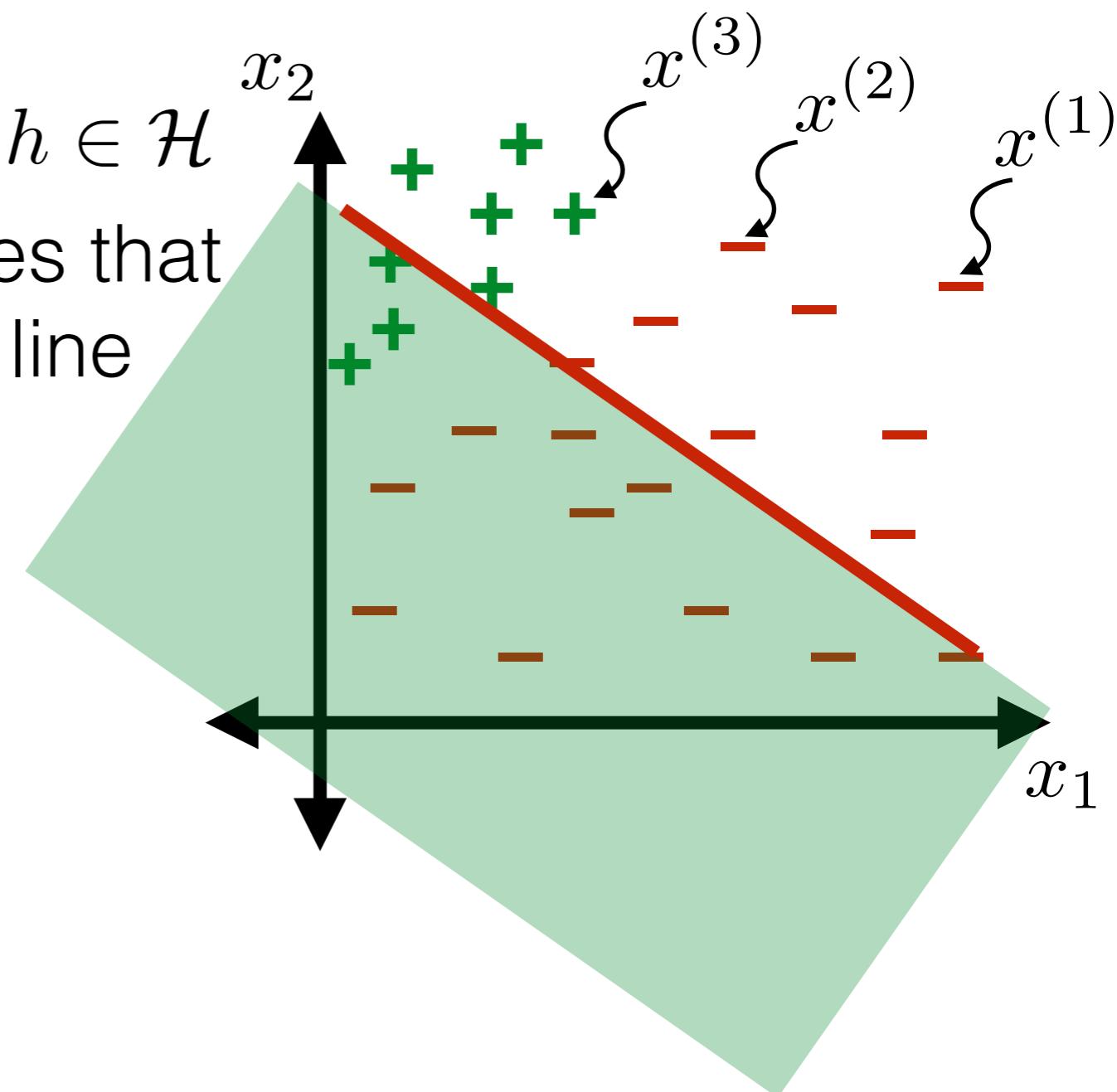
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



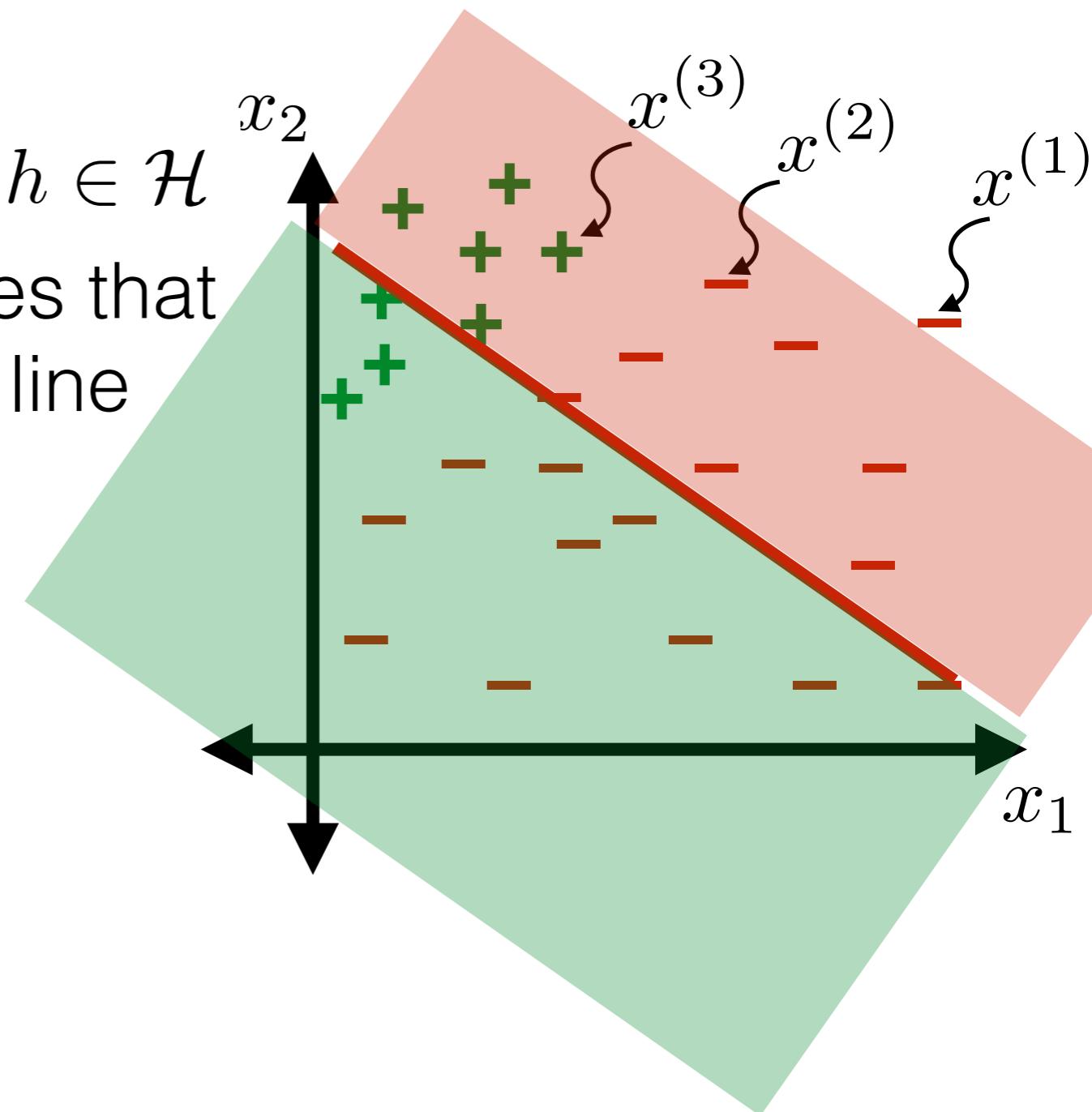
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



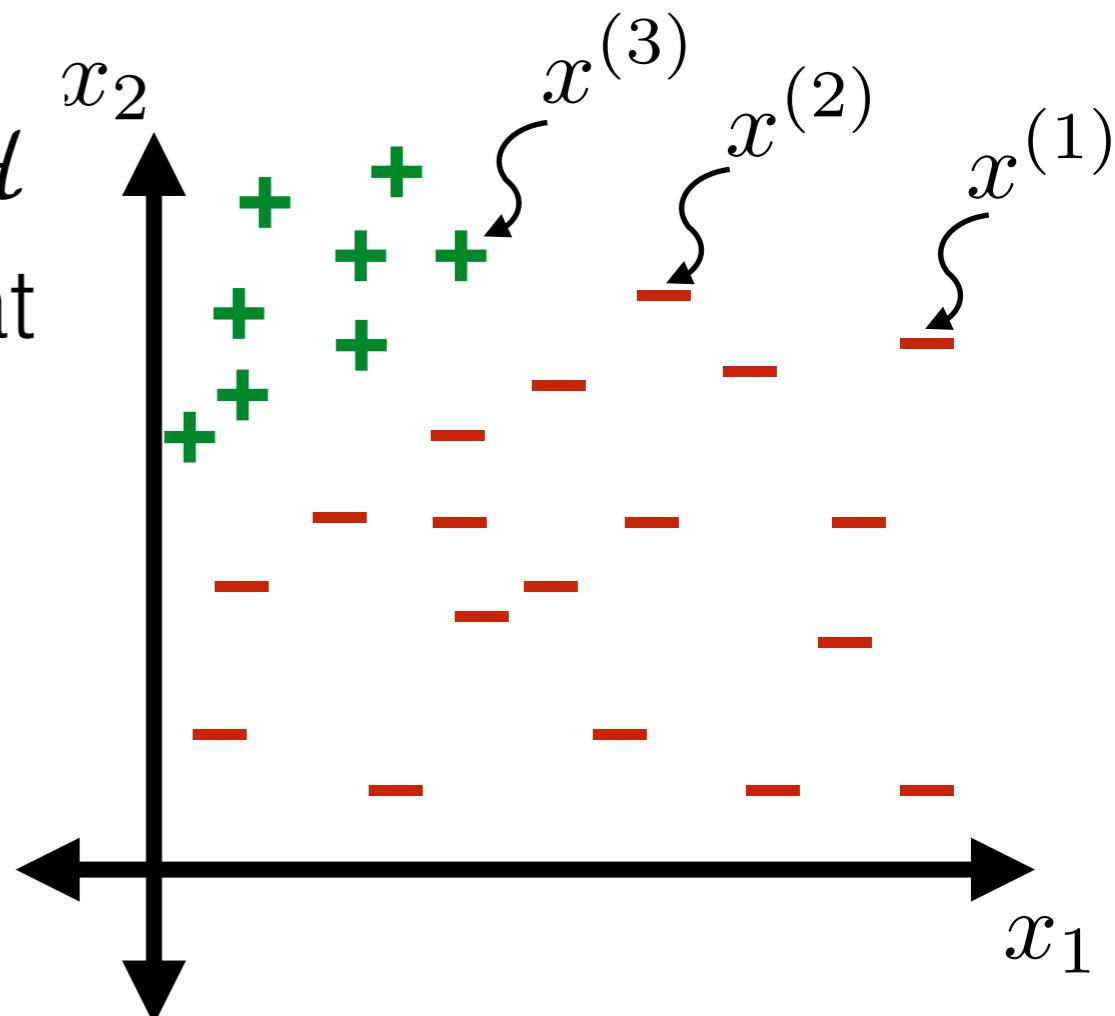
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



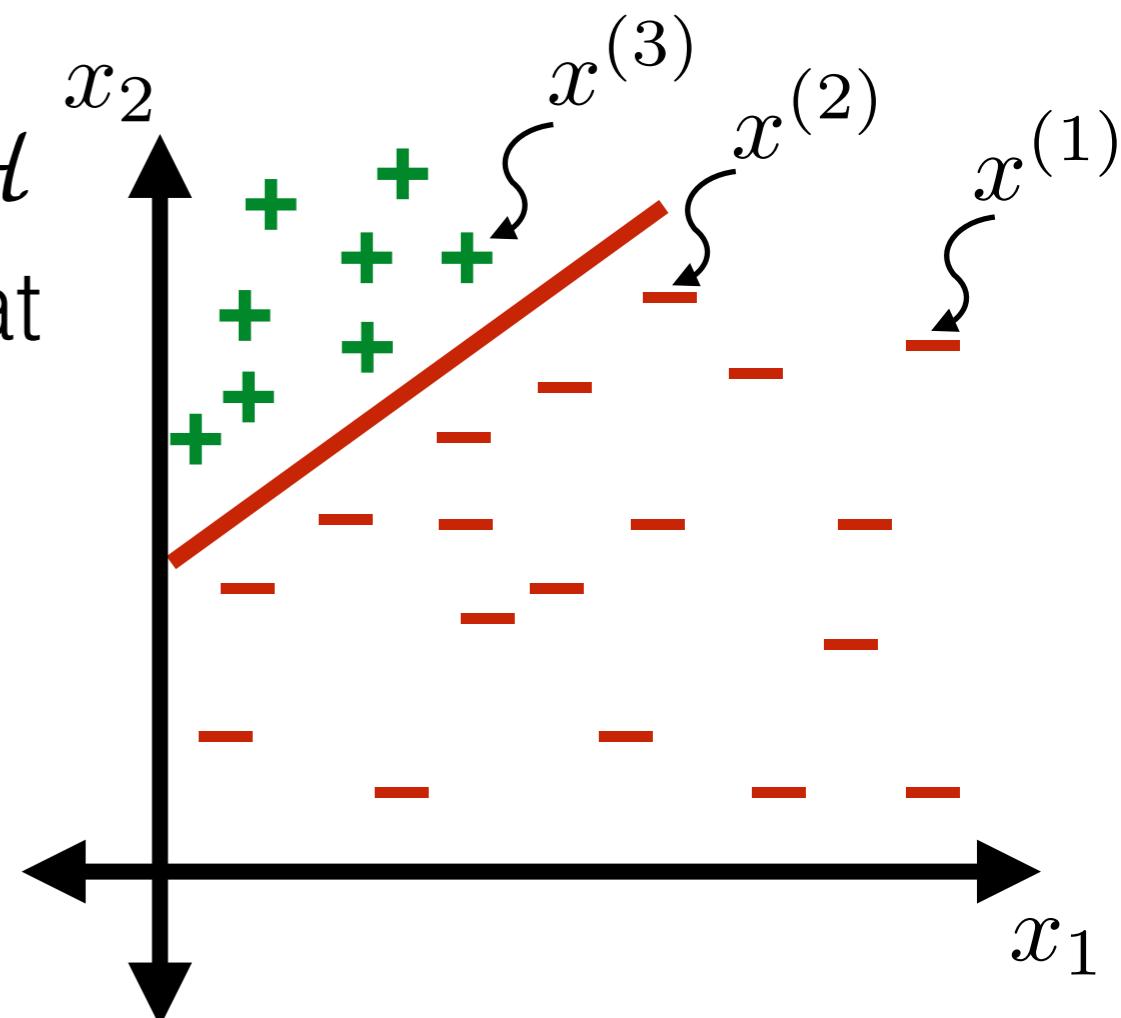
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



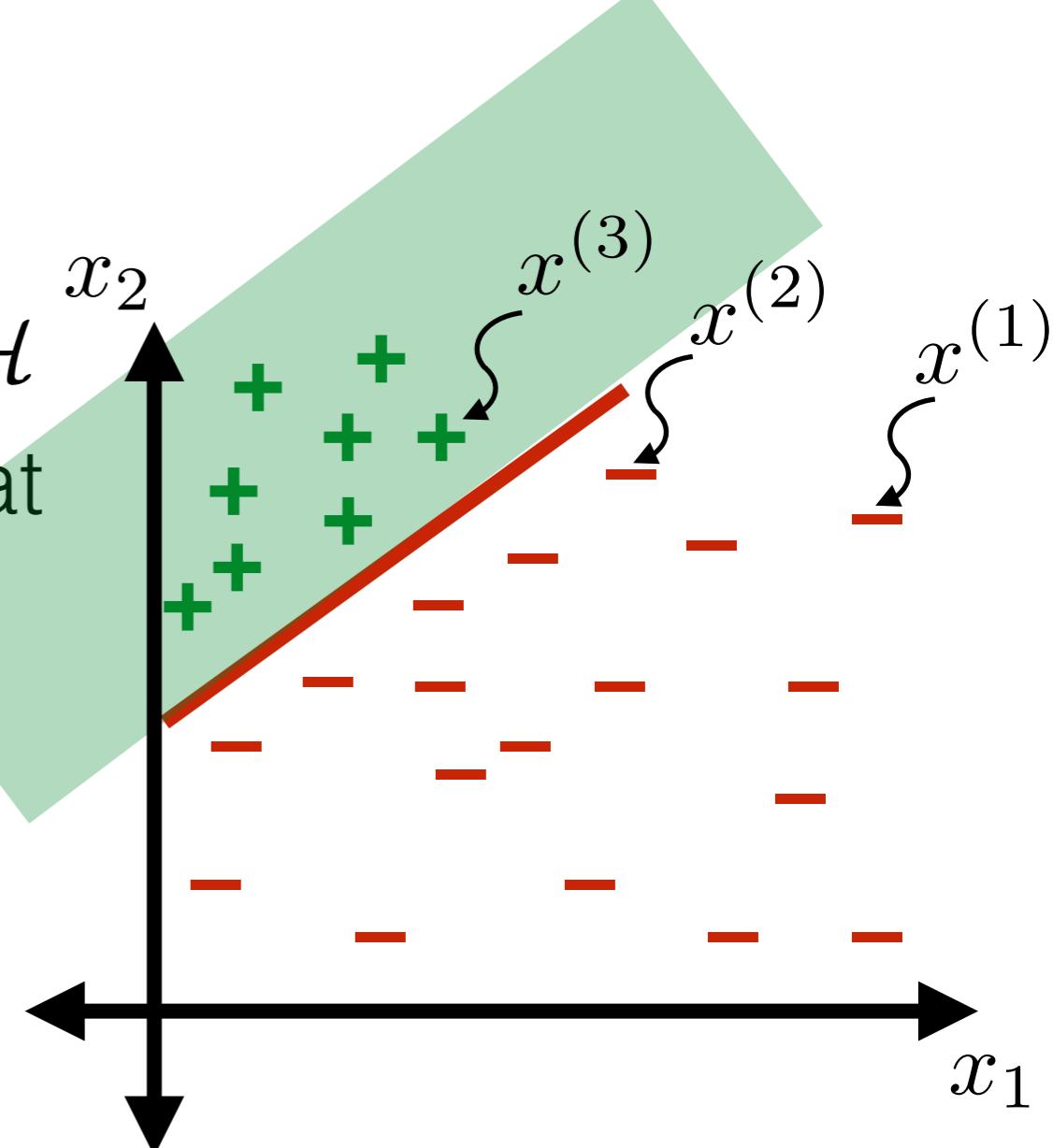
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



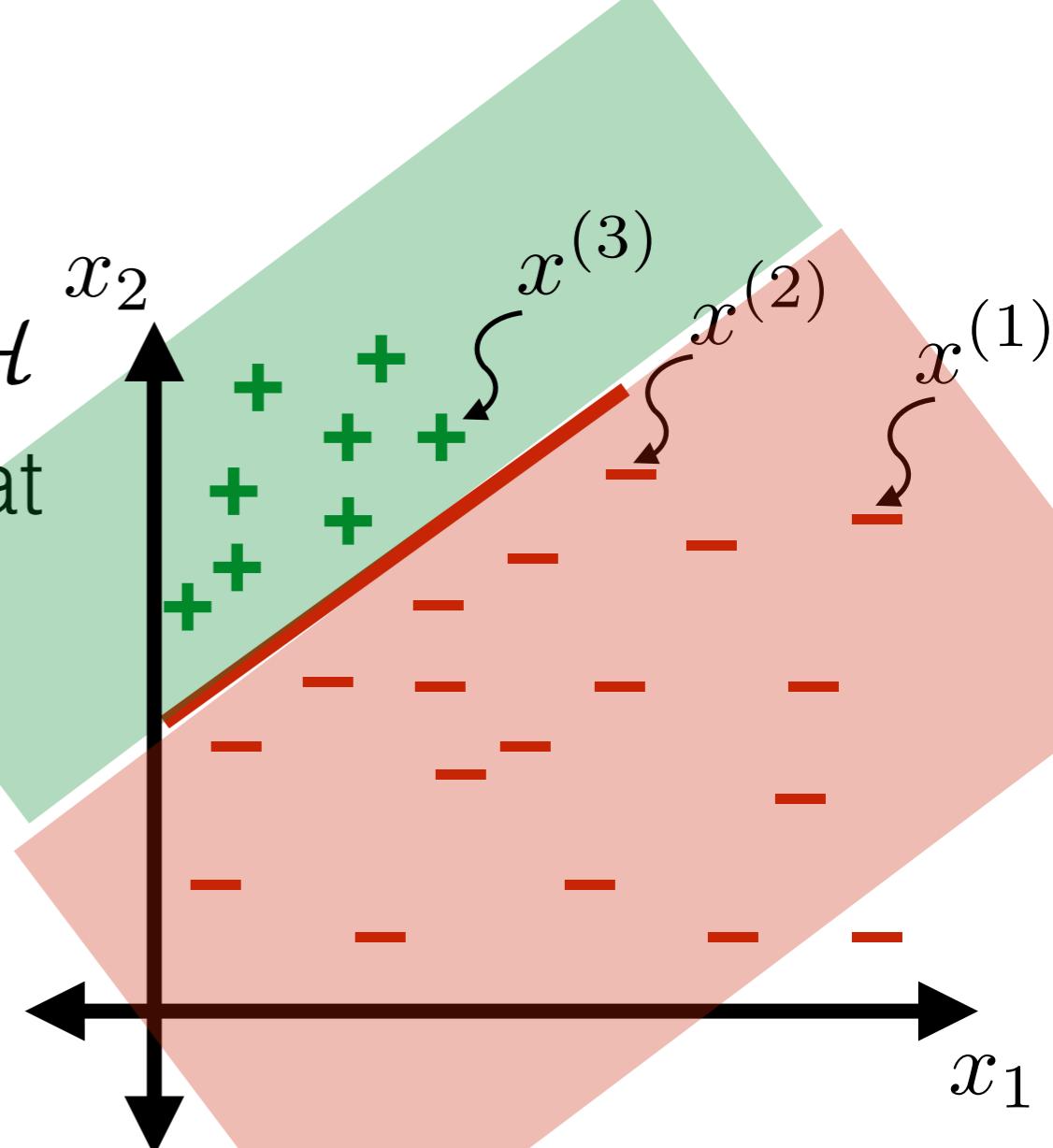
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



Linear classifiers

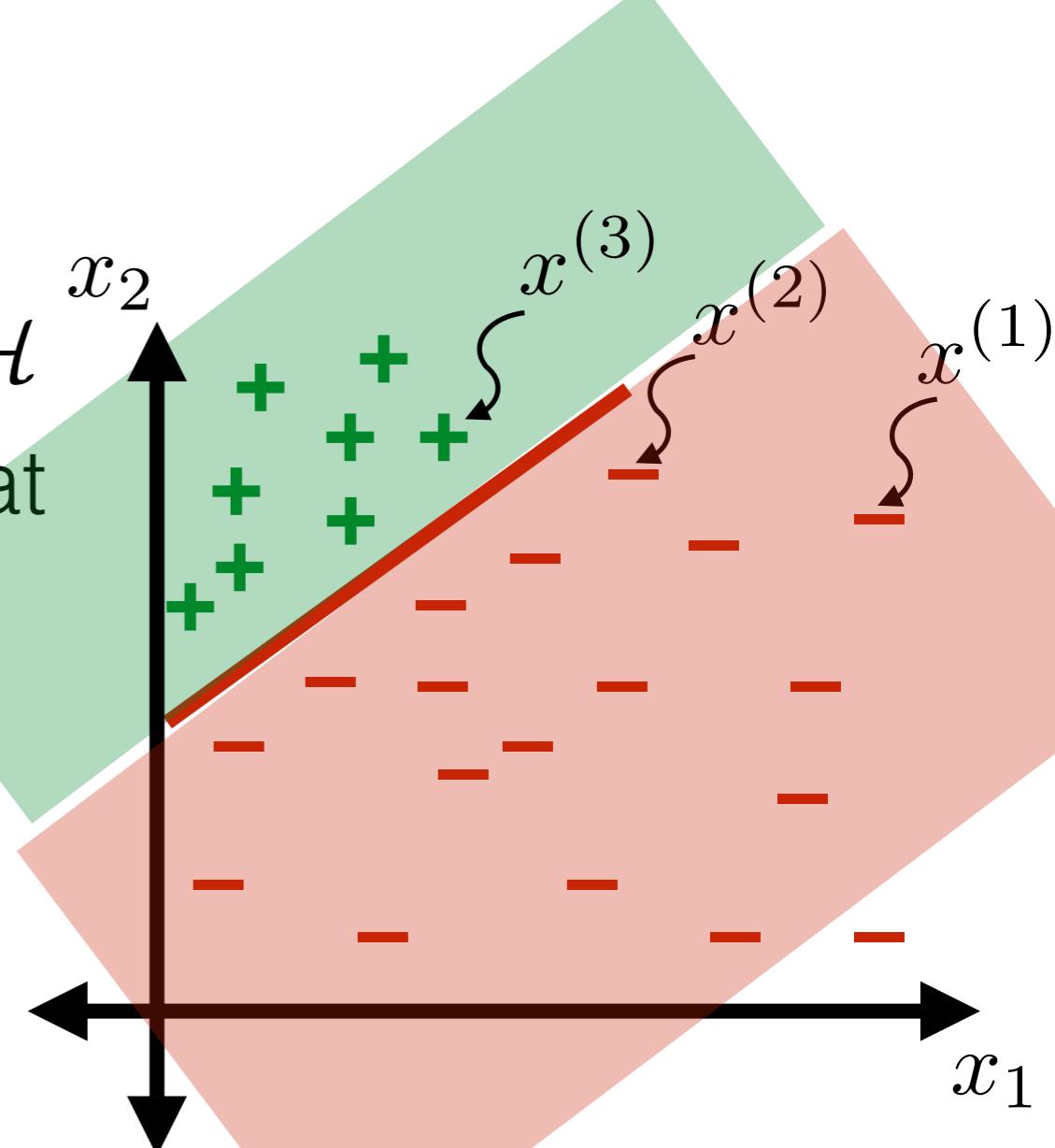
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side

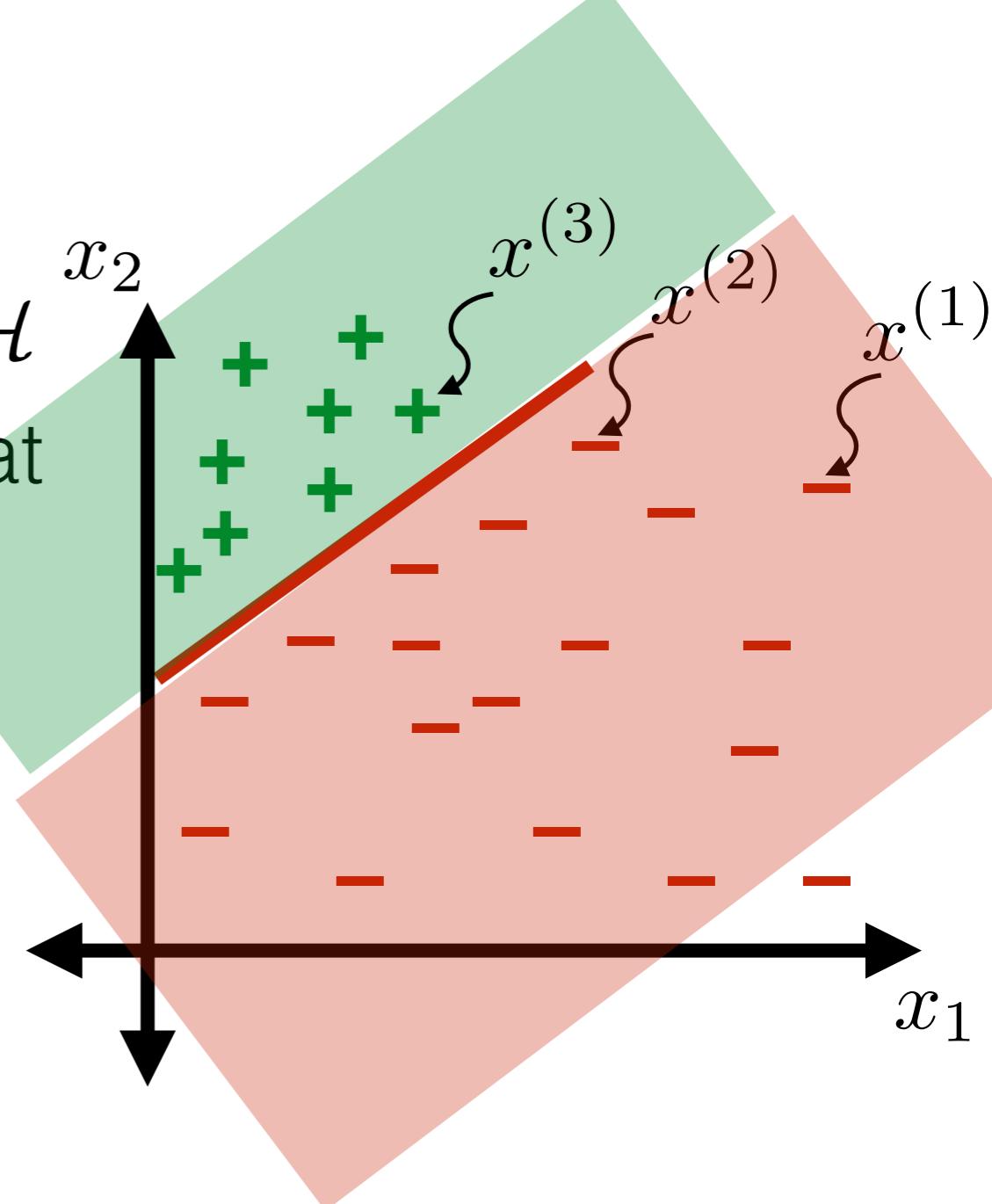
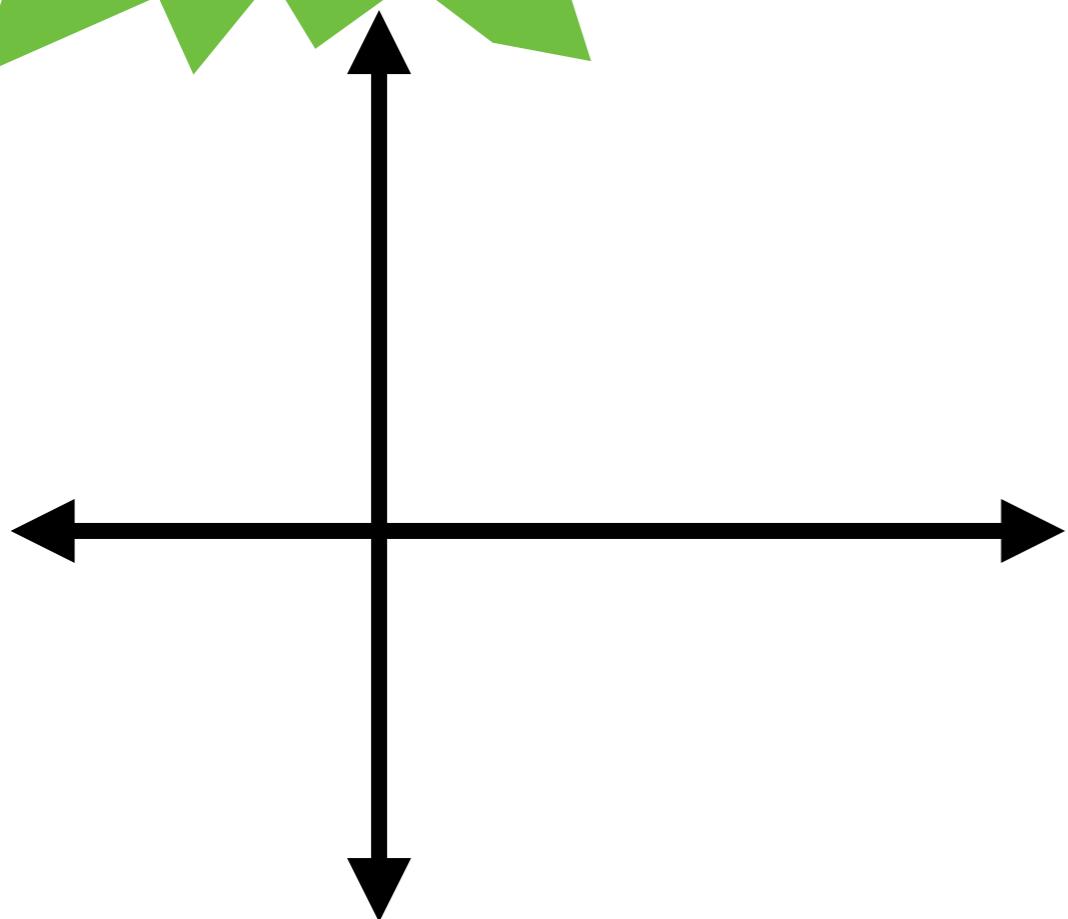
Math facts!



Linear classifiers

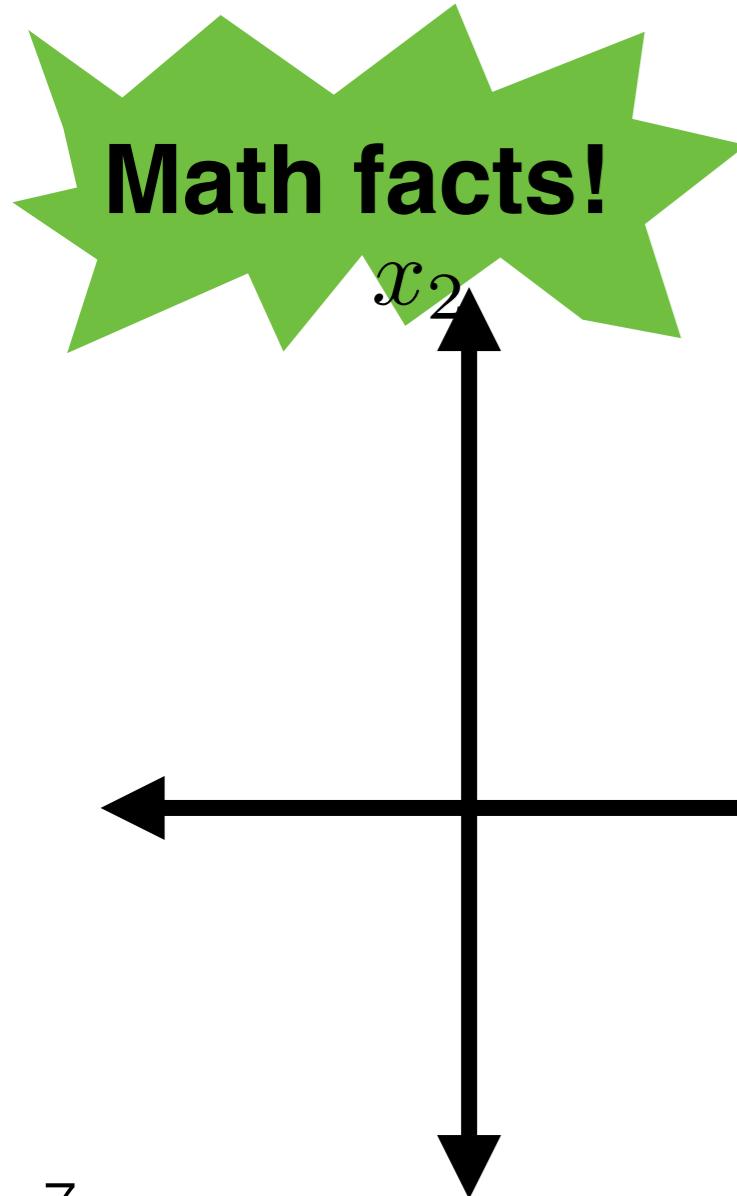
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side

Math facts!

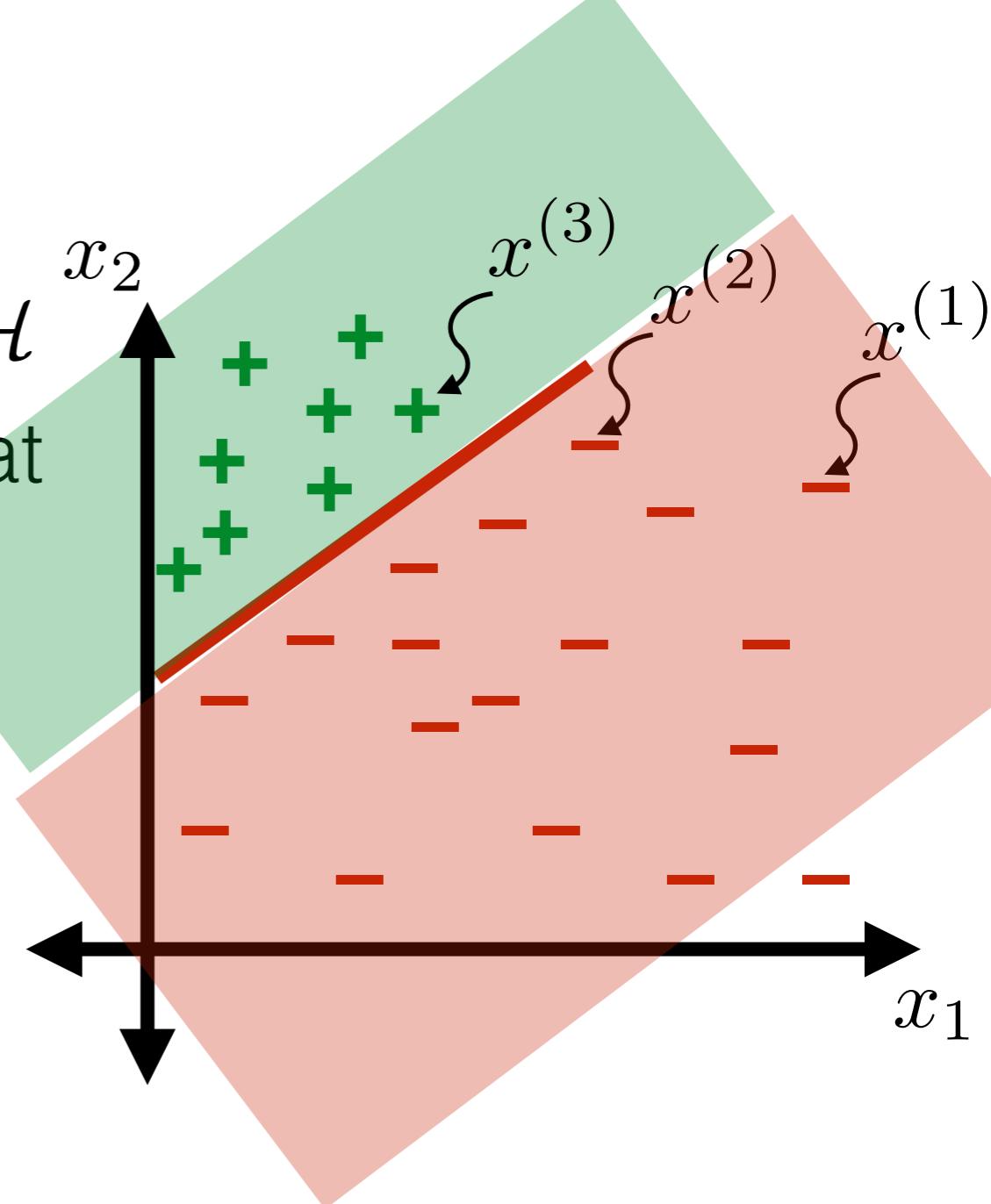


Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side

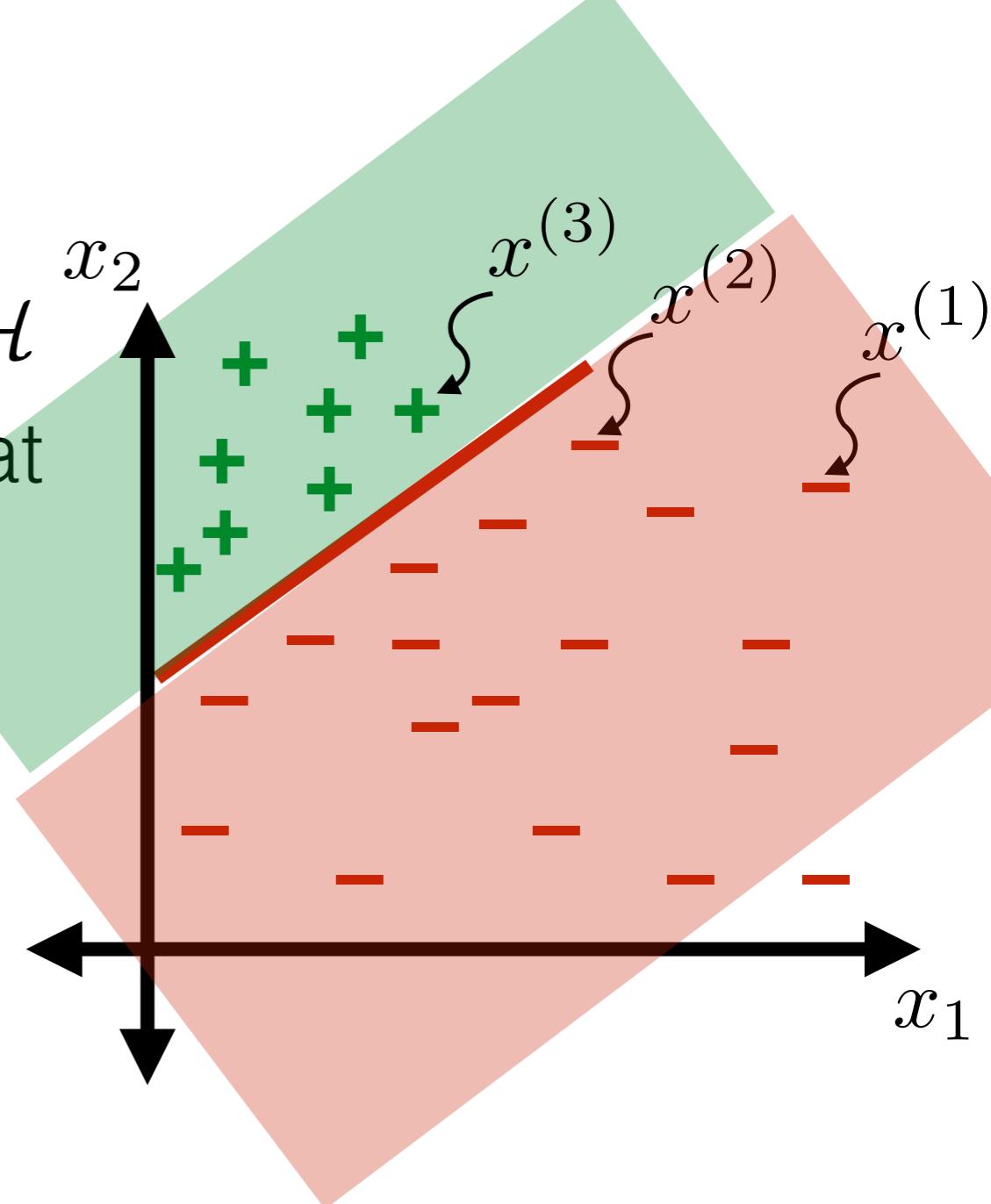
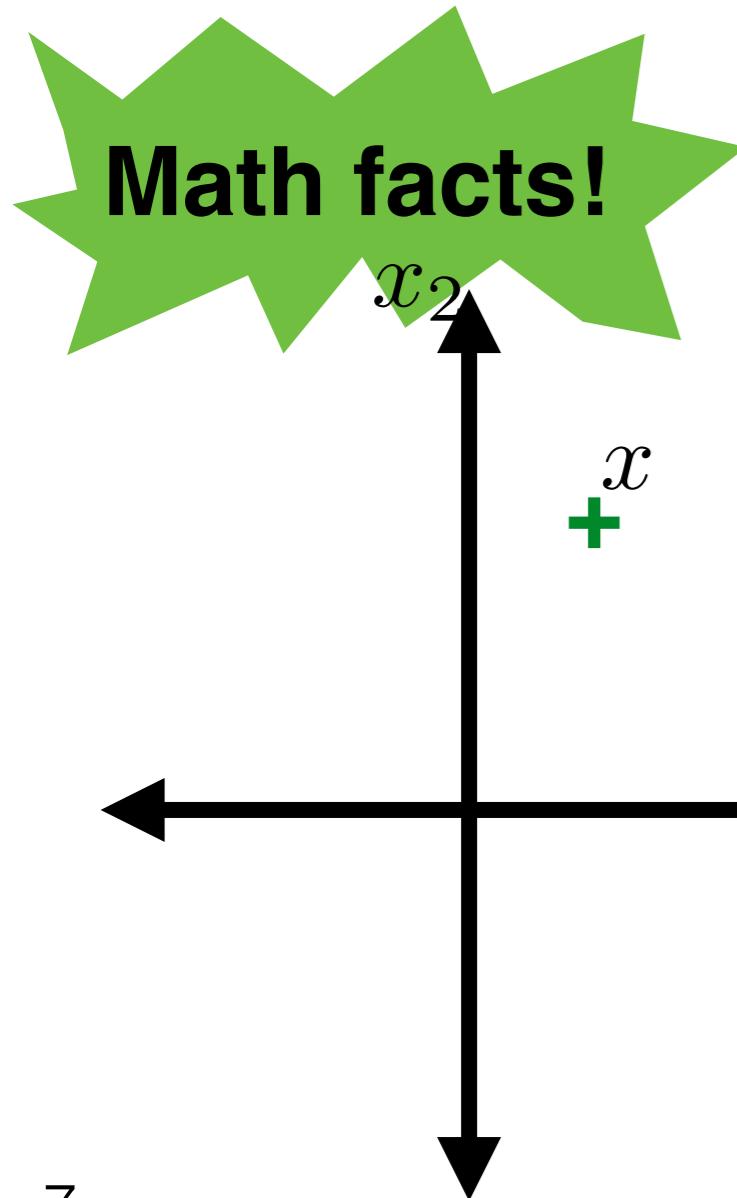


Math facts!



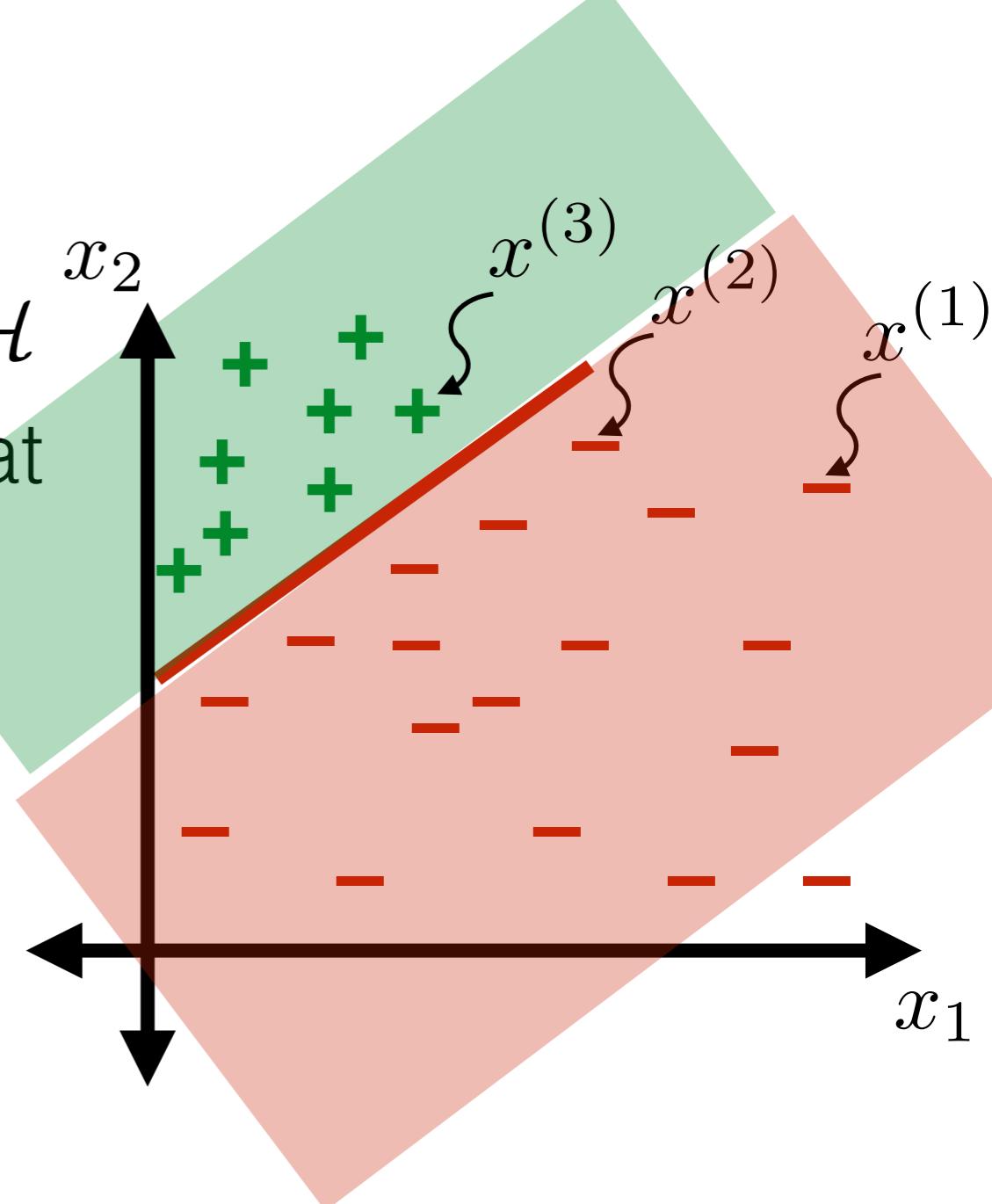
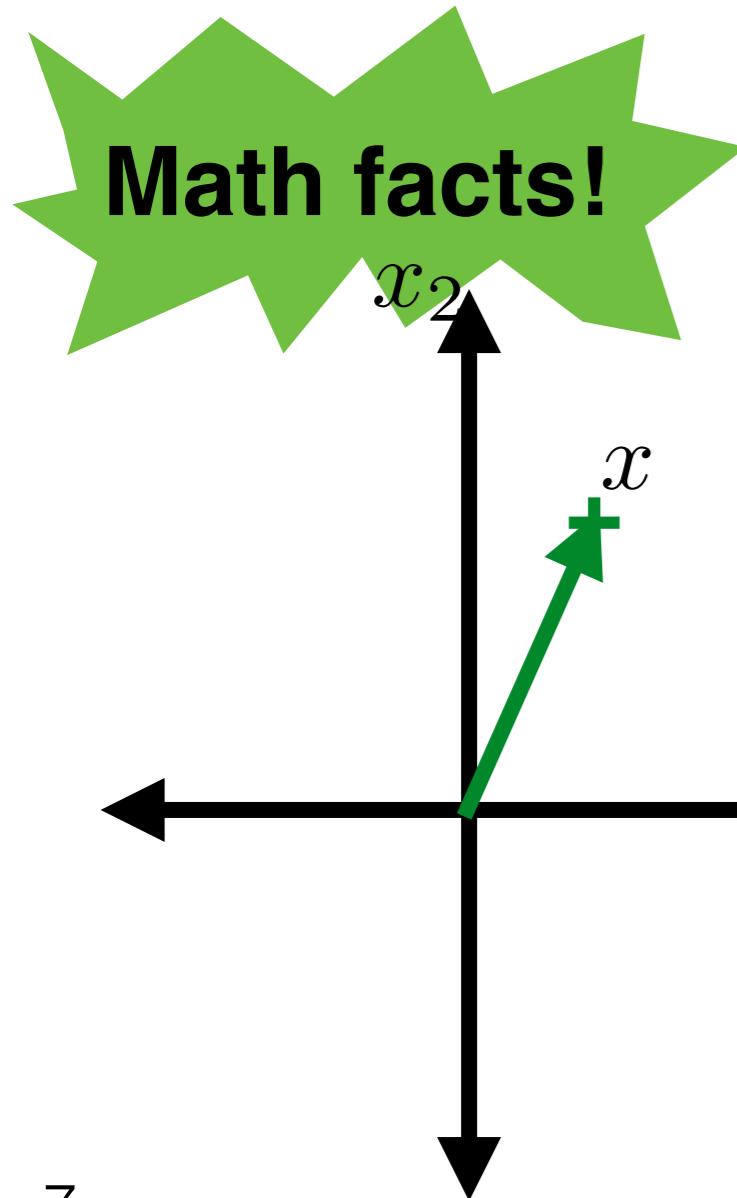
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



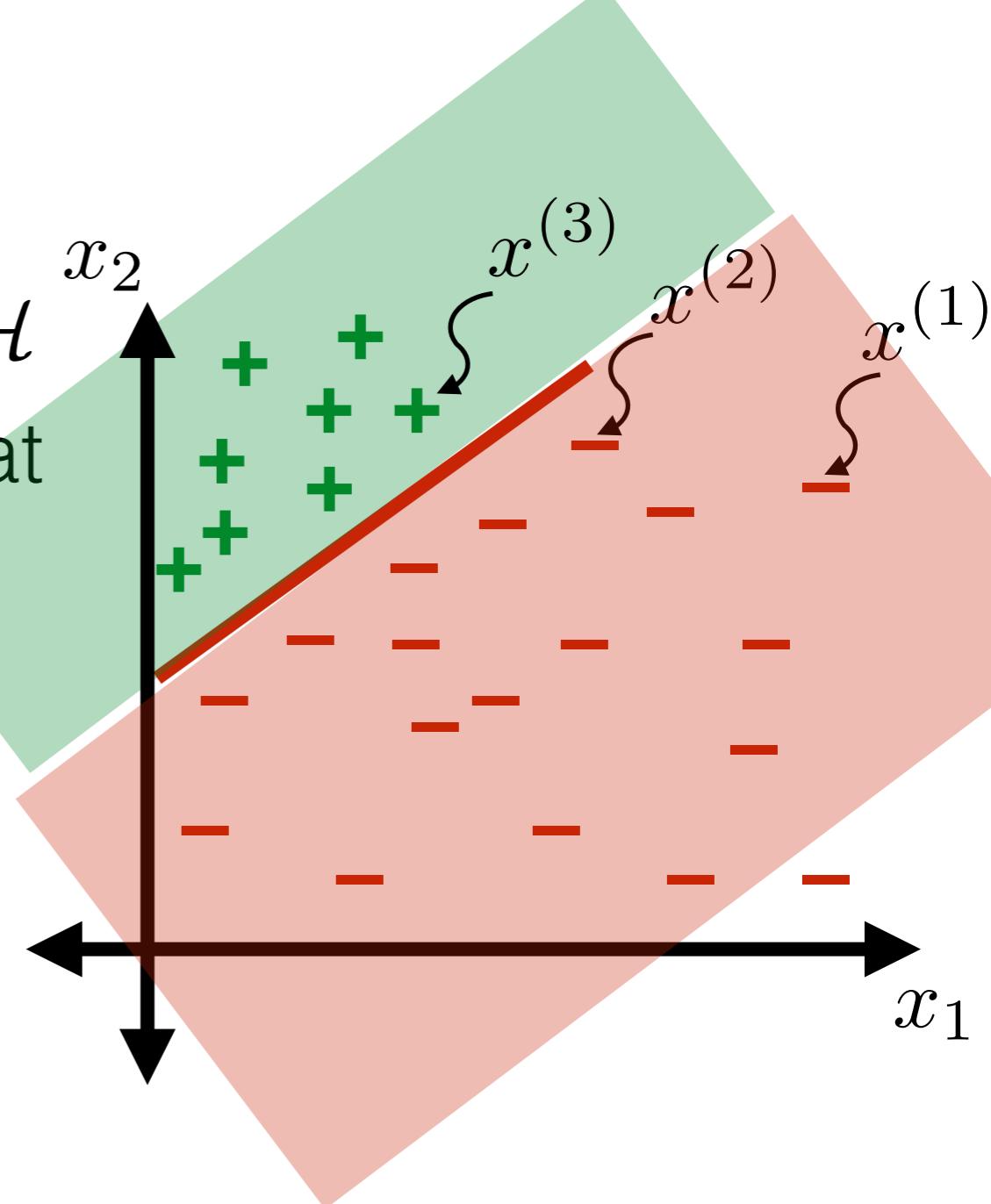
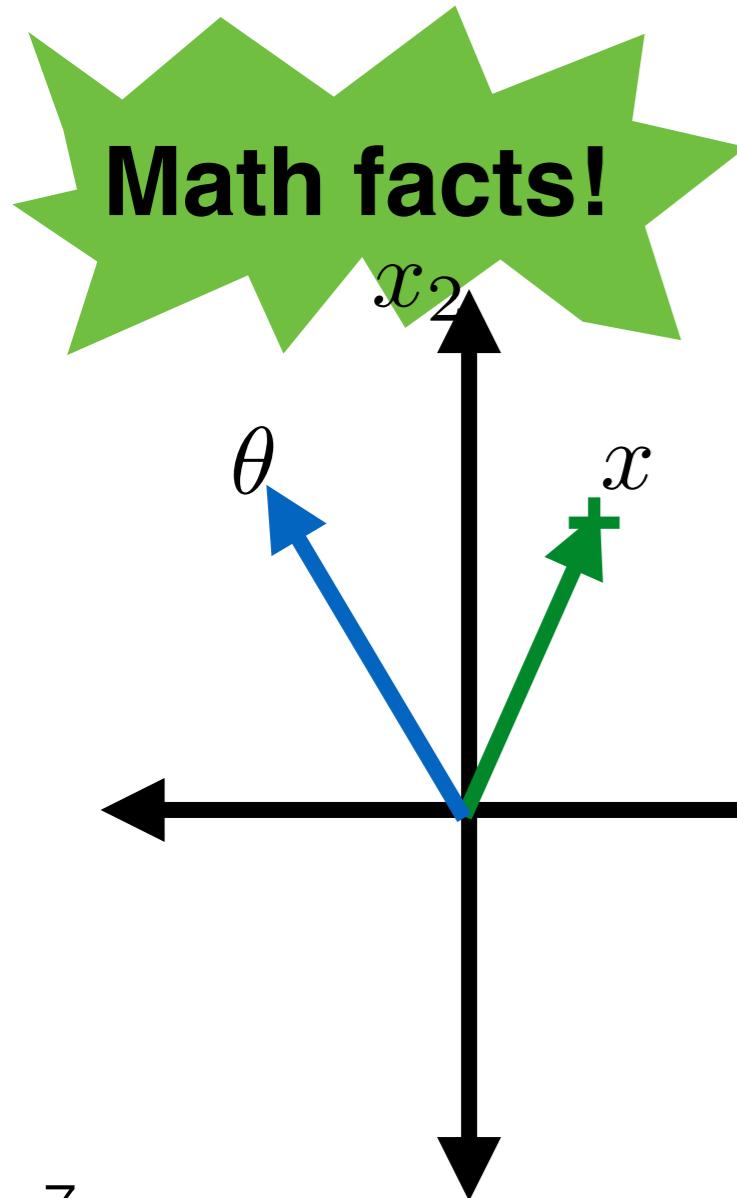
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



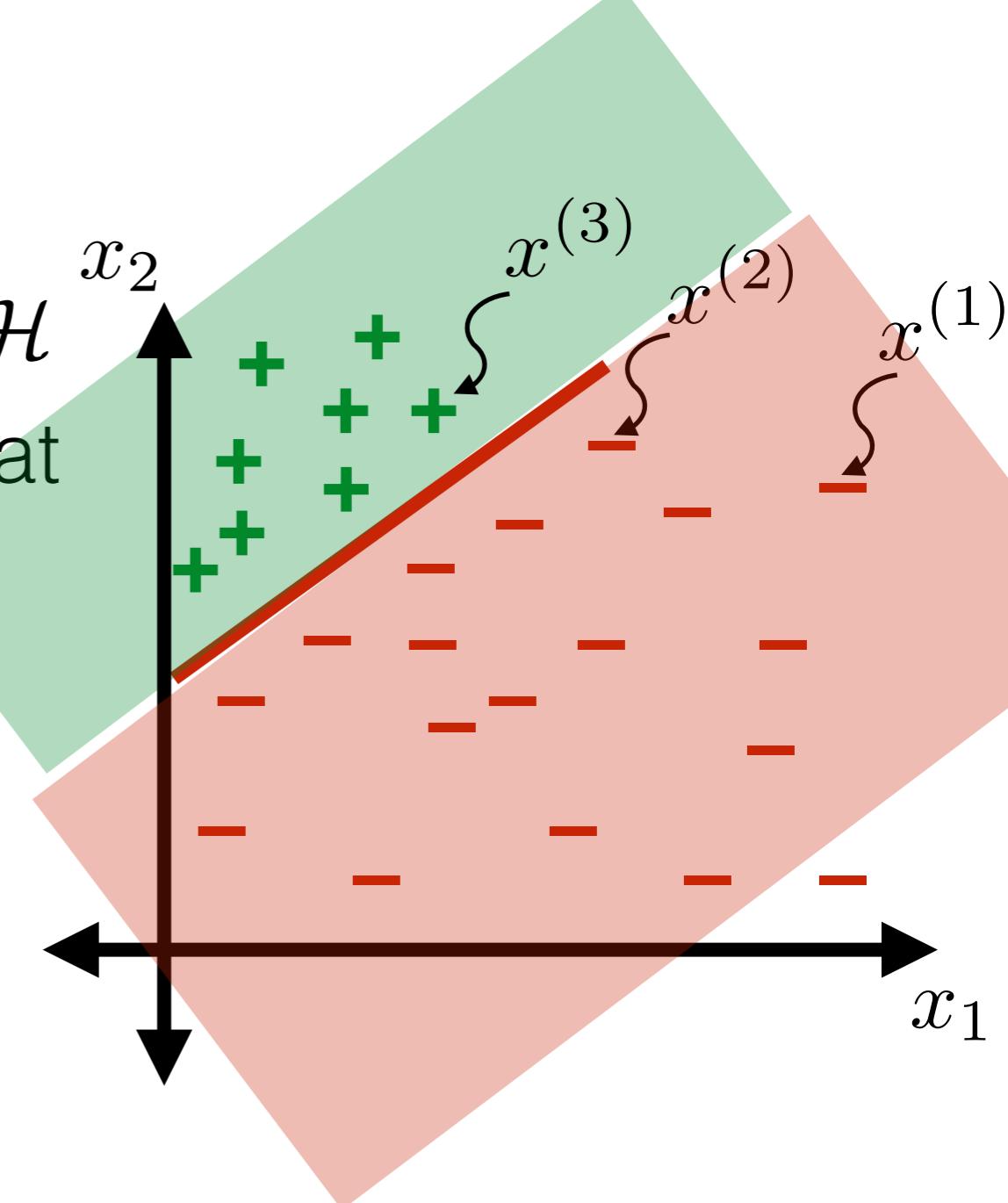
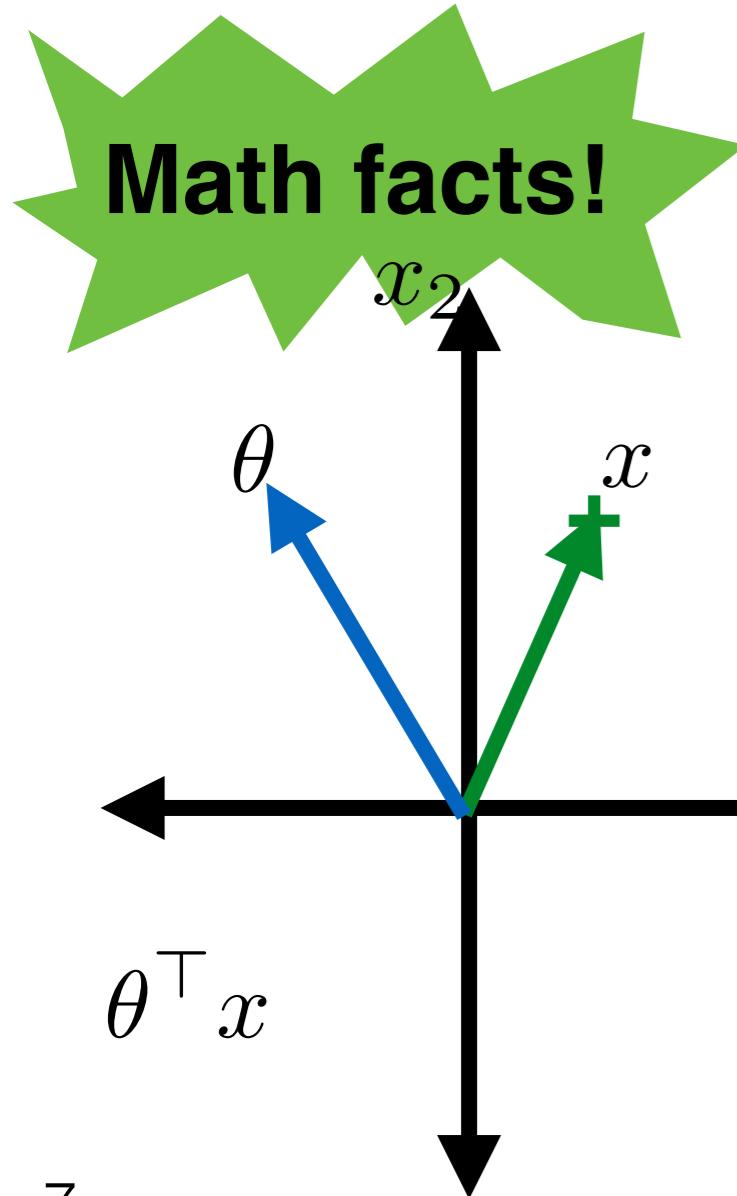
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



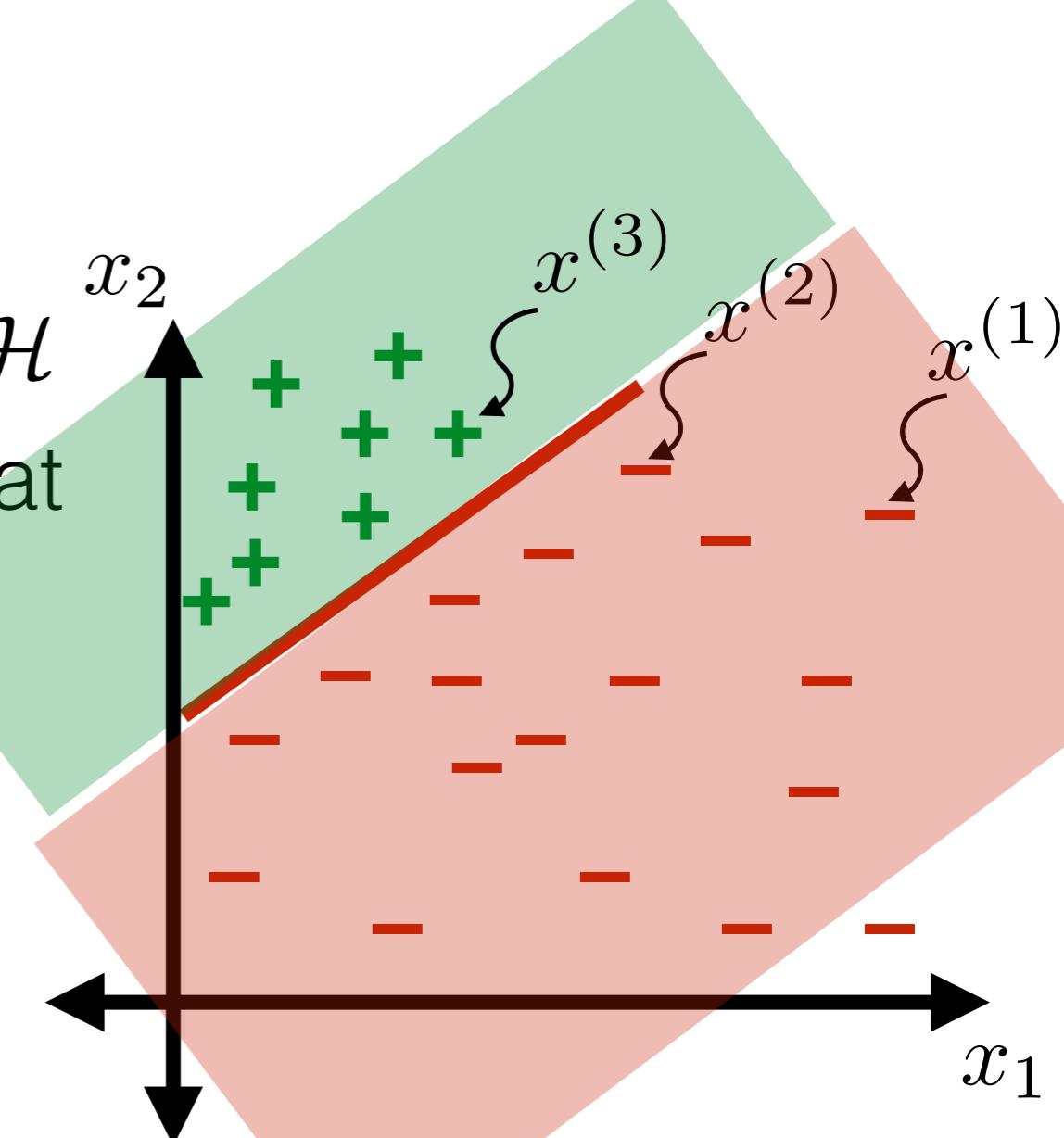
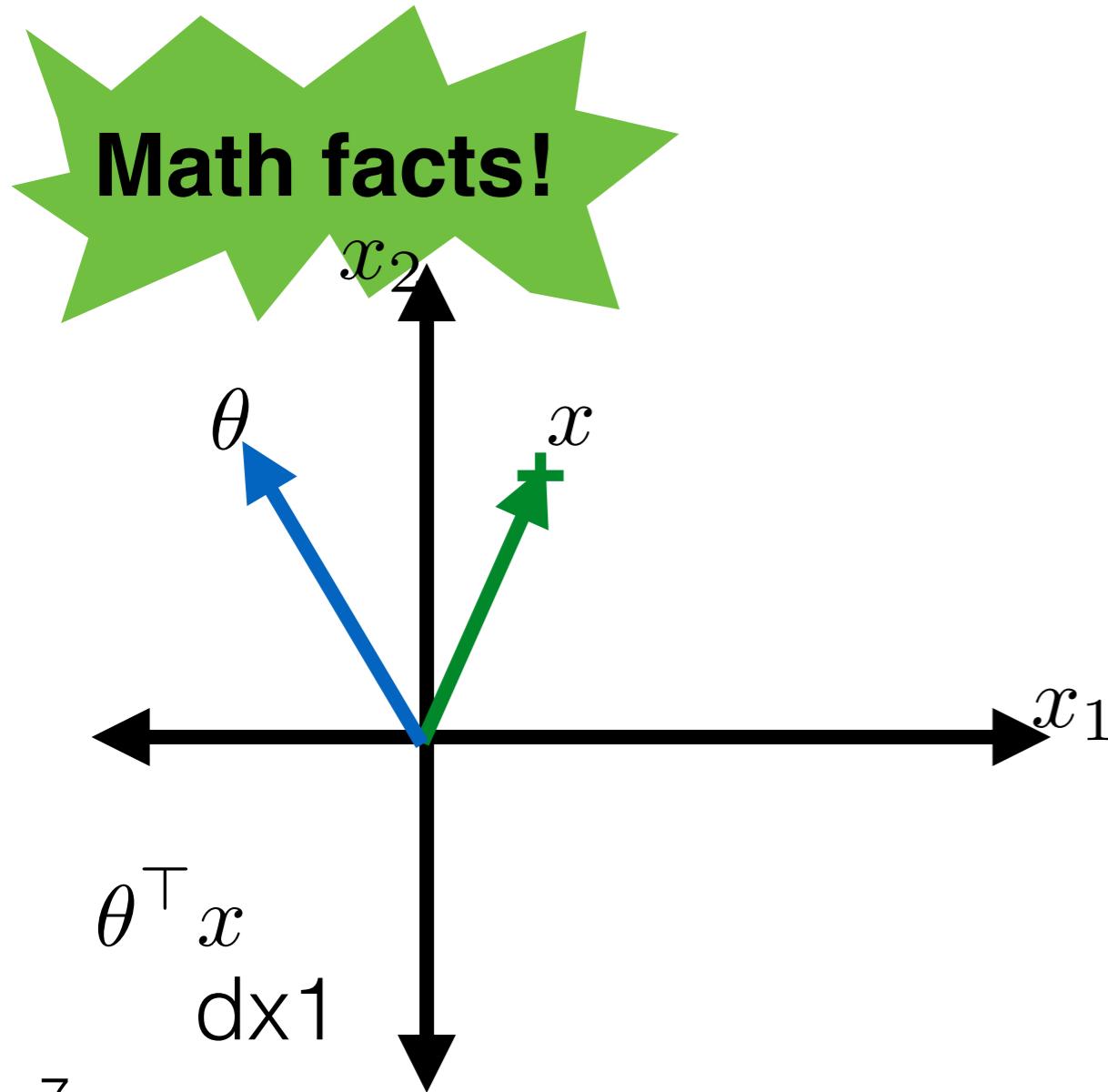
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



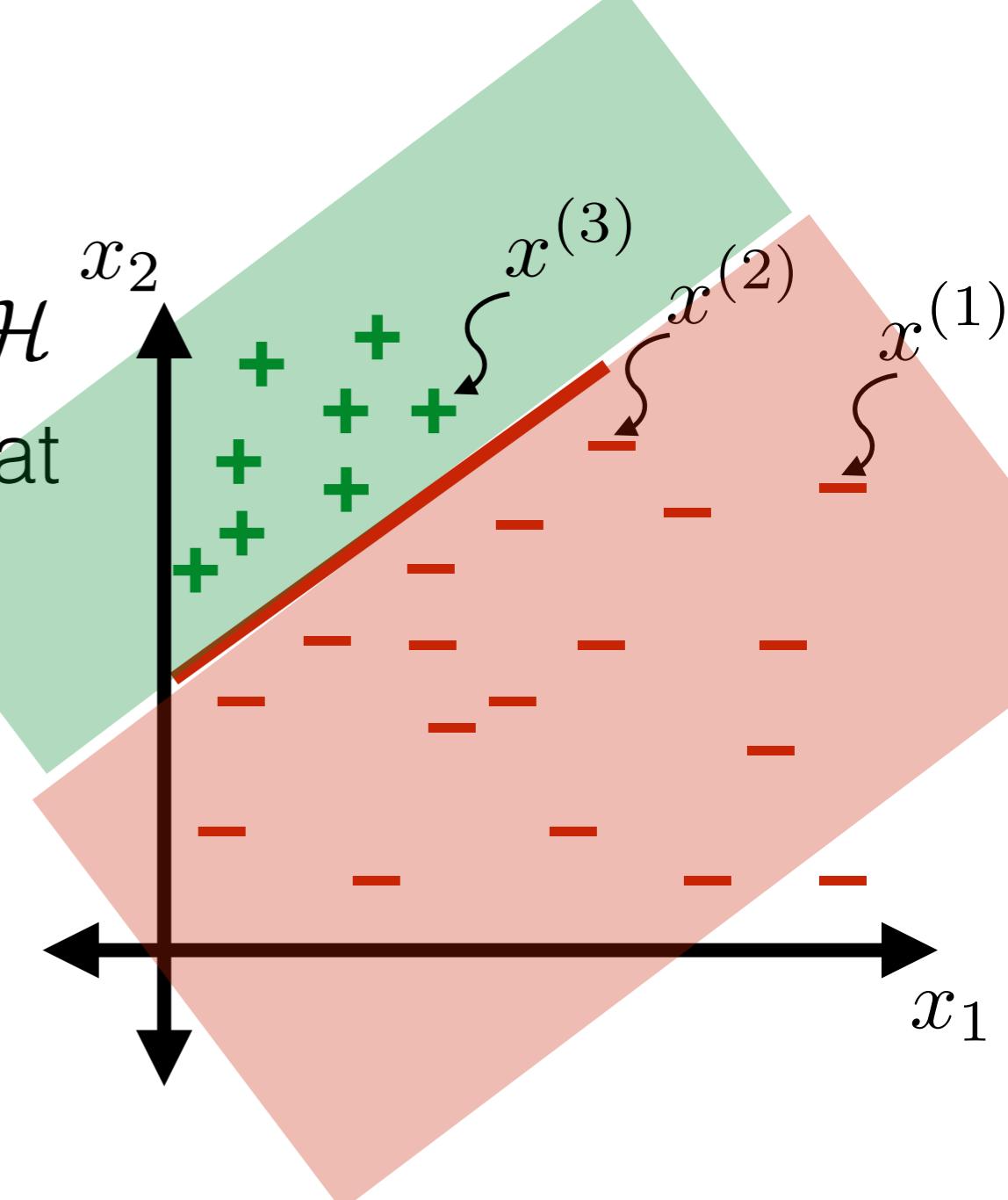
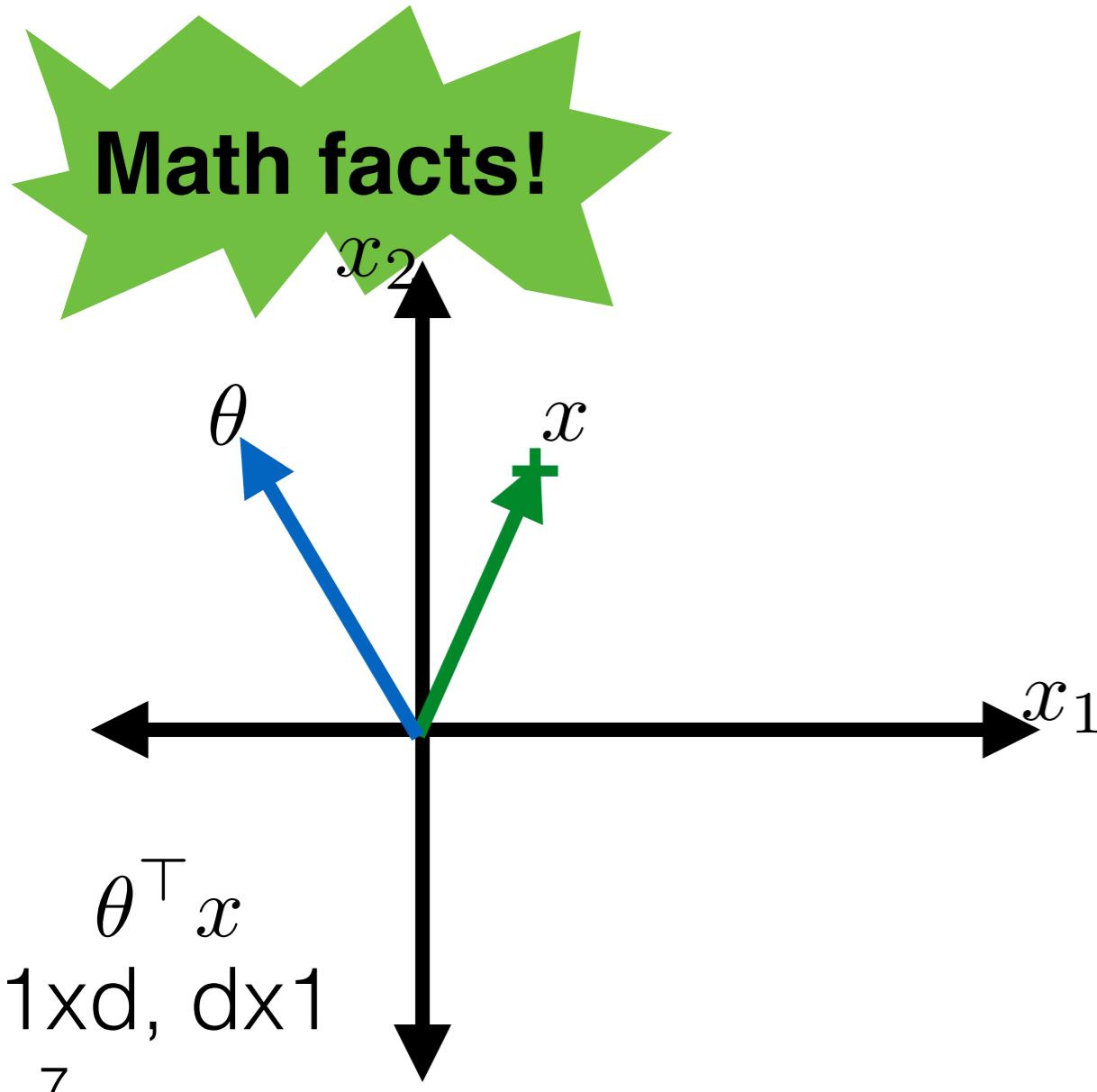
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



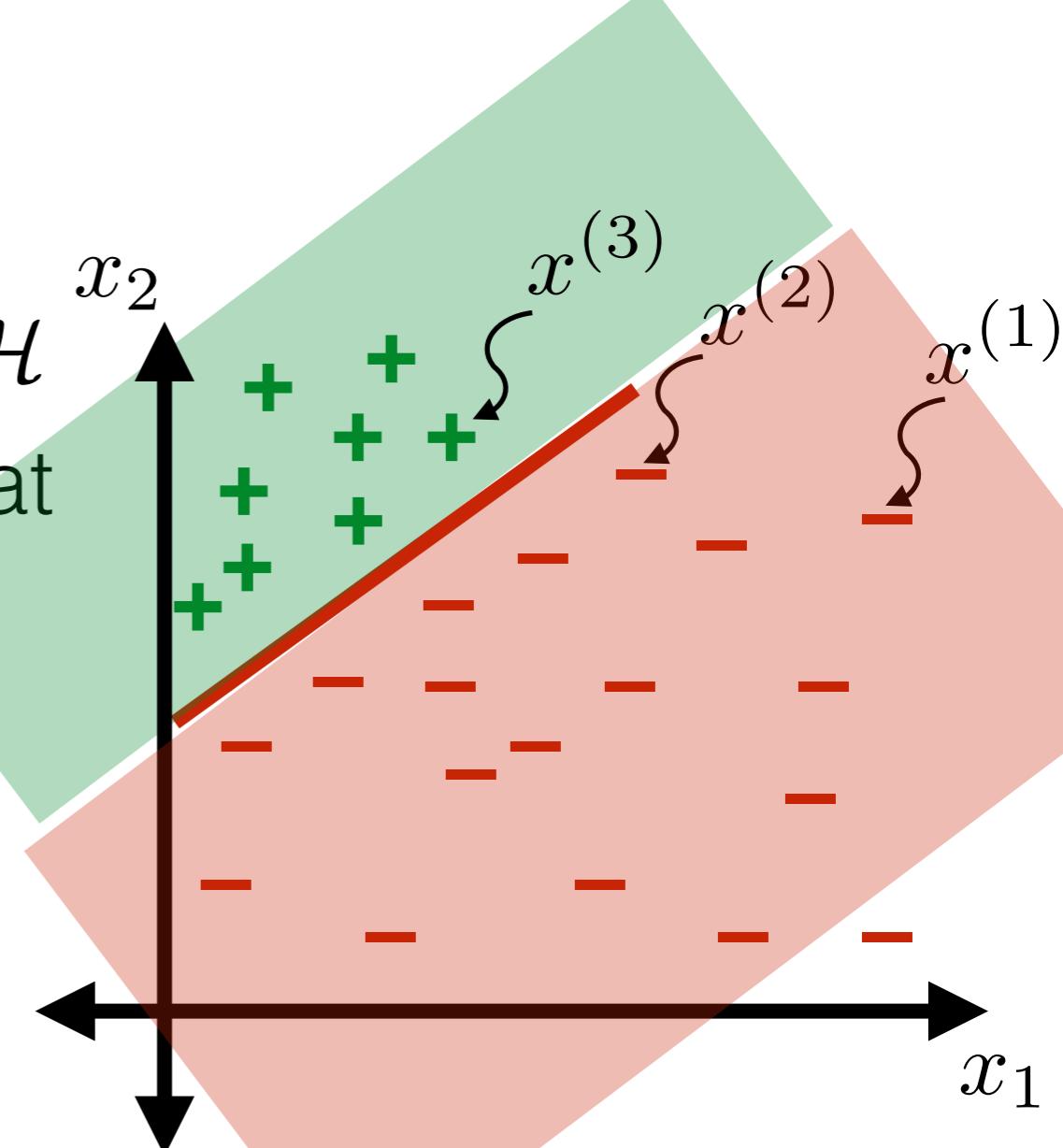
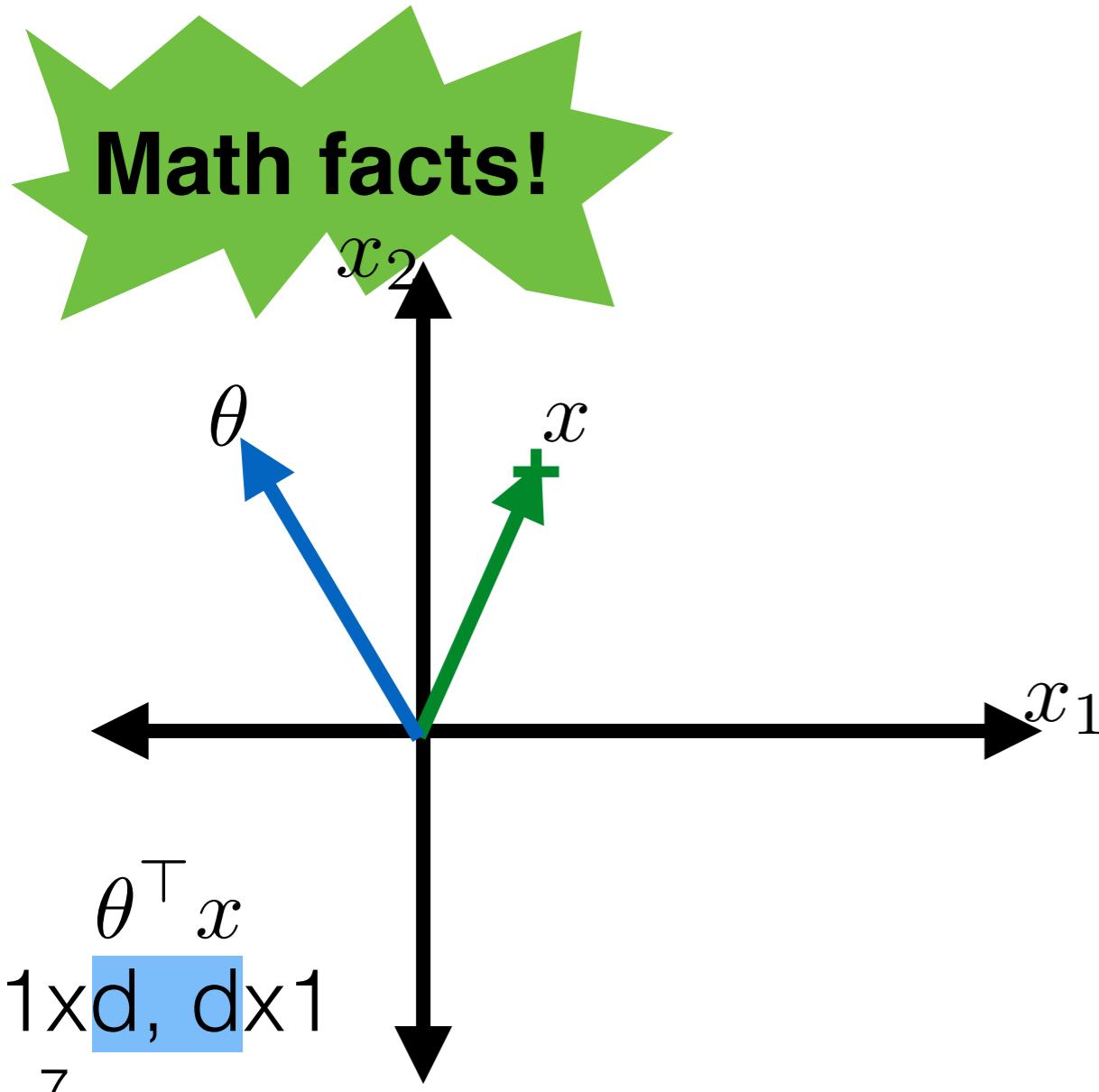
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



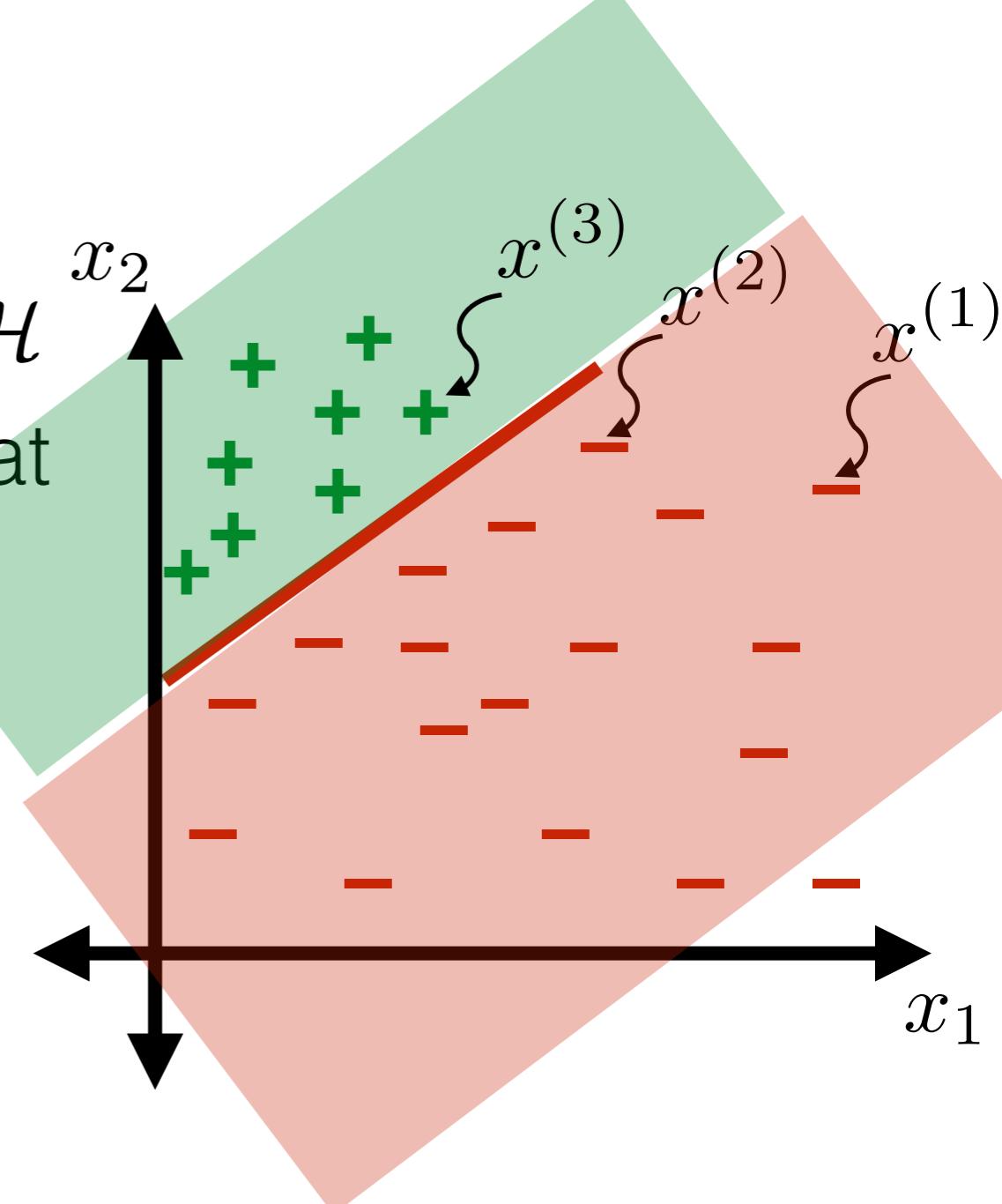
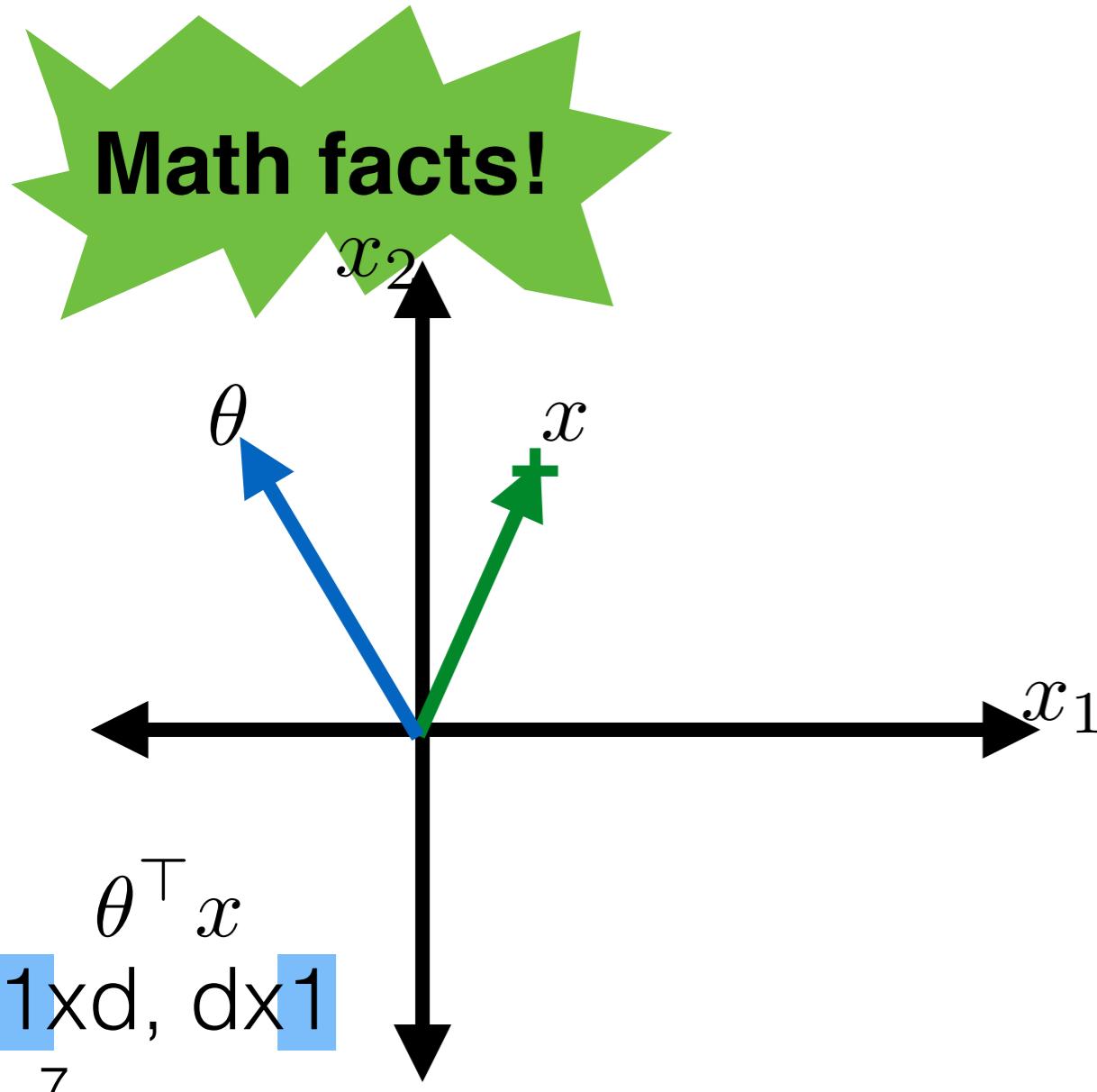
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



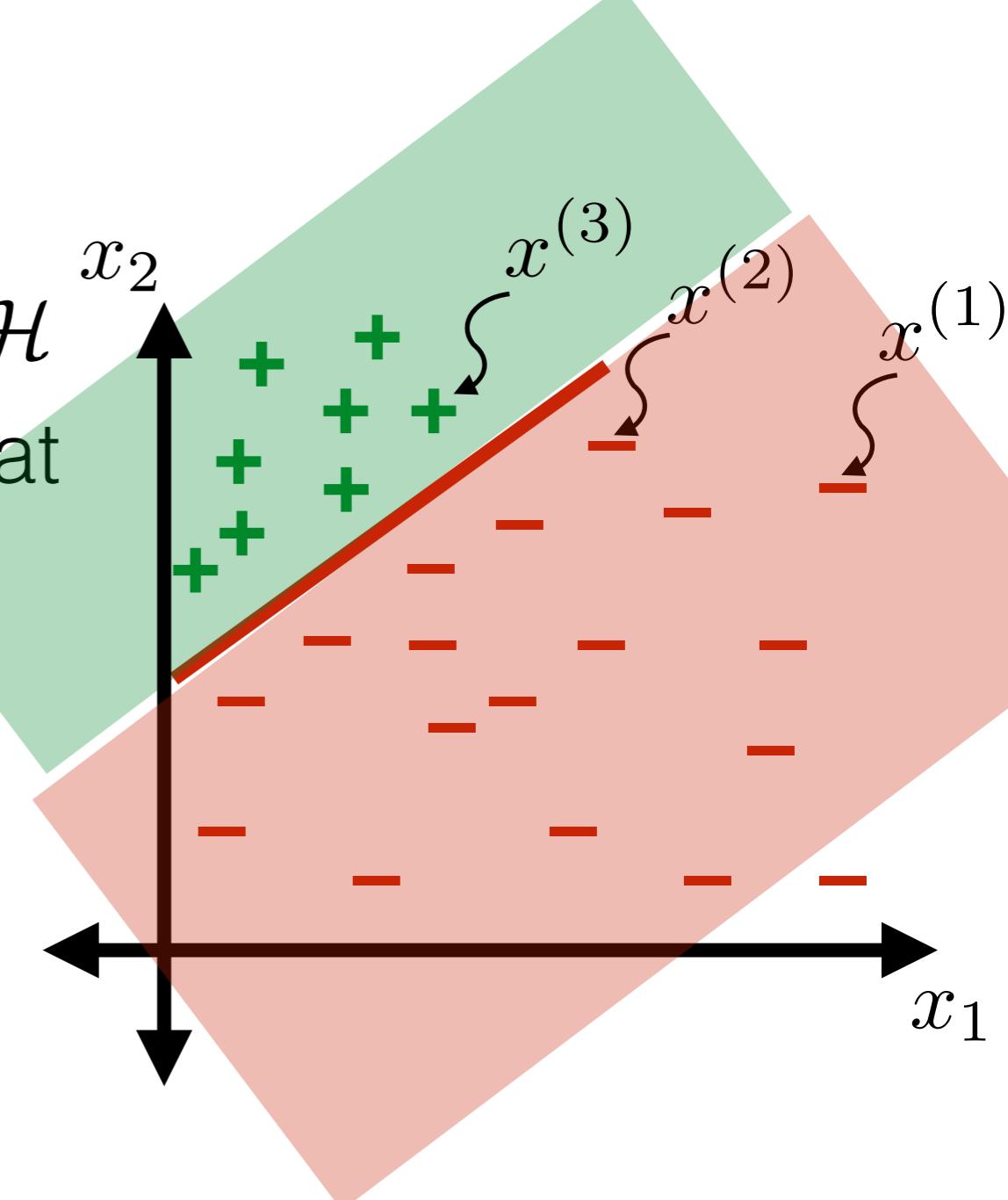
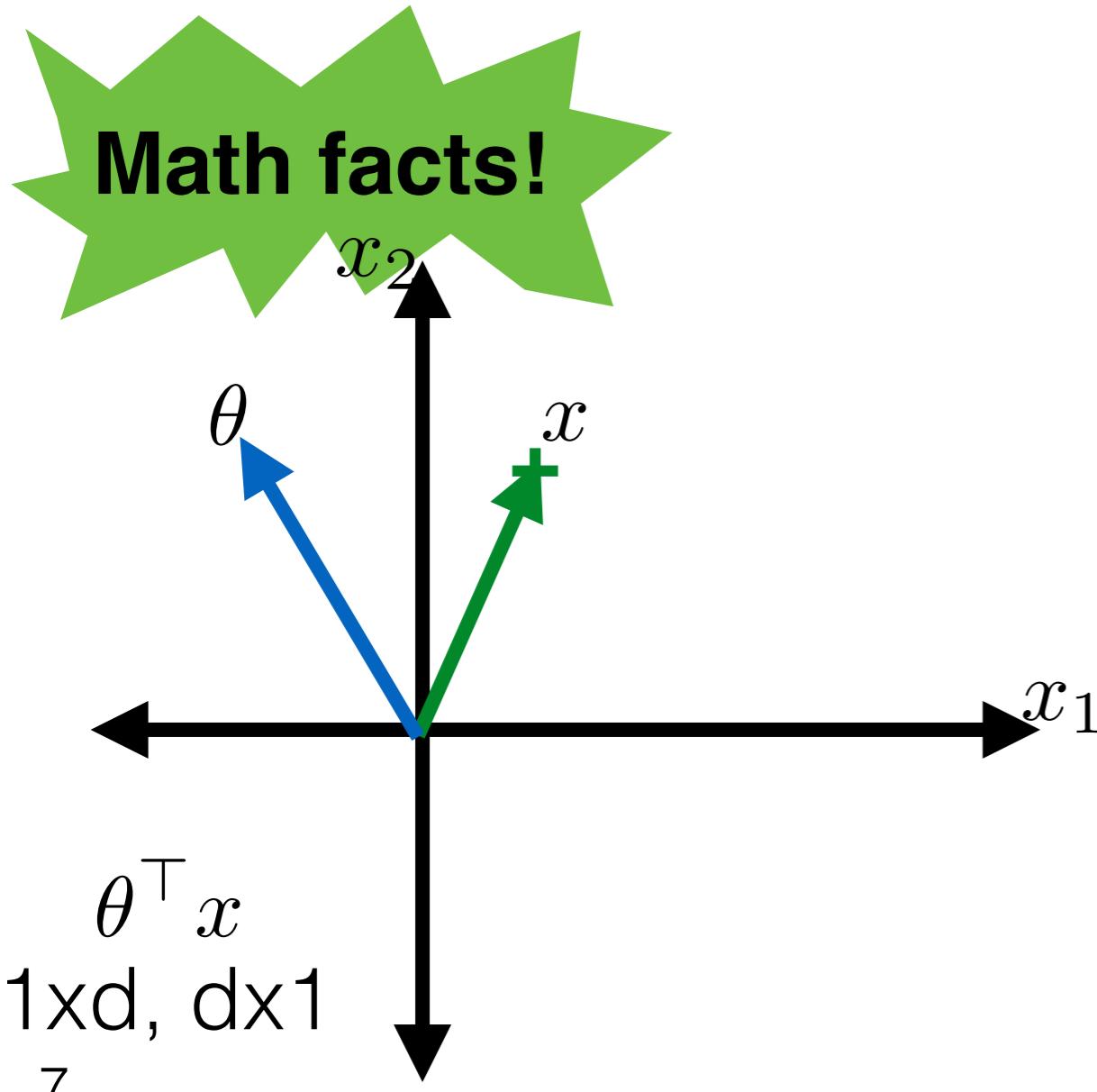
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



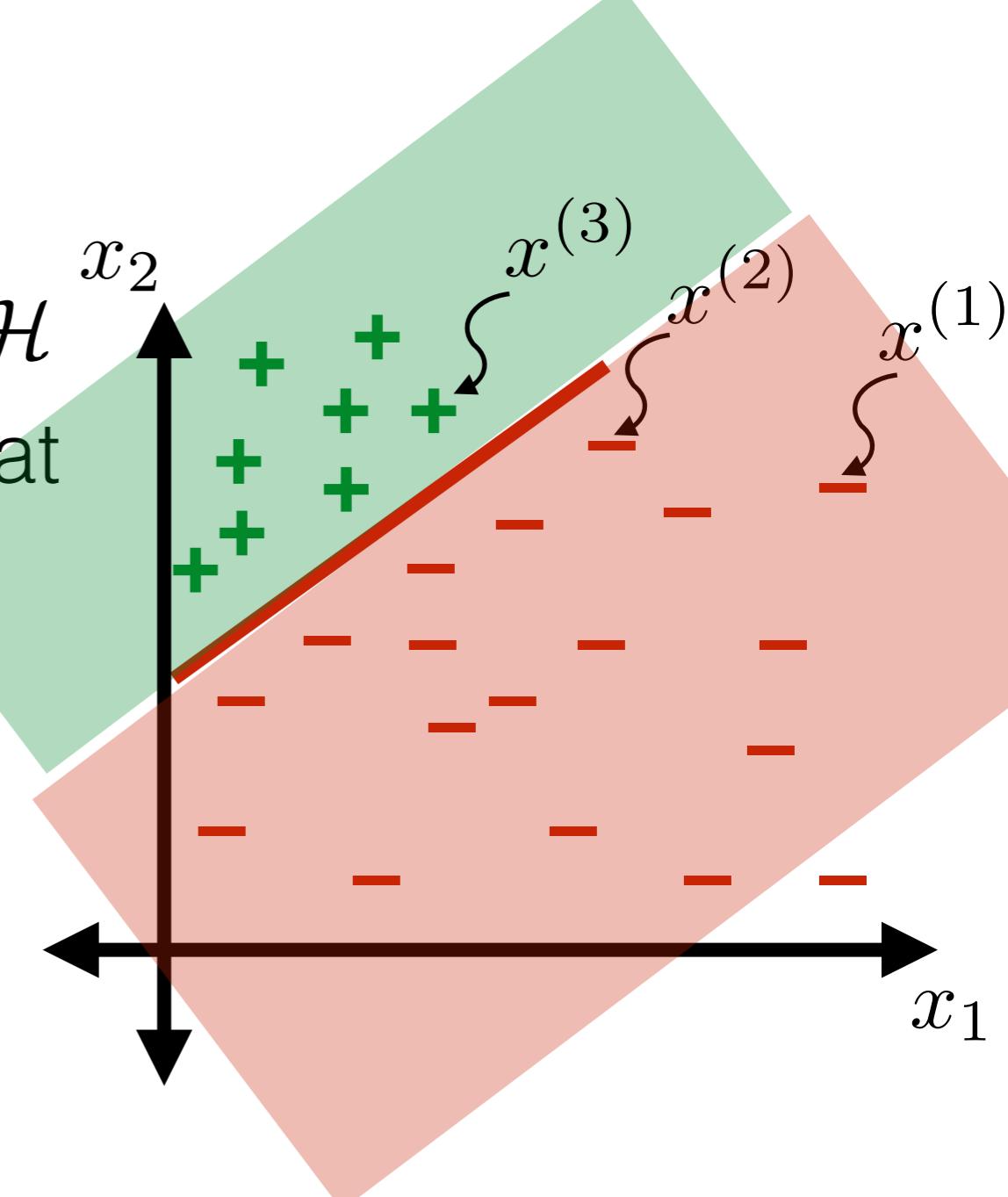
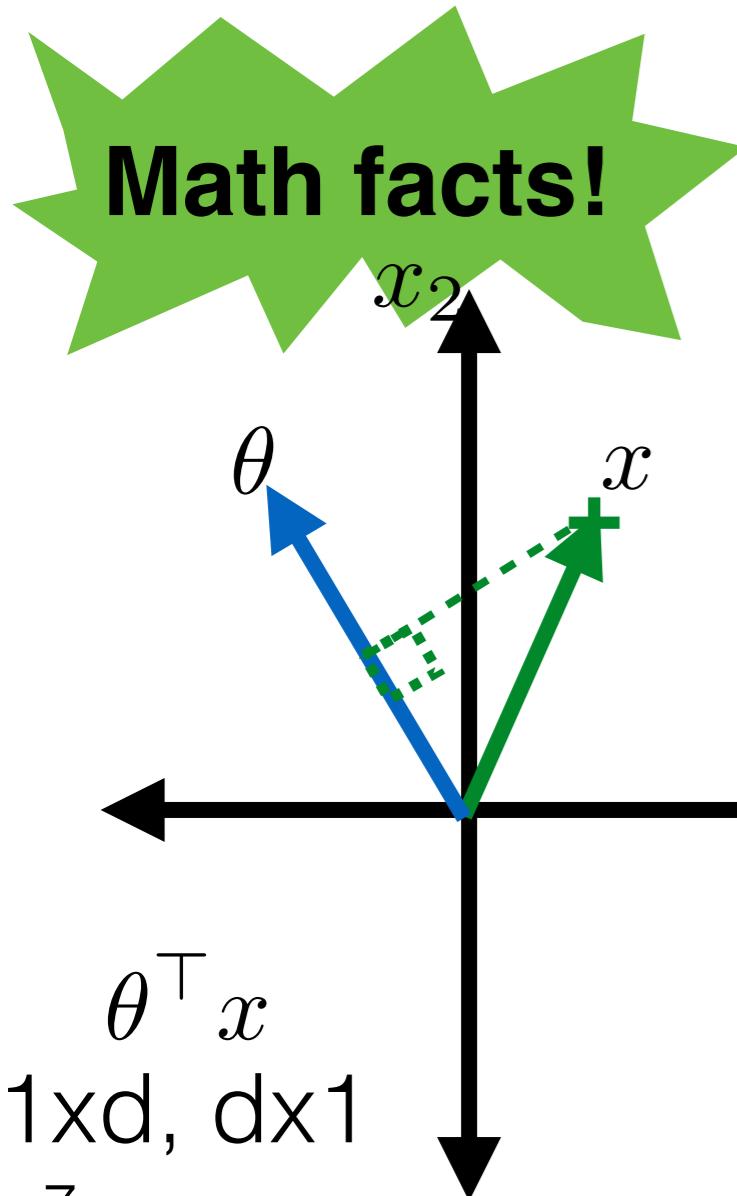
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



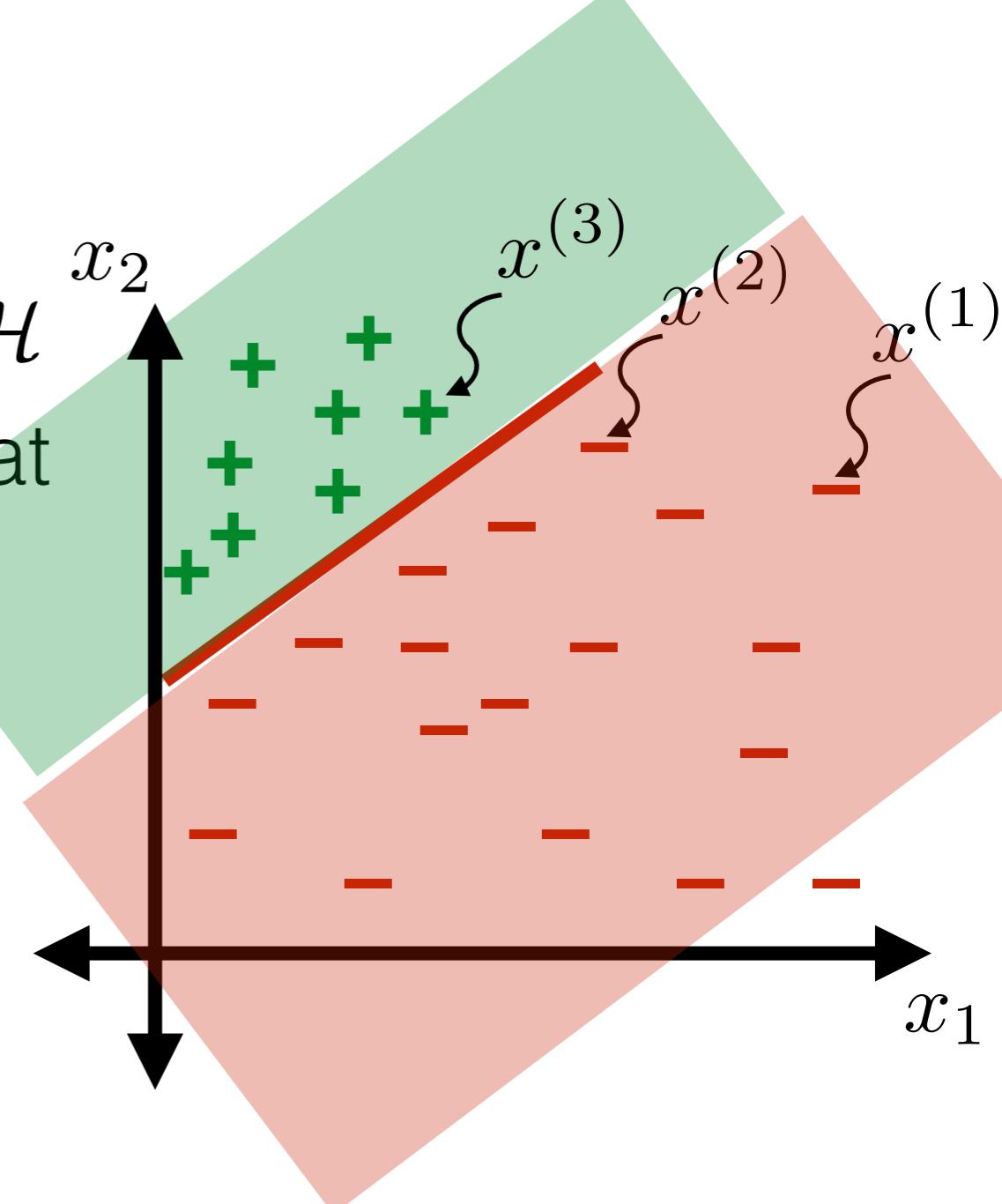
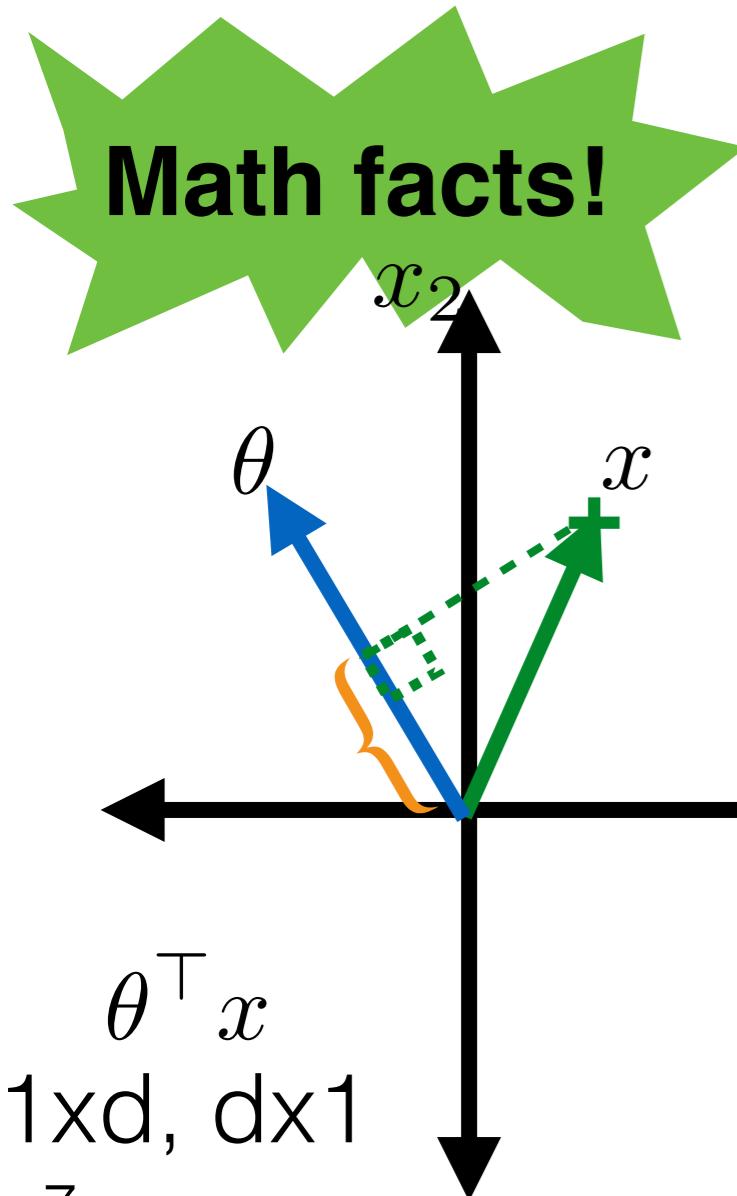
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



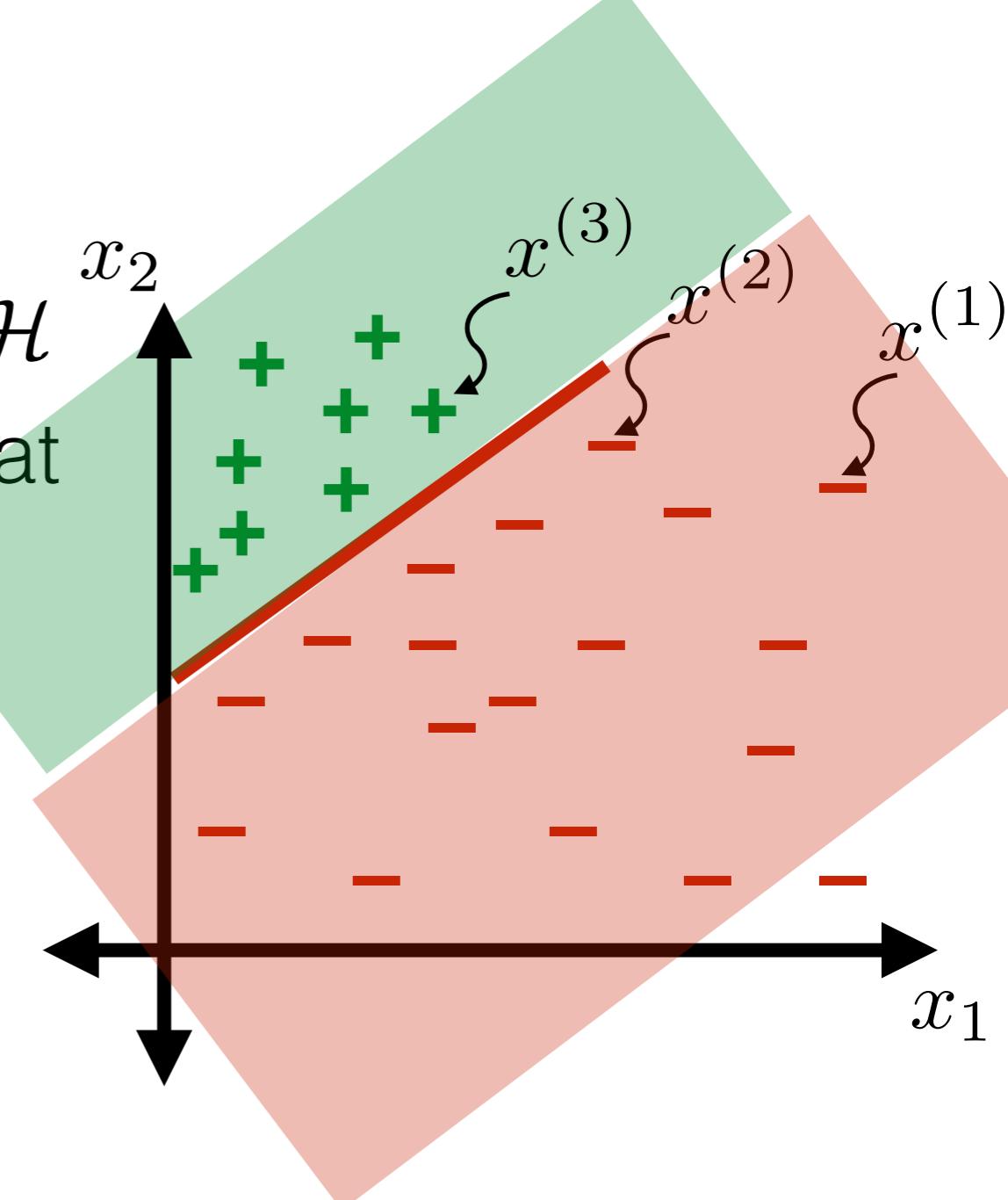
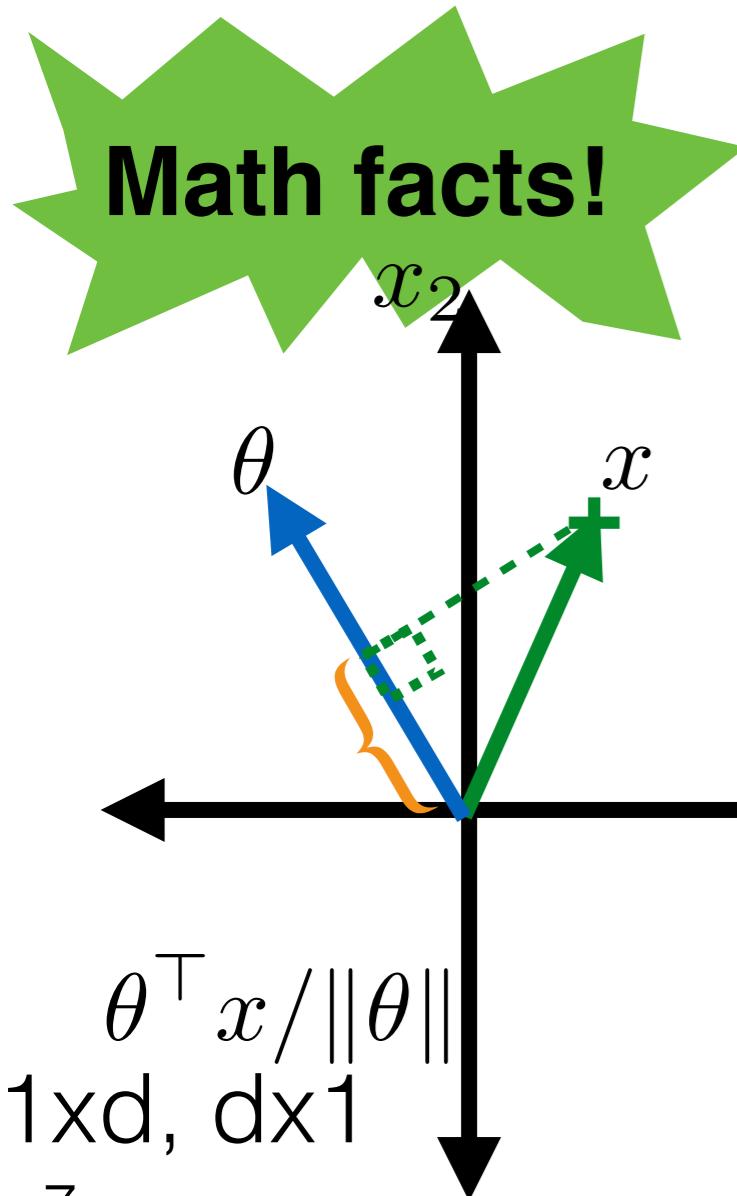
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



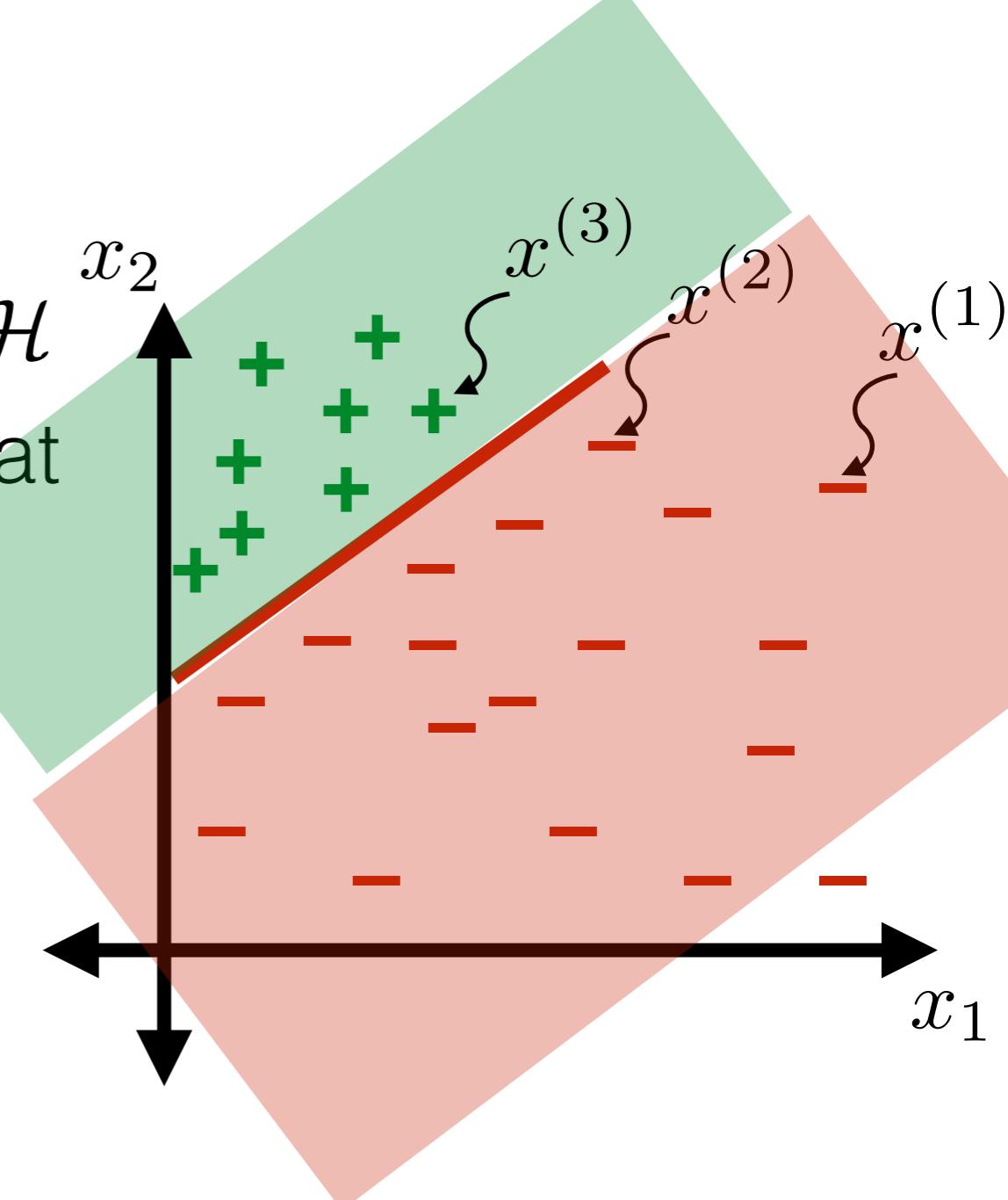
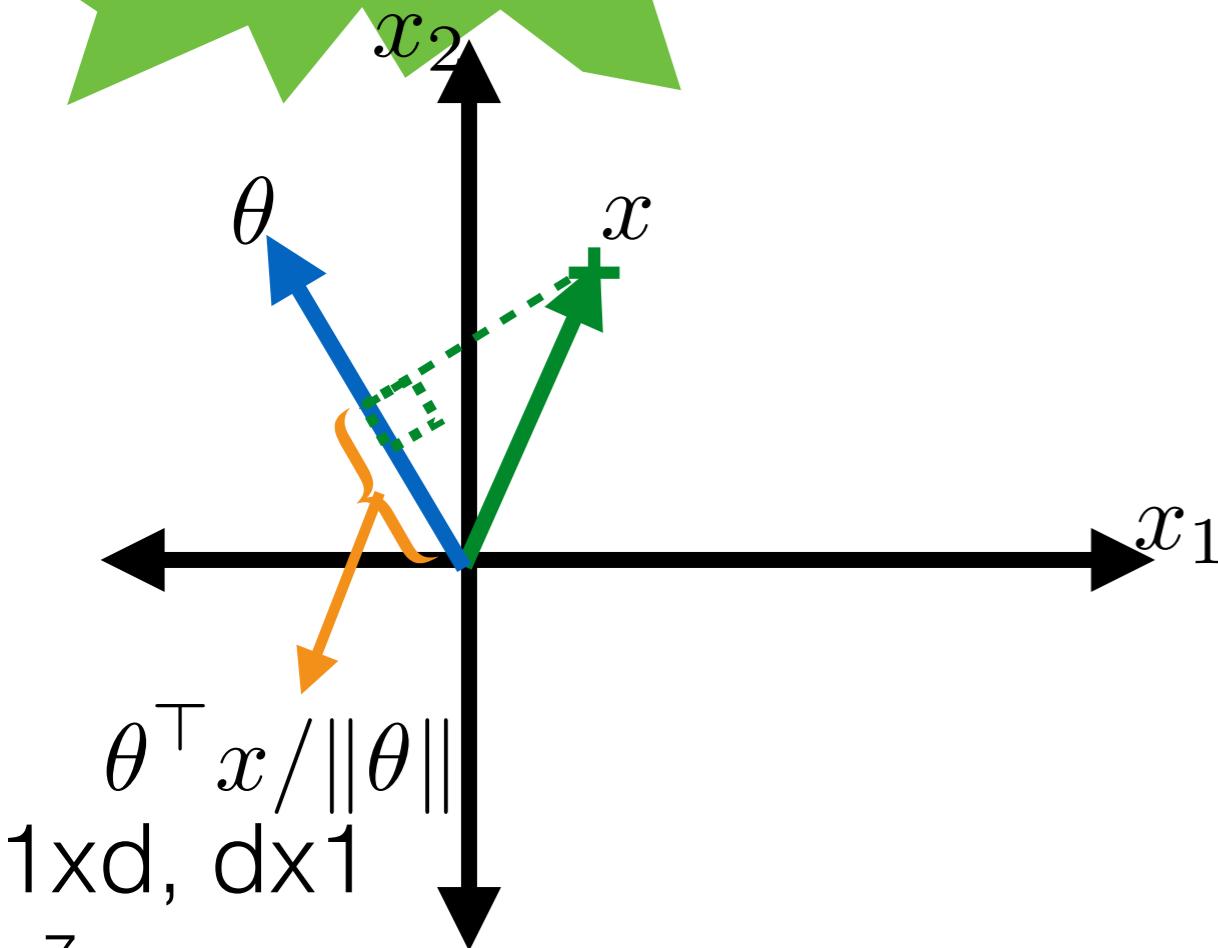
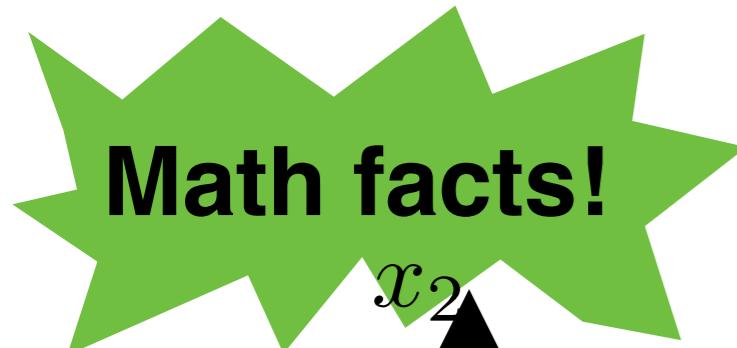
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



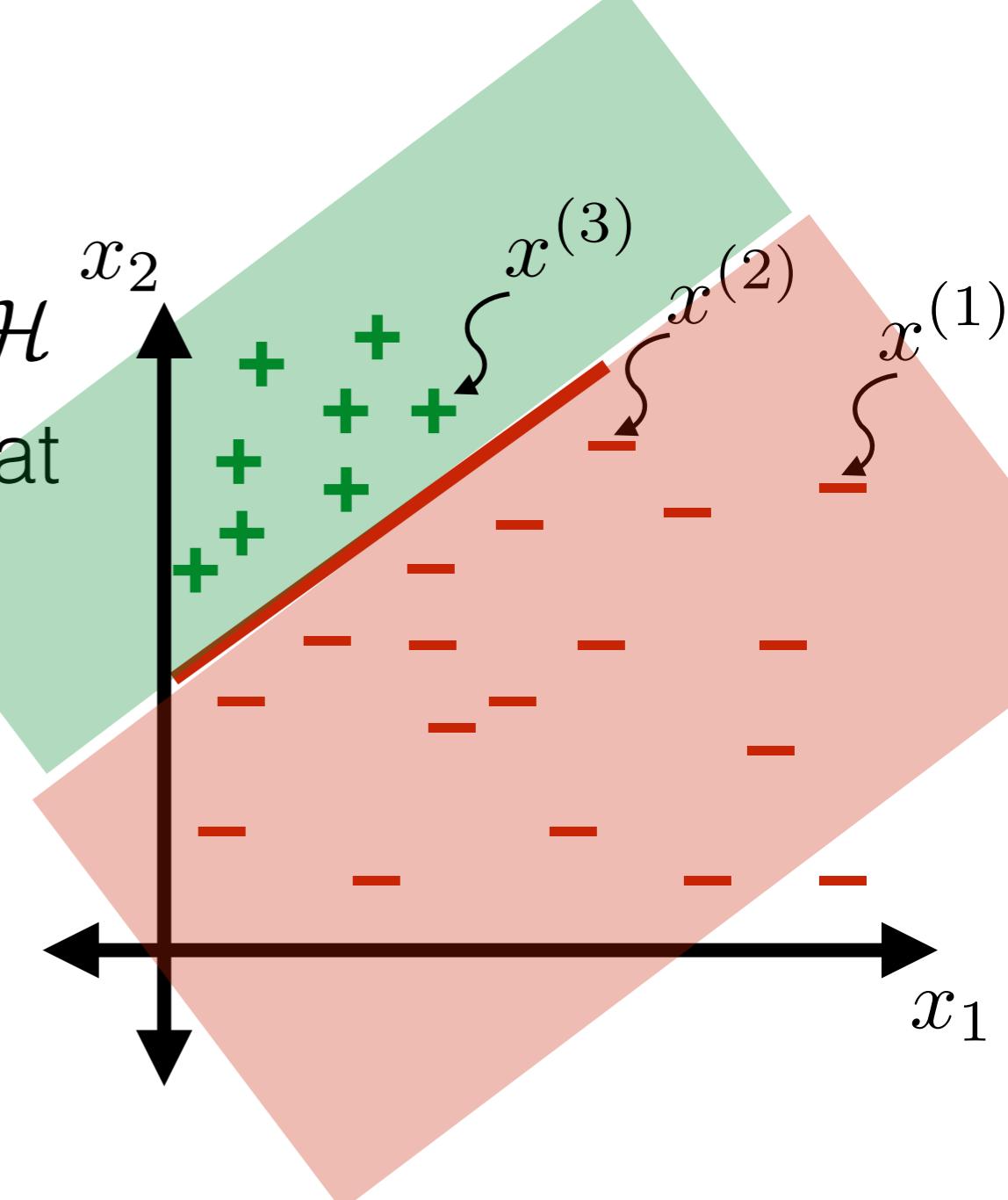
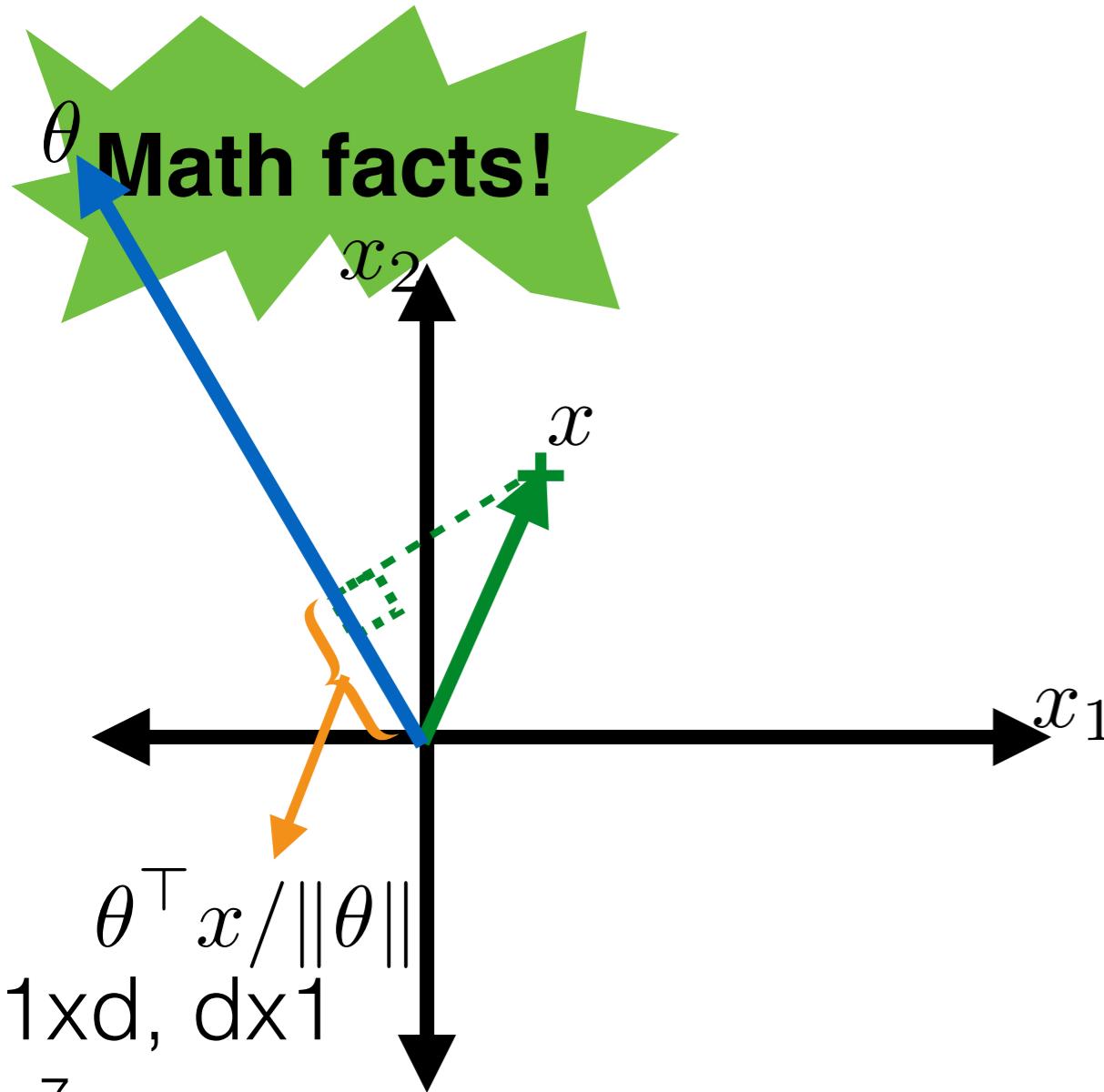
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



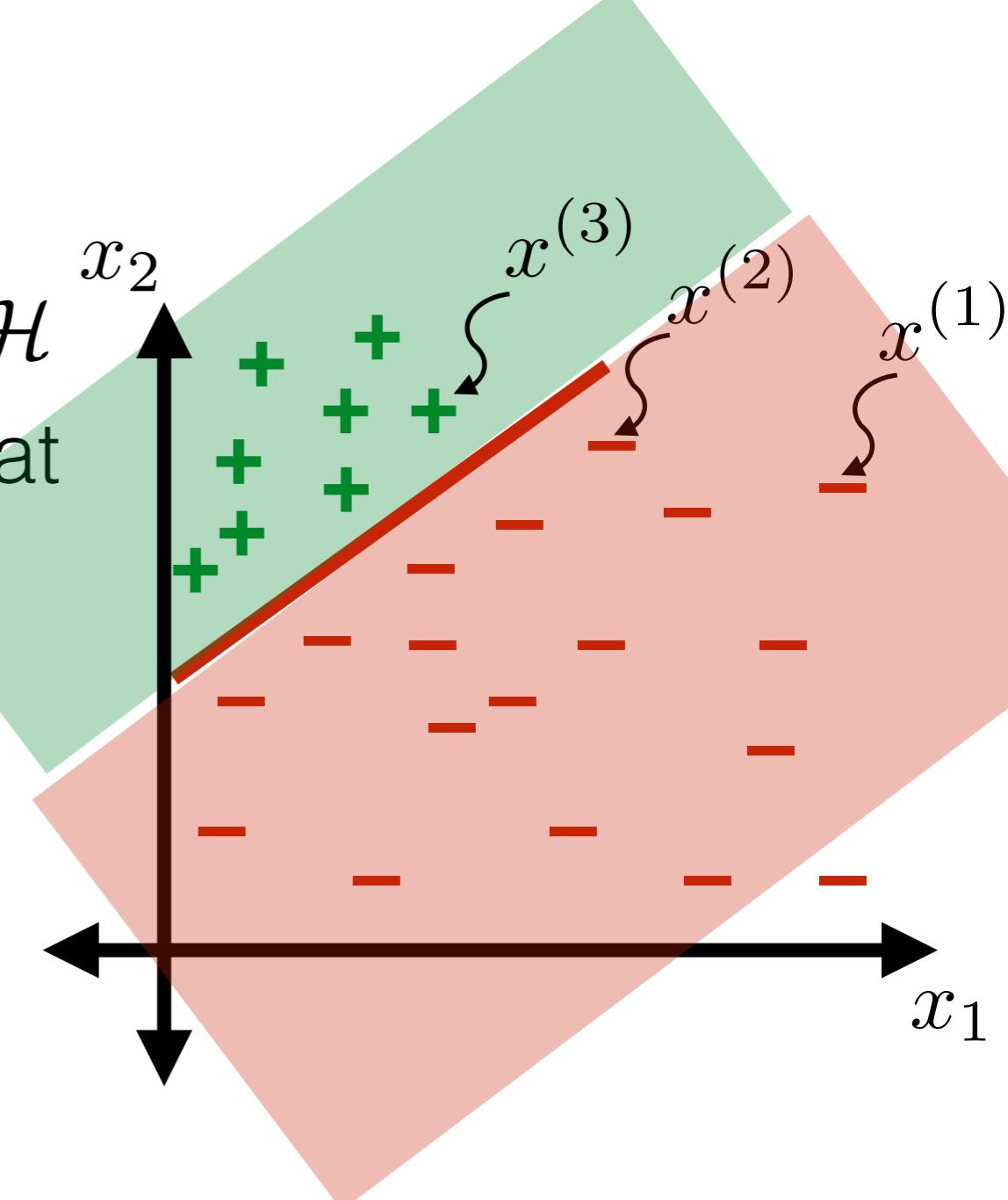
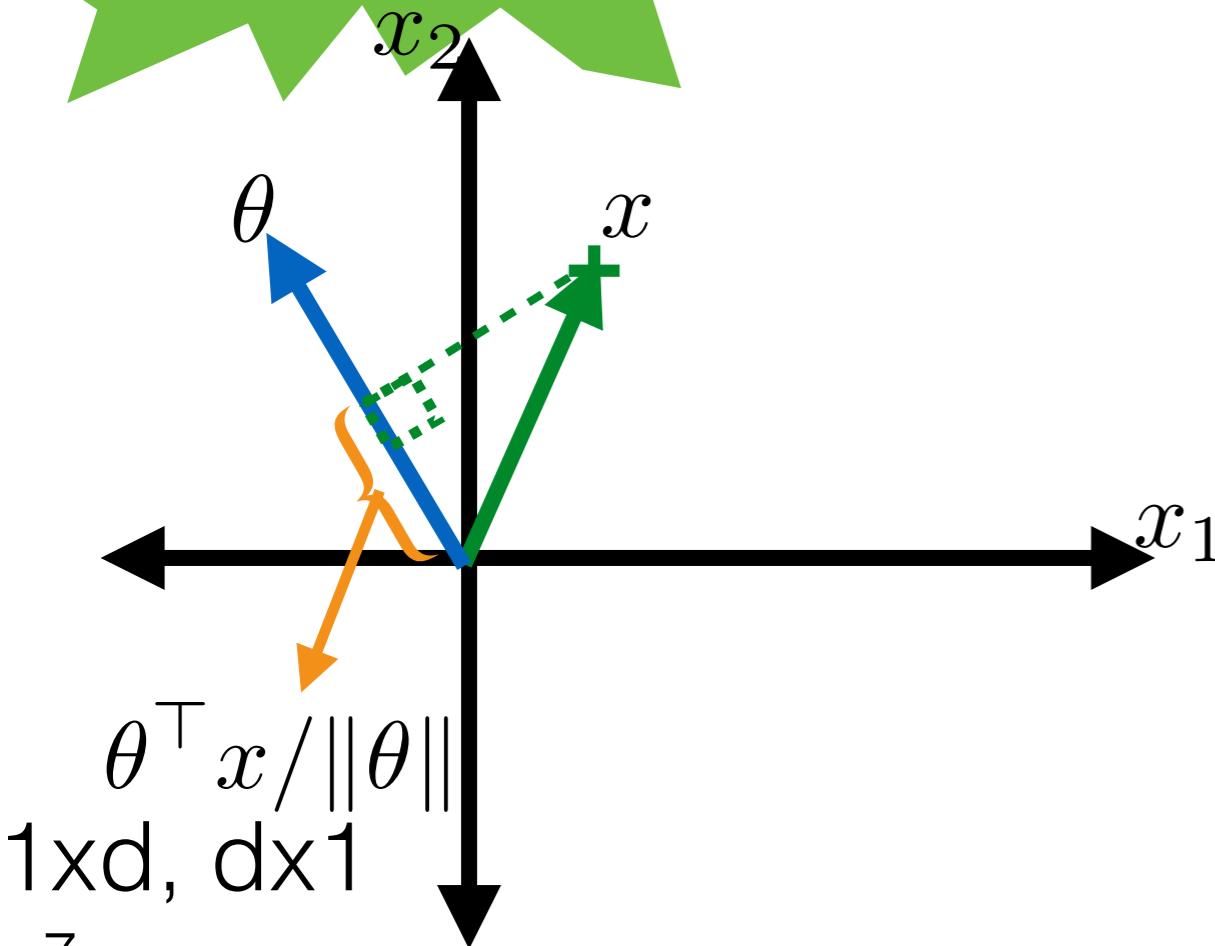
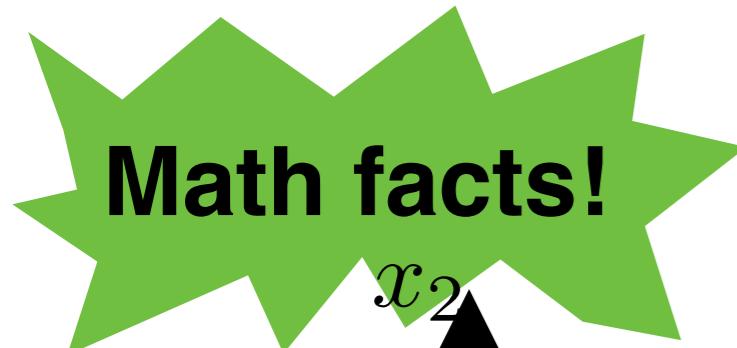
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



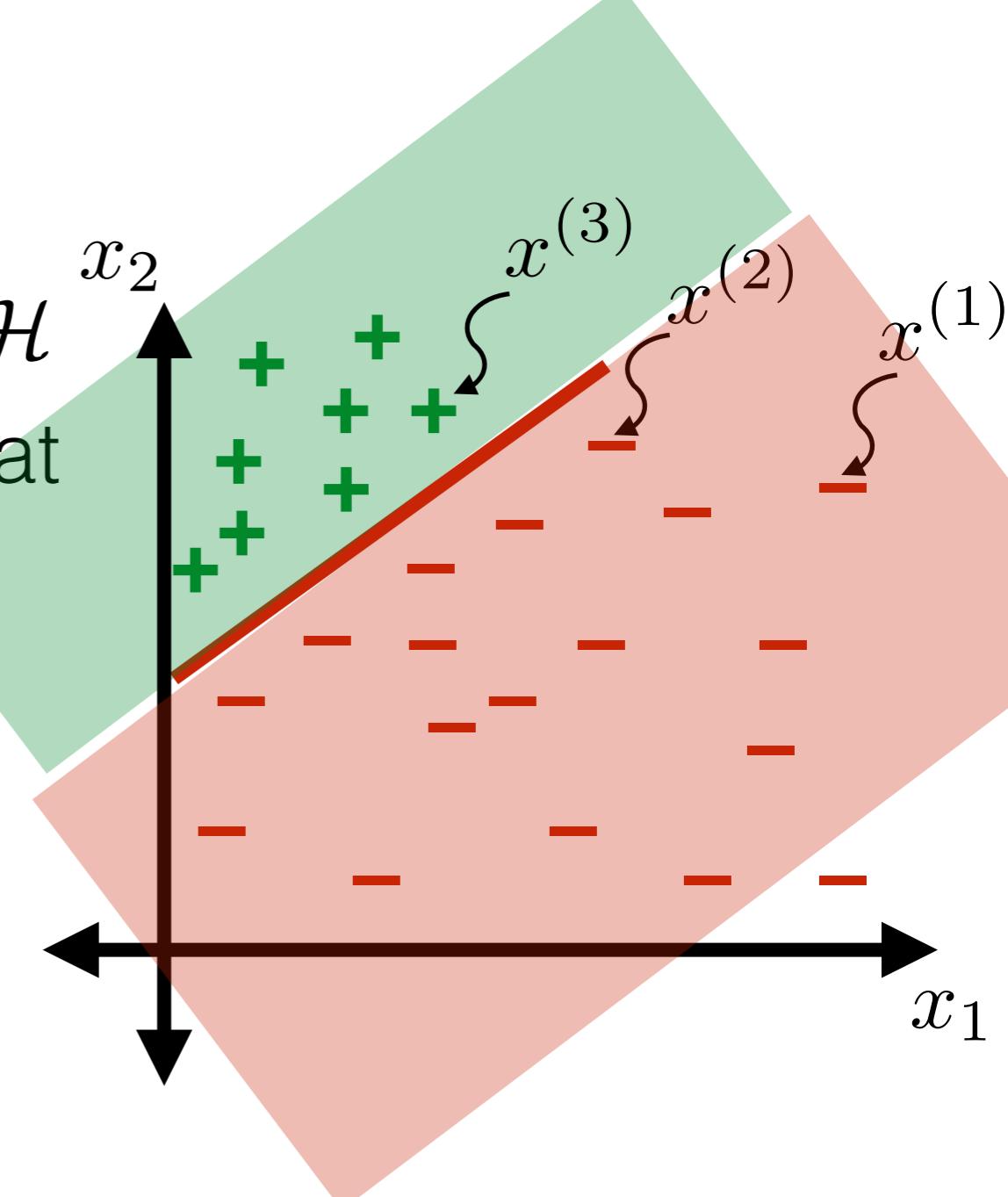
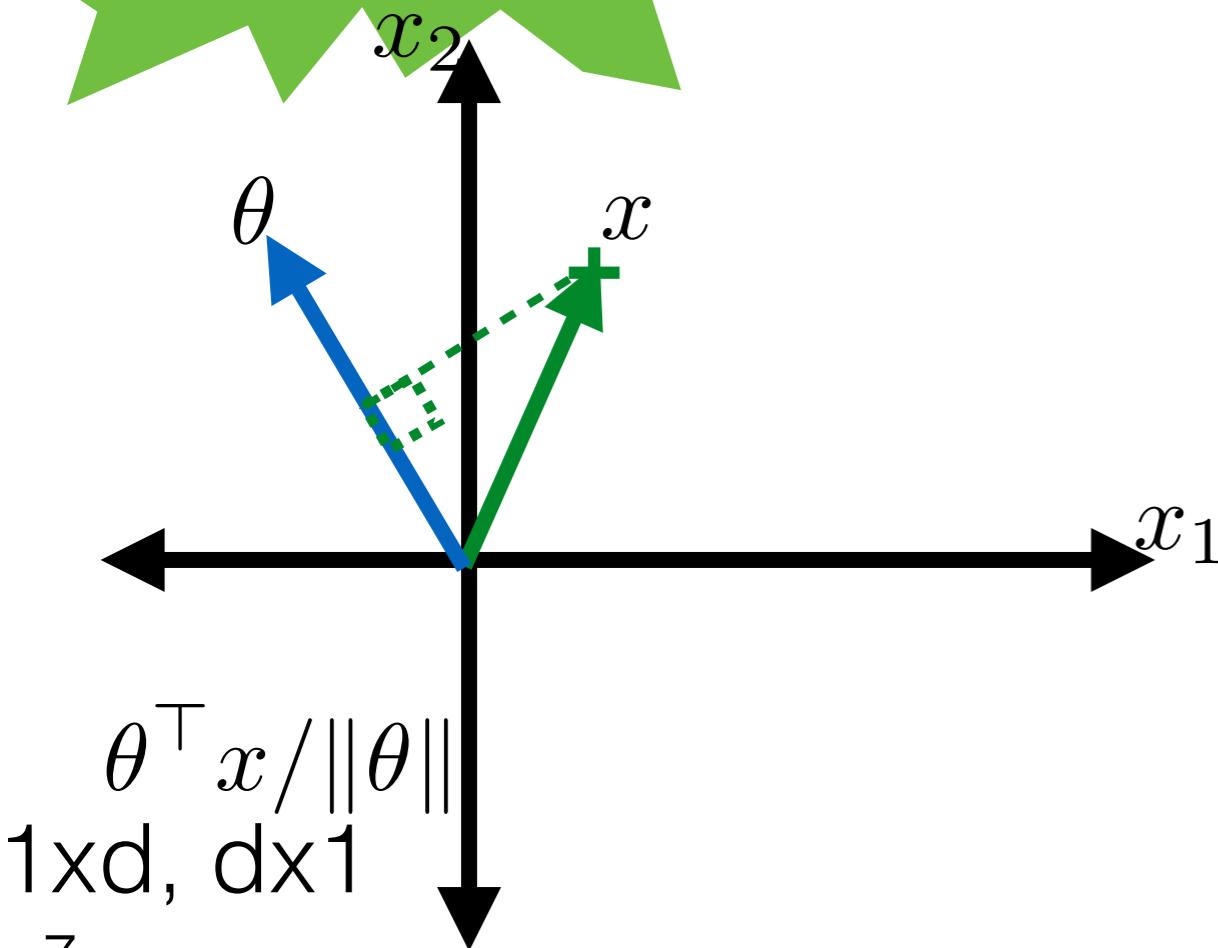
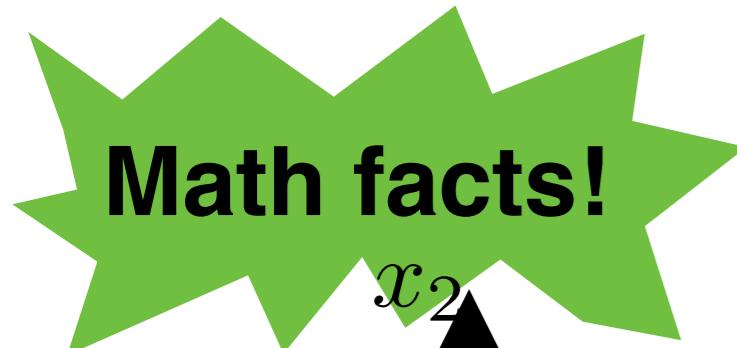
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



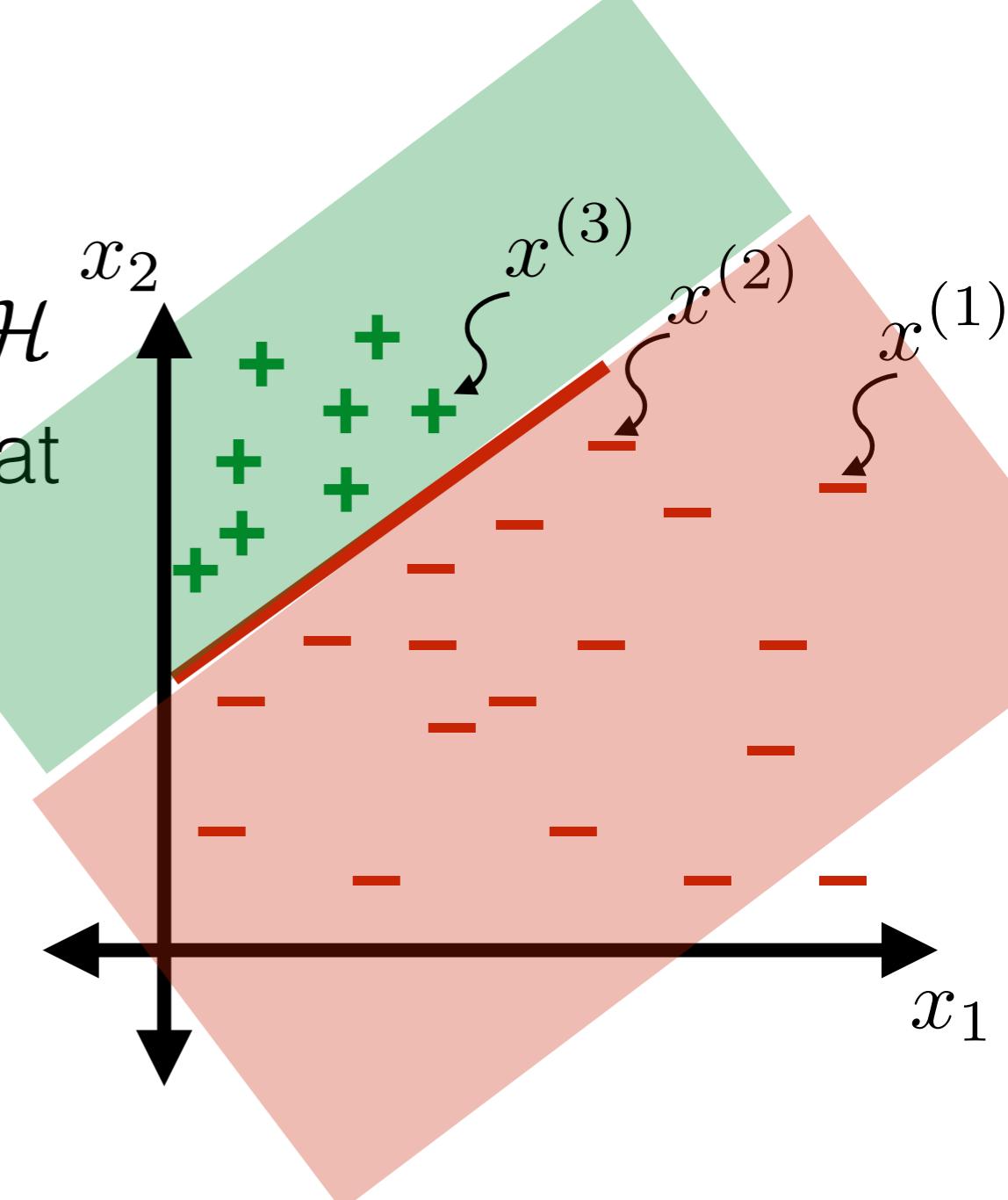
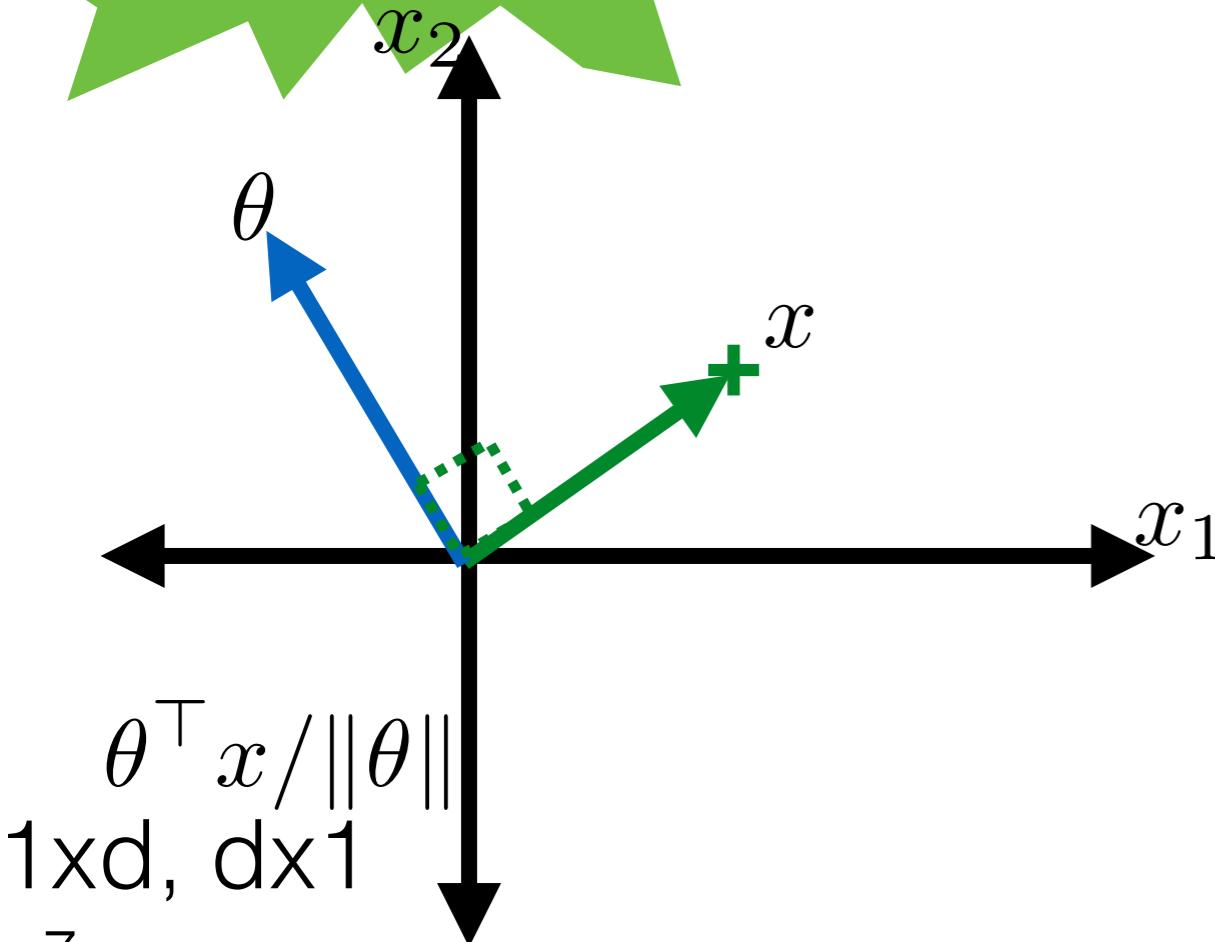
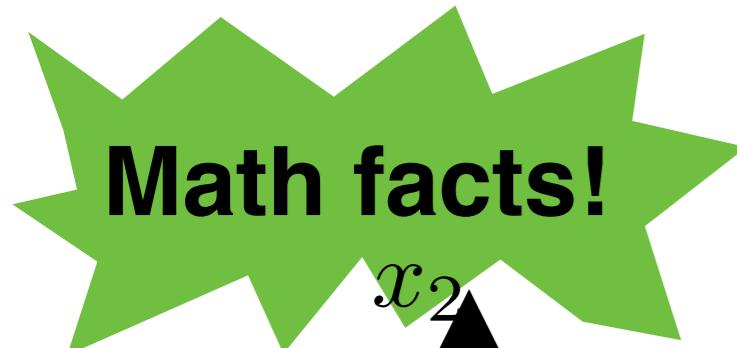
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



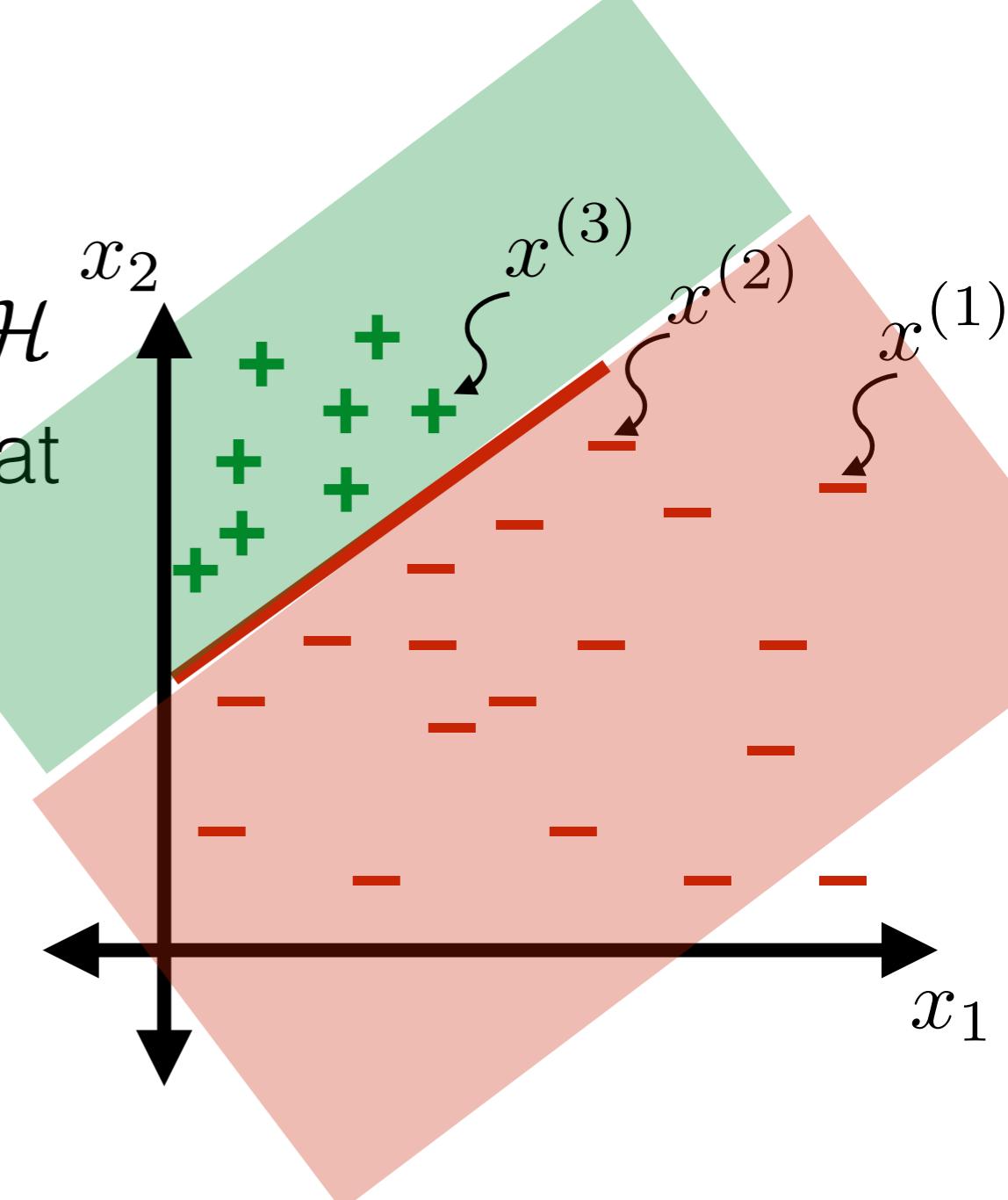
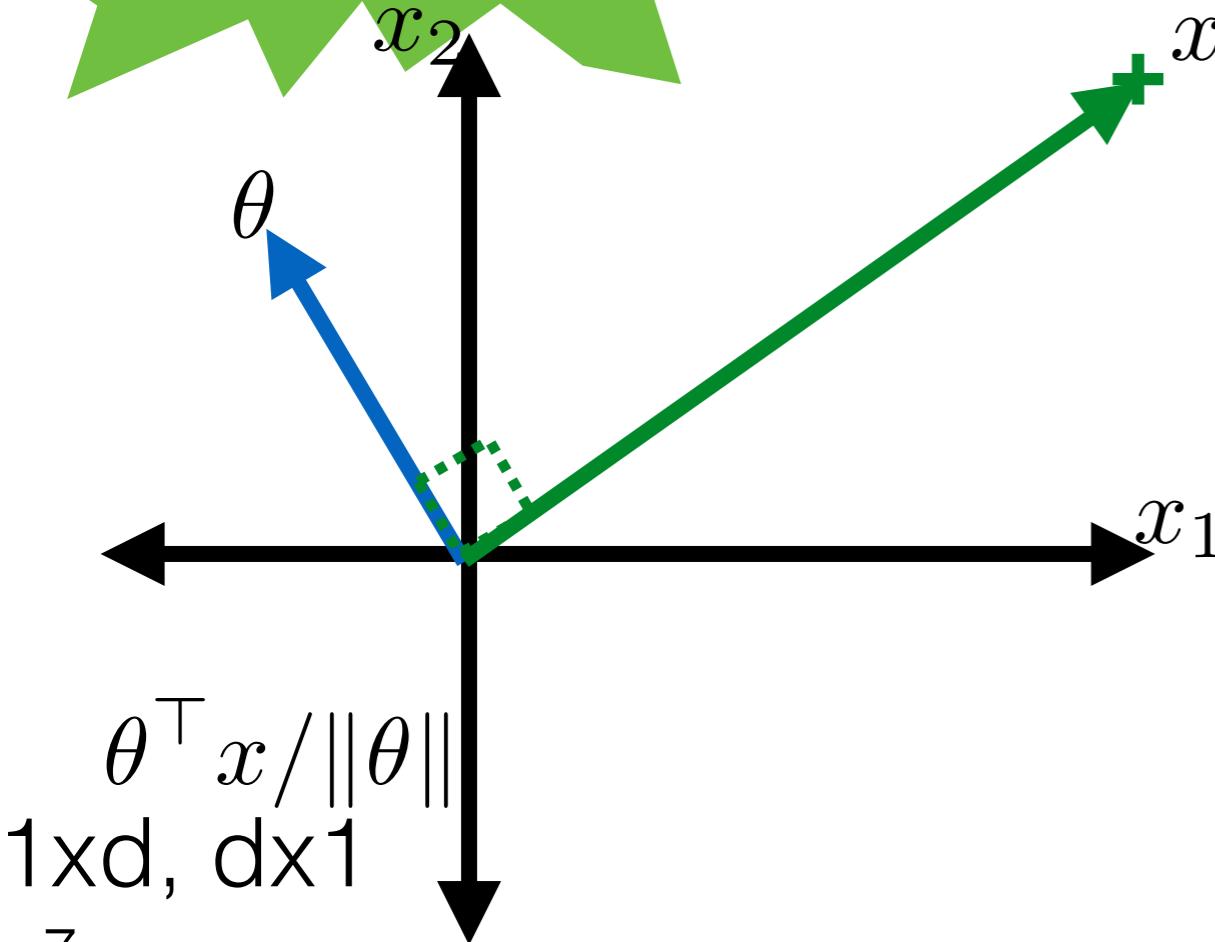
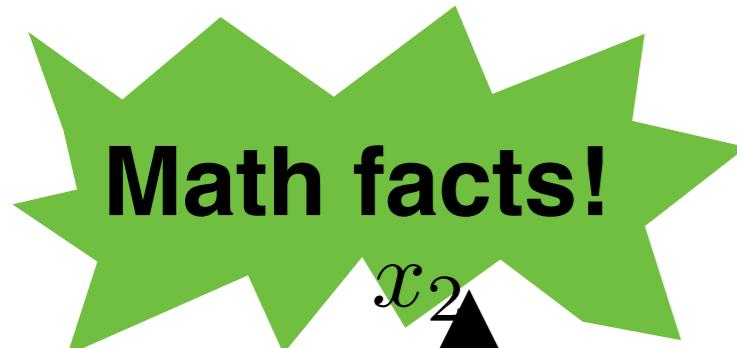
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



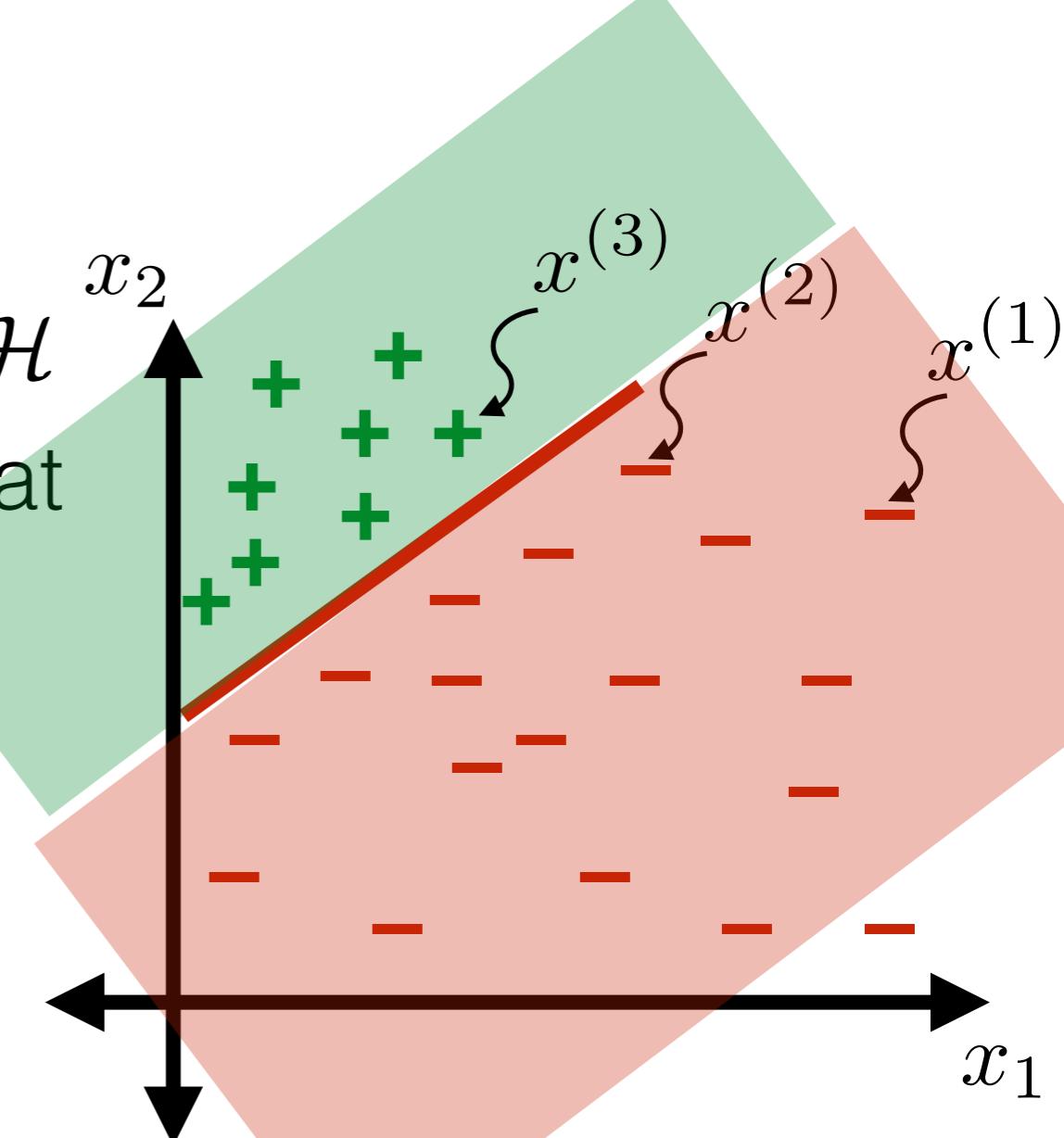
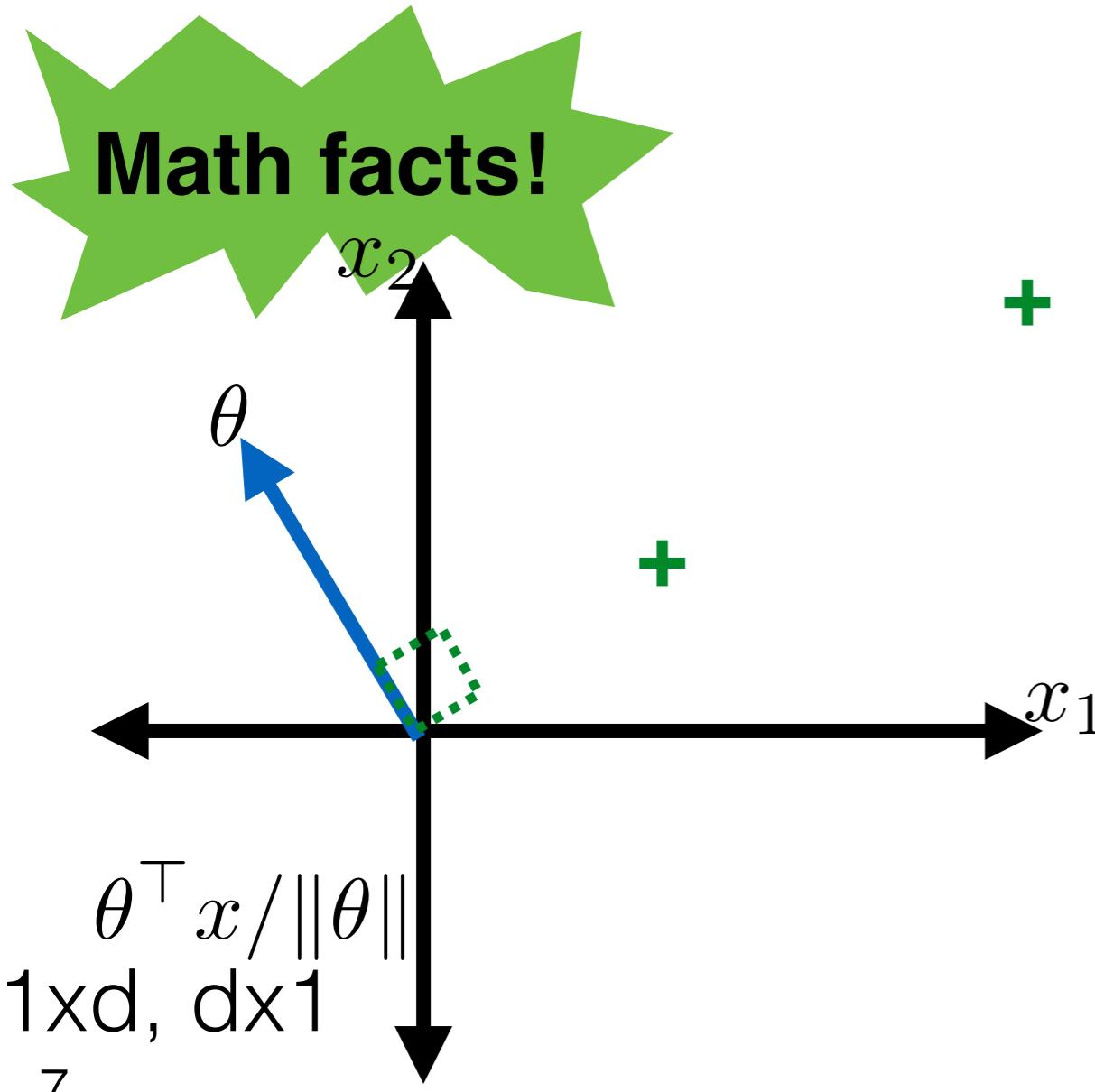
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



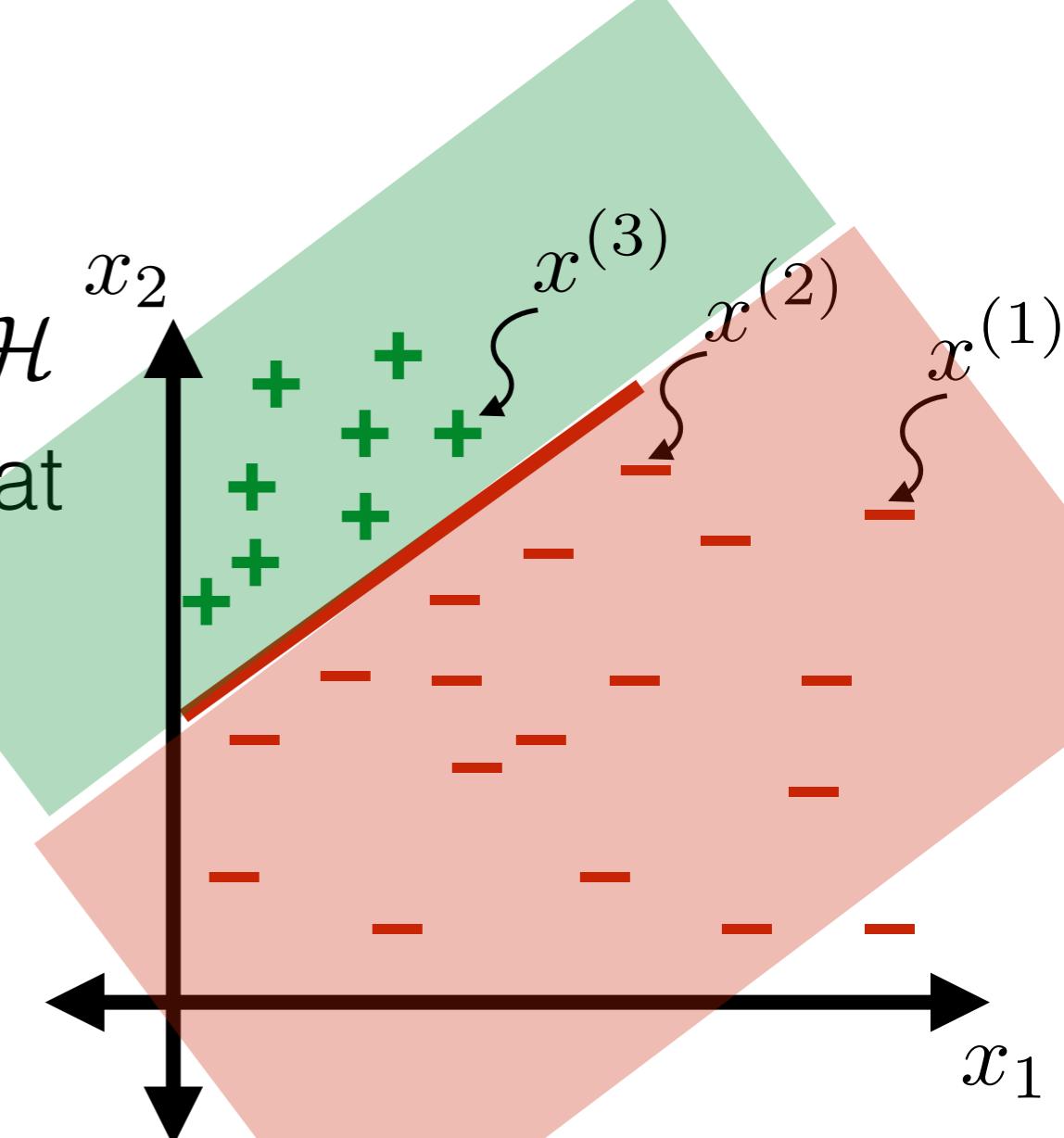
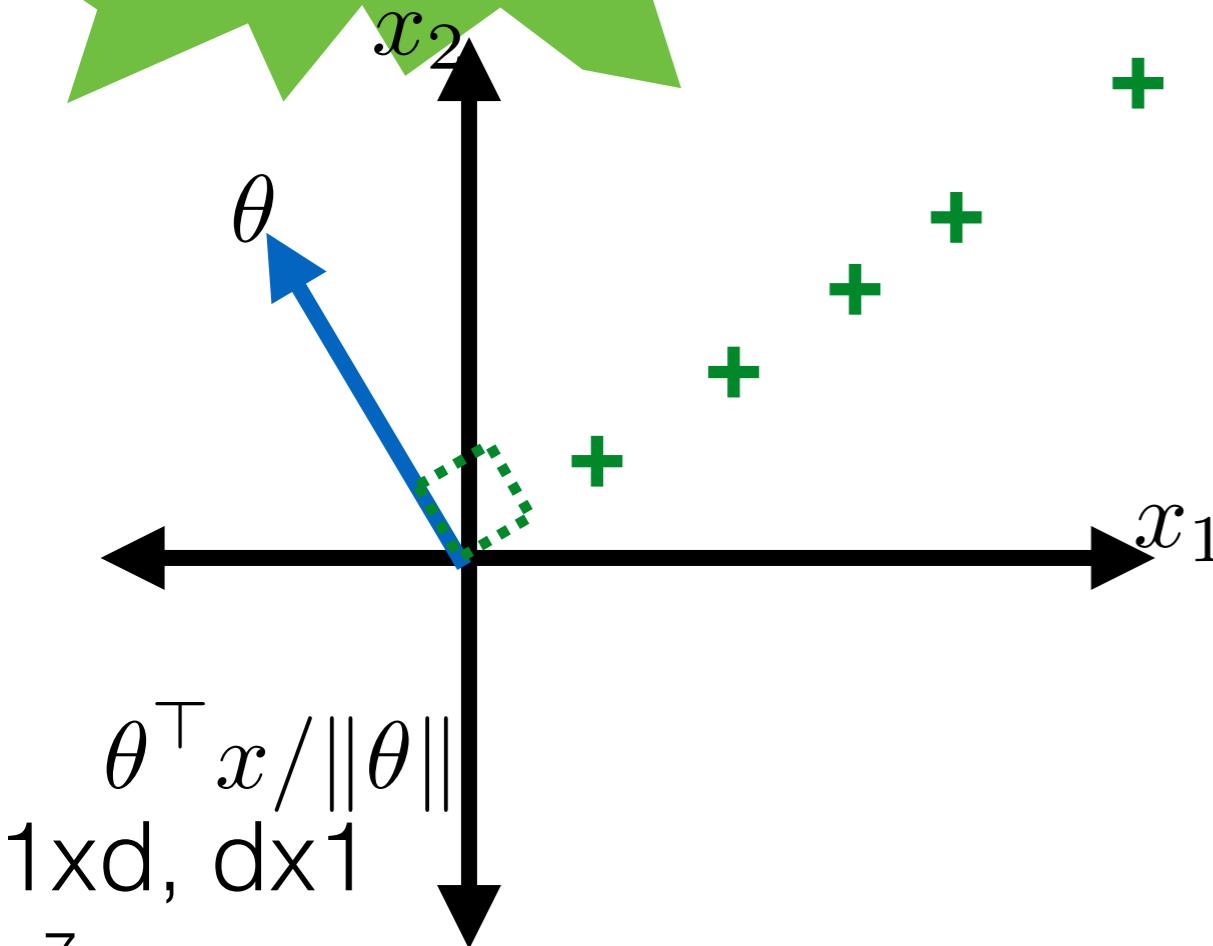
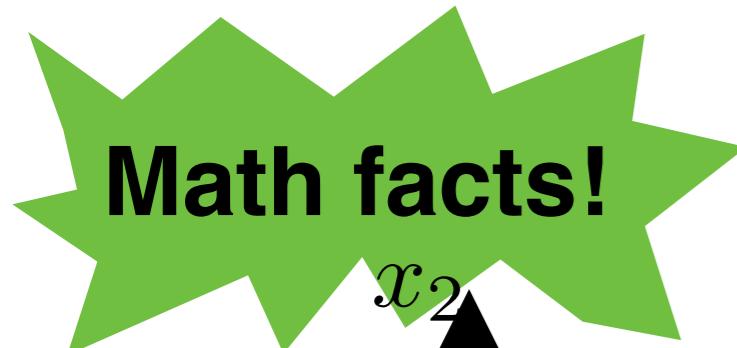
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



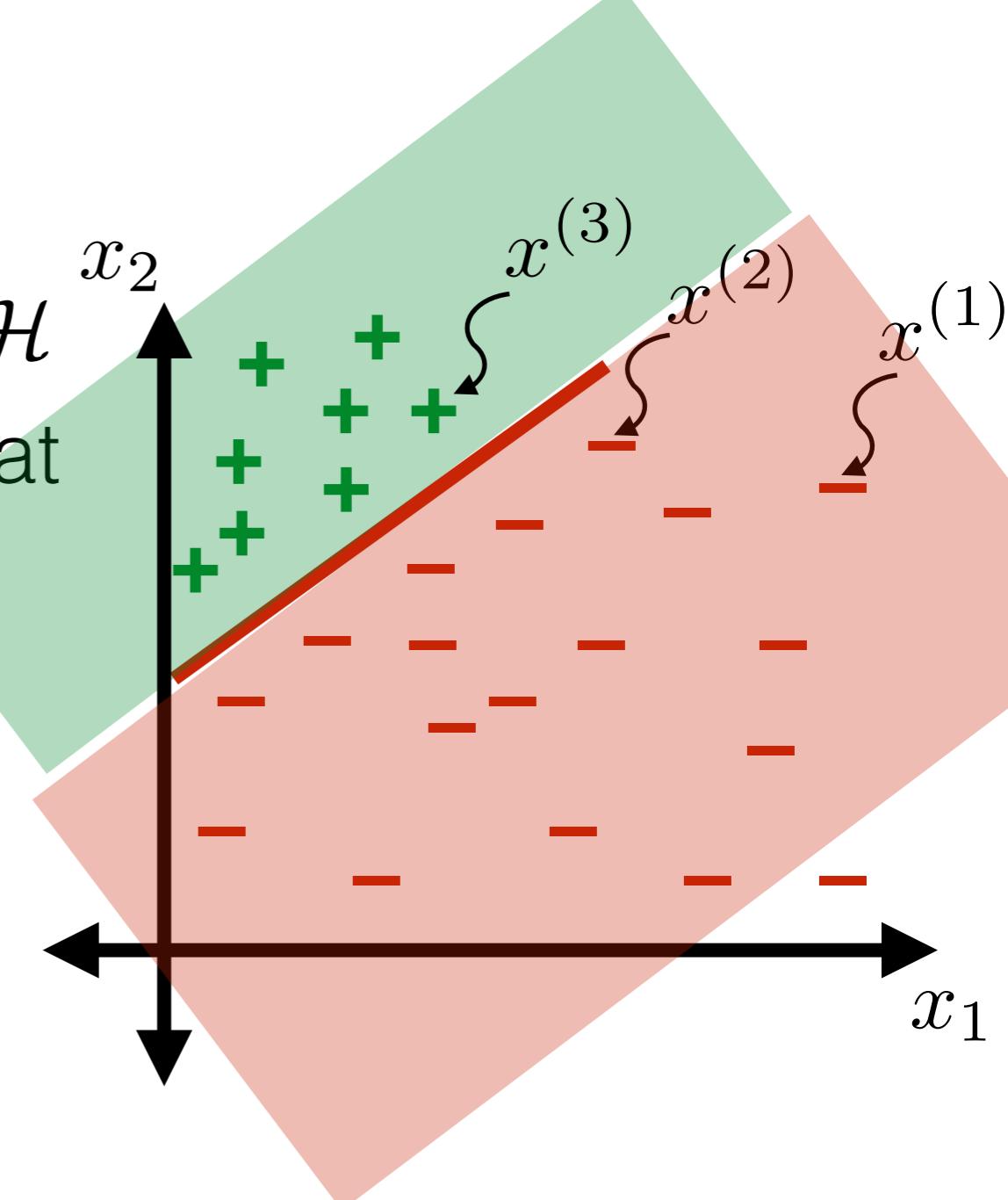
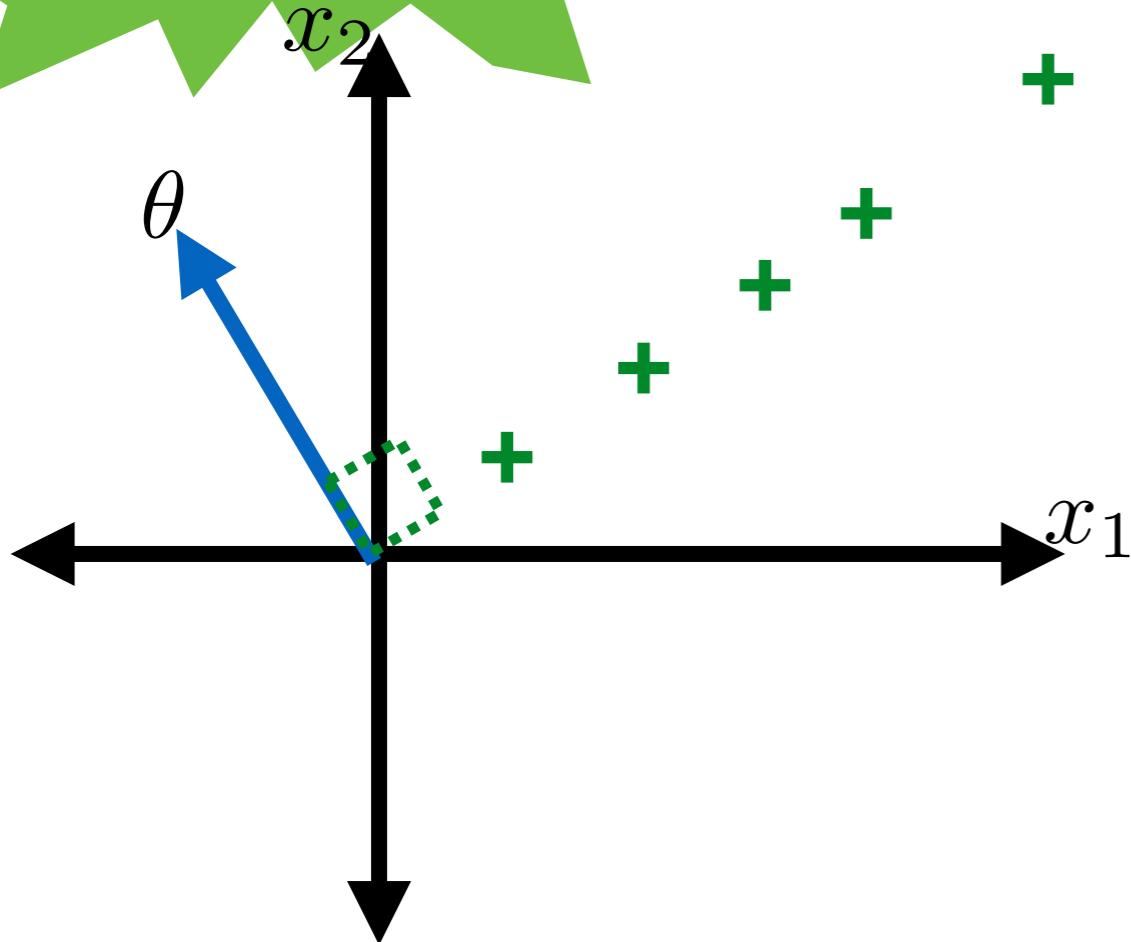
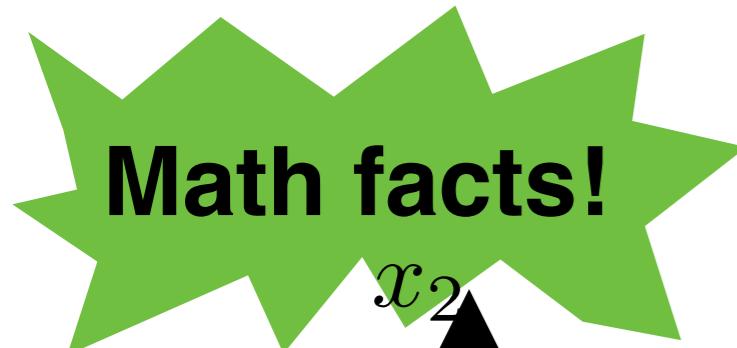
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



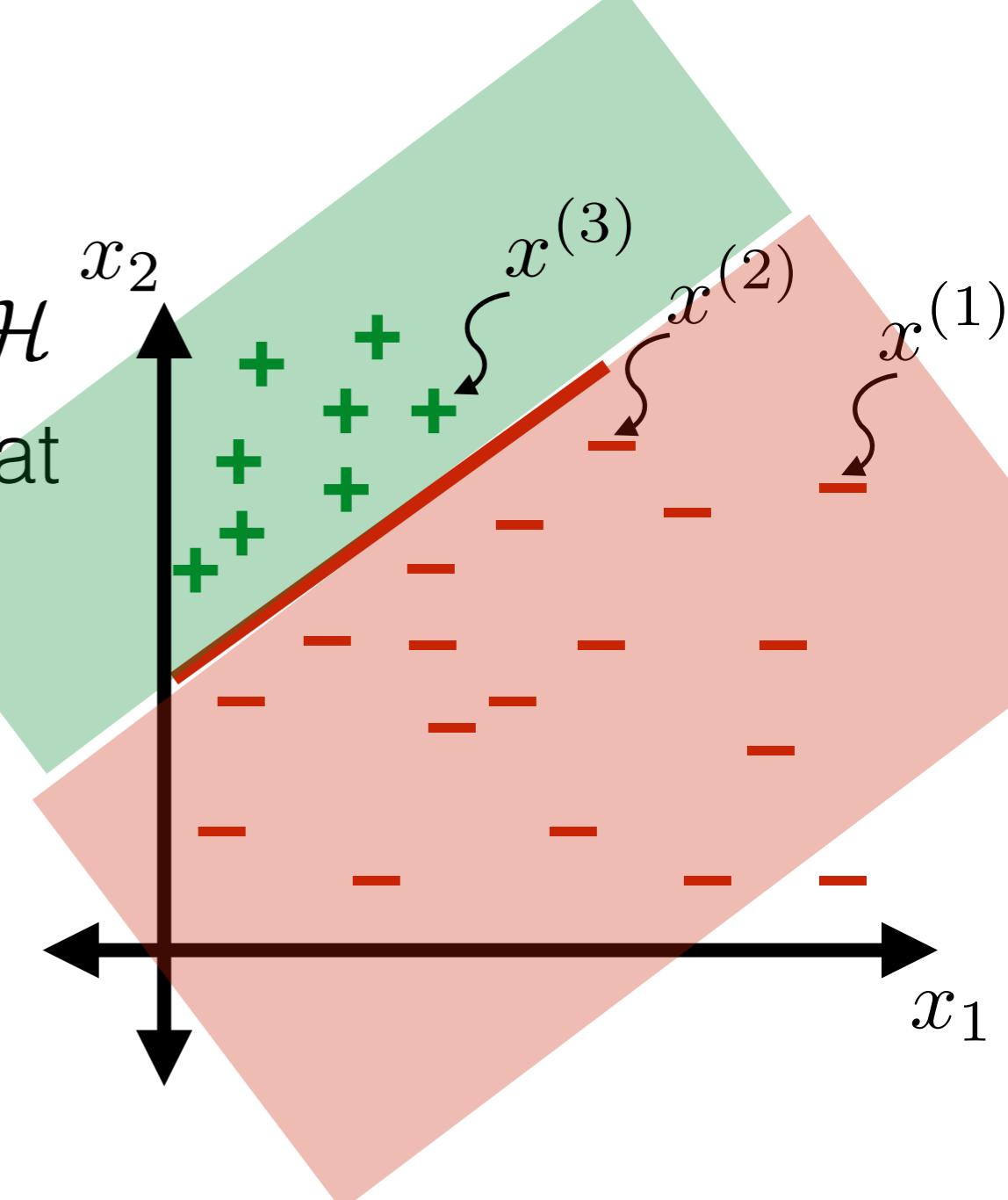
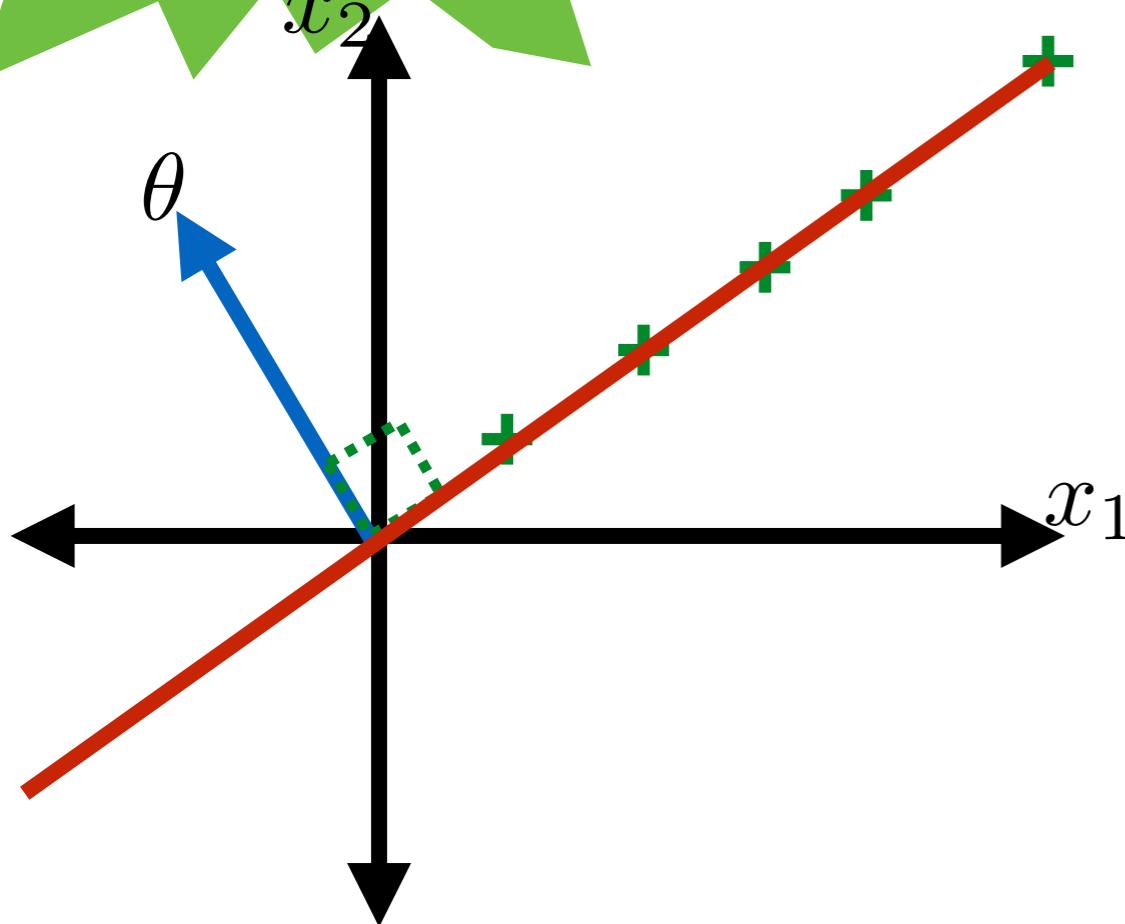
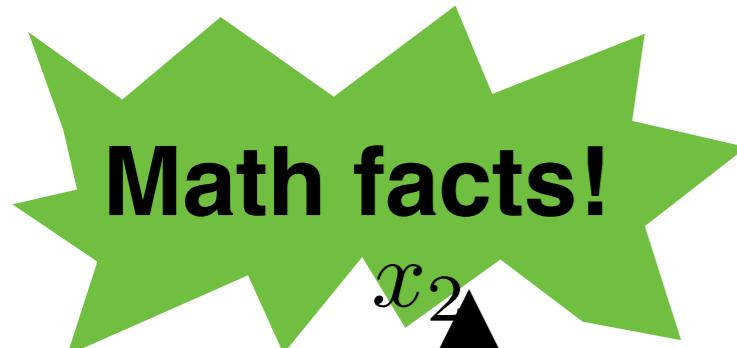
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



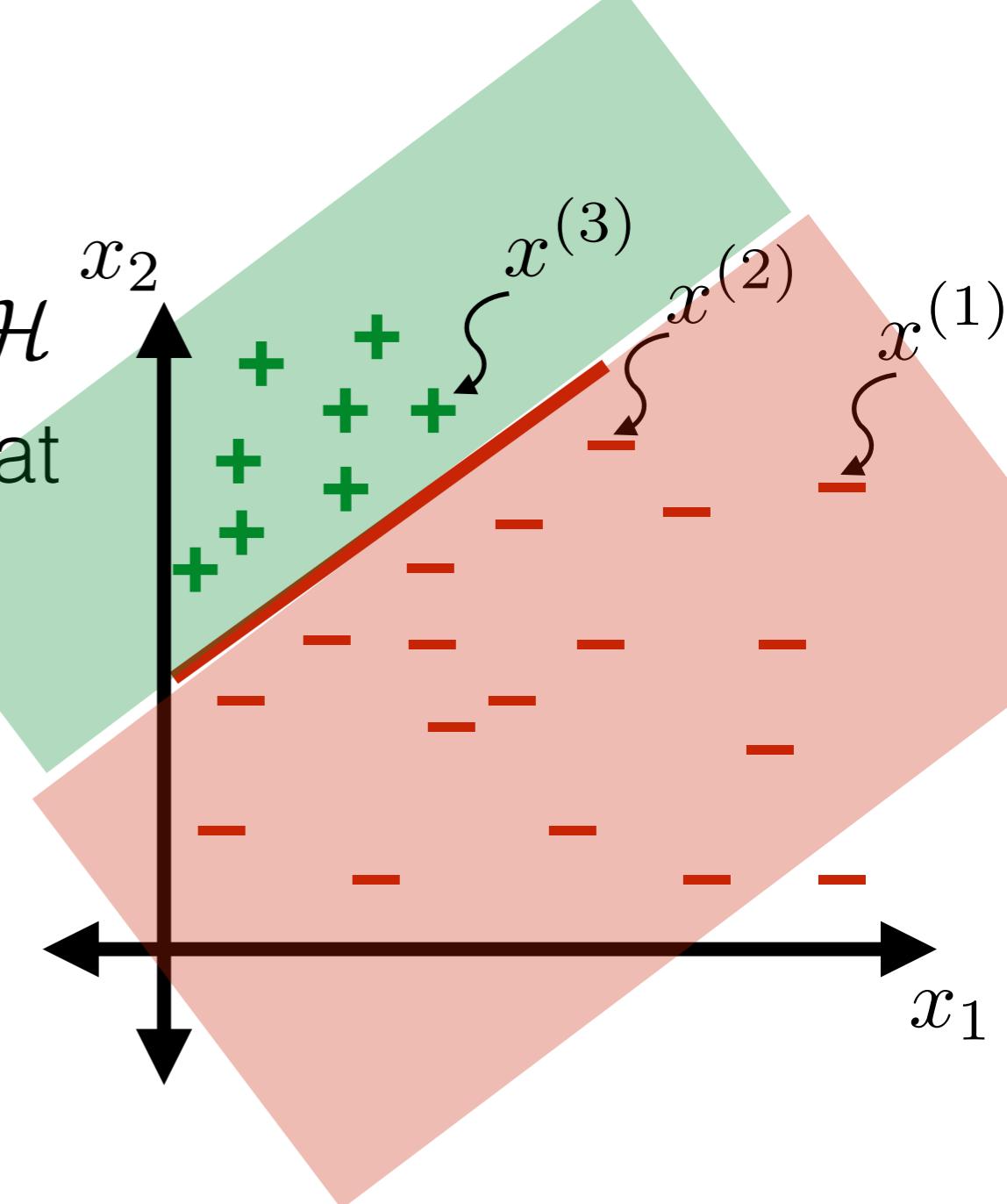
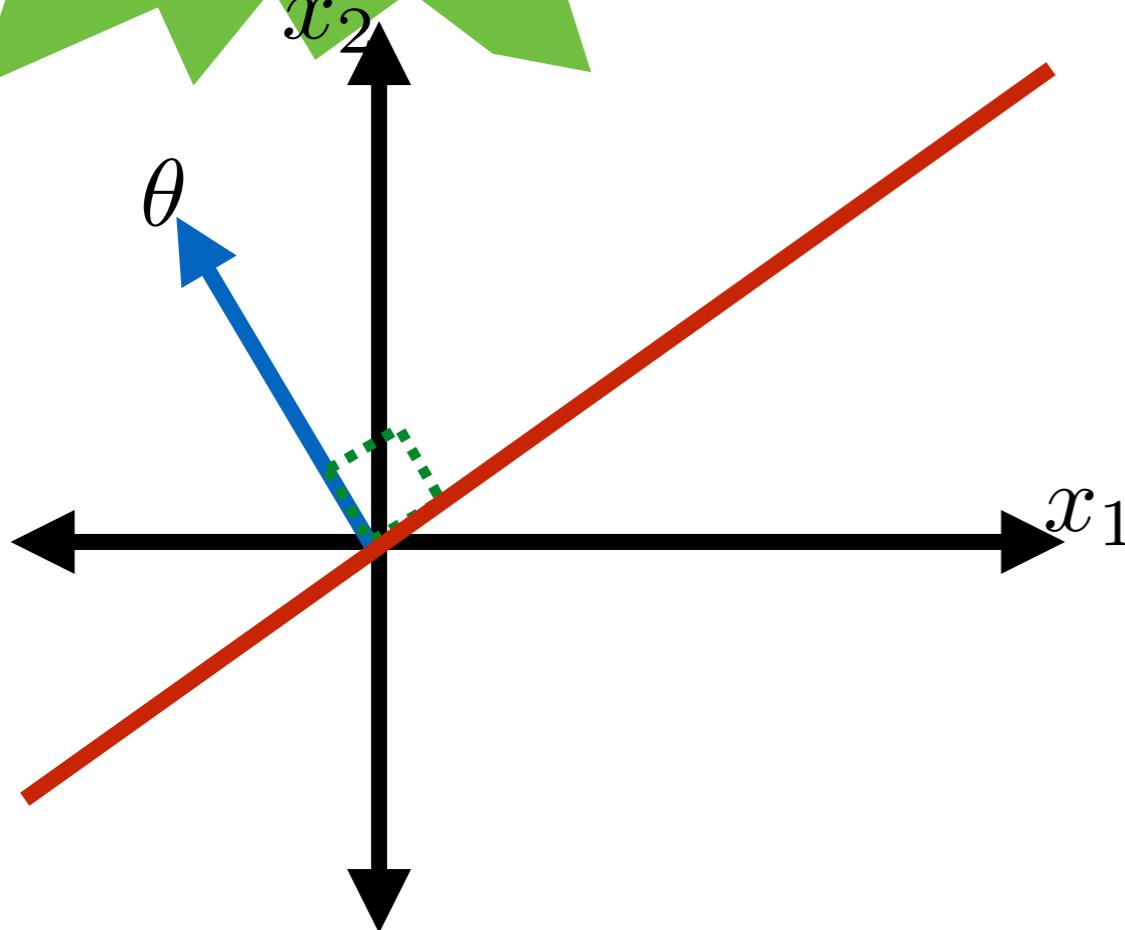
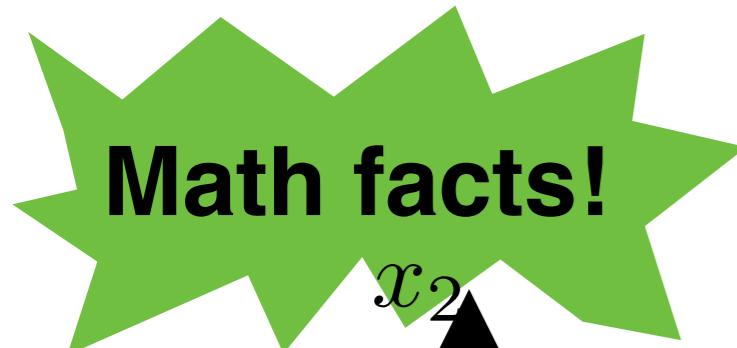
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



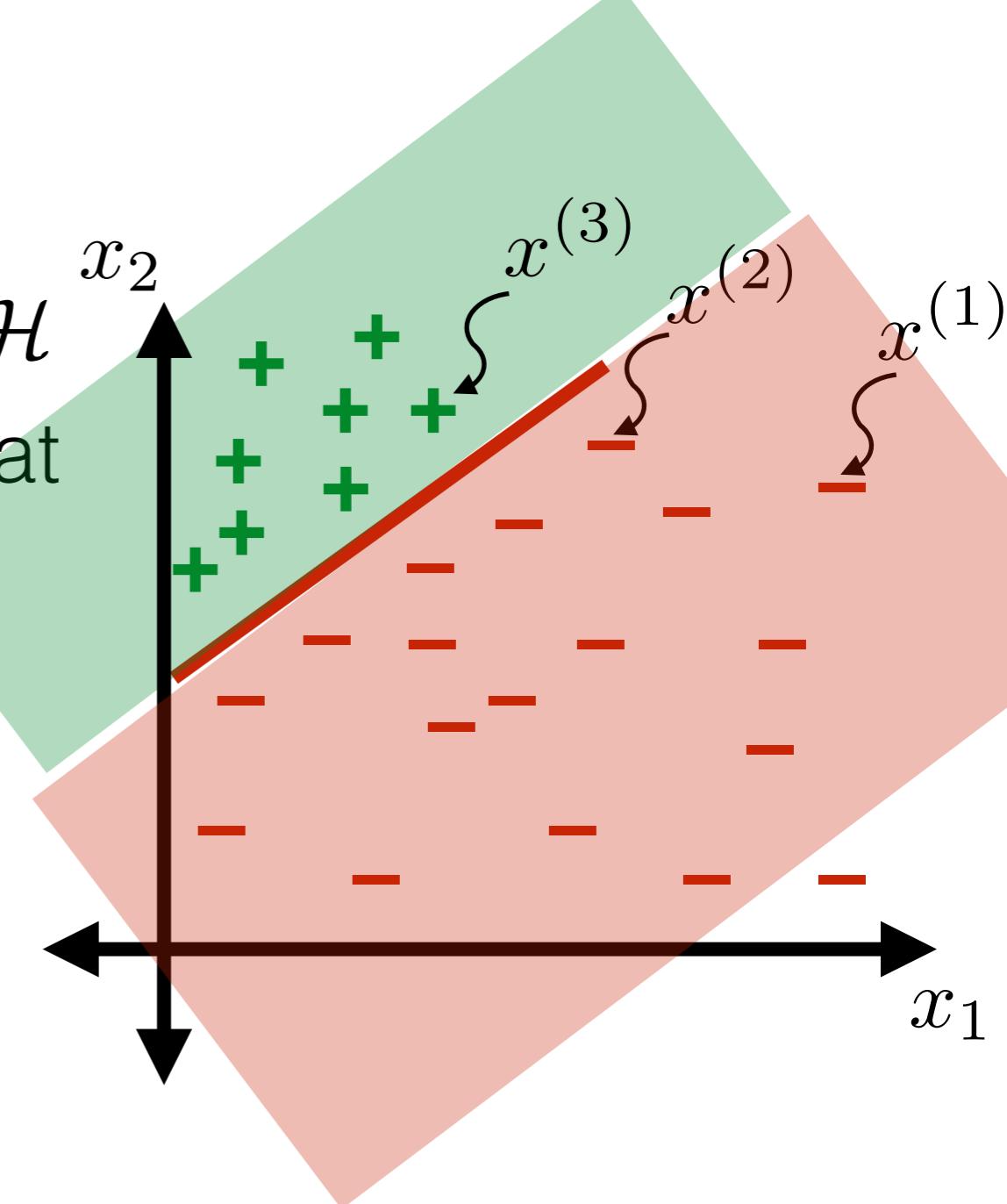
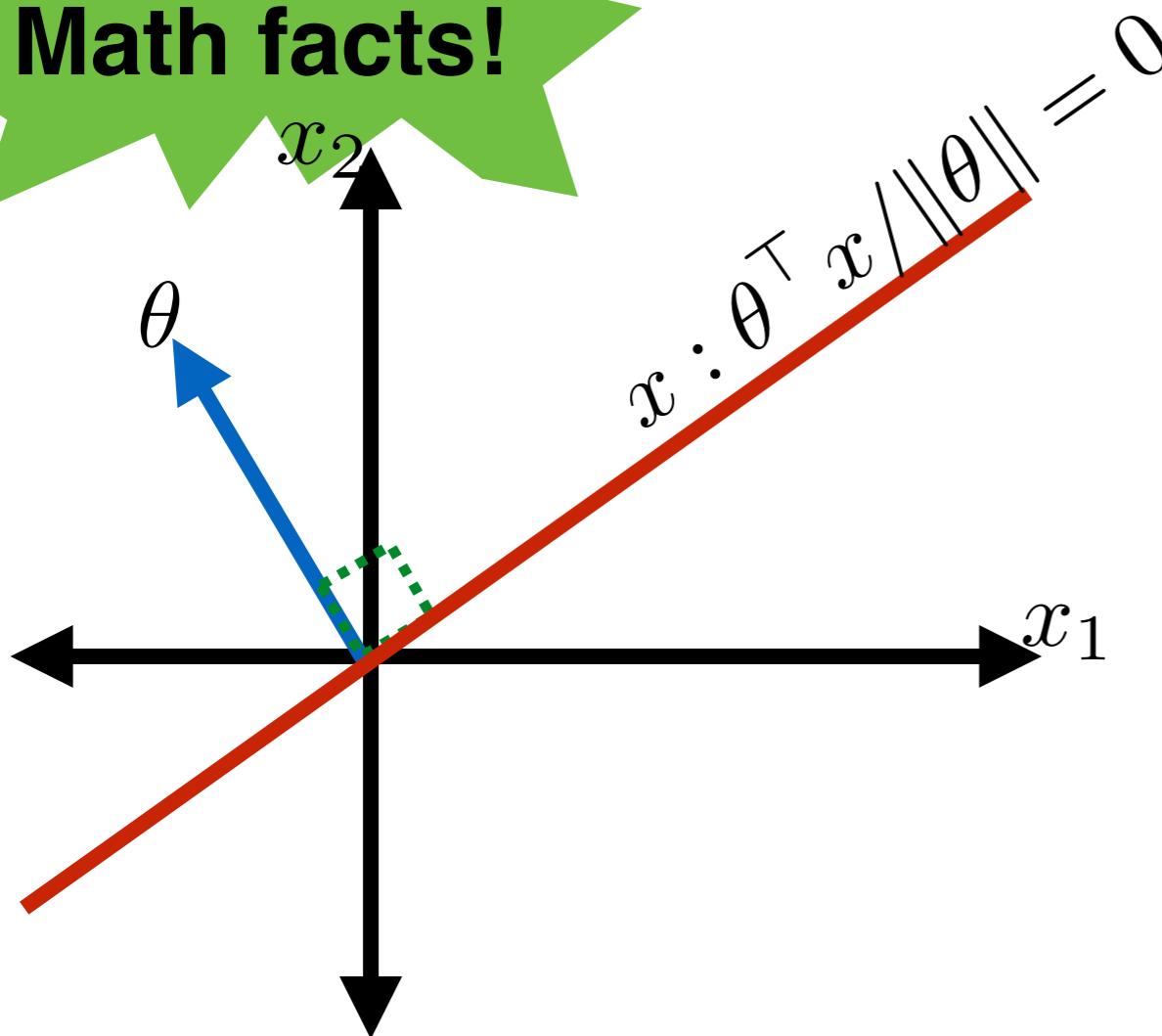
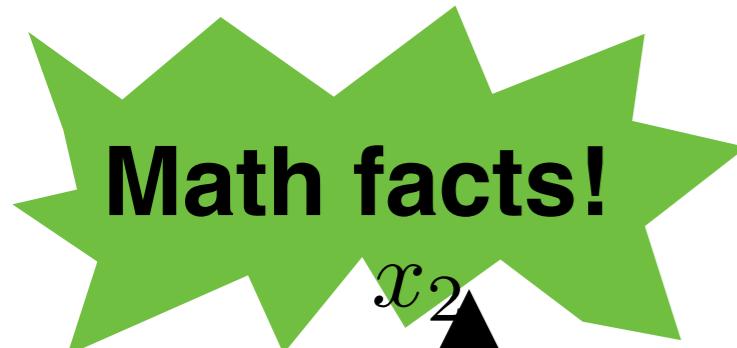
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



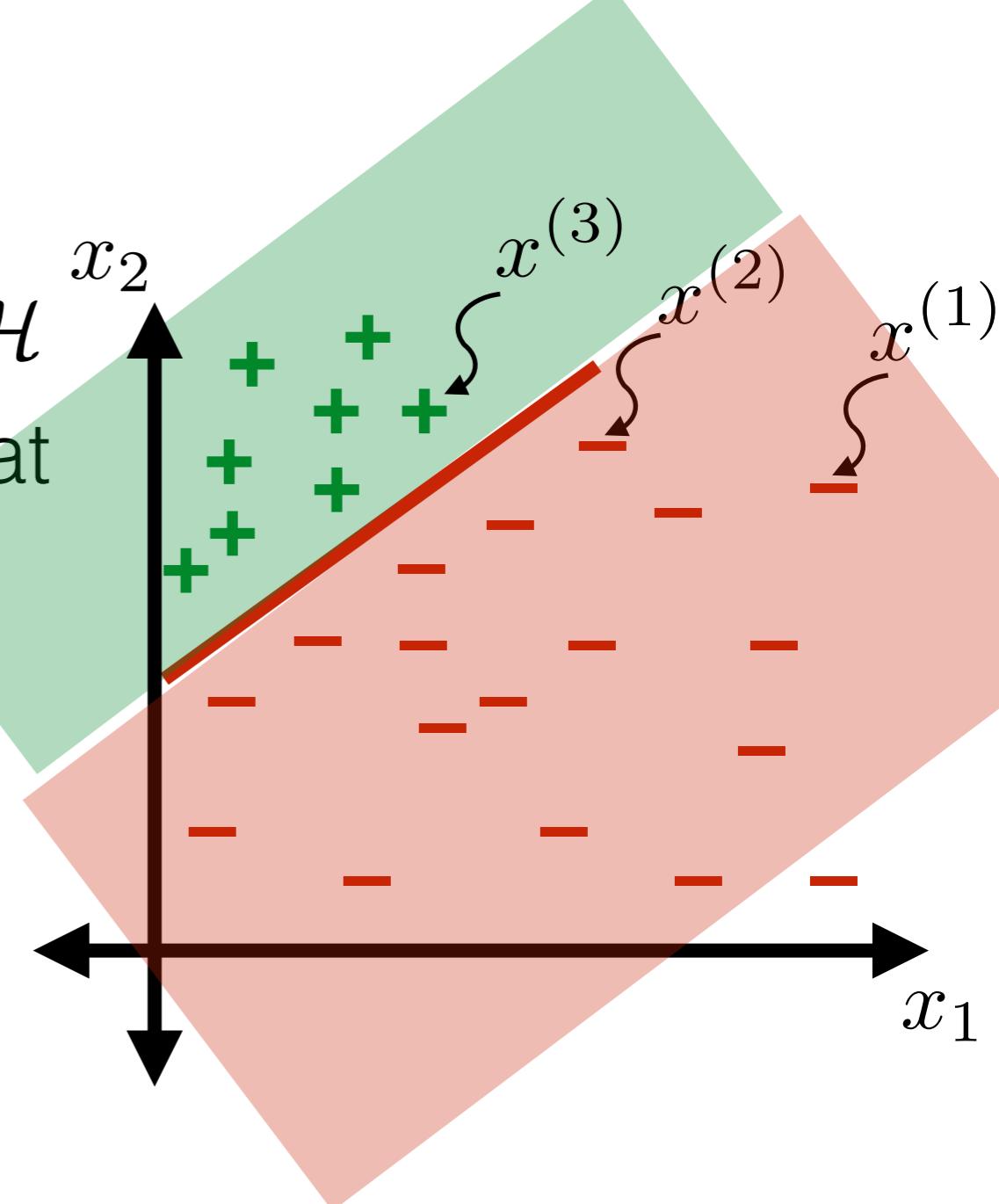
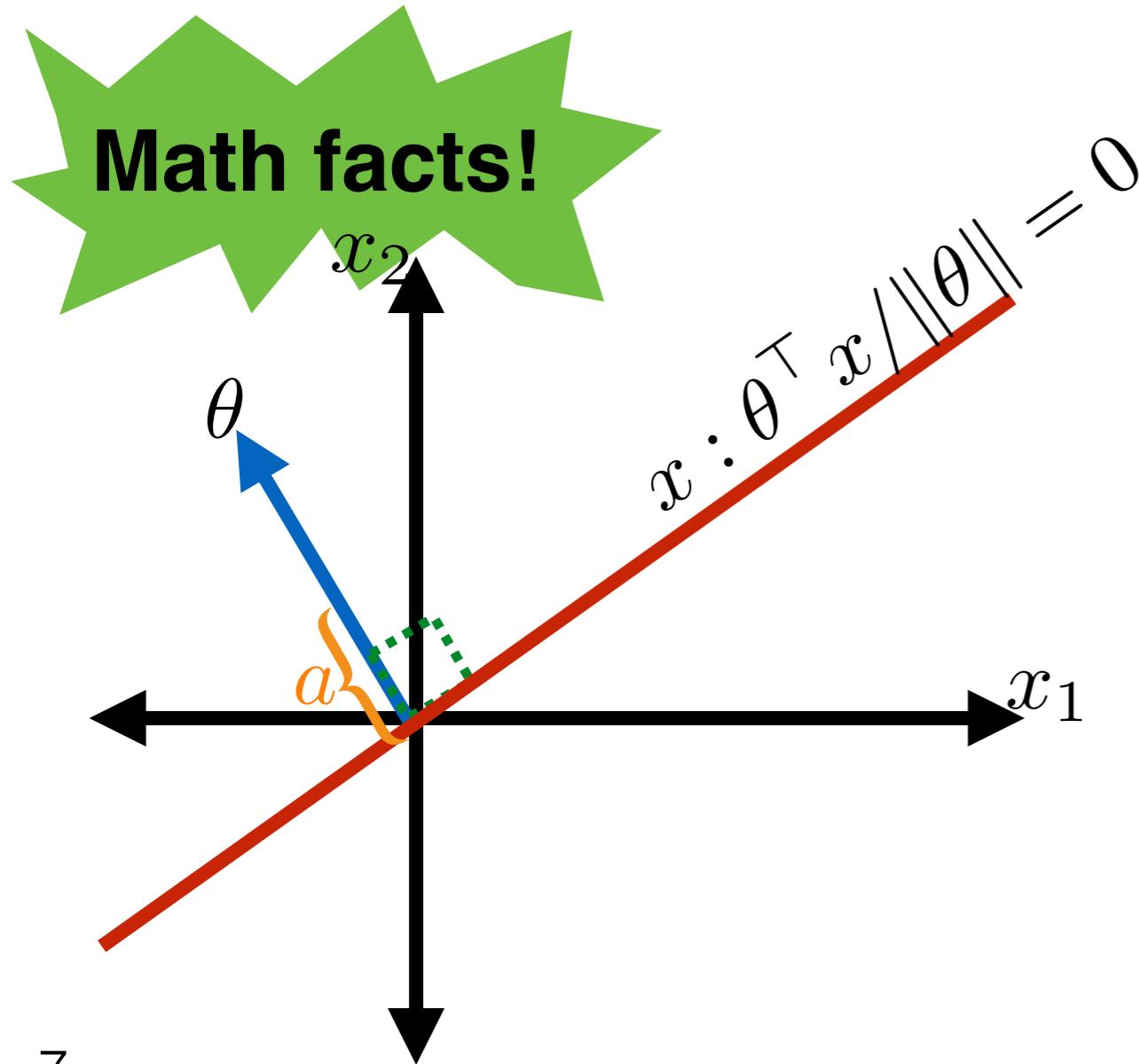
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



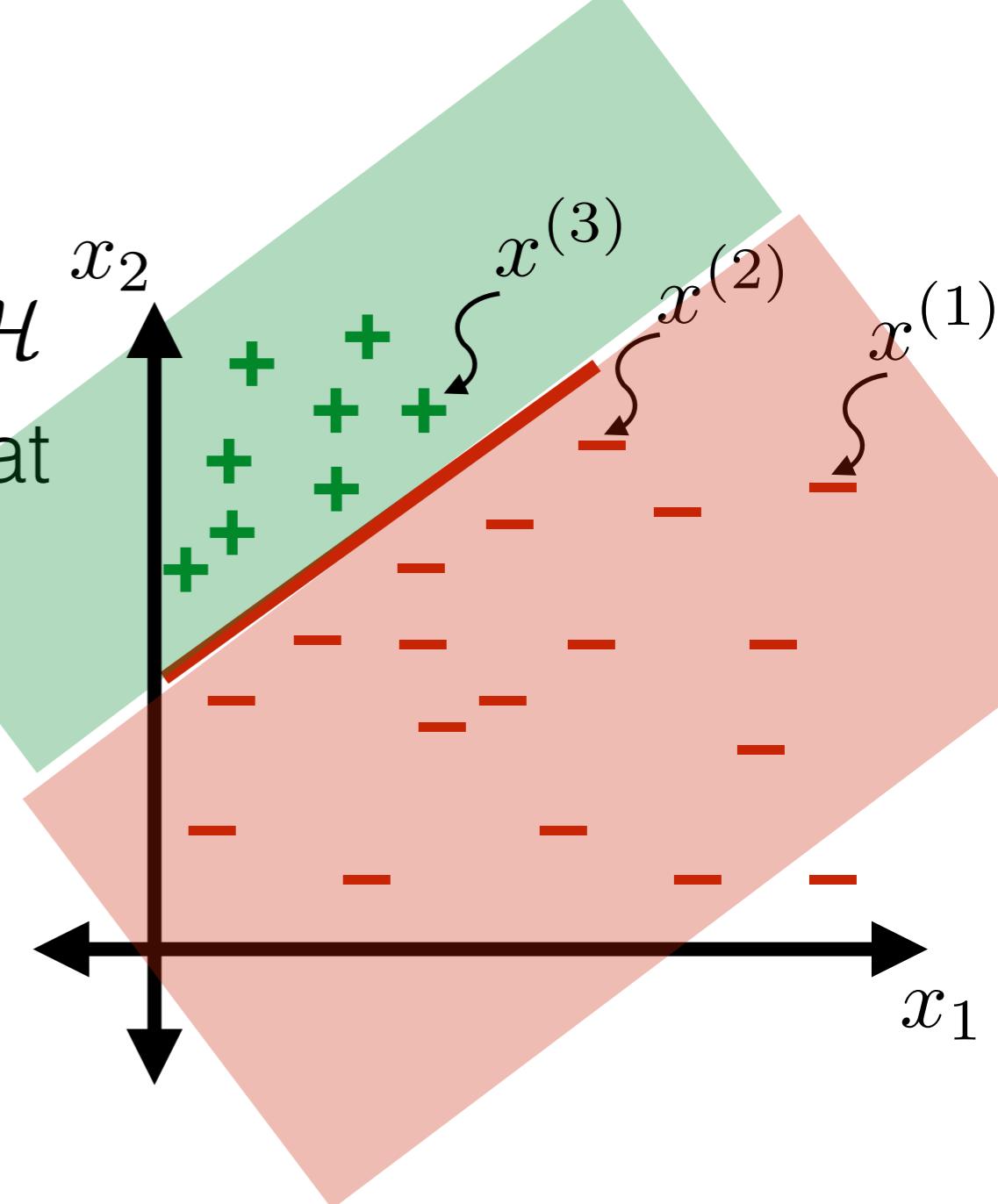
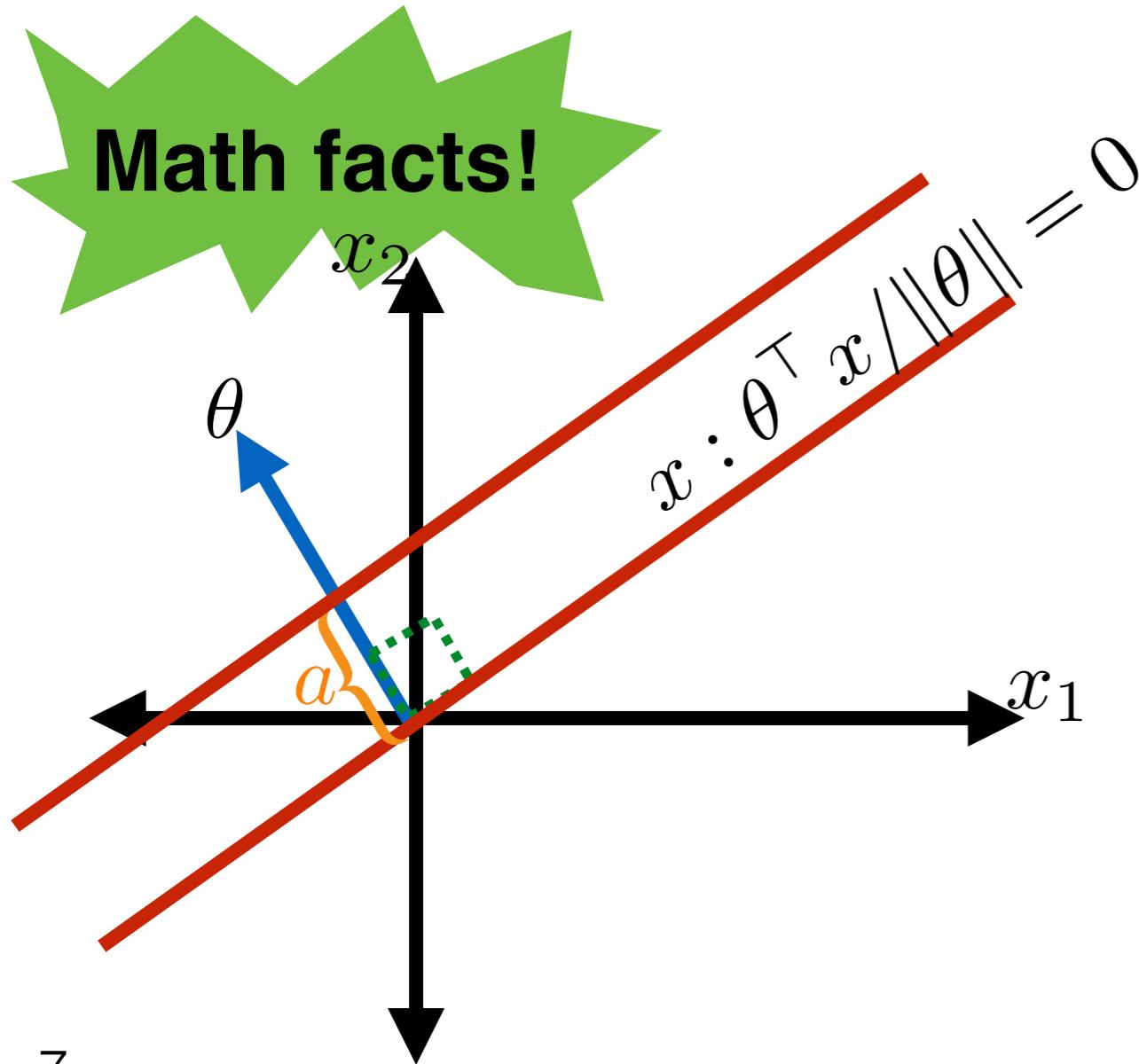
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



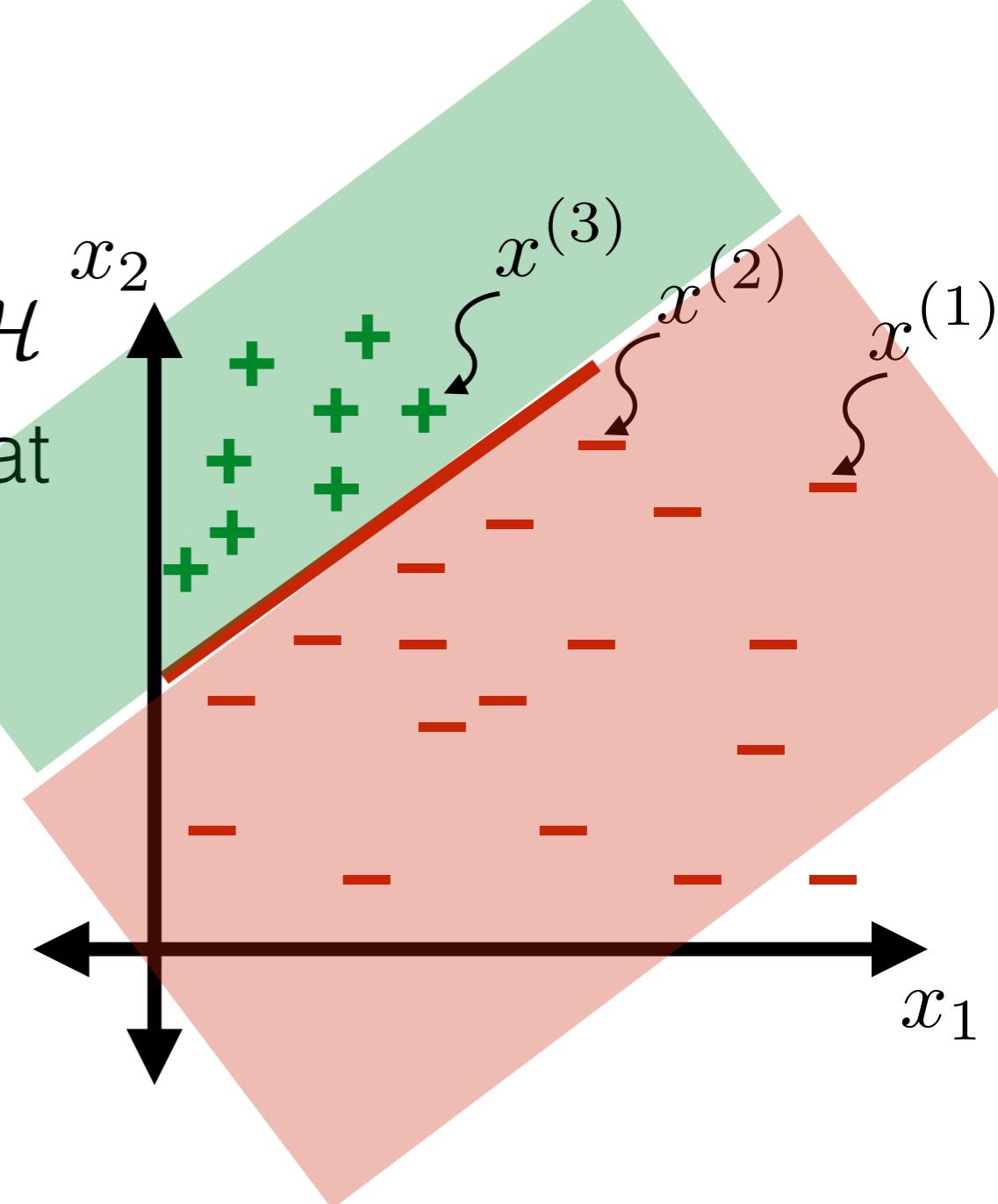
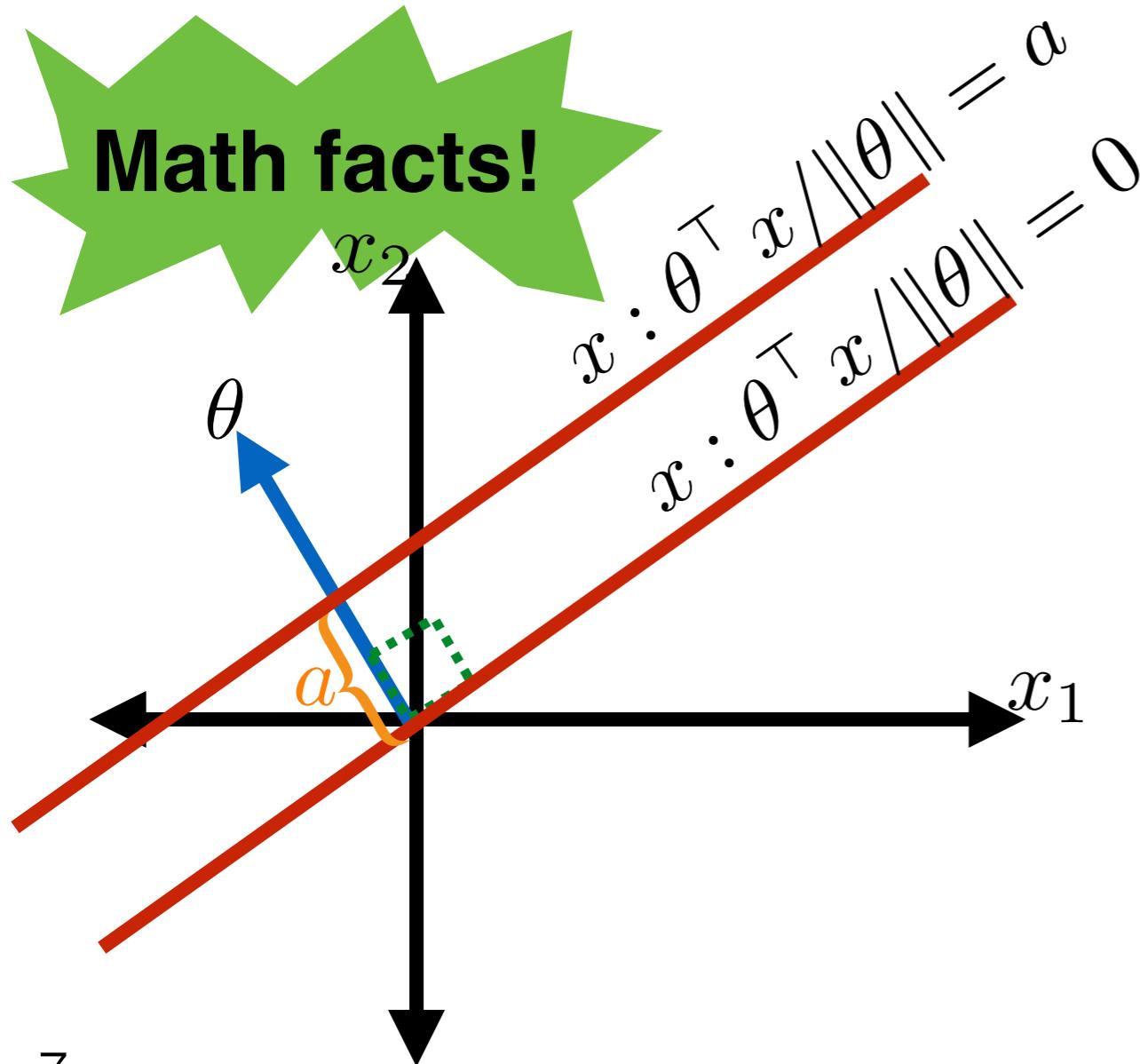
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



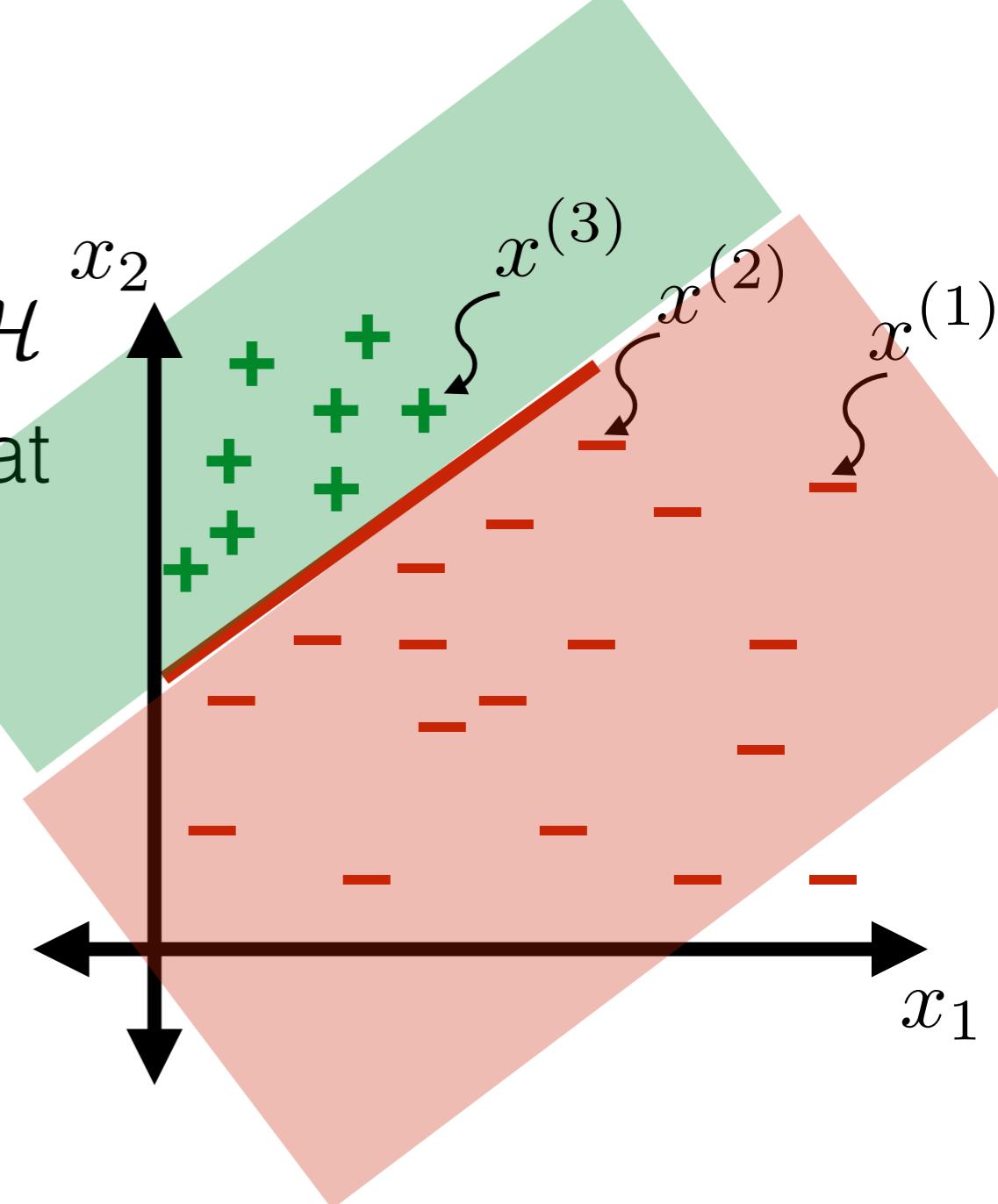
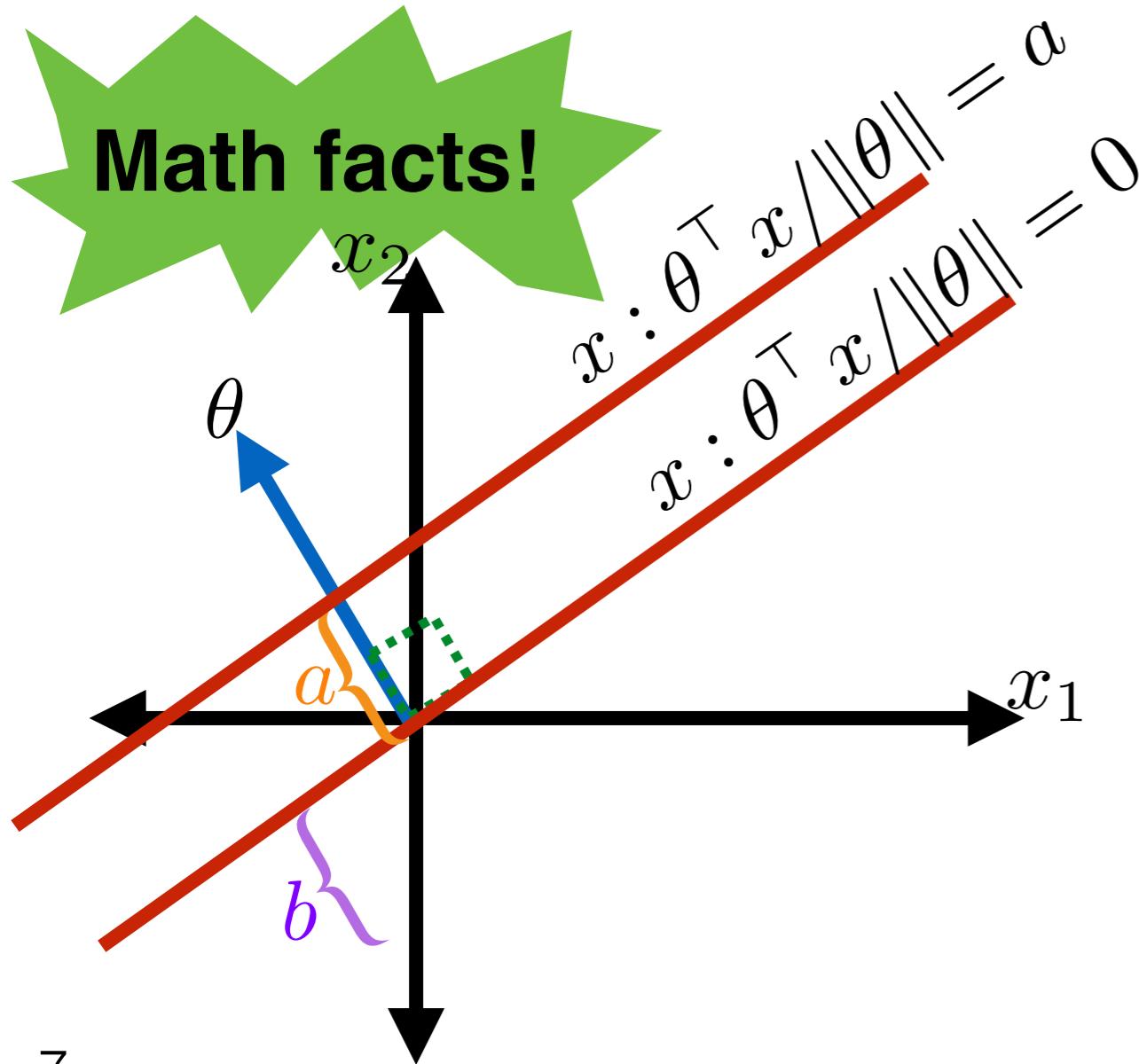
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



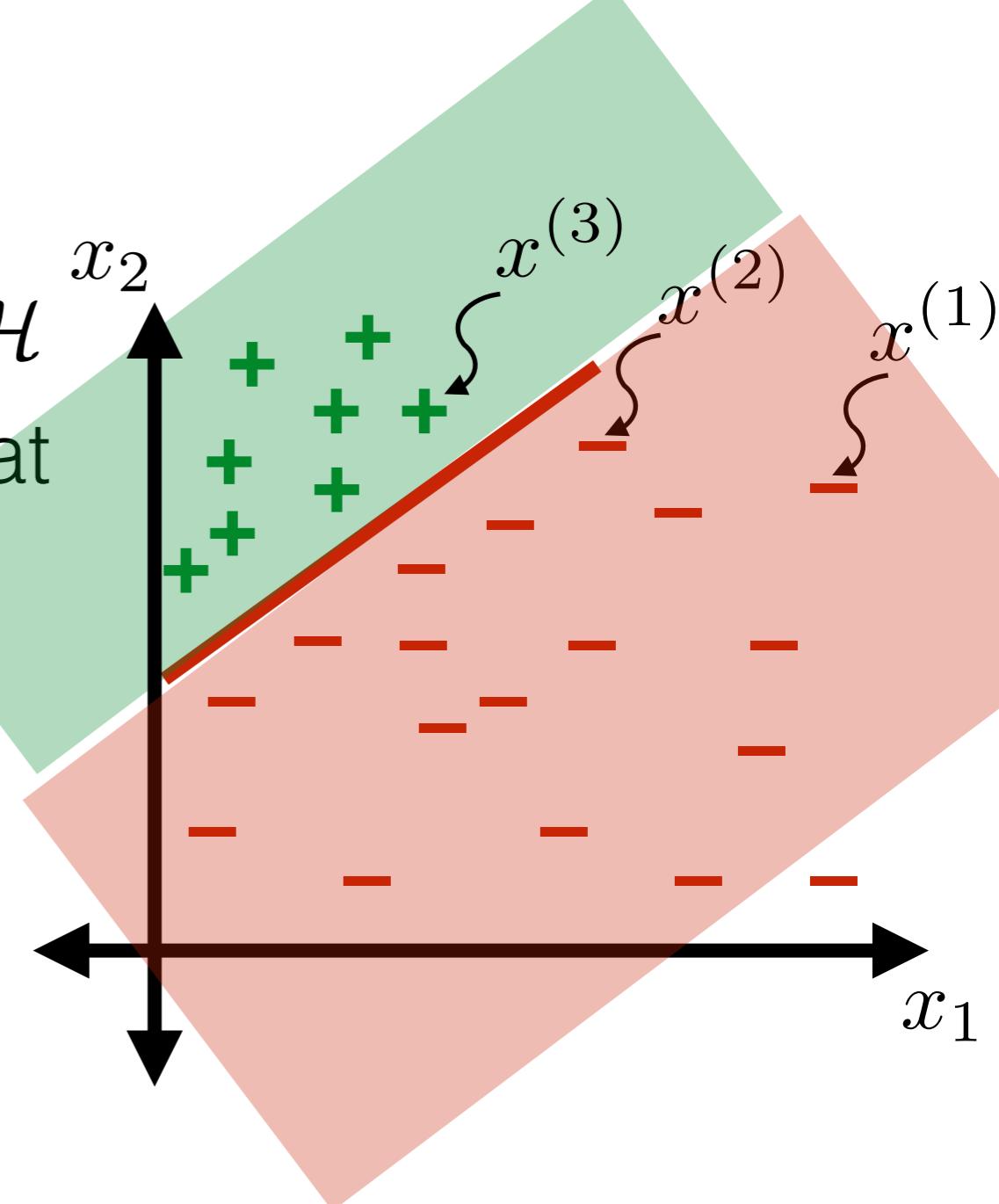
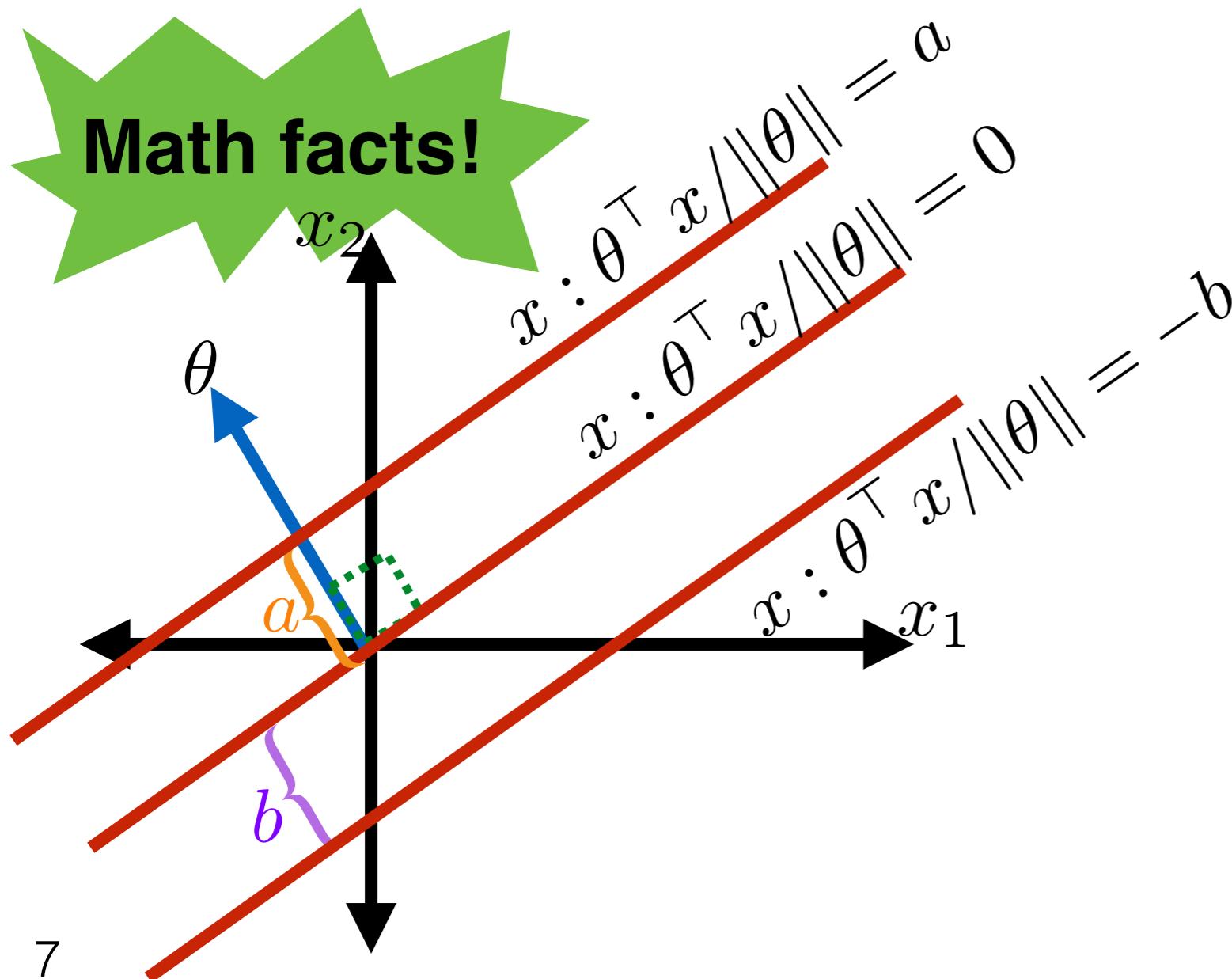
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



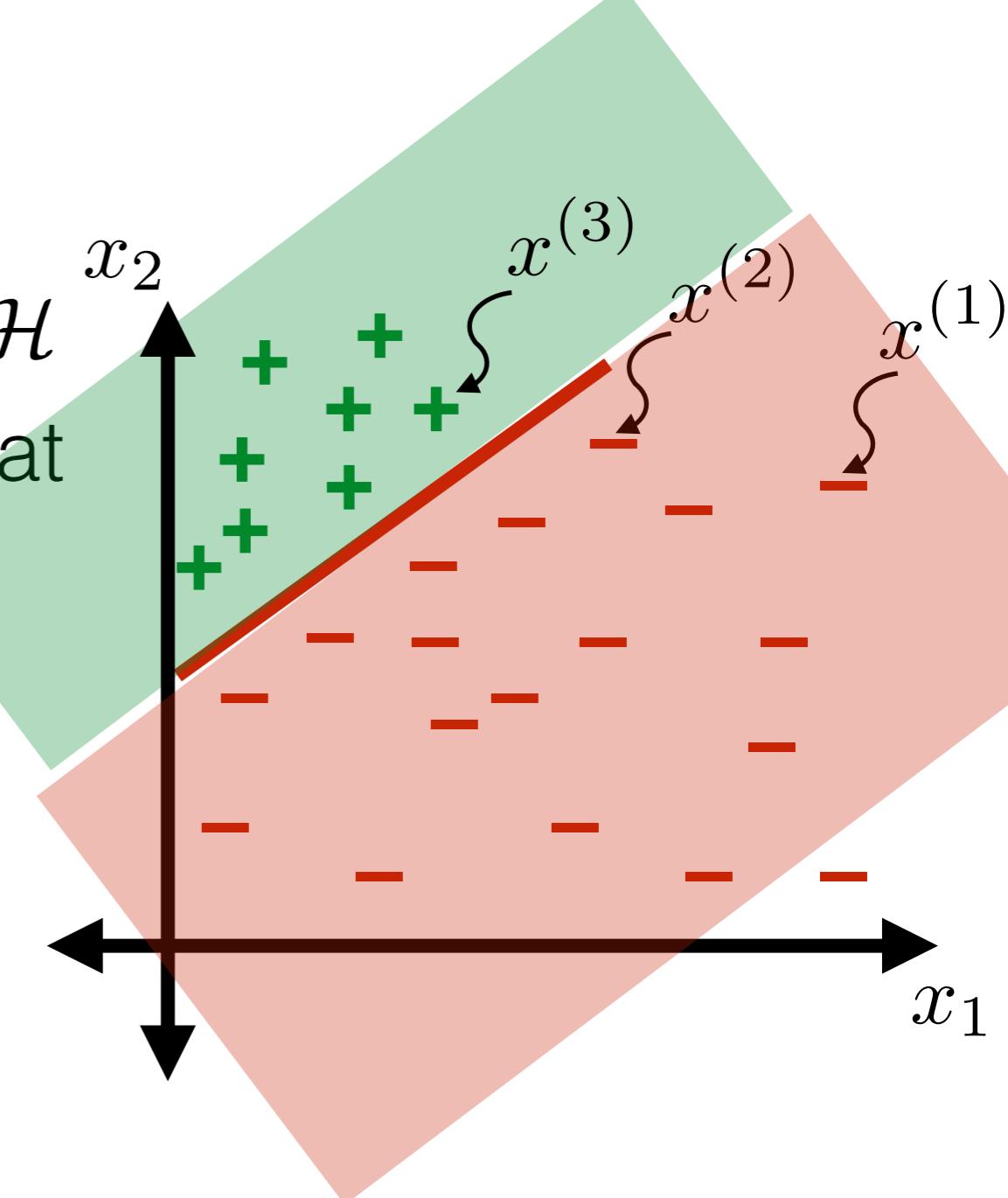
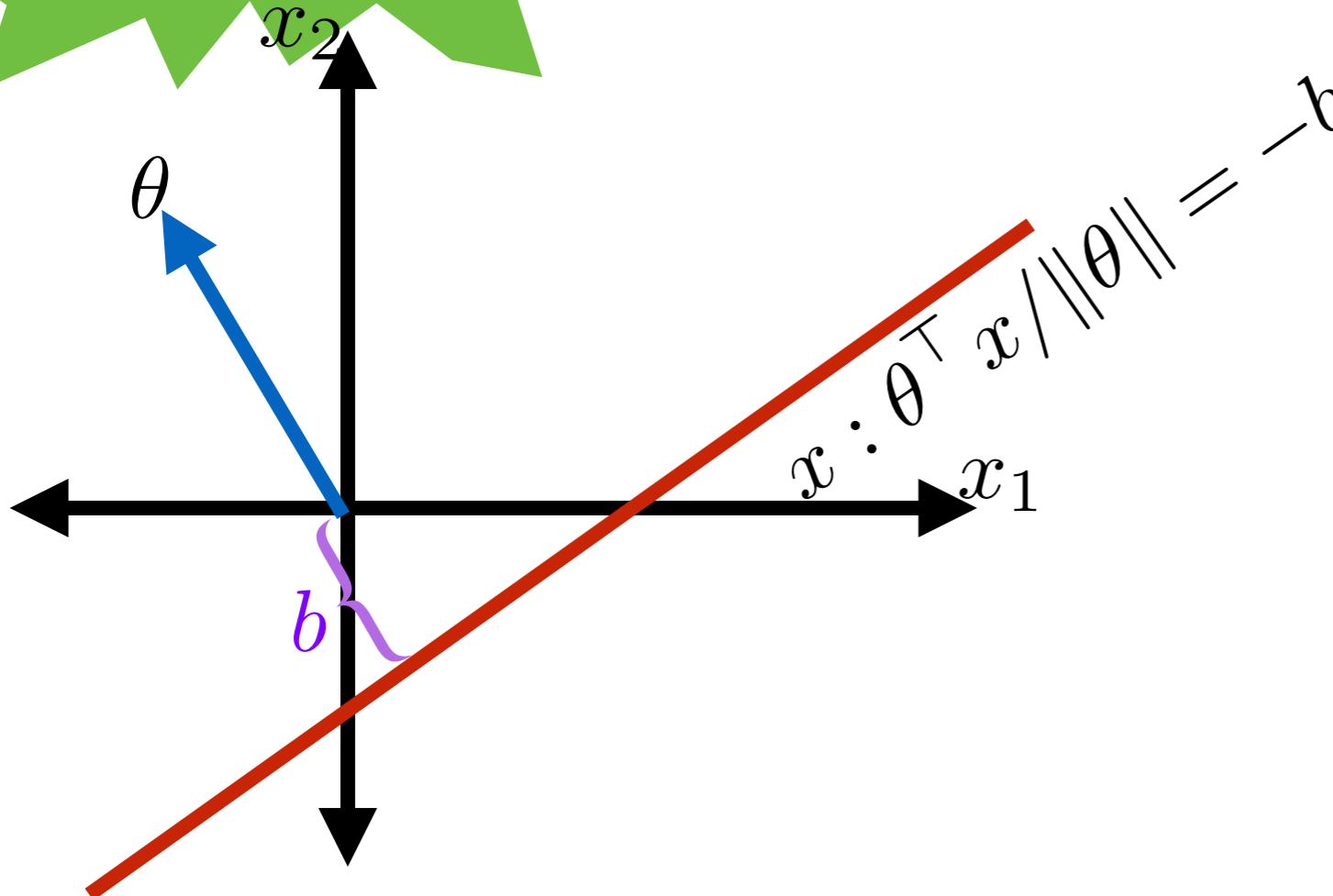
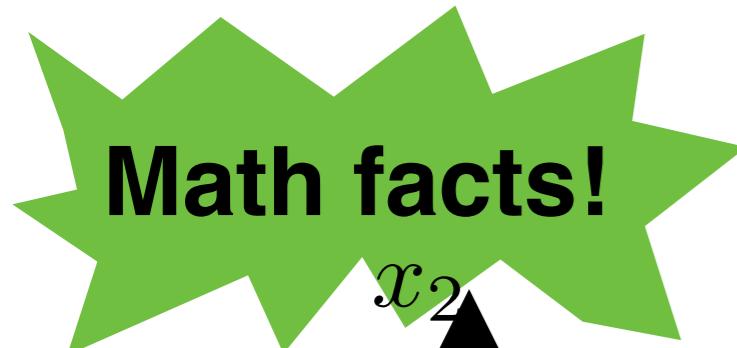
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



Linear classifiers

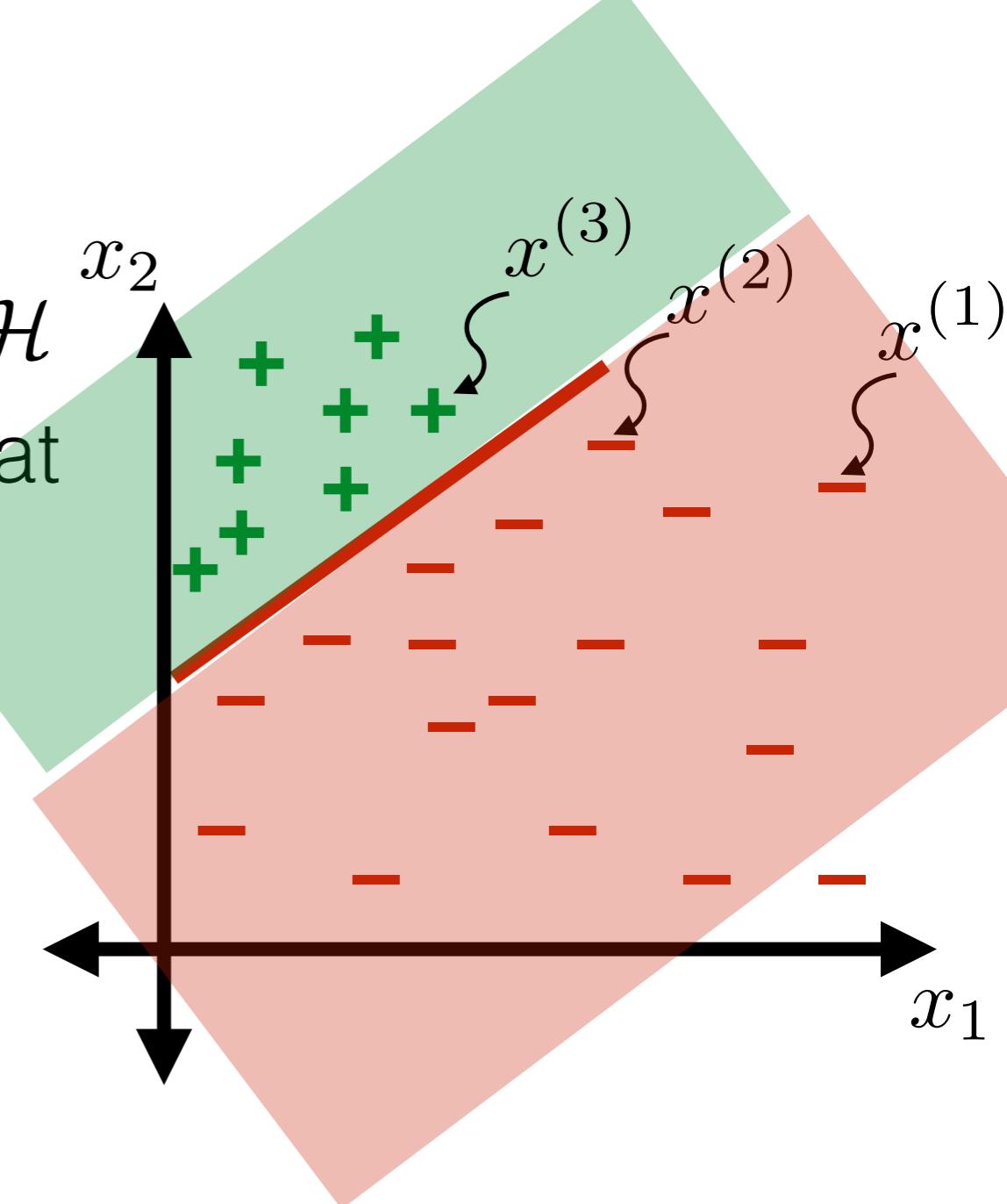
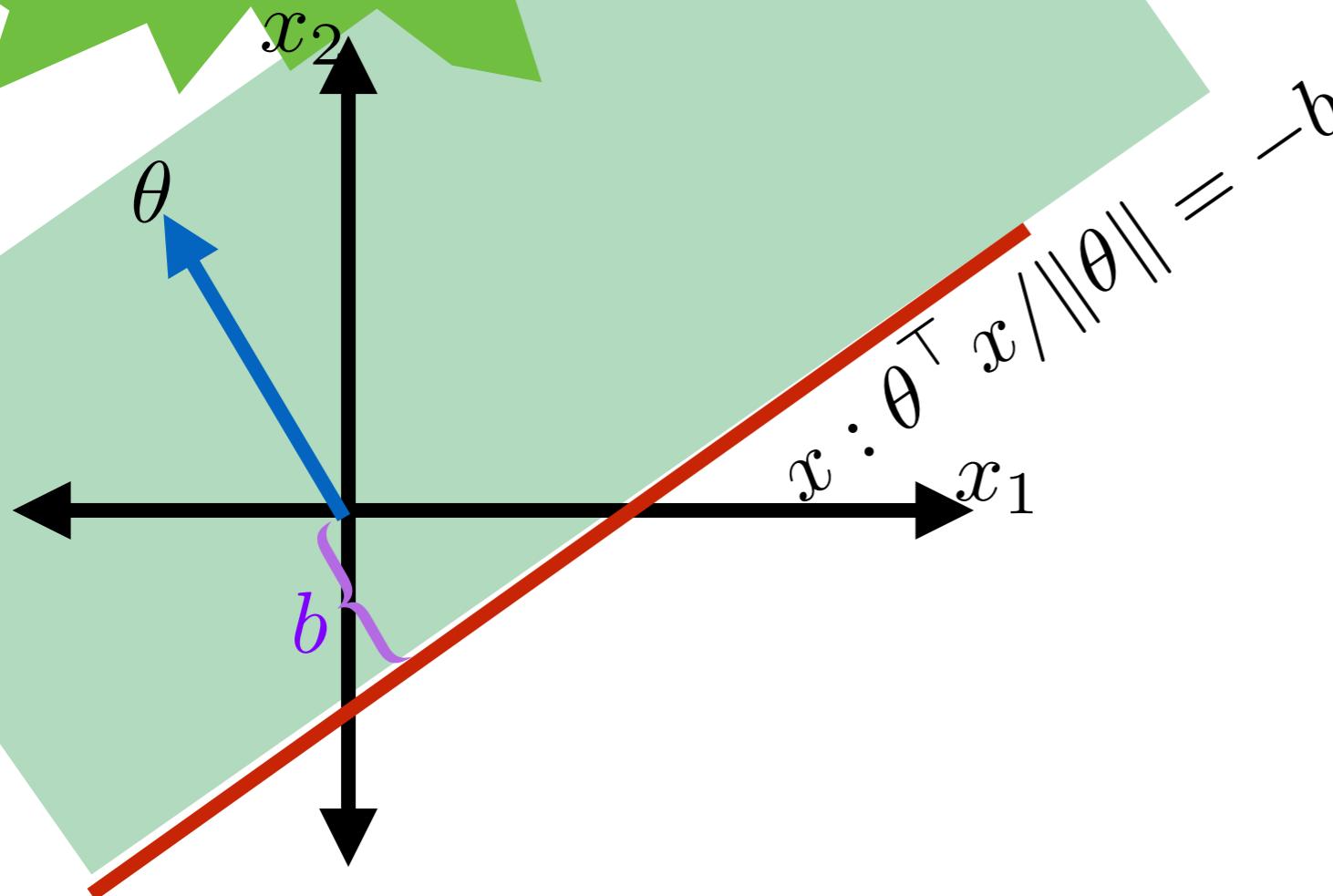
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



Linear classifiers

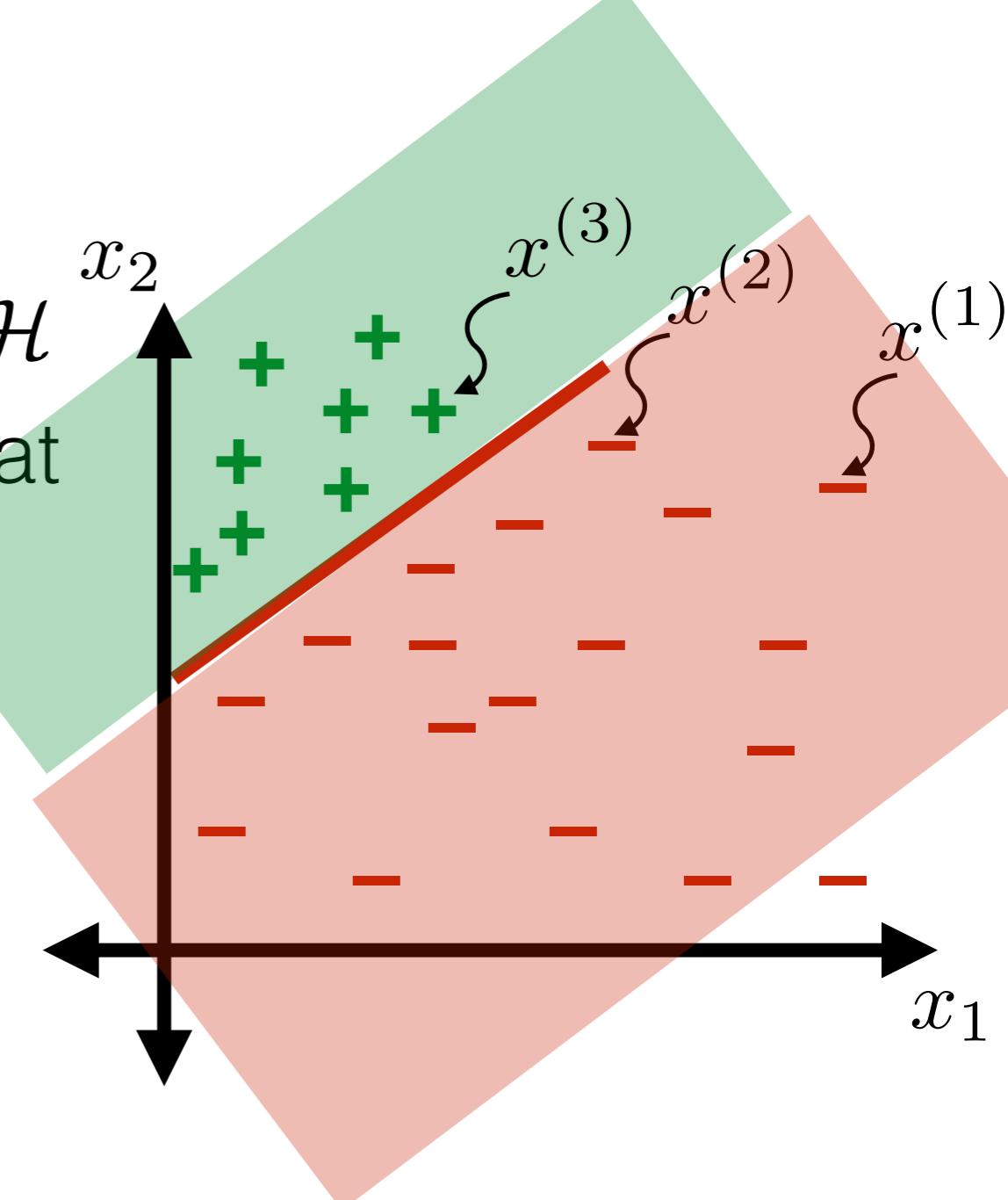
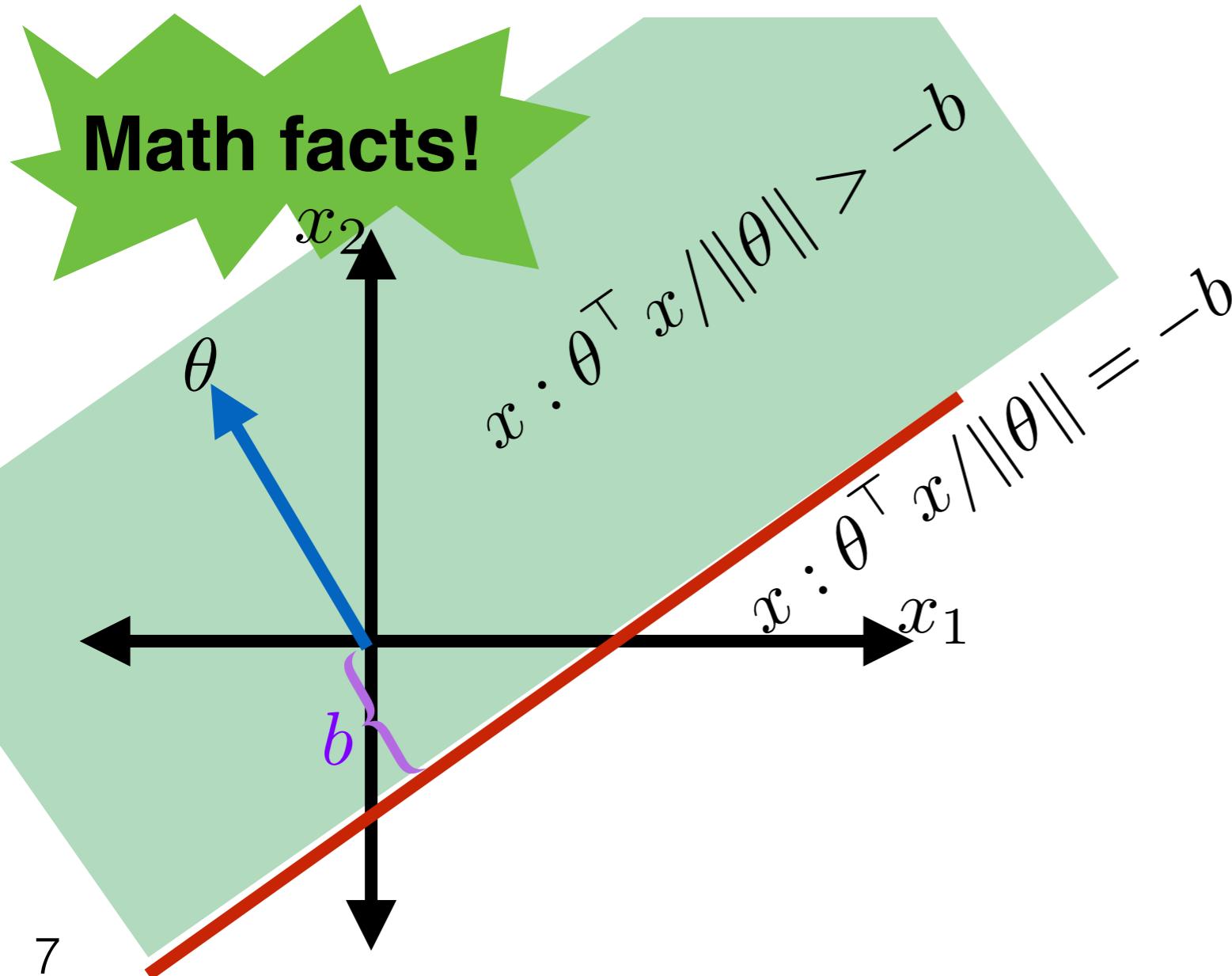
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side

Math facts!



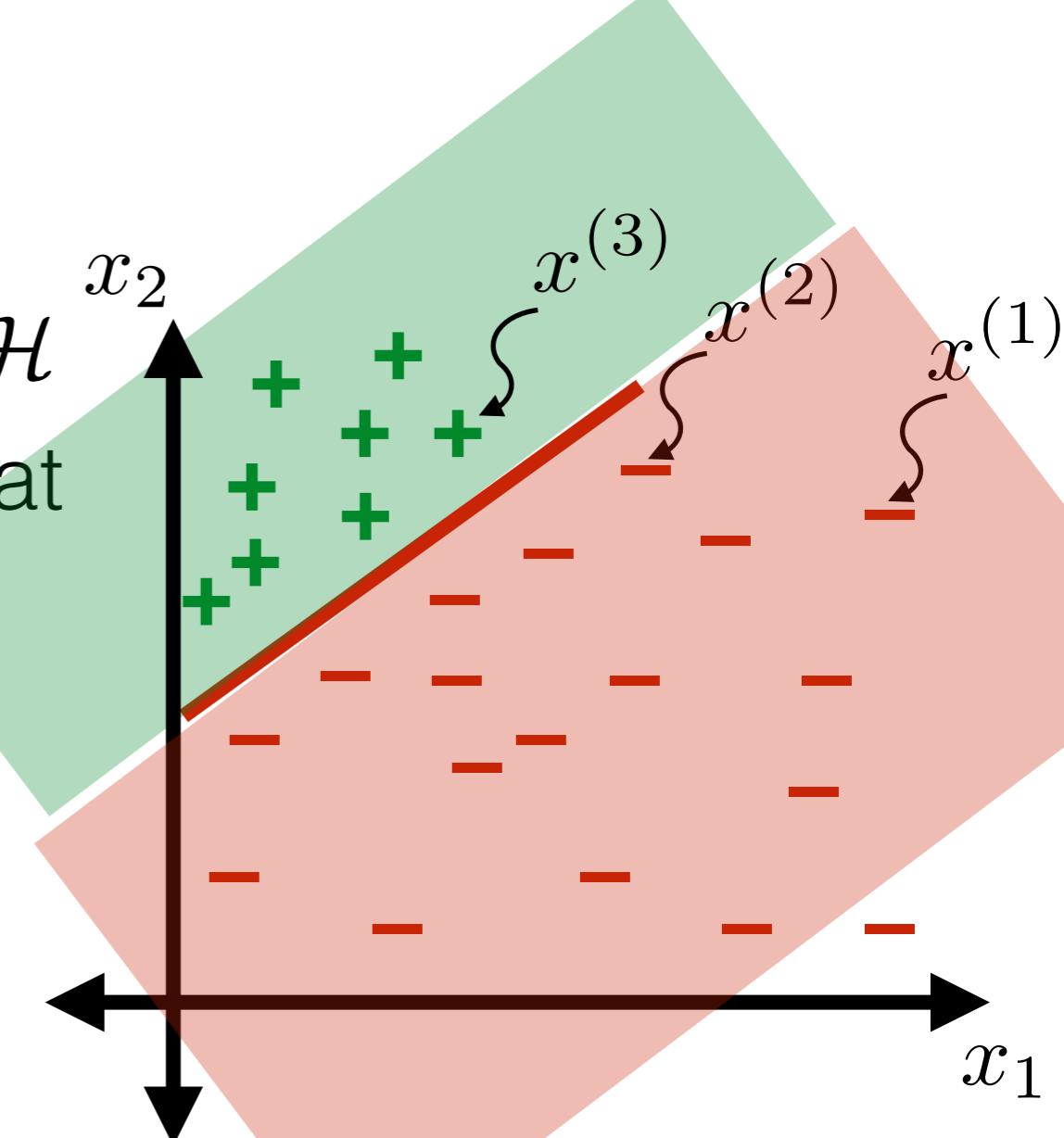
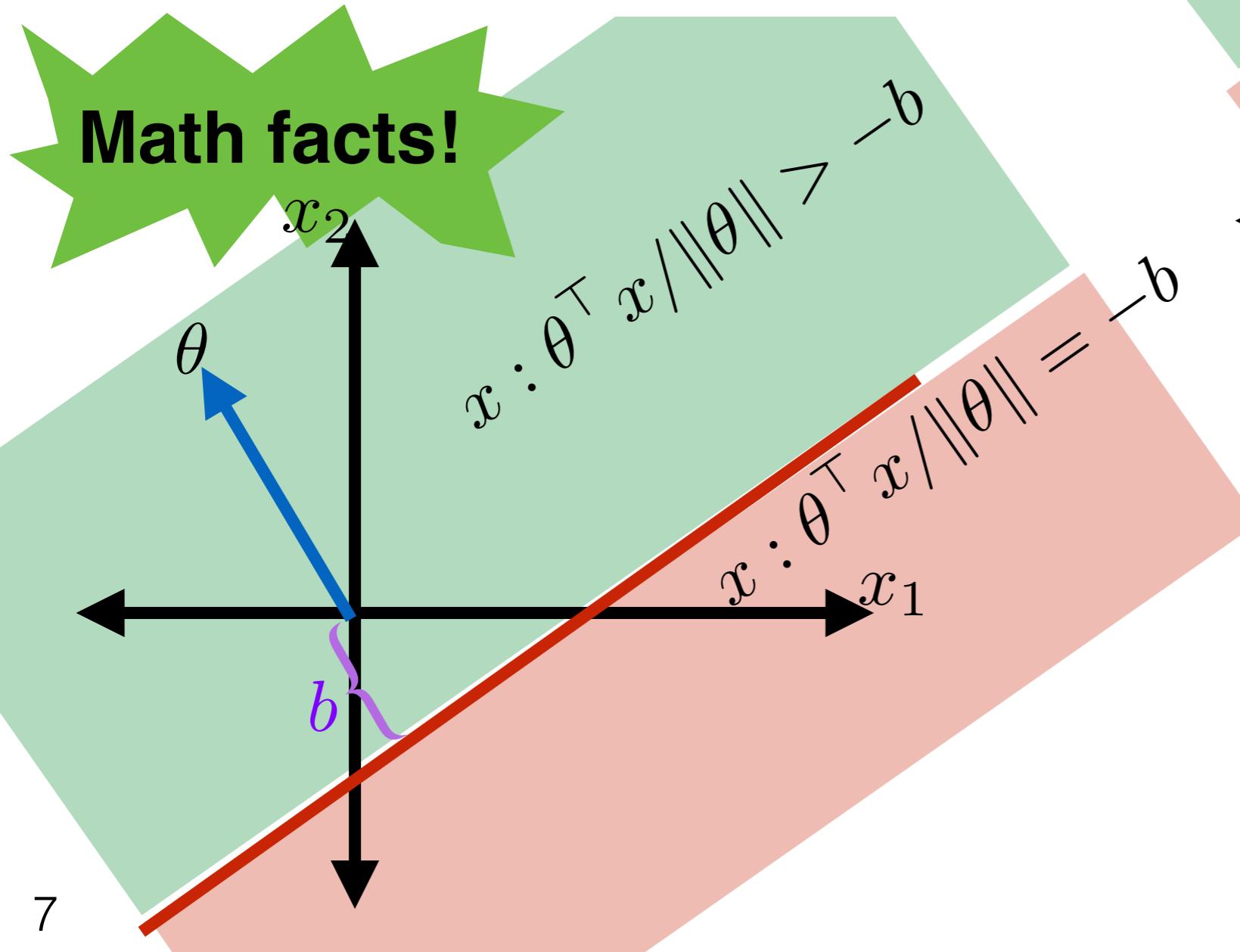
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



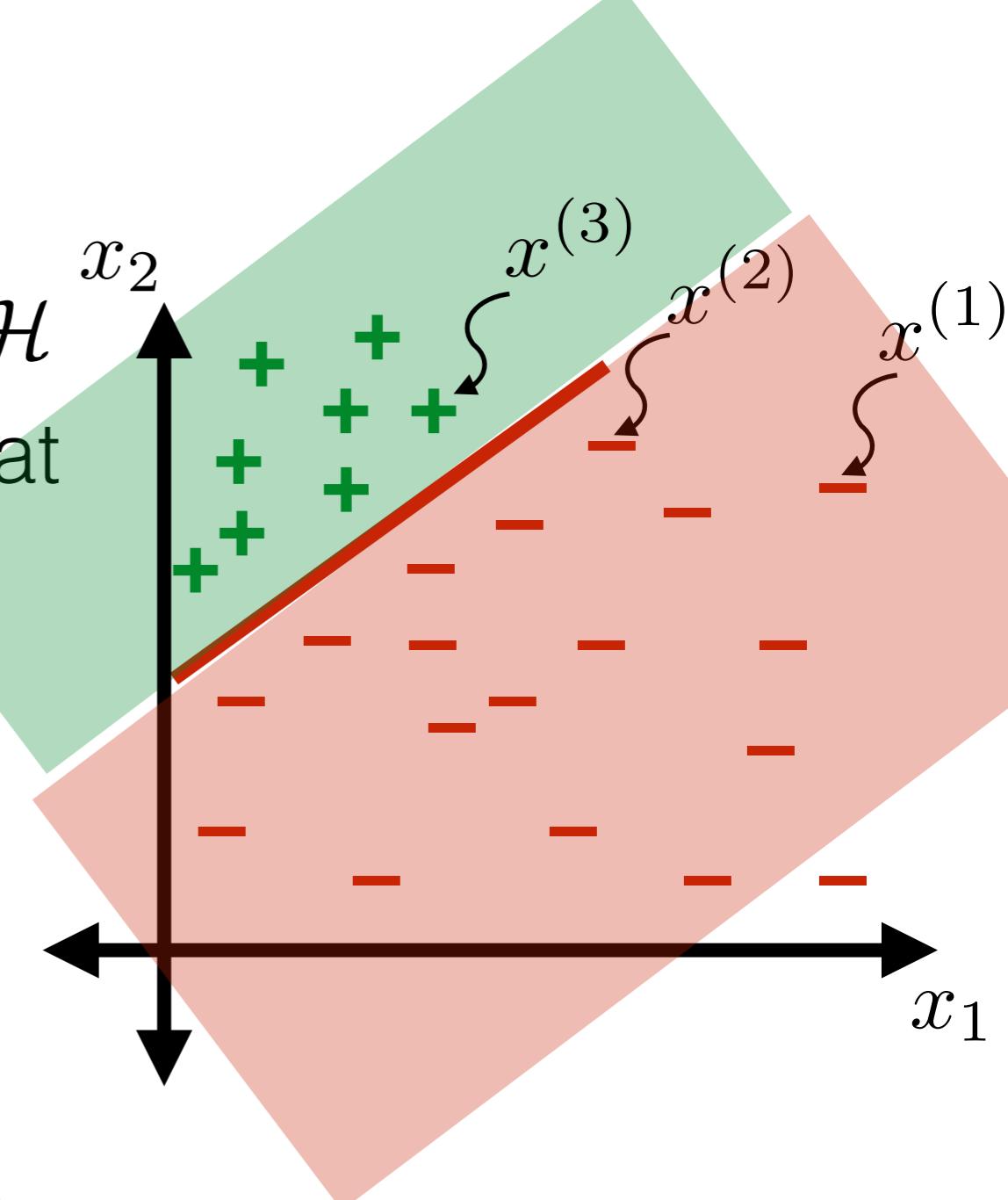
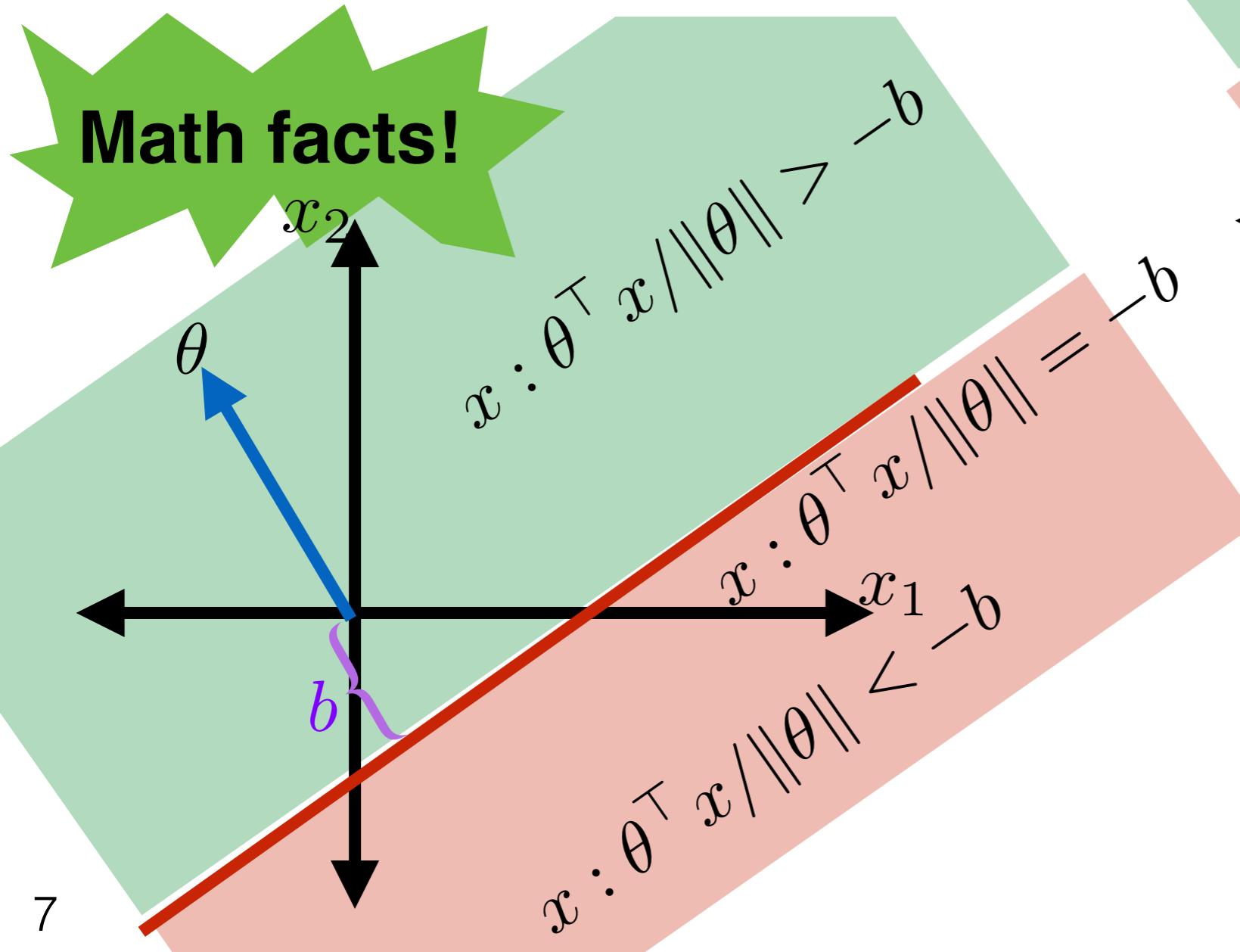
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



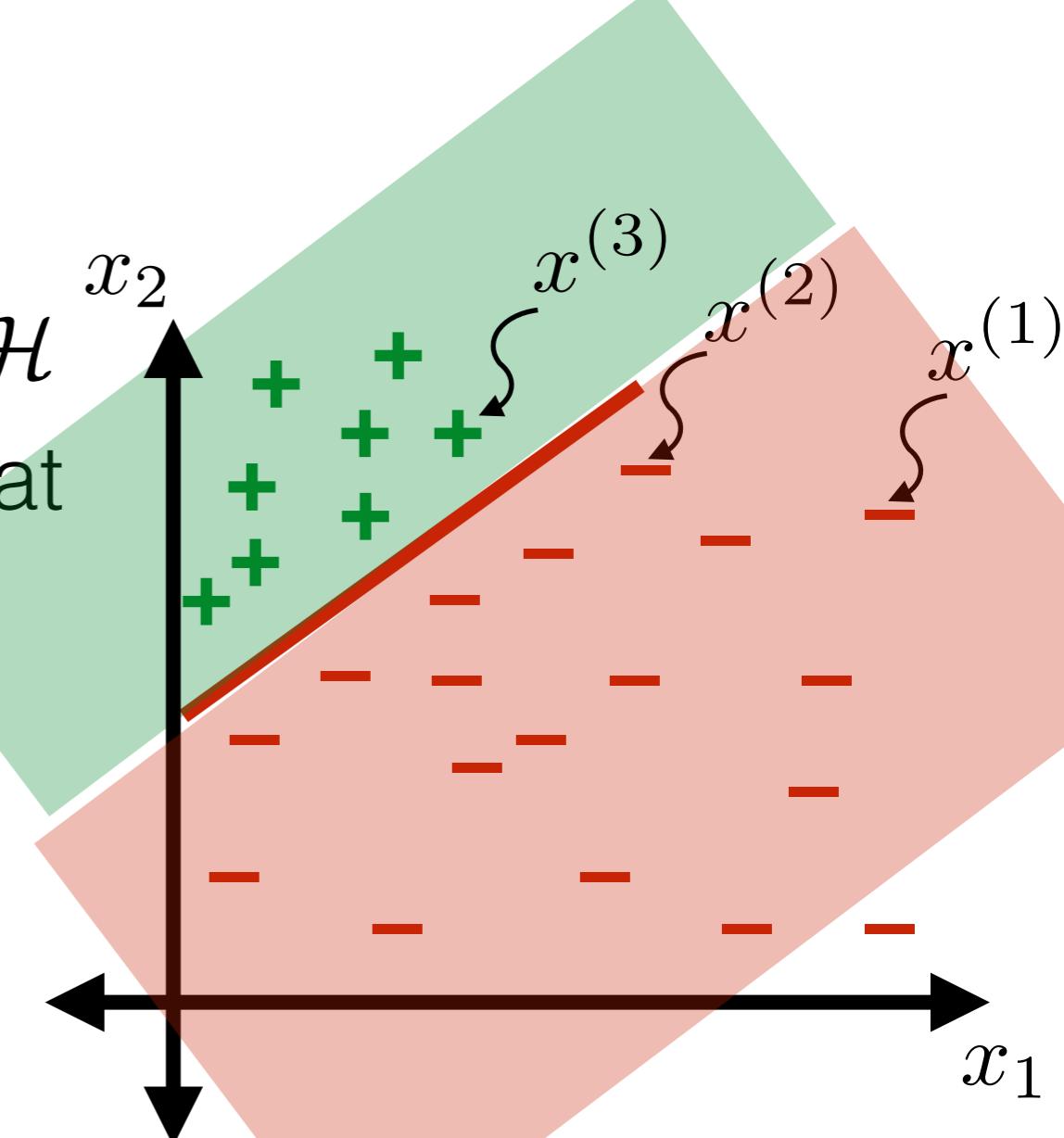
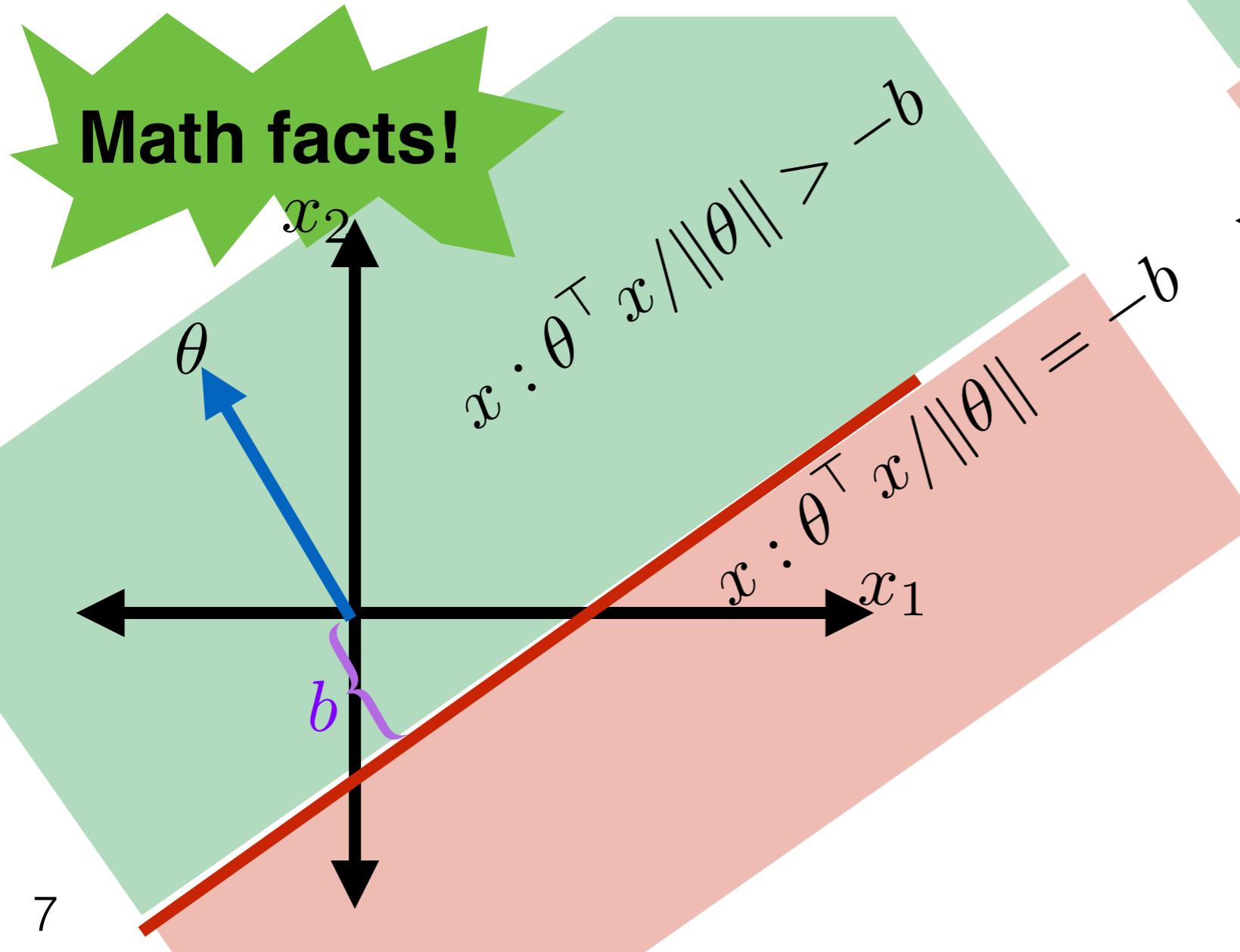
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



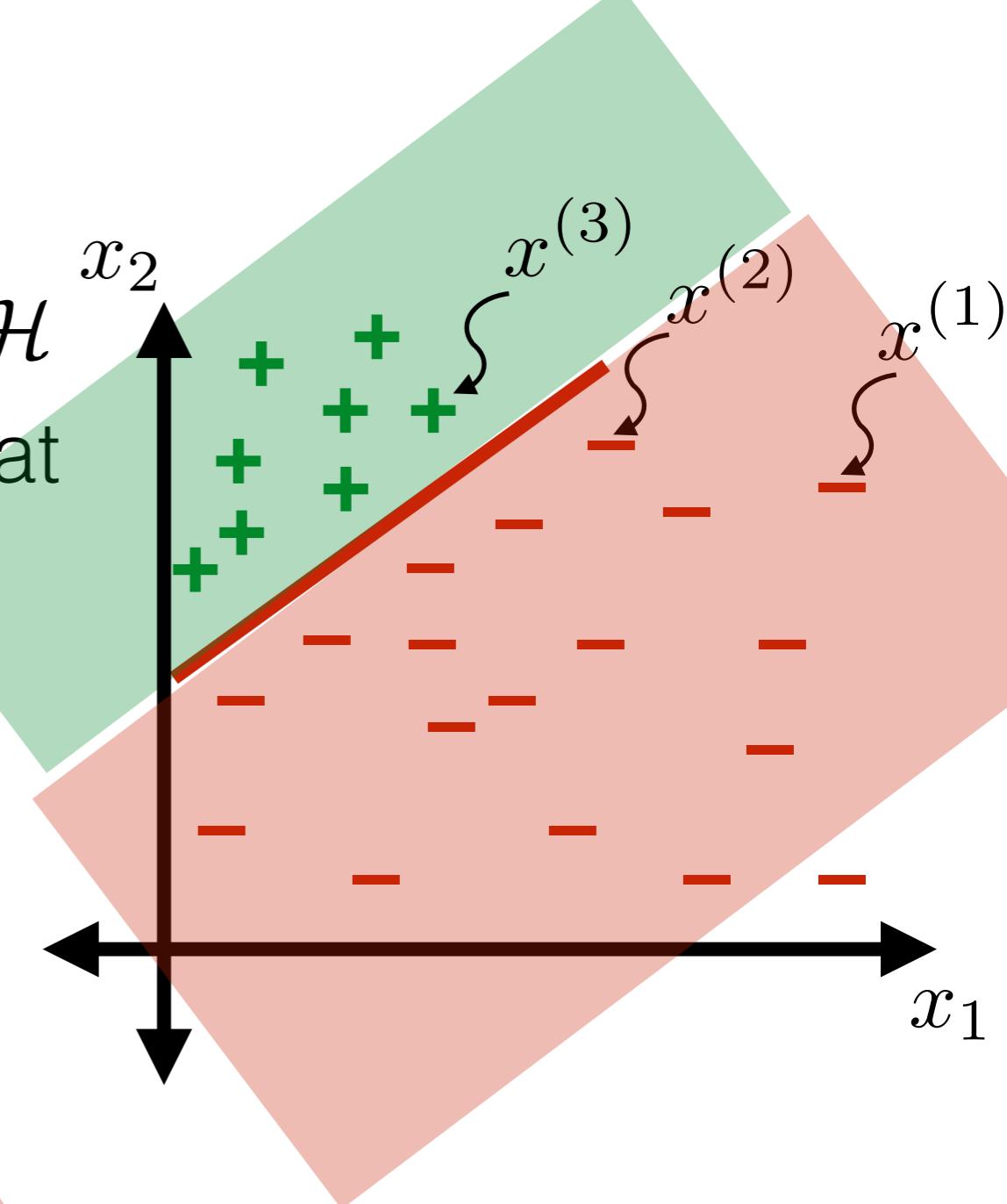
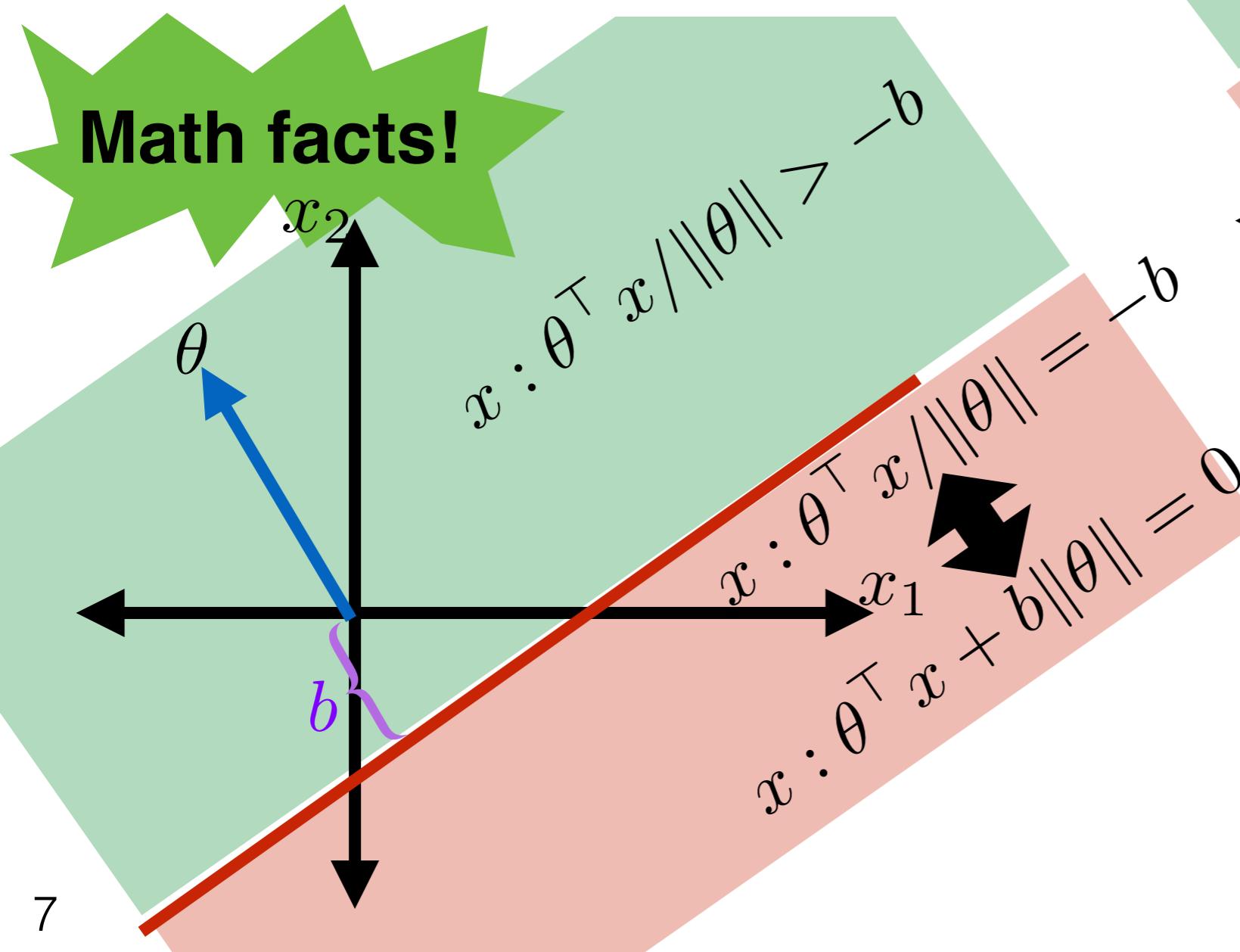
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



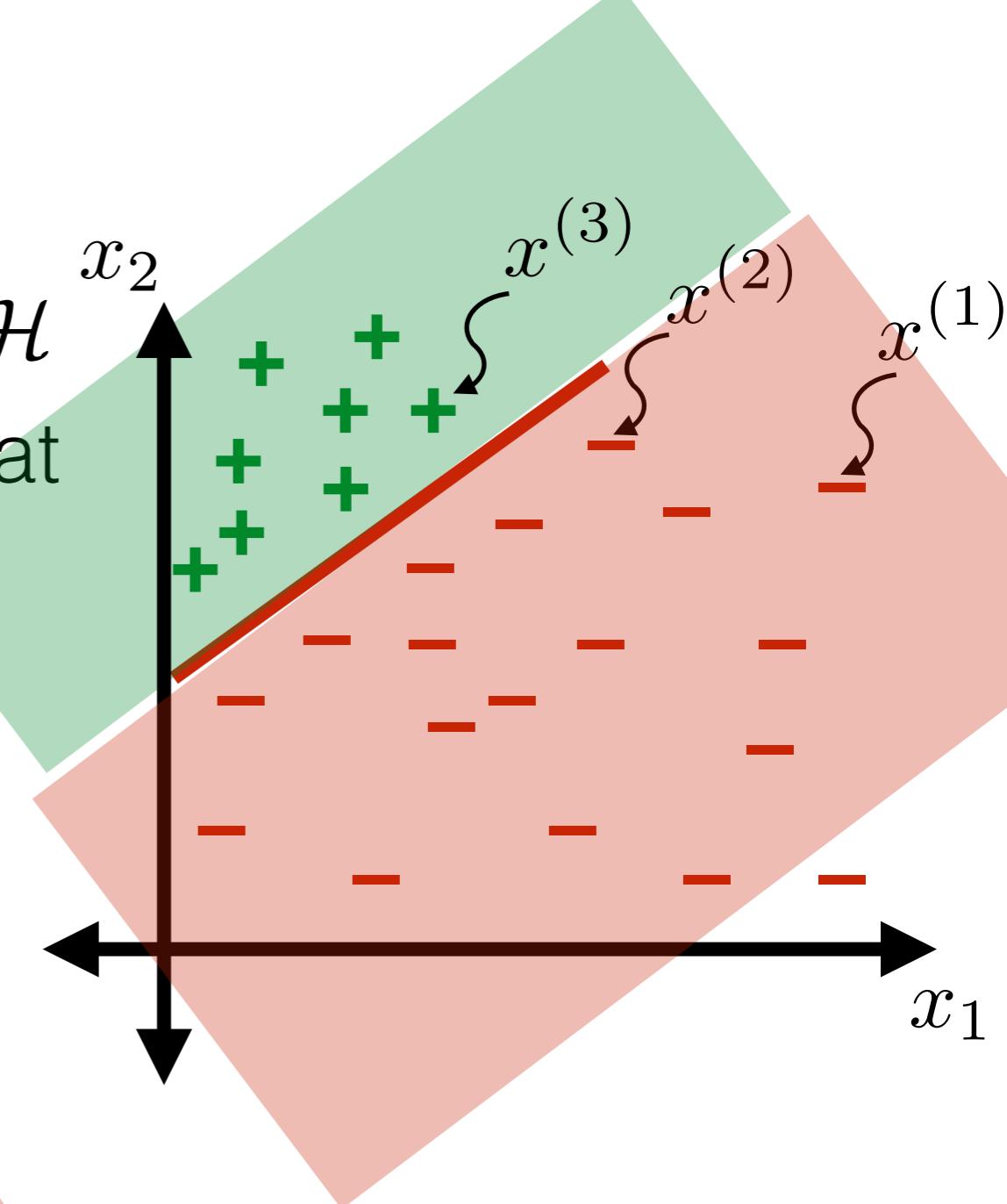
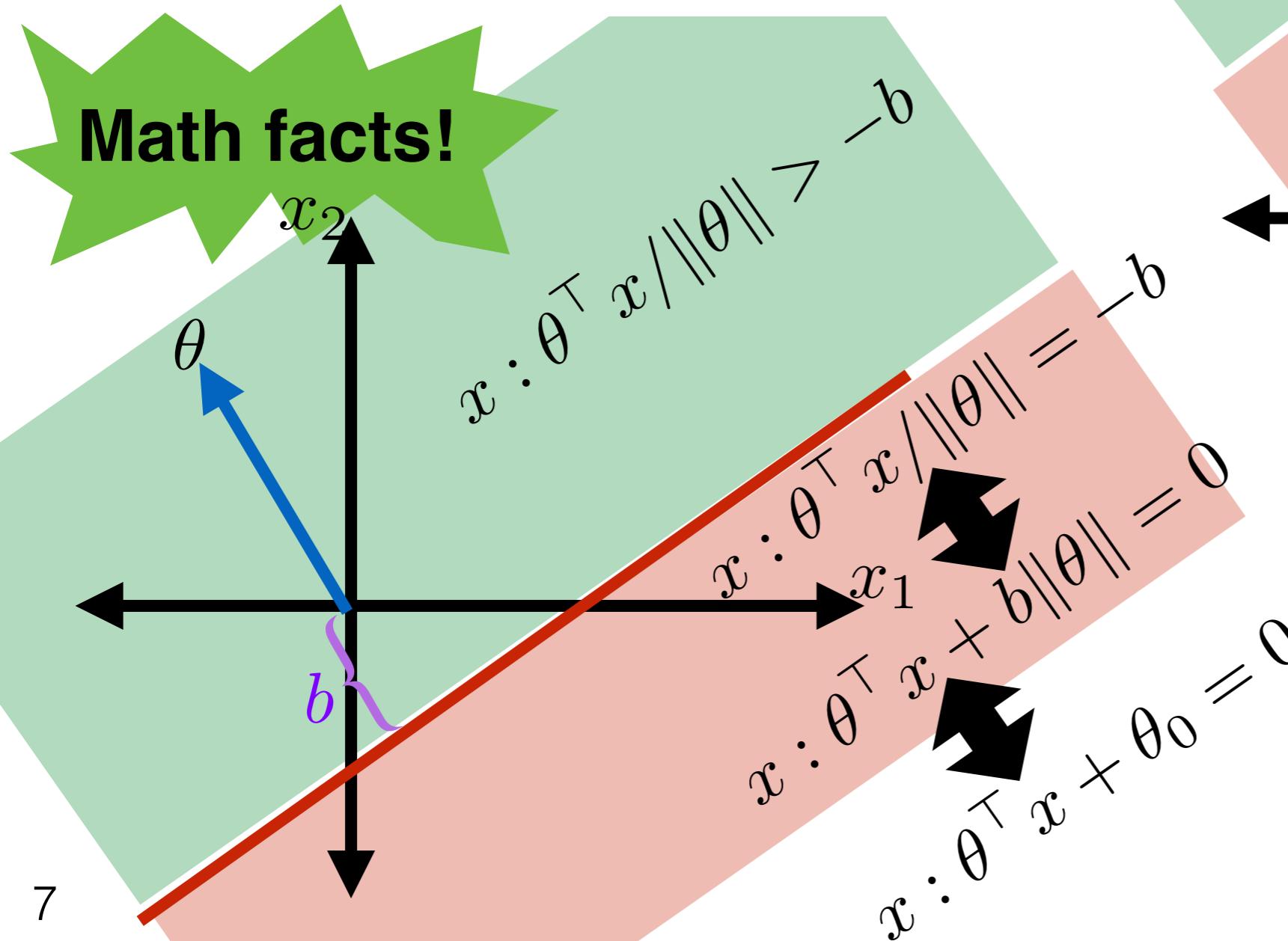
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



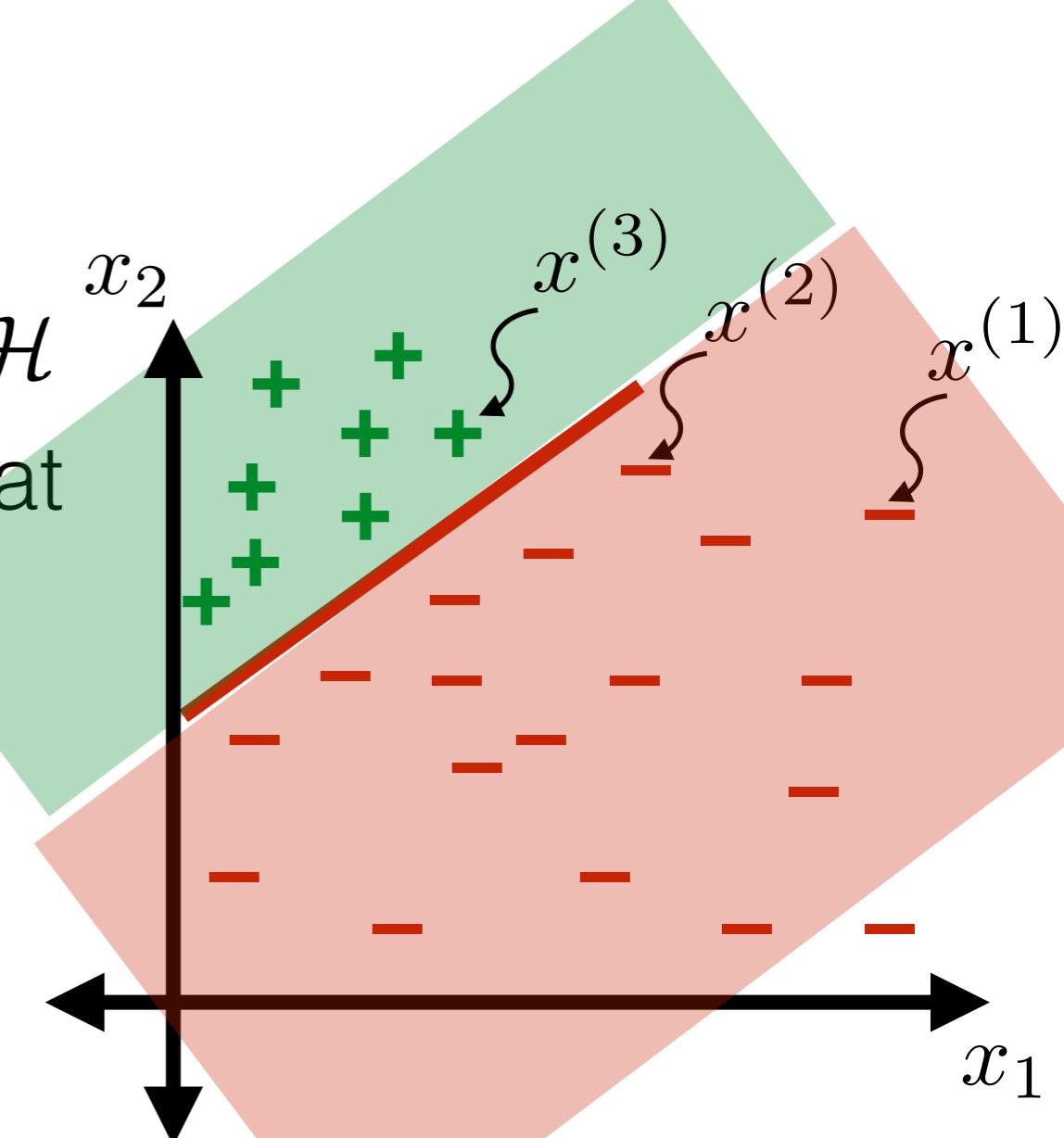
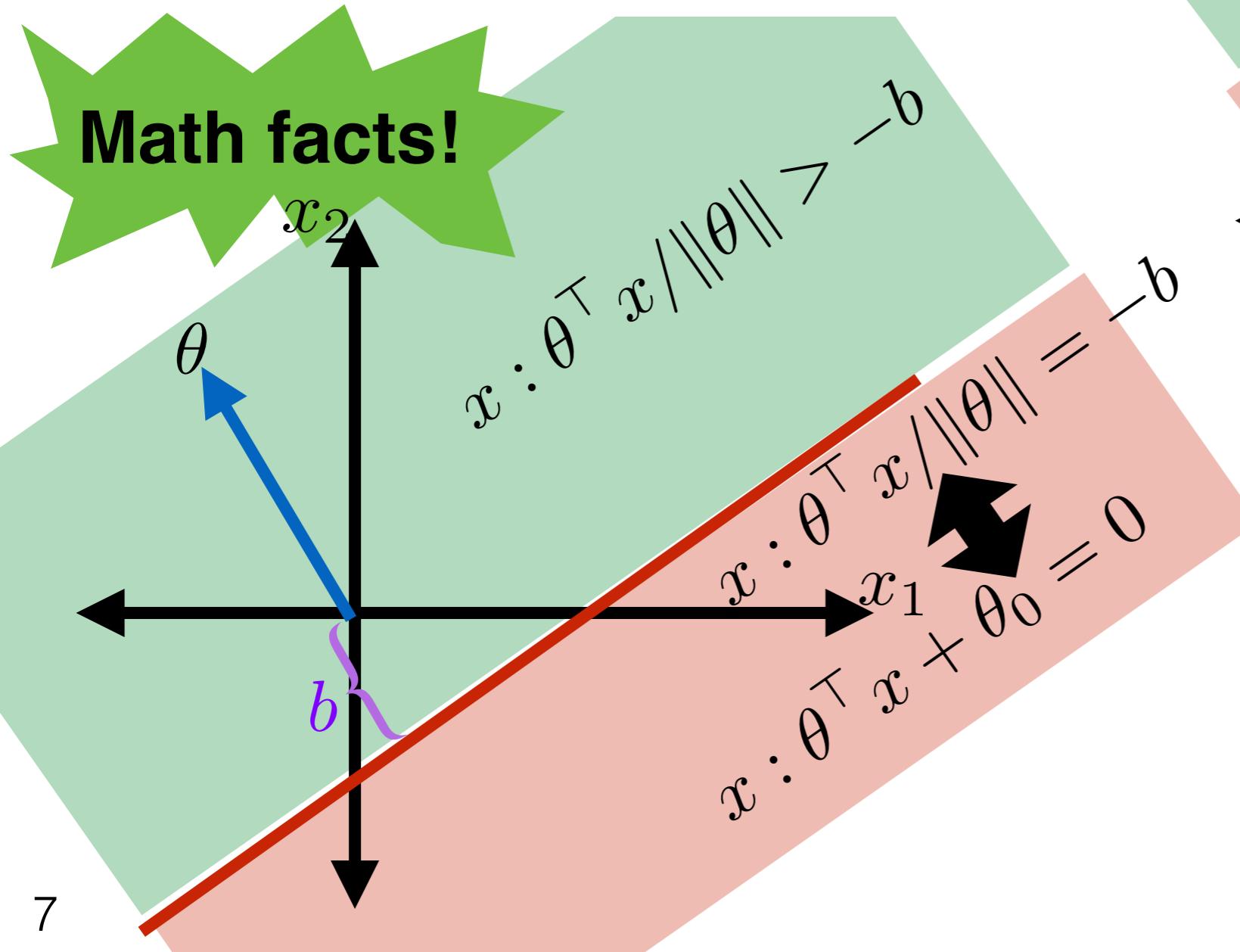
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



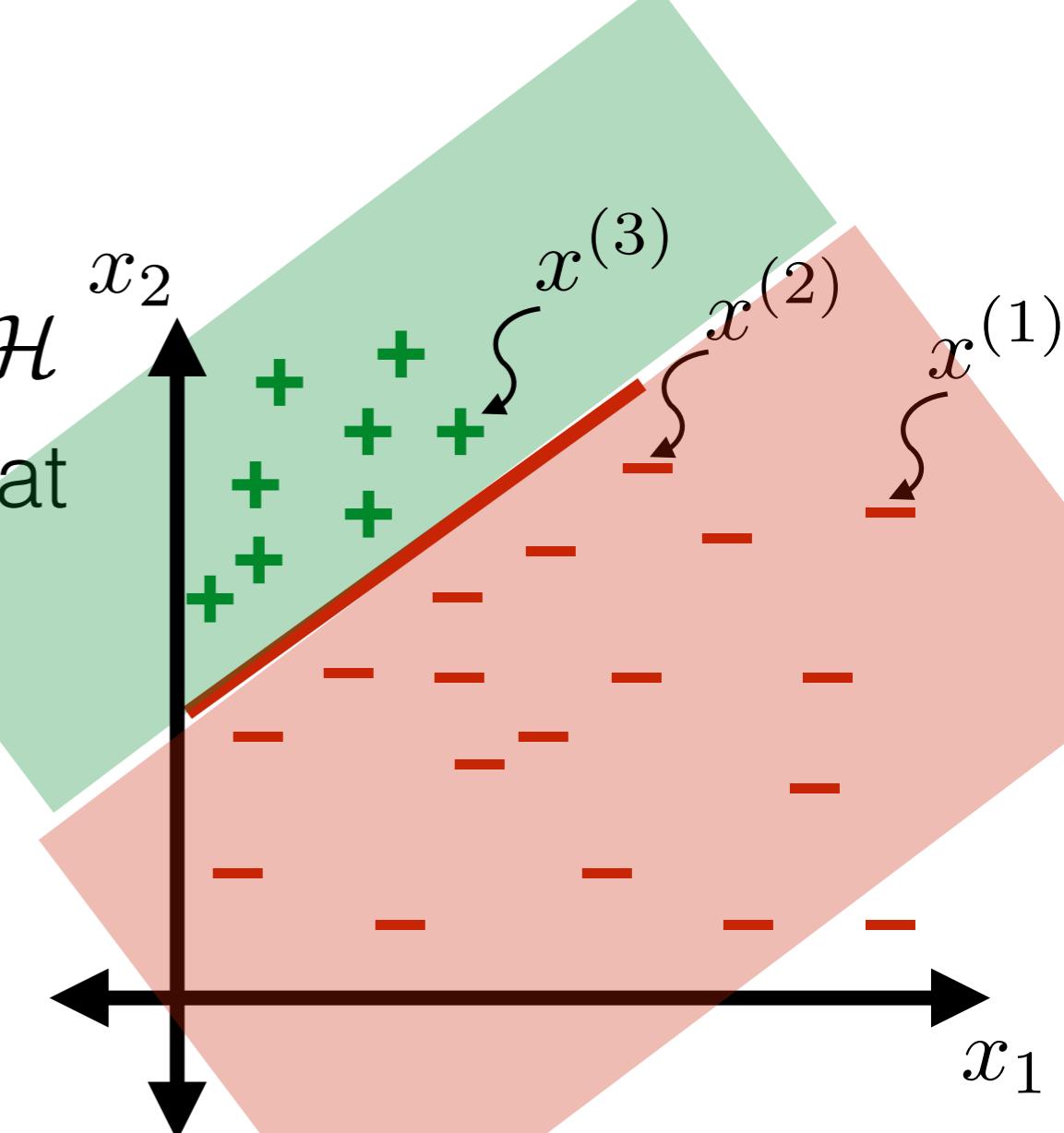
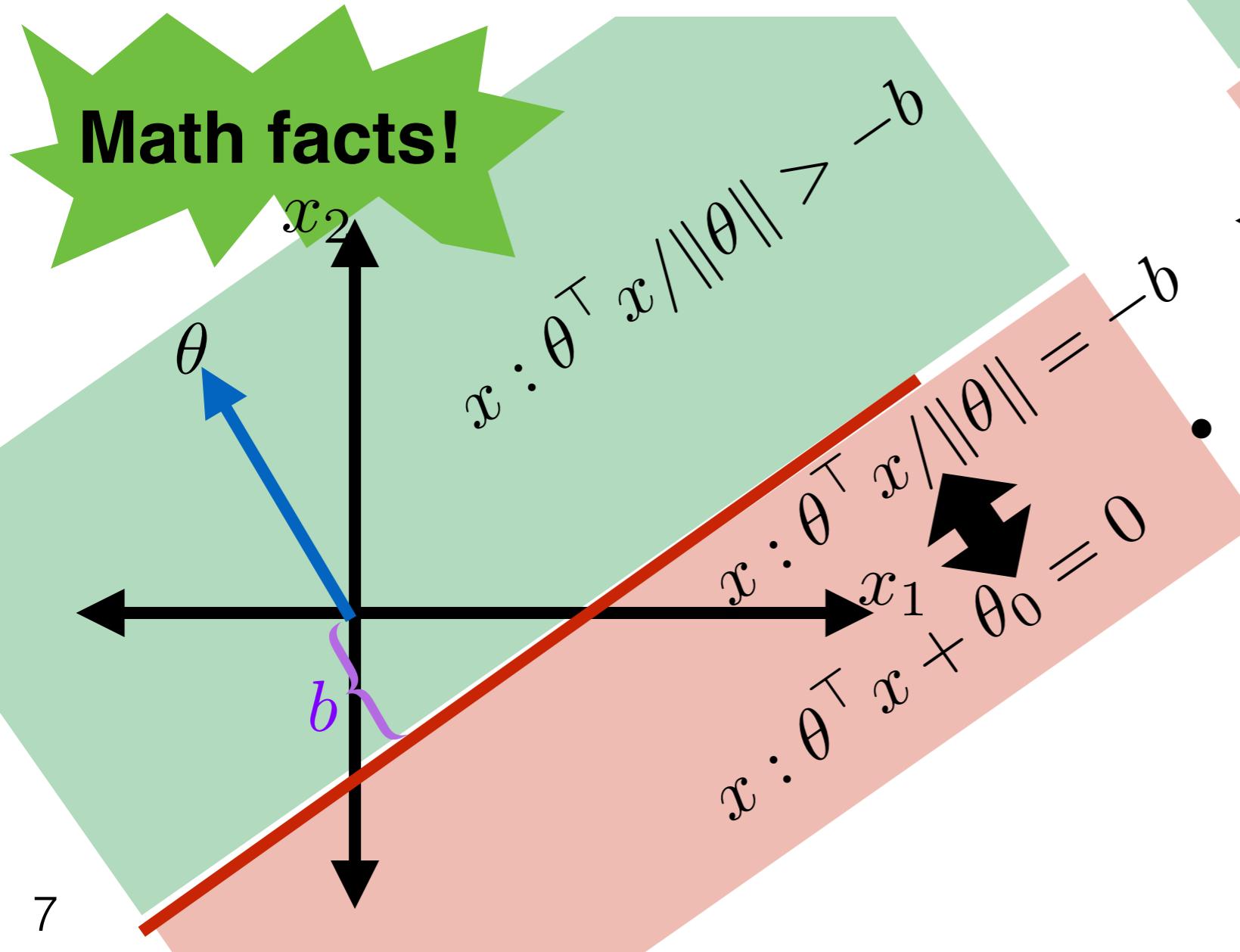
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



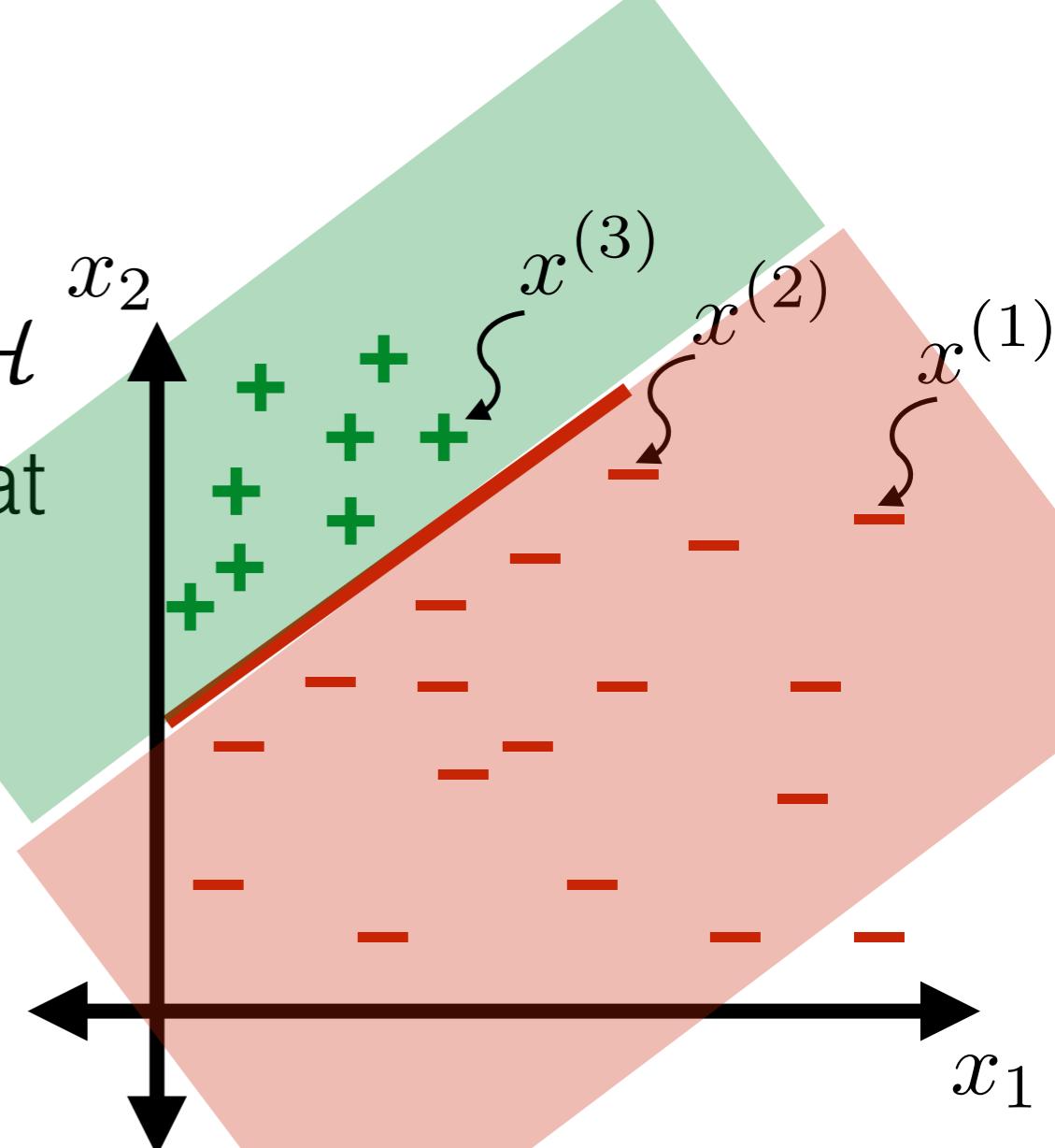
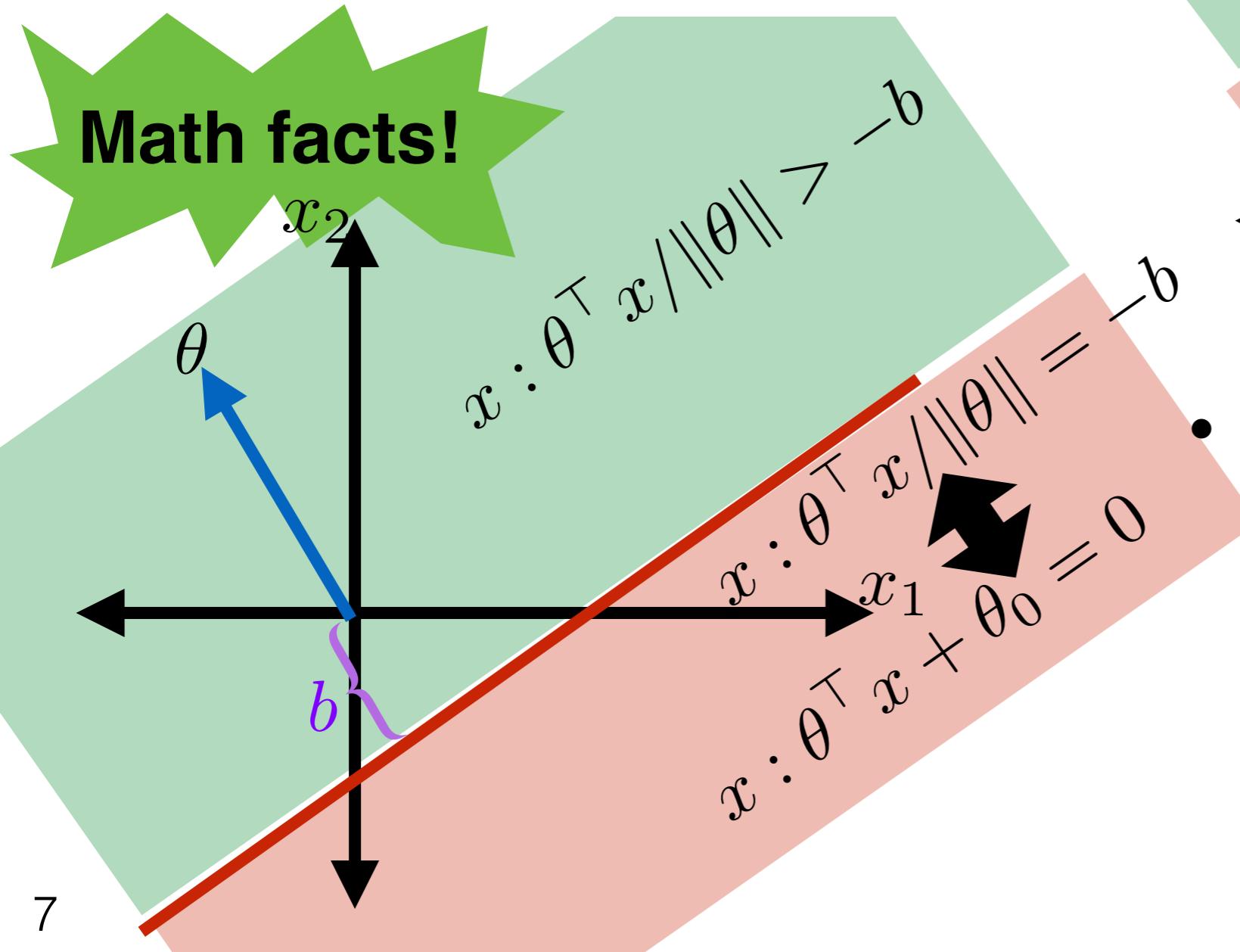
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



Linear classifiers

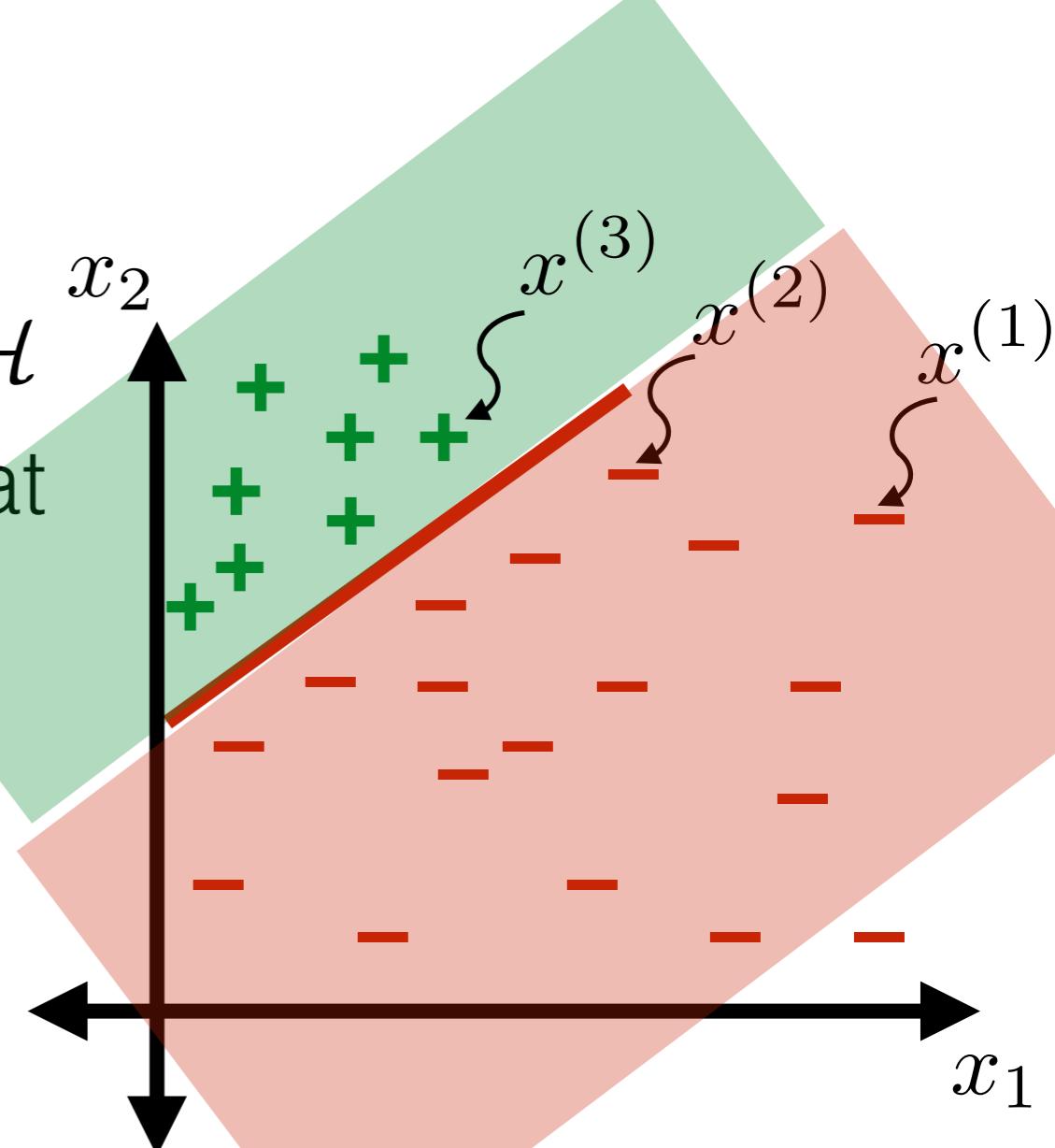
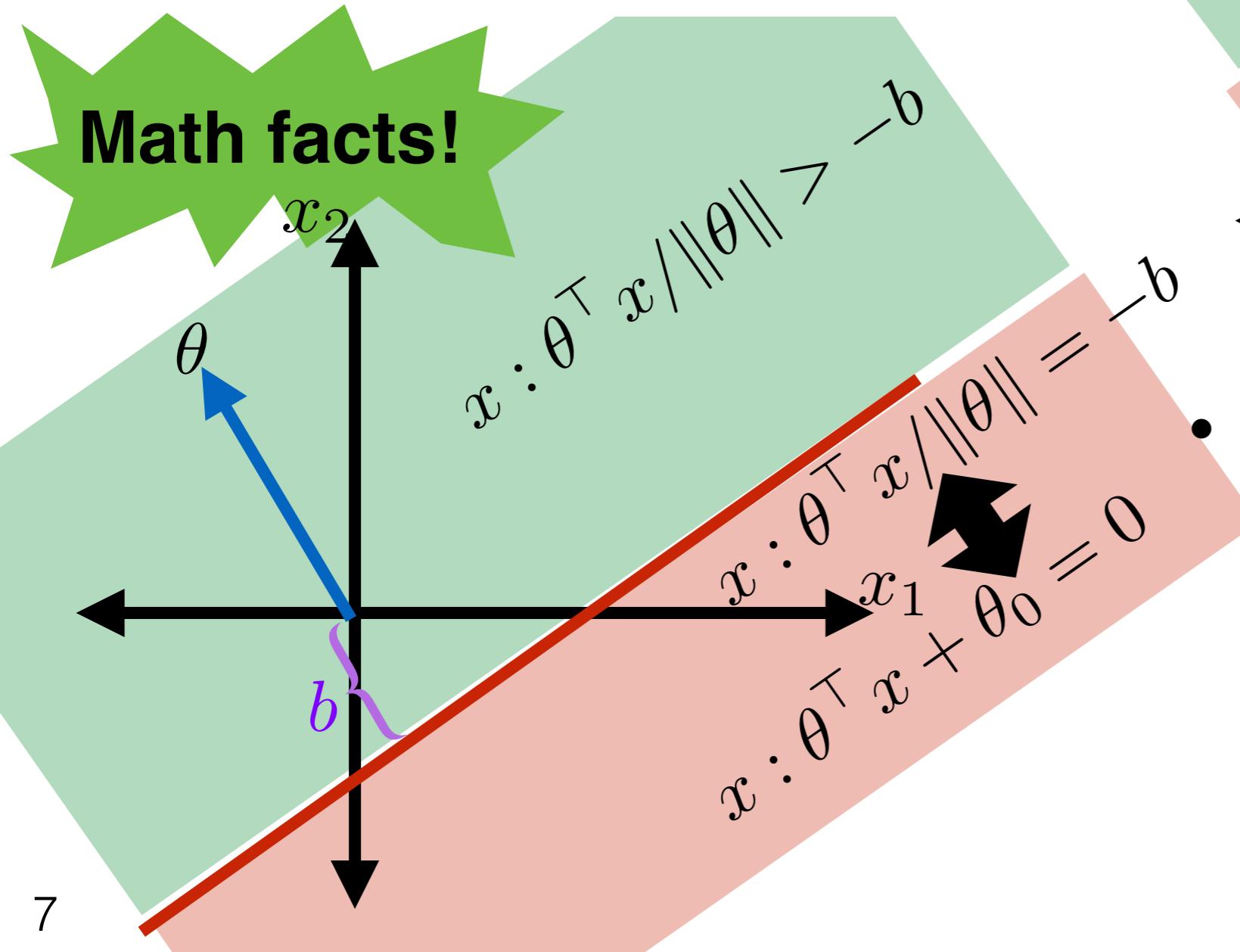
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x) = \text{sign}(\theta^\top x + \theta_0)$$

Linear classifiers

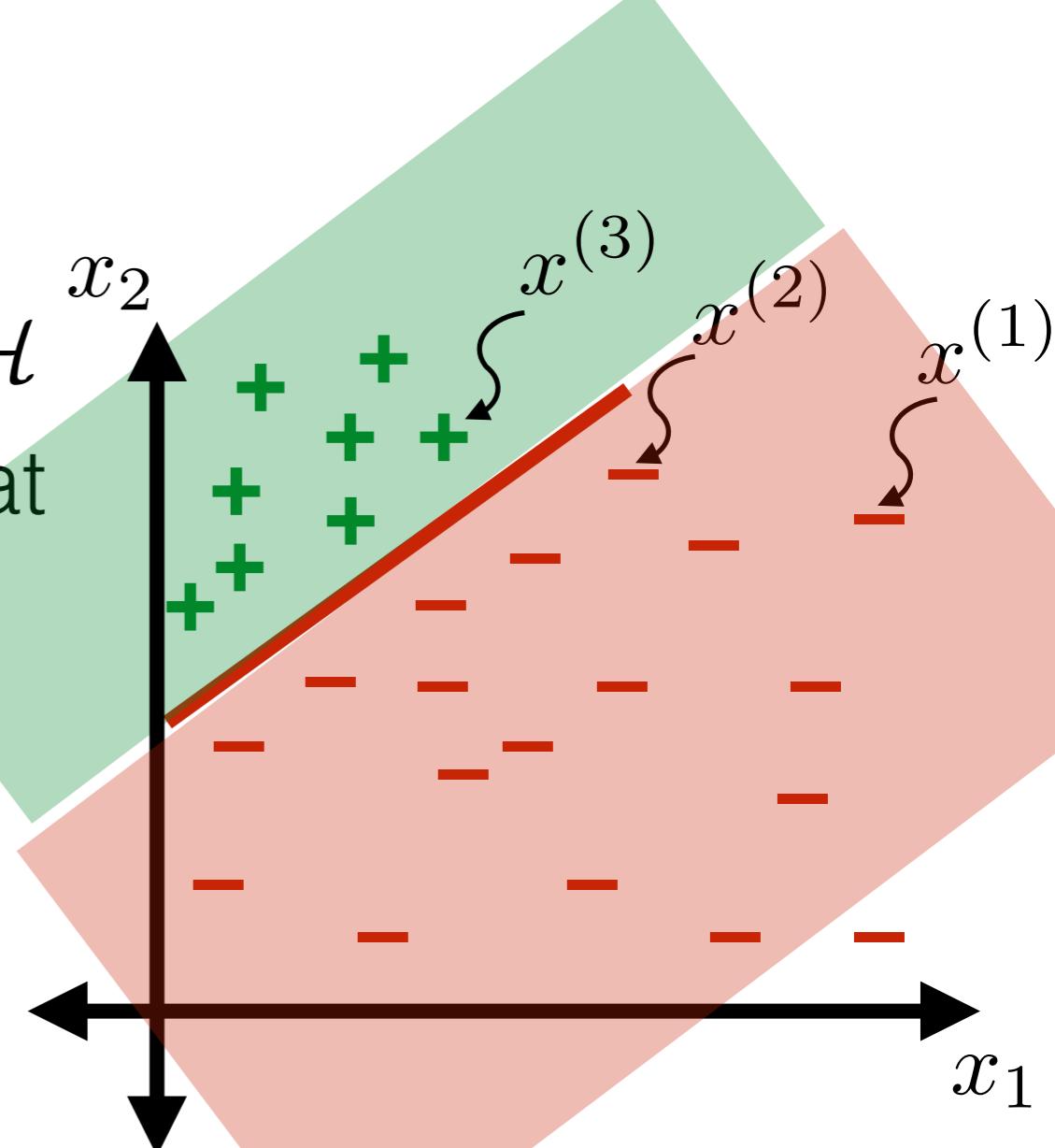
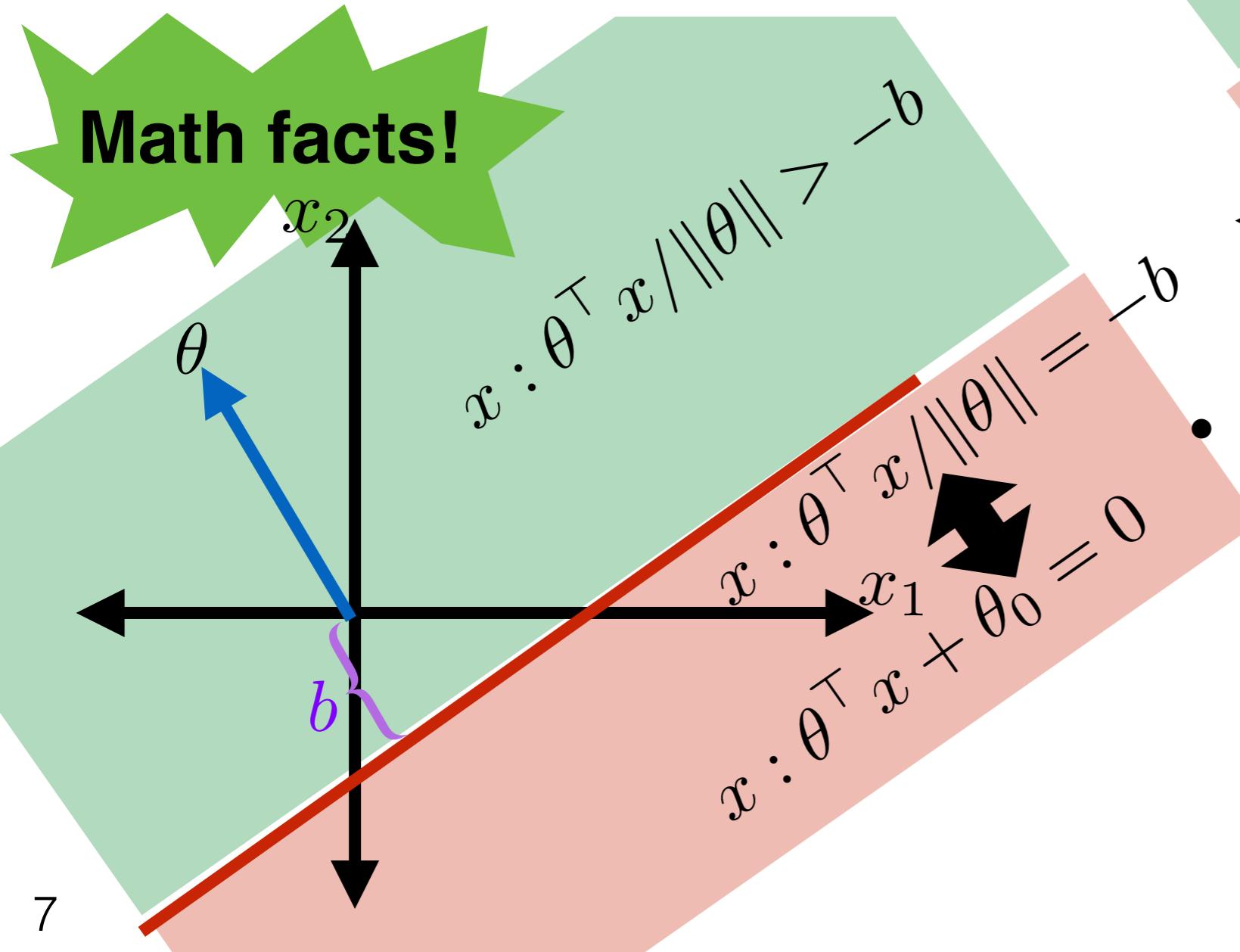
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 < 0 \end{cases}$$

Linear classifiers

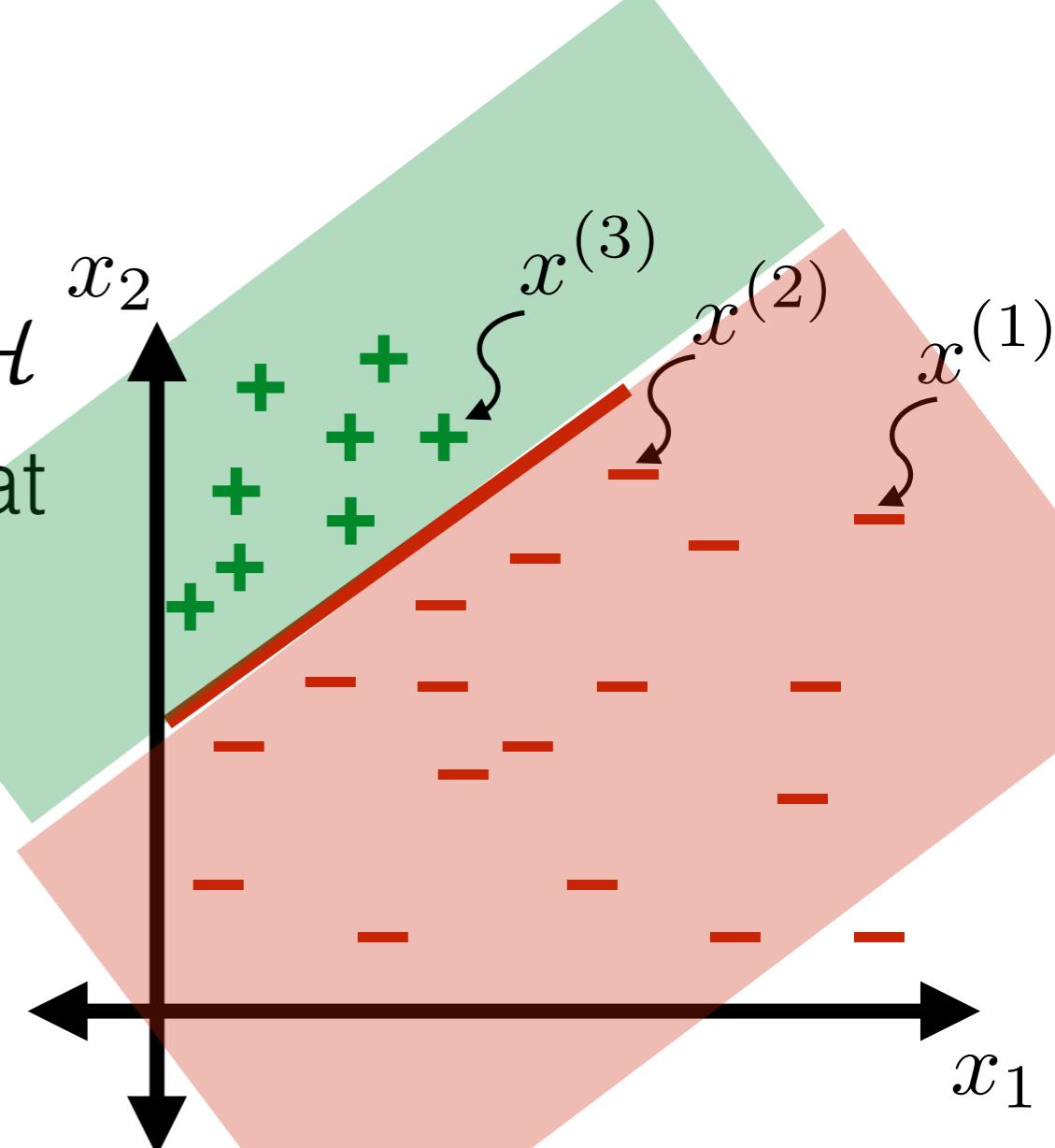
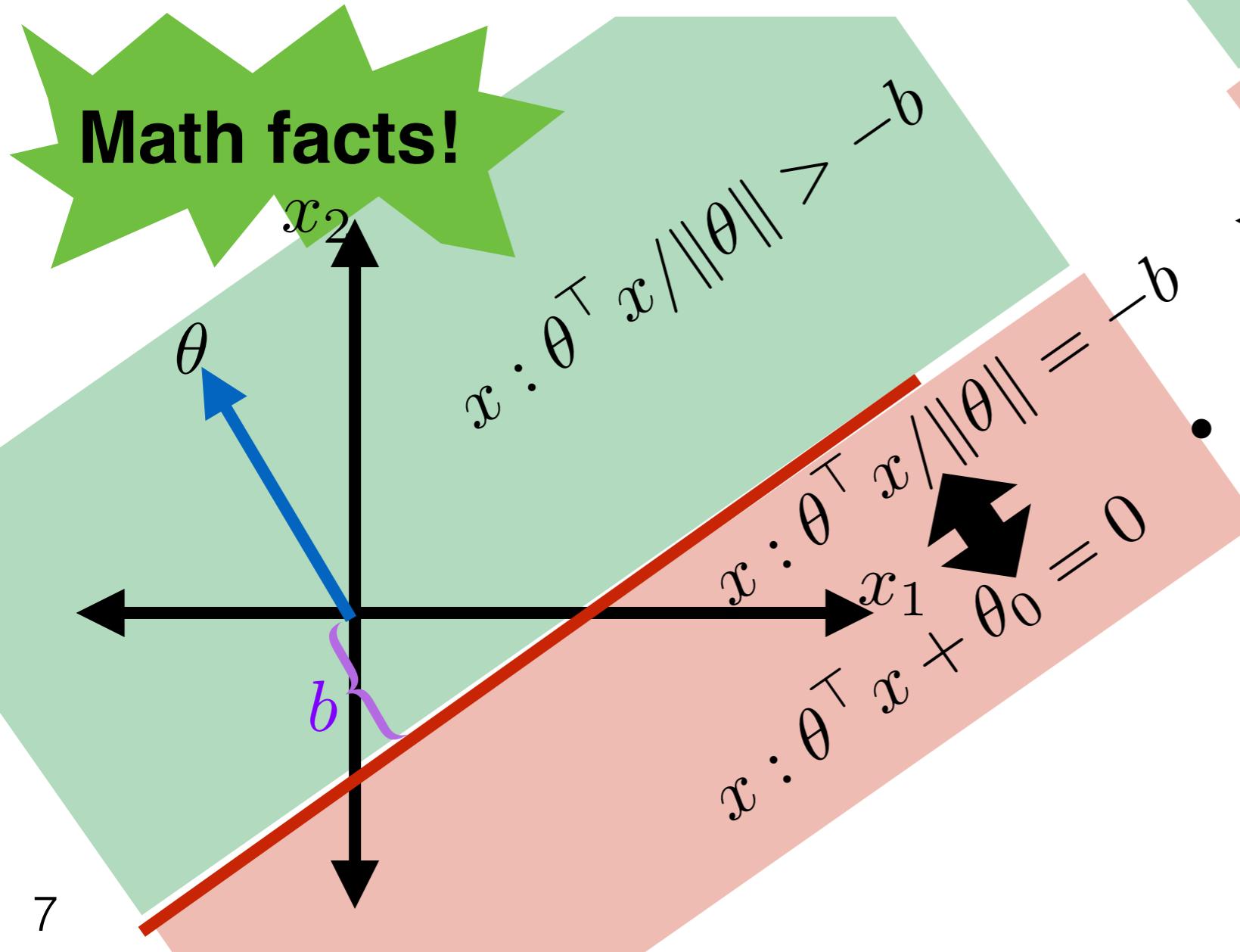
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 < 0 \end{cases}$$

Linear classifiers

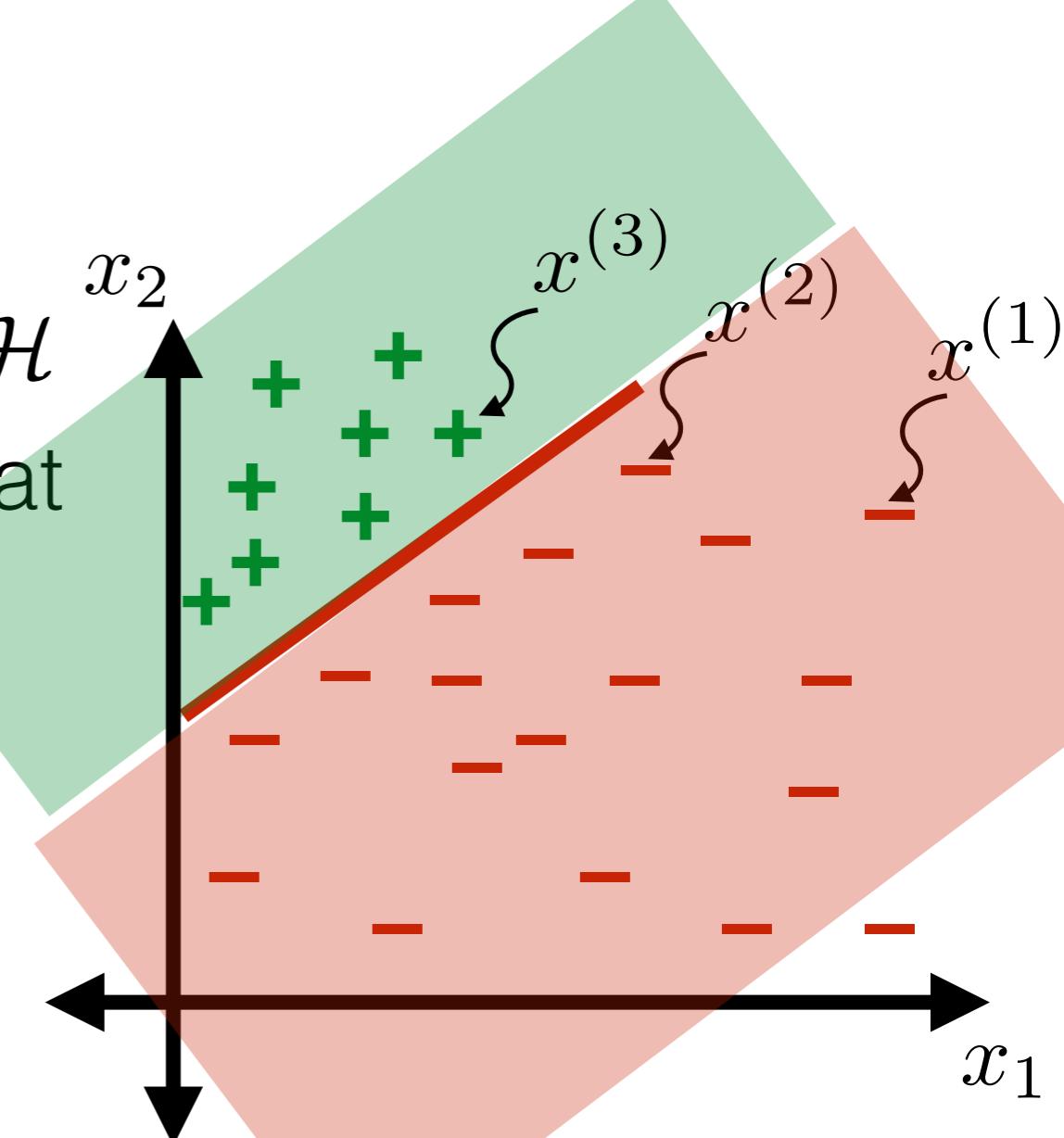
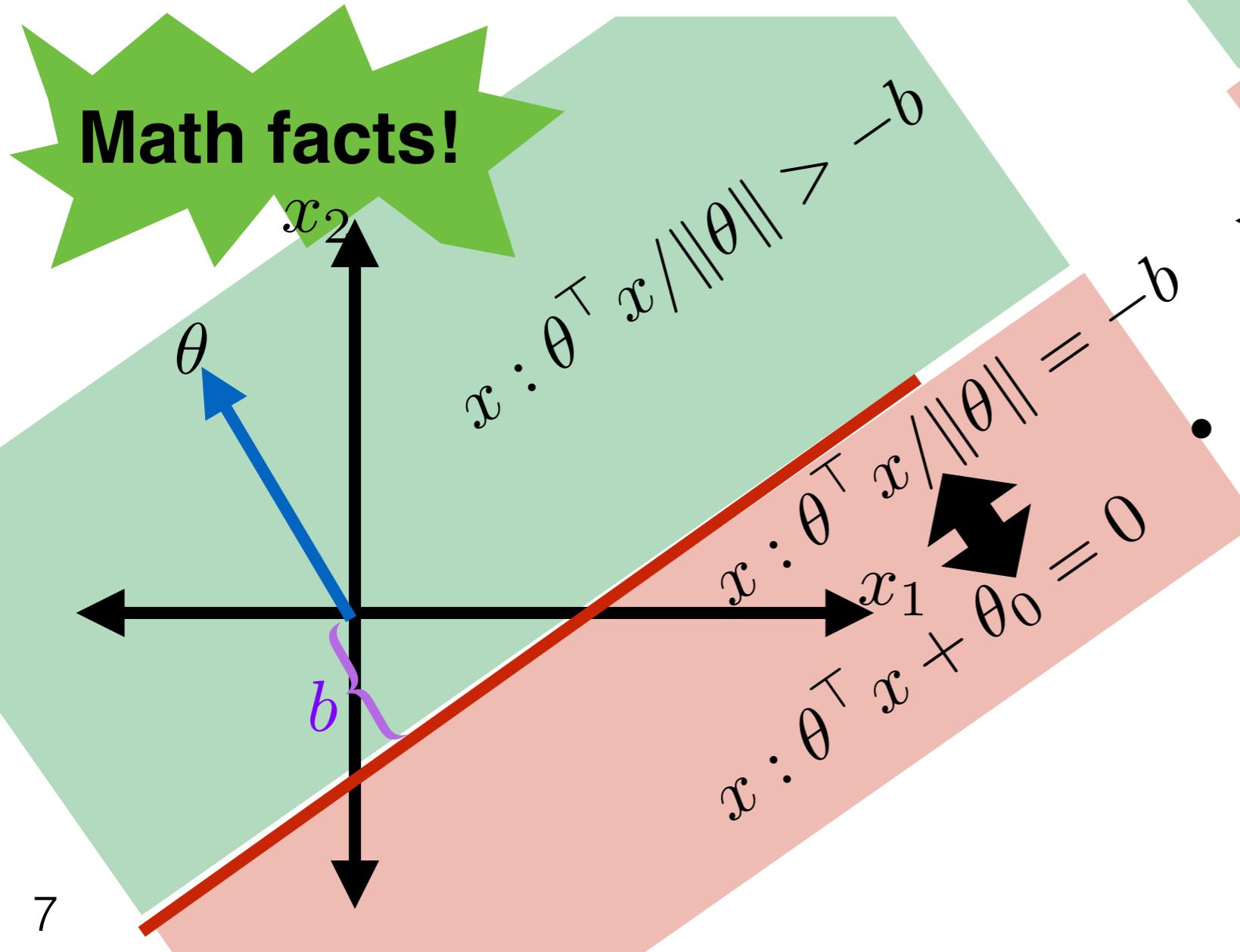
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$

Linear classifiers

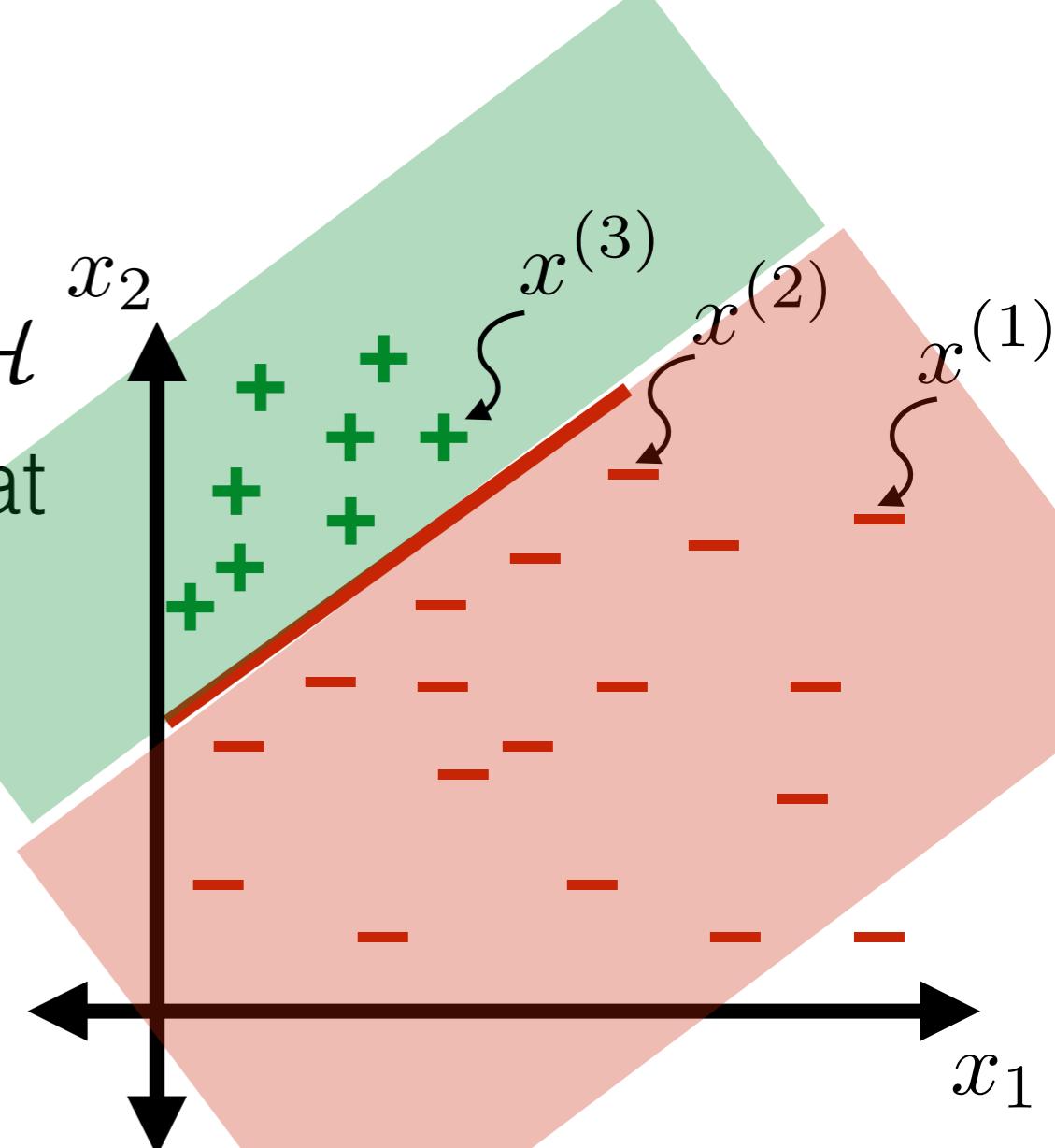
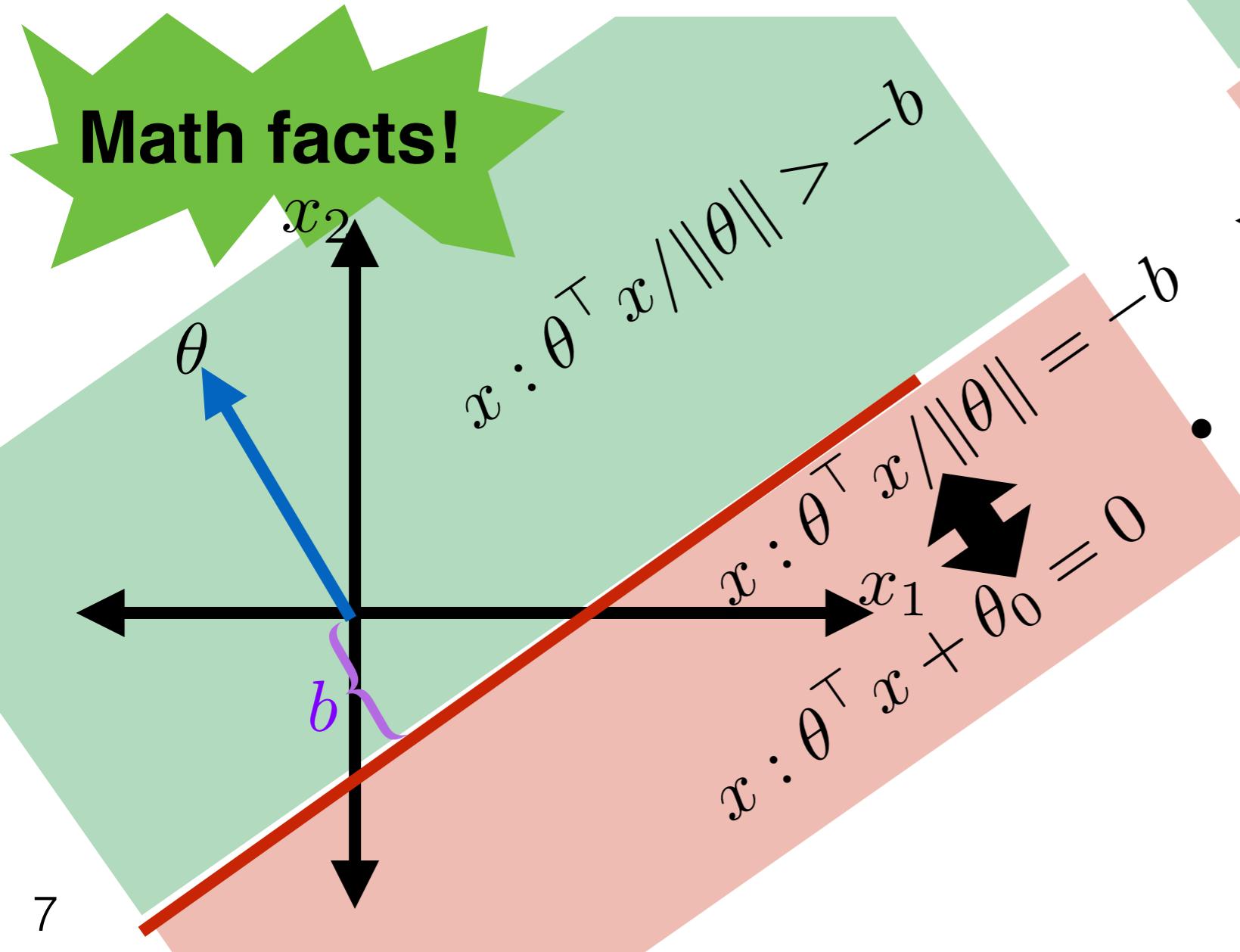
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$

Linear classifiers

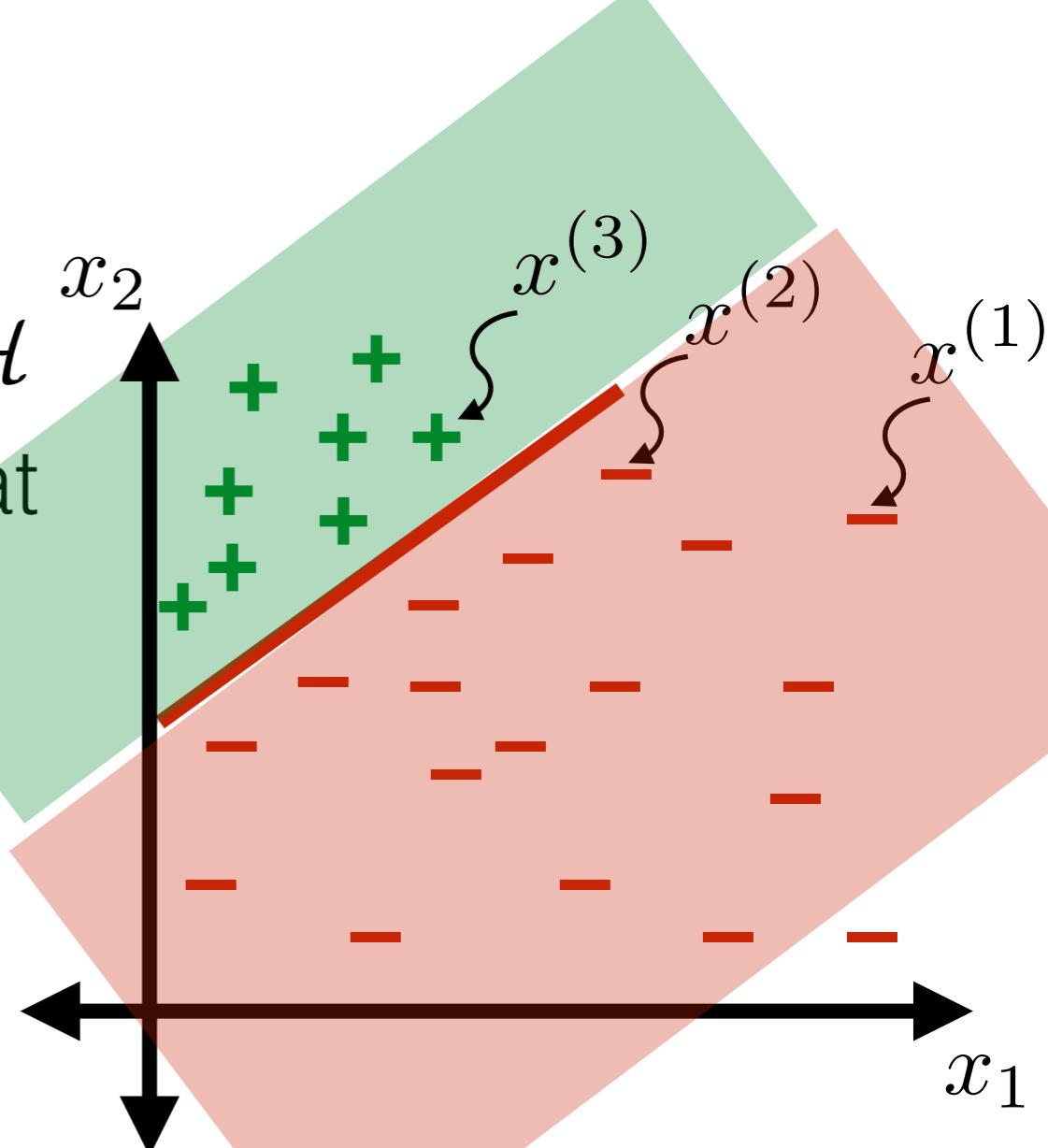
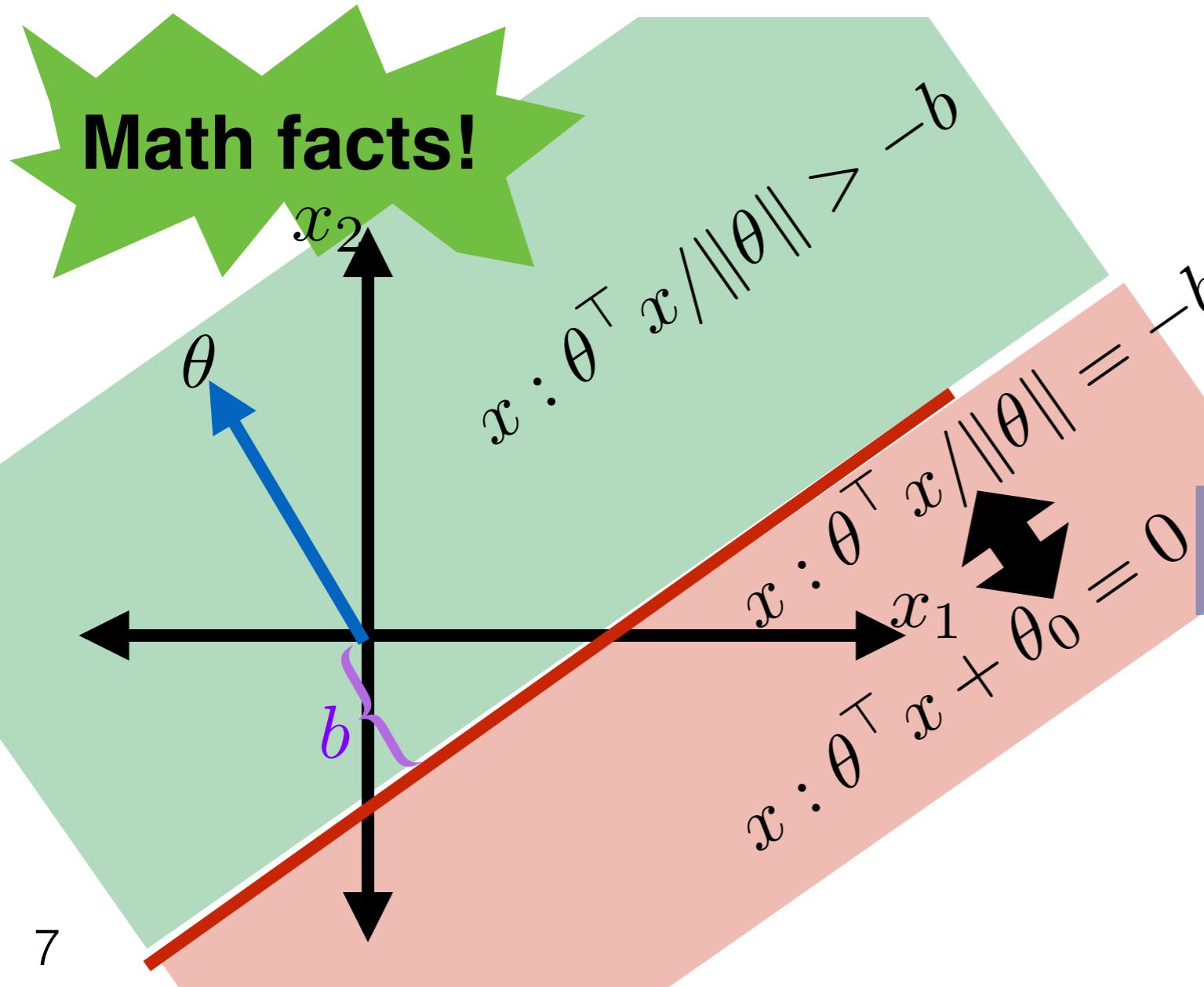
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$

Linear classifiers

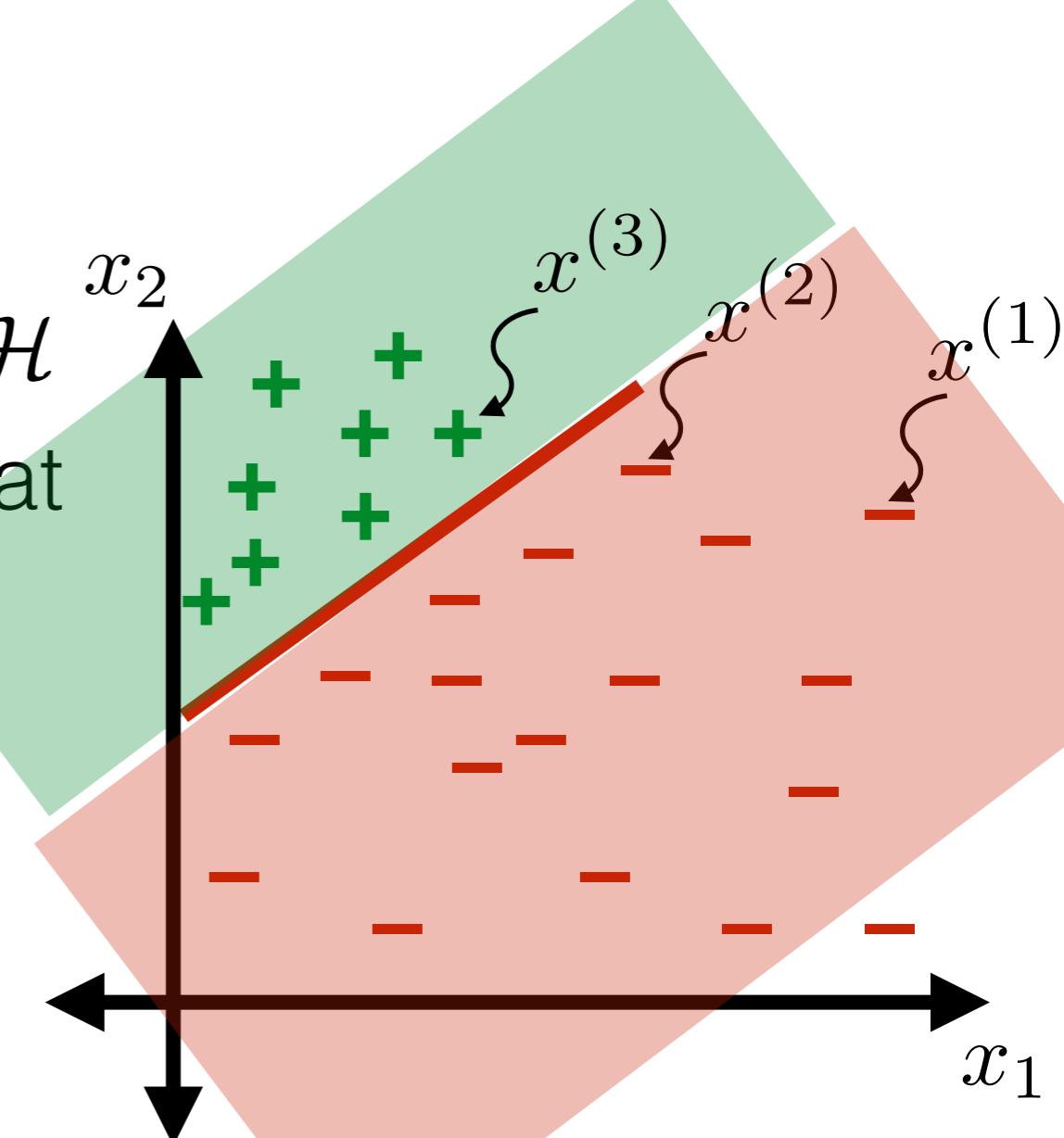
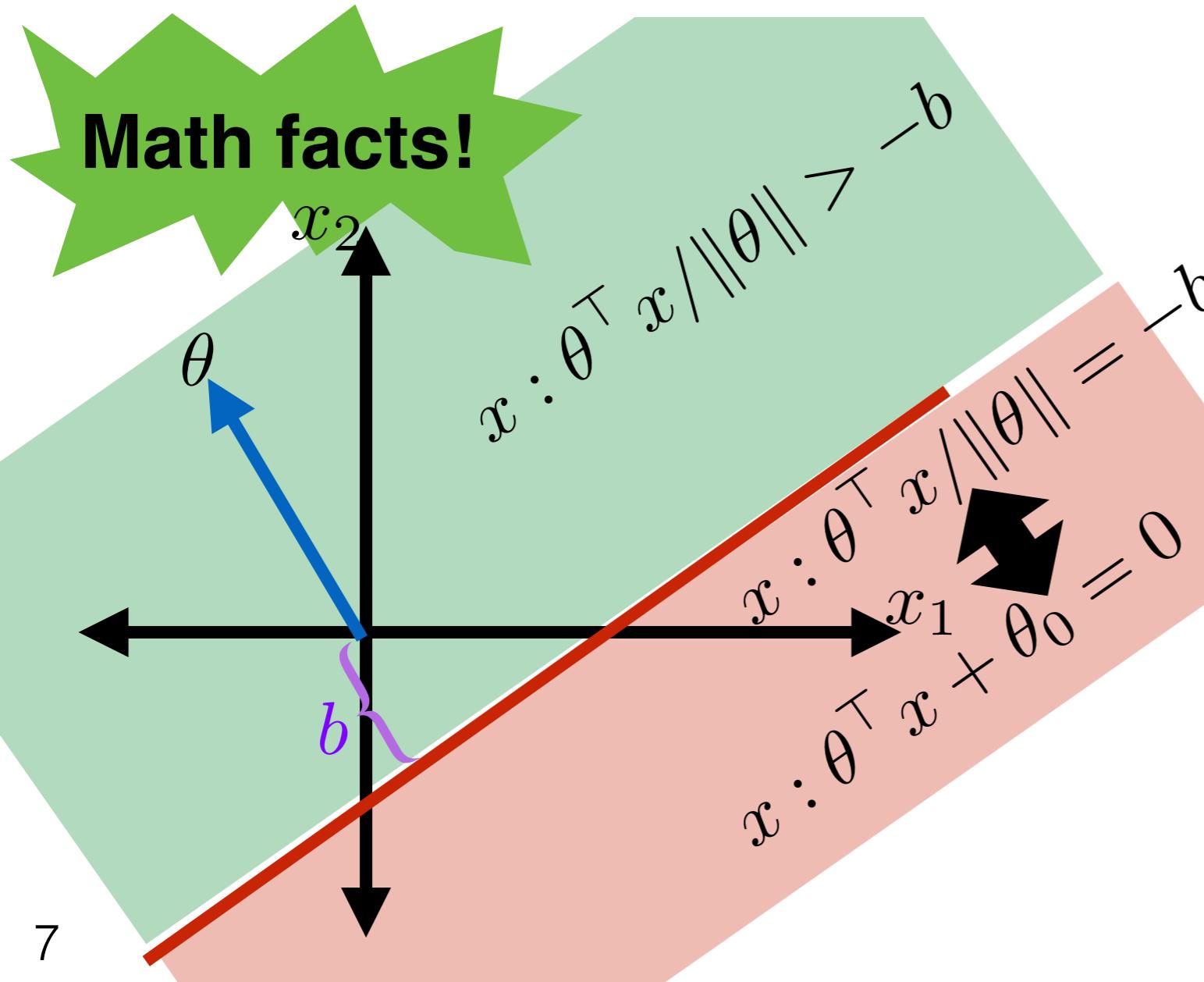
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x; \theta, \theta_0) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$

Linear classifiers

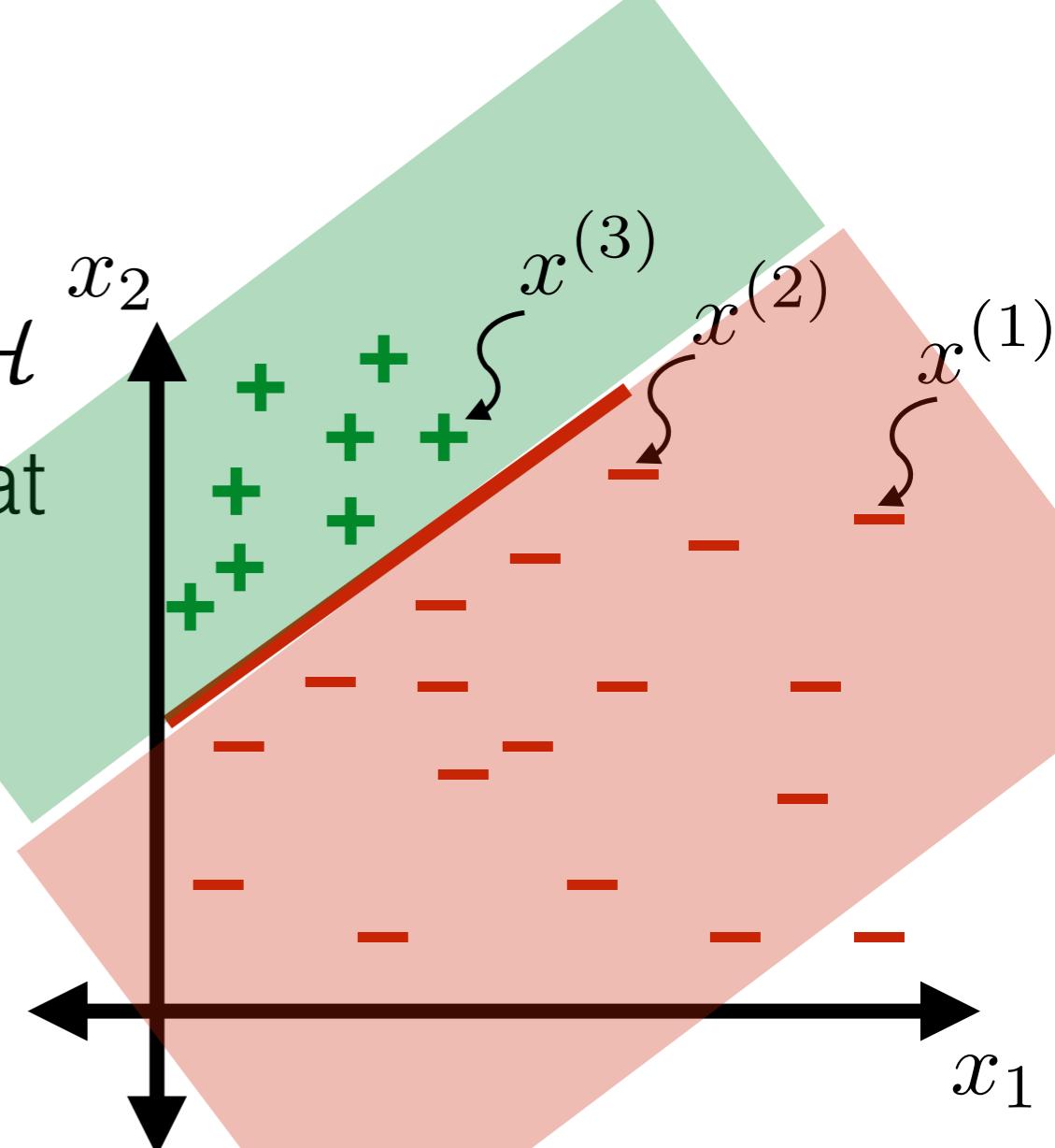
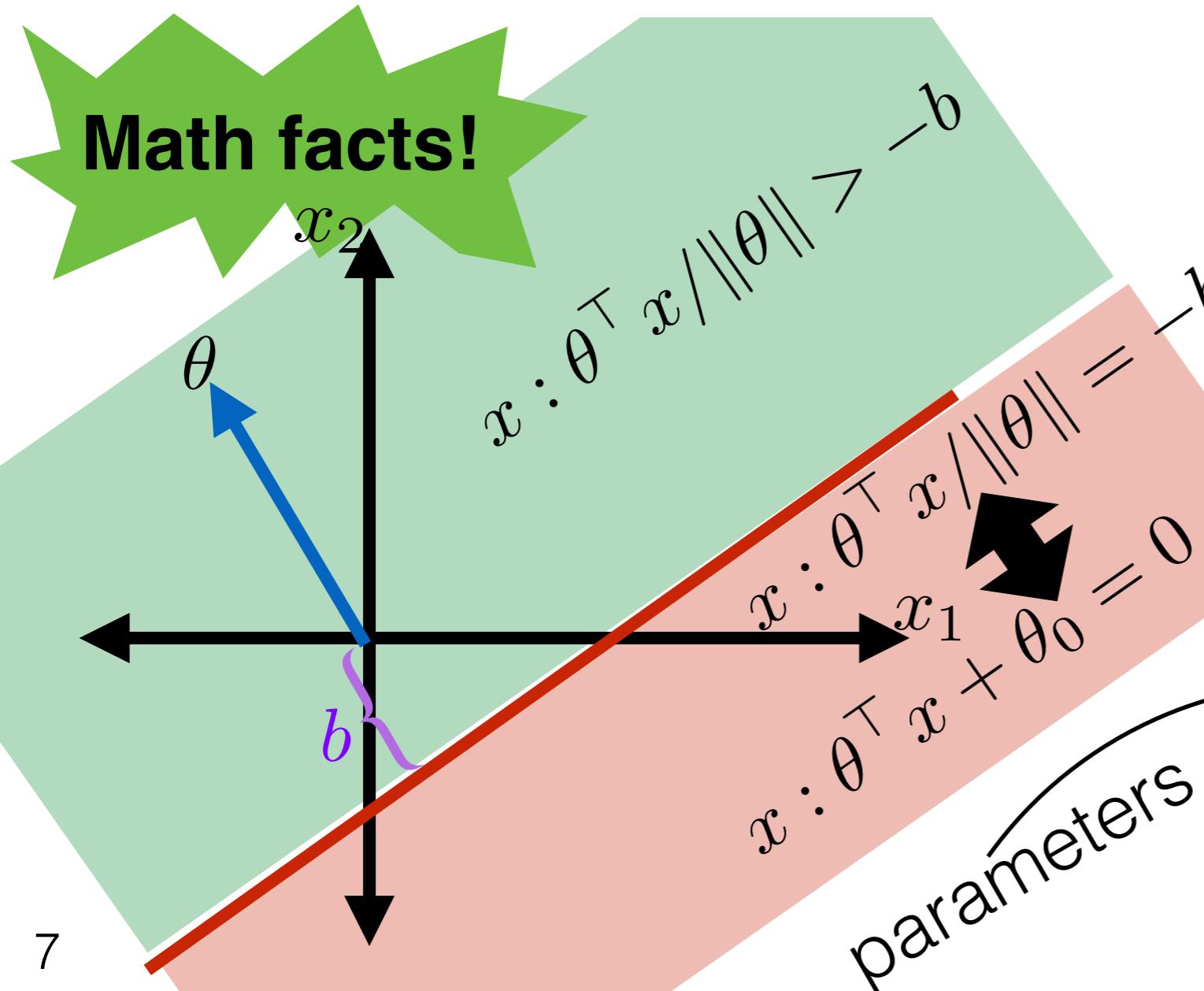
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x; \theta, \theta_0) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$

Linear classifiers

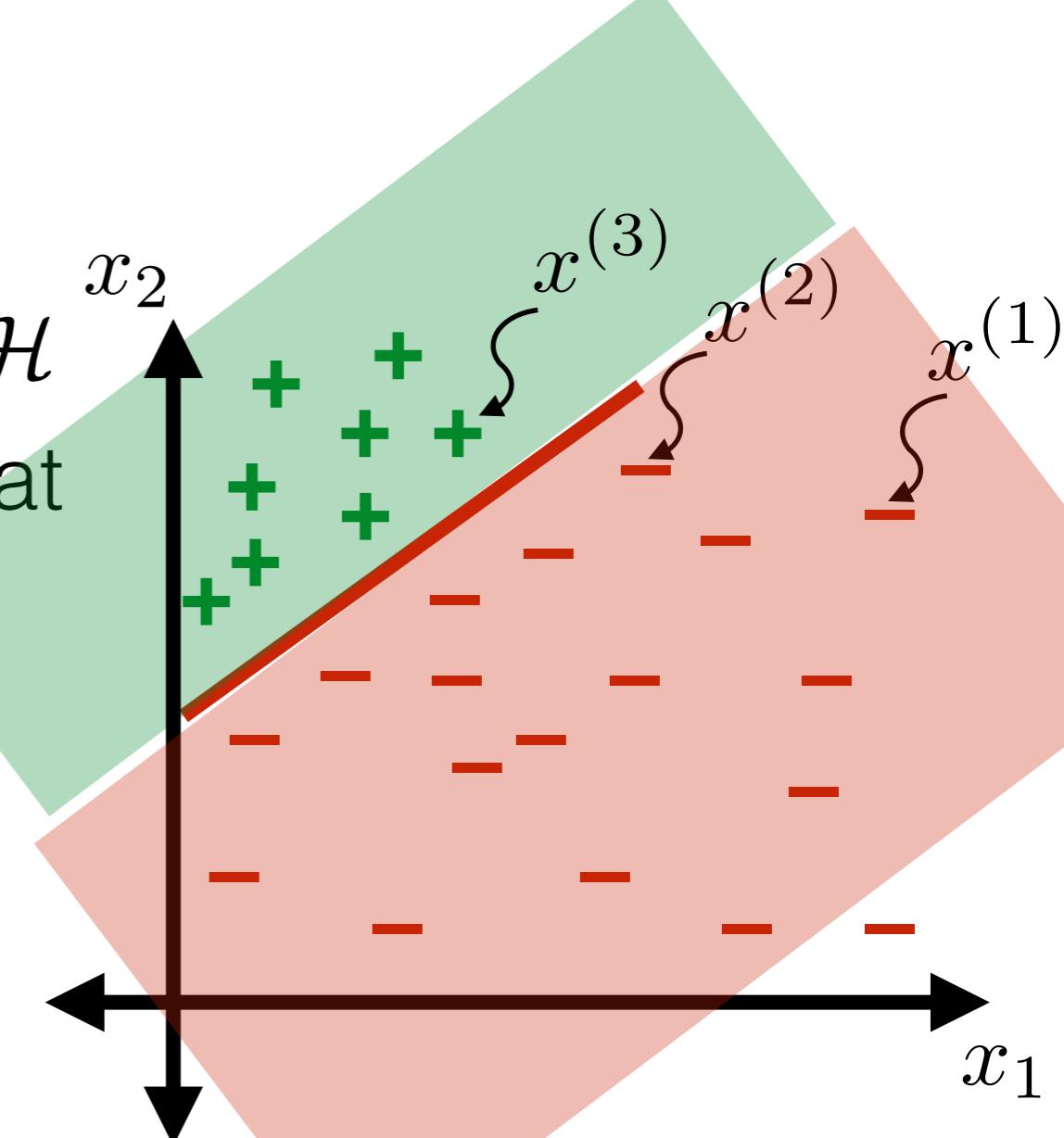
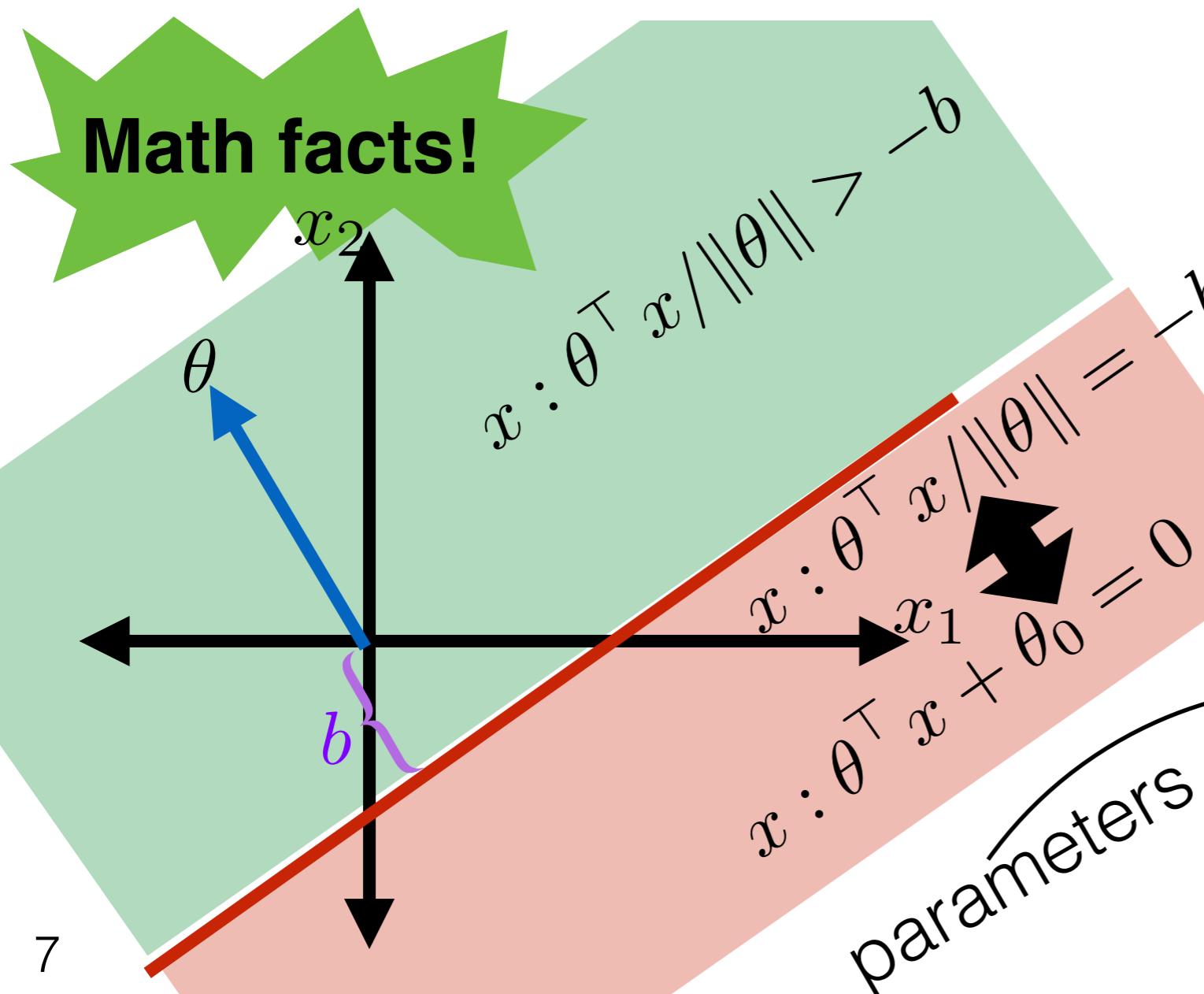
- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



- Linear classifier:
$$h(x; \theta, \theta_0) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$

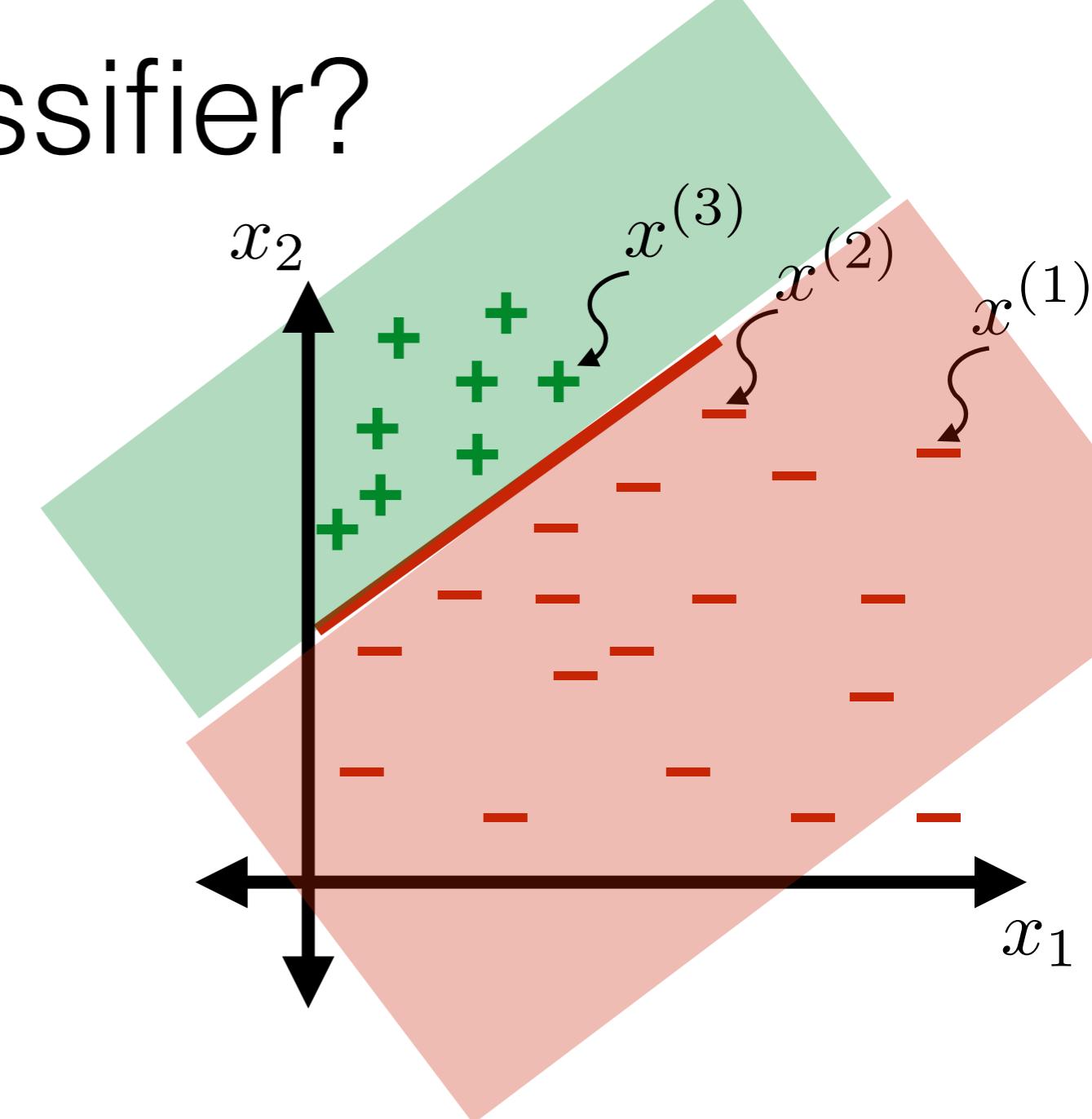
Linear classifiers

- Hypothesis class \mathcal{H} : set of $h \in \mathcal{H}$
 - Example \mathcal{H} : All hypotheses that label +1 on one side of a line and -1 on the other side



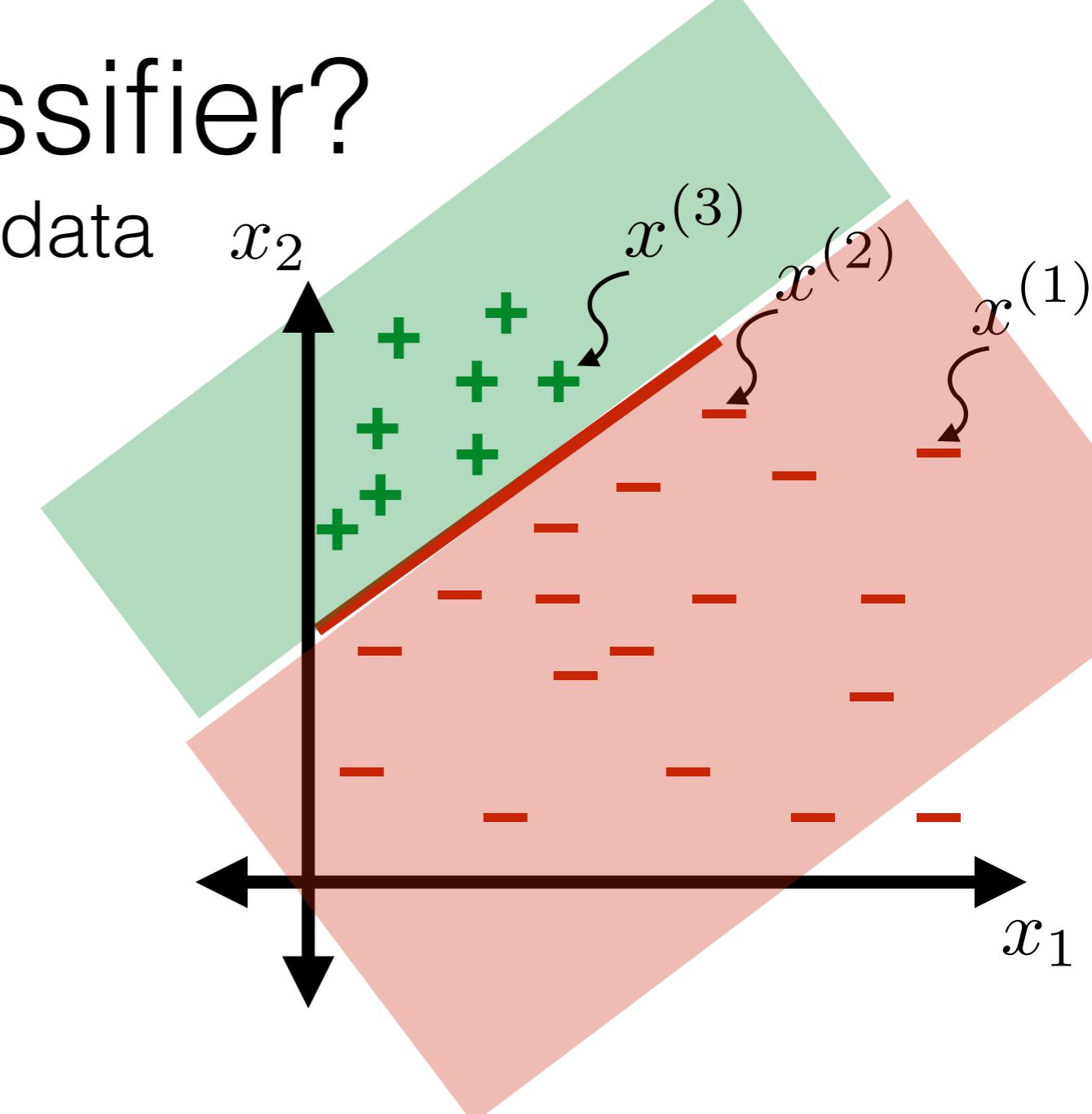
- Linear classifier:
$$h(x; \theta, \theta_0) = \text{sign}(\theta^\top x + \theta_0)$$
$$= \begin{cases} +1 & \text{if } \theta^\top x + \theta_0 > 0 \\ -1 & \text{if } \theta^\top x + \theta_0 \leq 0 \end{cases}$$
- $\mathcal{H} = \text{set of all such } h$

How good is a classifier?



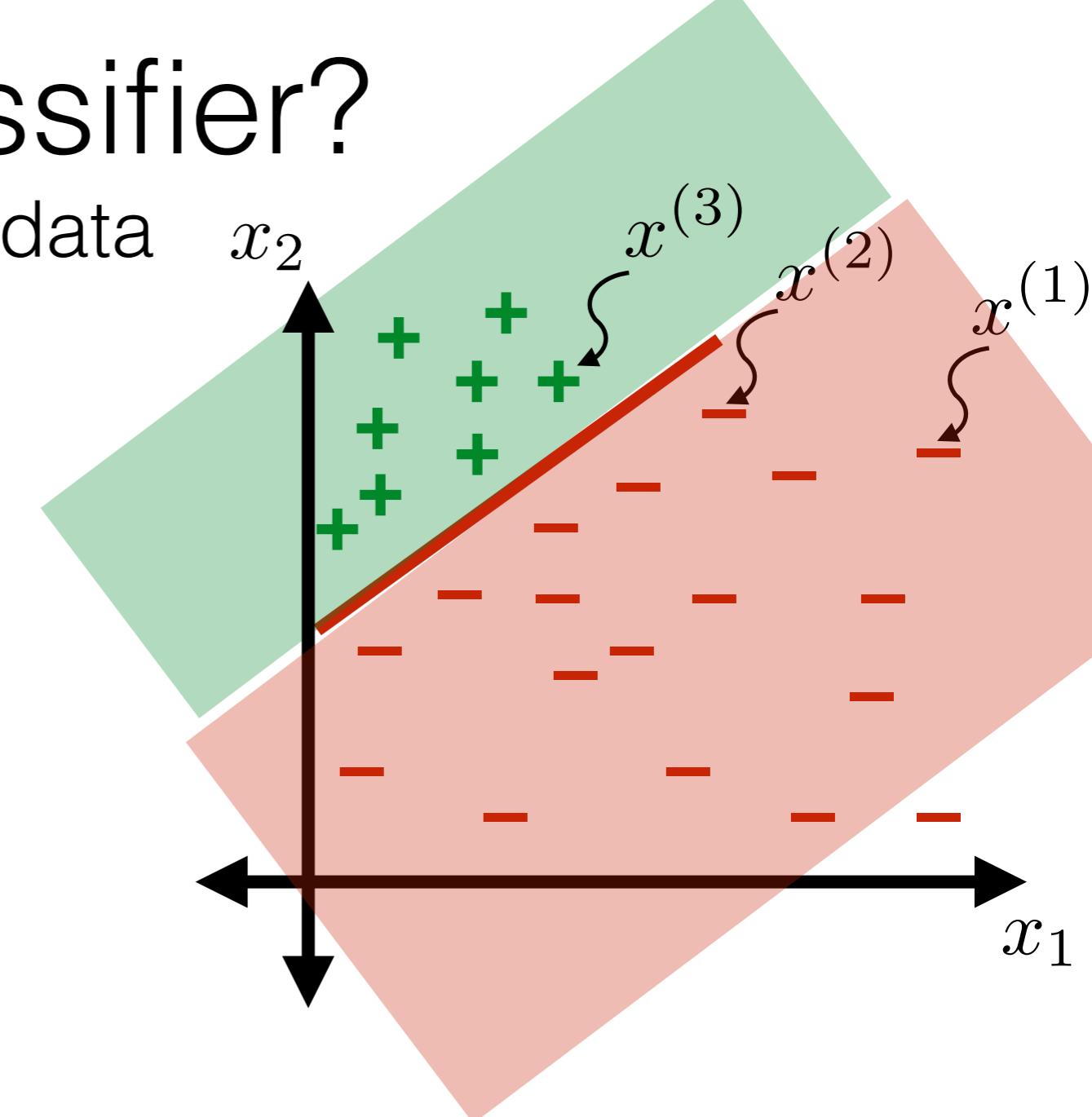
How good is a classifier?

- Should predict well on future data



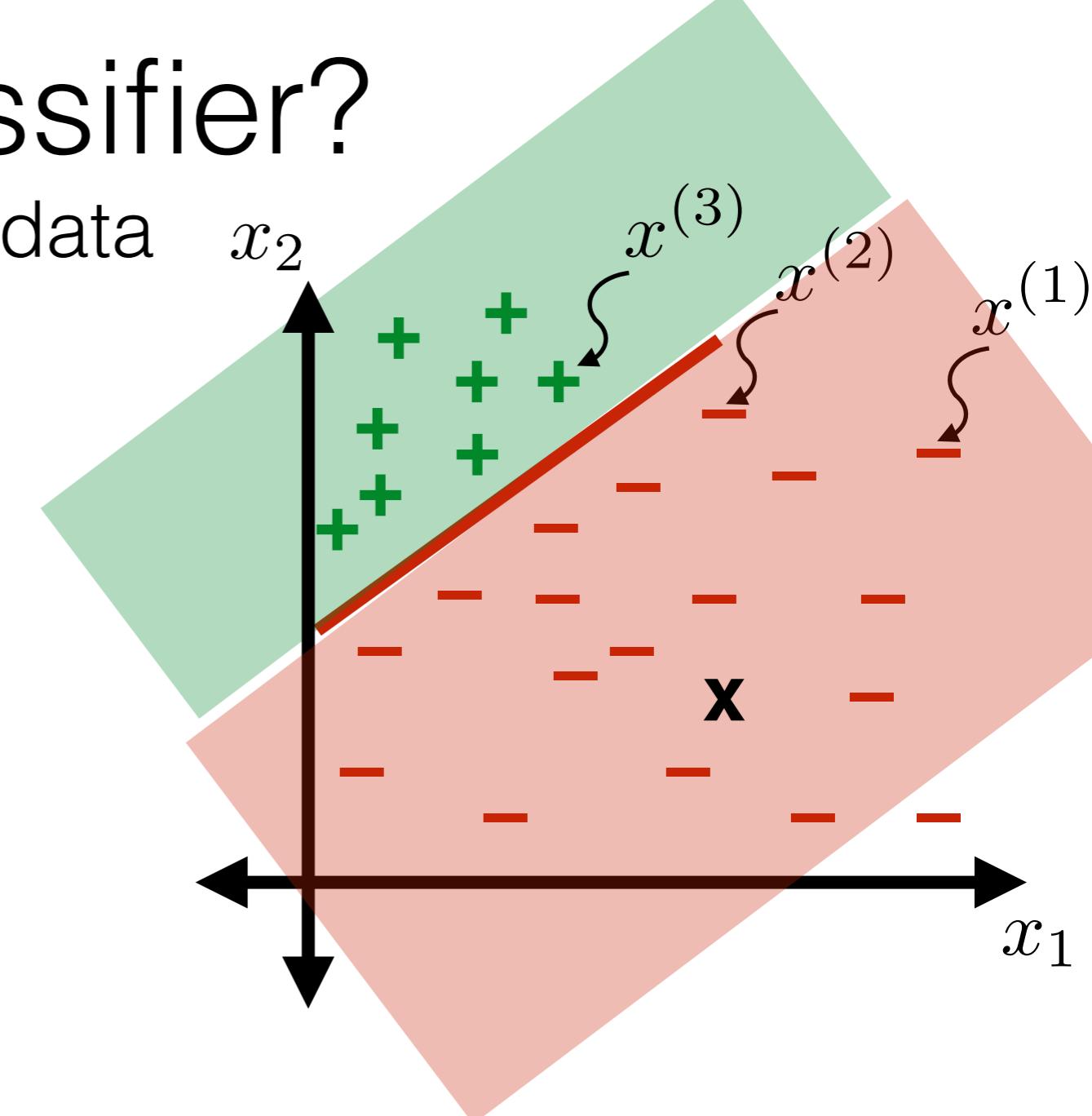
How good is a classifier?

- Should predict well on future data
- How good is a classifier at a single point?



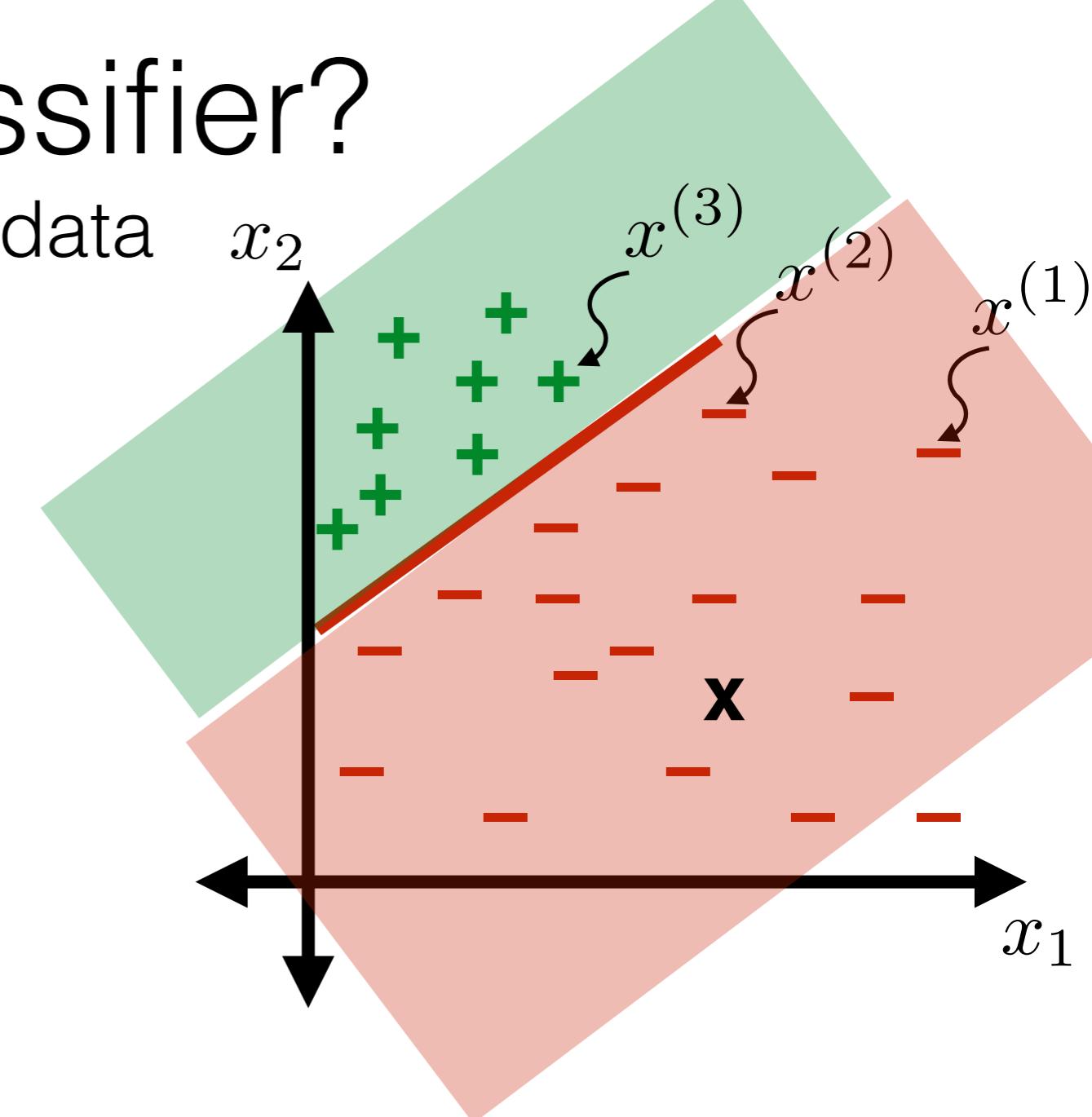
How good is a classifier?

- Should predict well on future data
- How good is a classifier at a single point?



How good is a classifier?

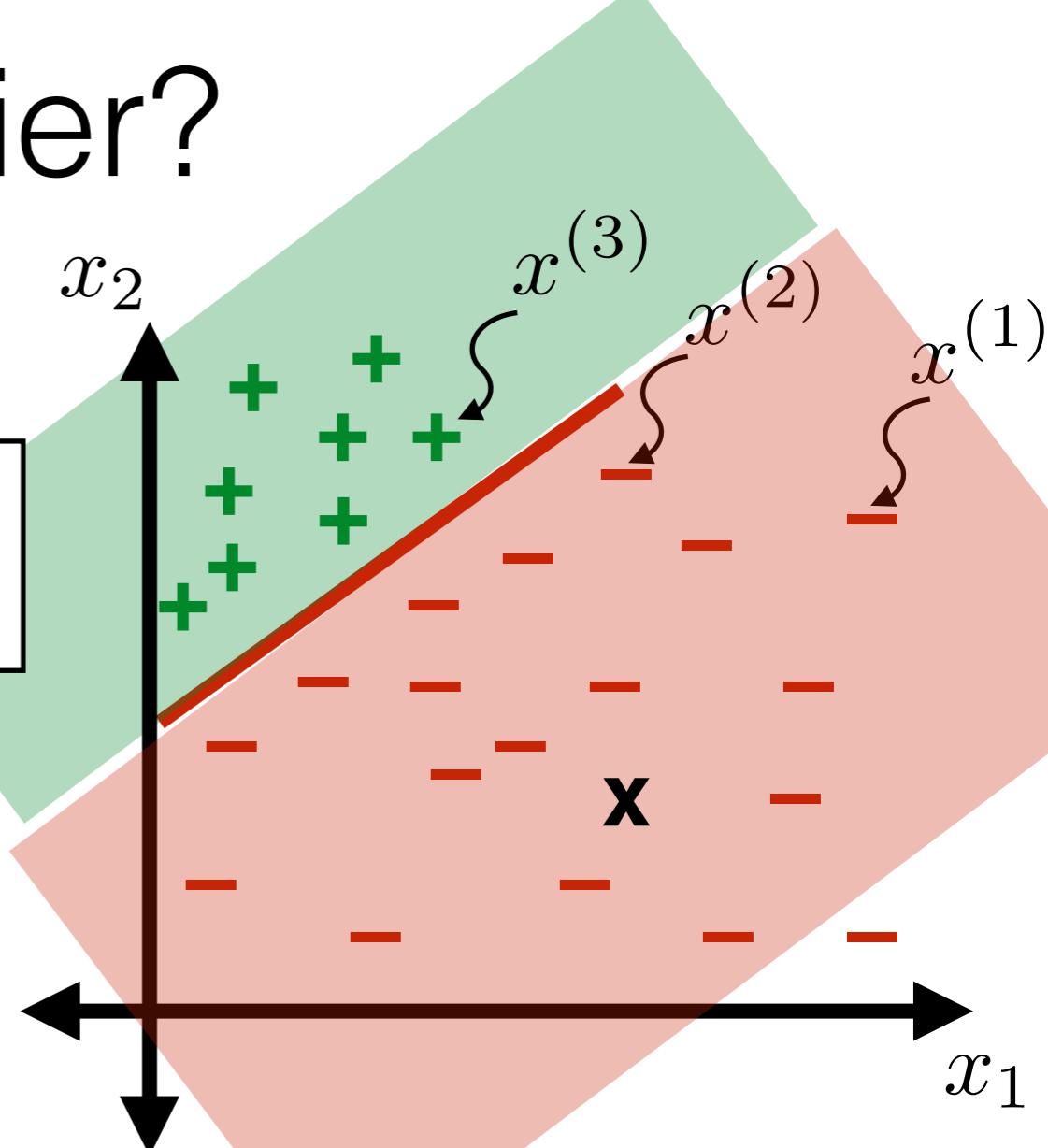
- Should predict well on future data
- How good is a classifier at a single point? Loss $L(g, a)$



How good is a classifier?

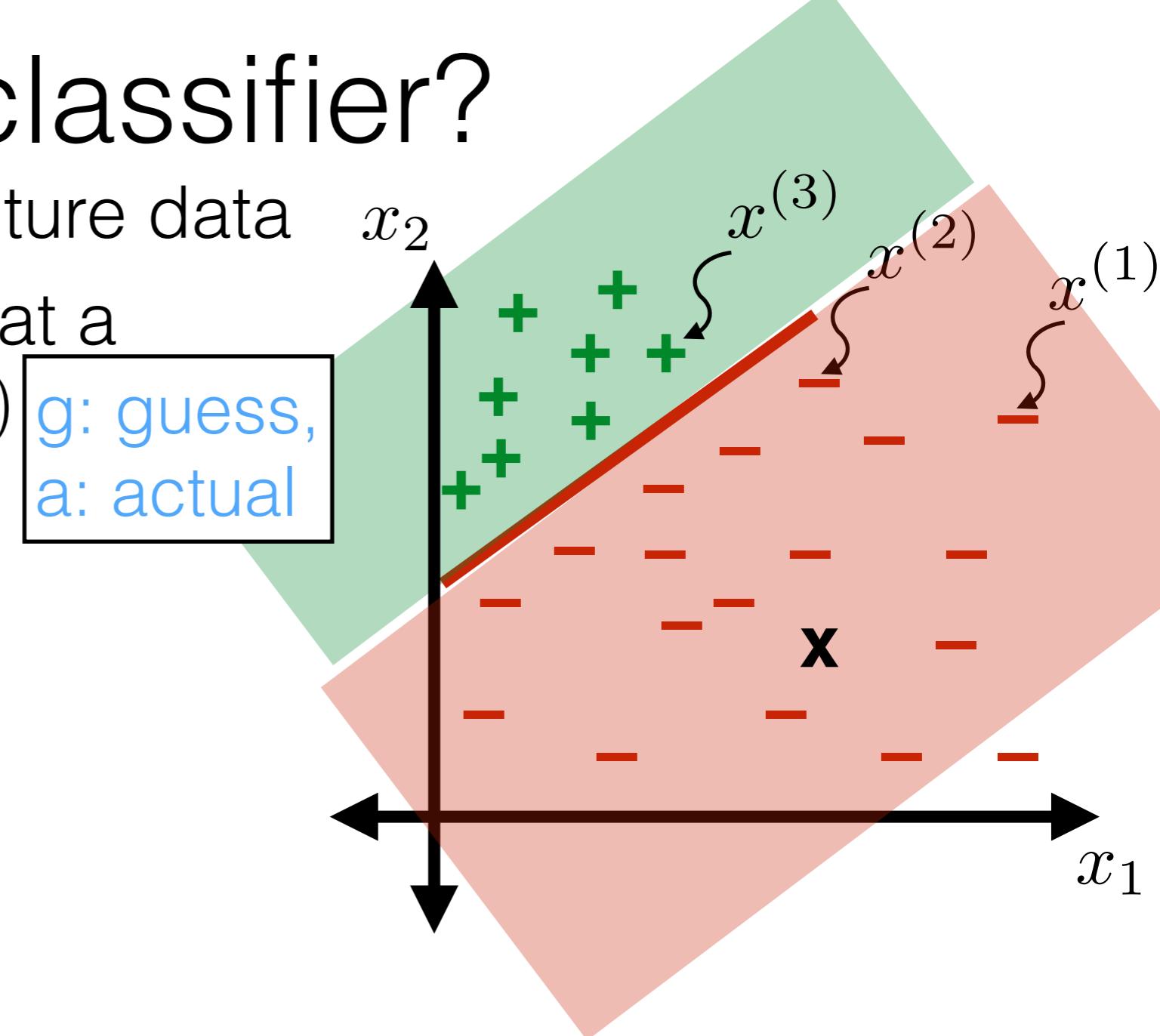
- Should predict well on future data
- How good is a classifier at a single point? Loss $L(g, a)$

g: guess,
a: actual



How good is a classifier?

- Should predict well on future data
- How good is a classifier at a single point? Loss $L(g, a)$
 - Example: 0-1 loss

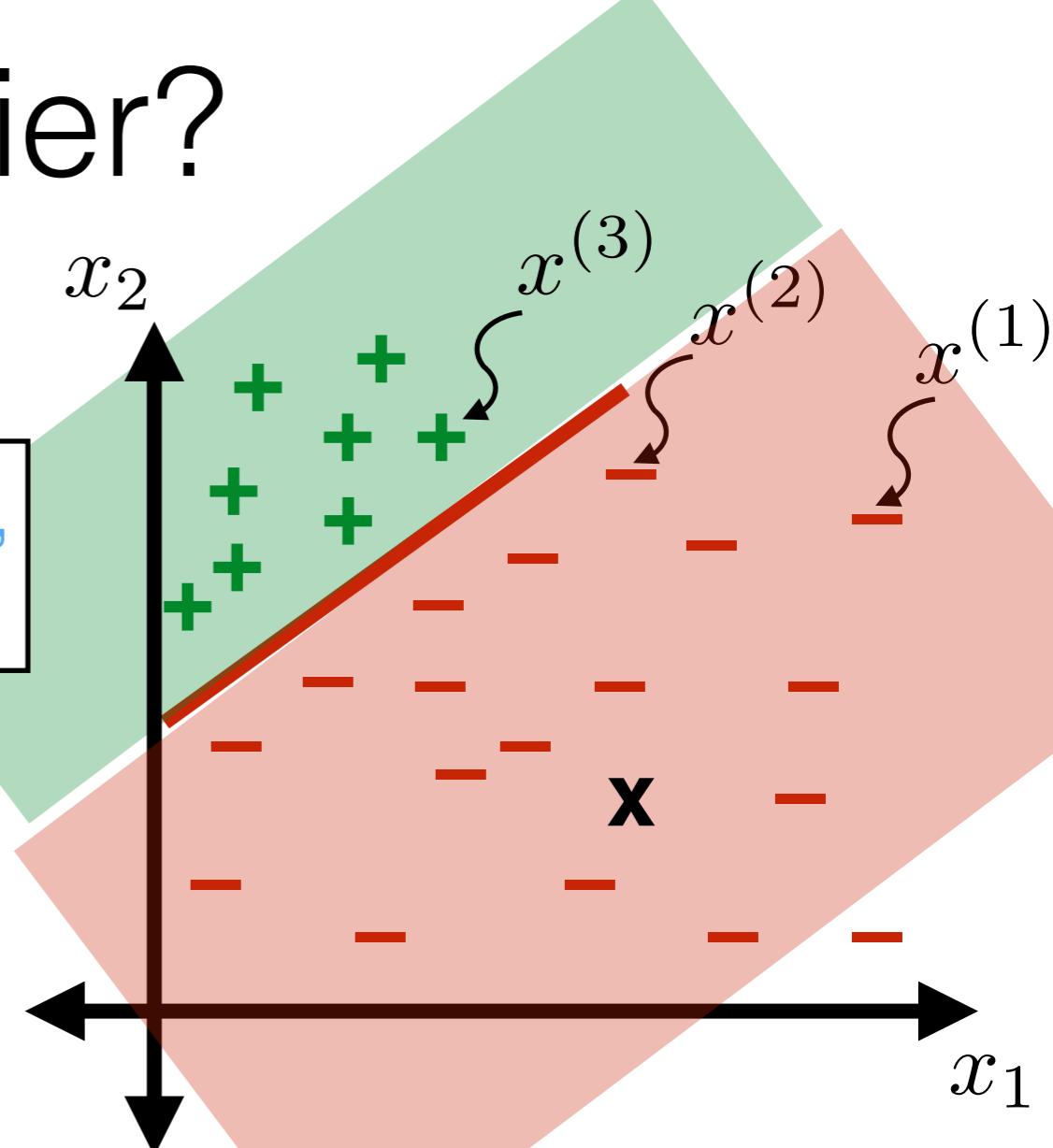


How good is a classifier?

- Should predict well on future data
- How good is a classifier at a single point? Loss $L(g, a)$
 - Example: 0-1 loss

$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

g: guess,
a: actual



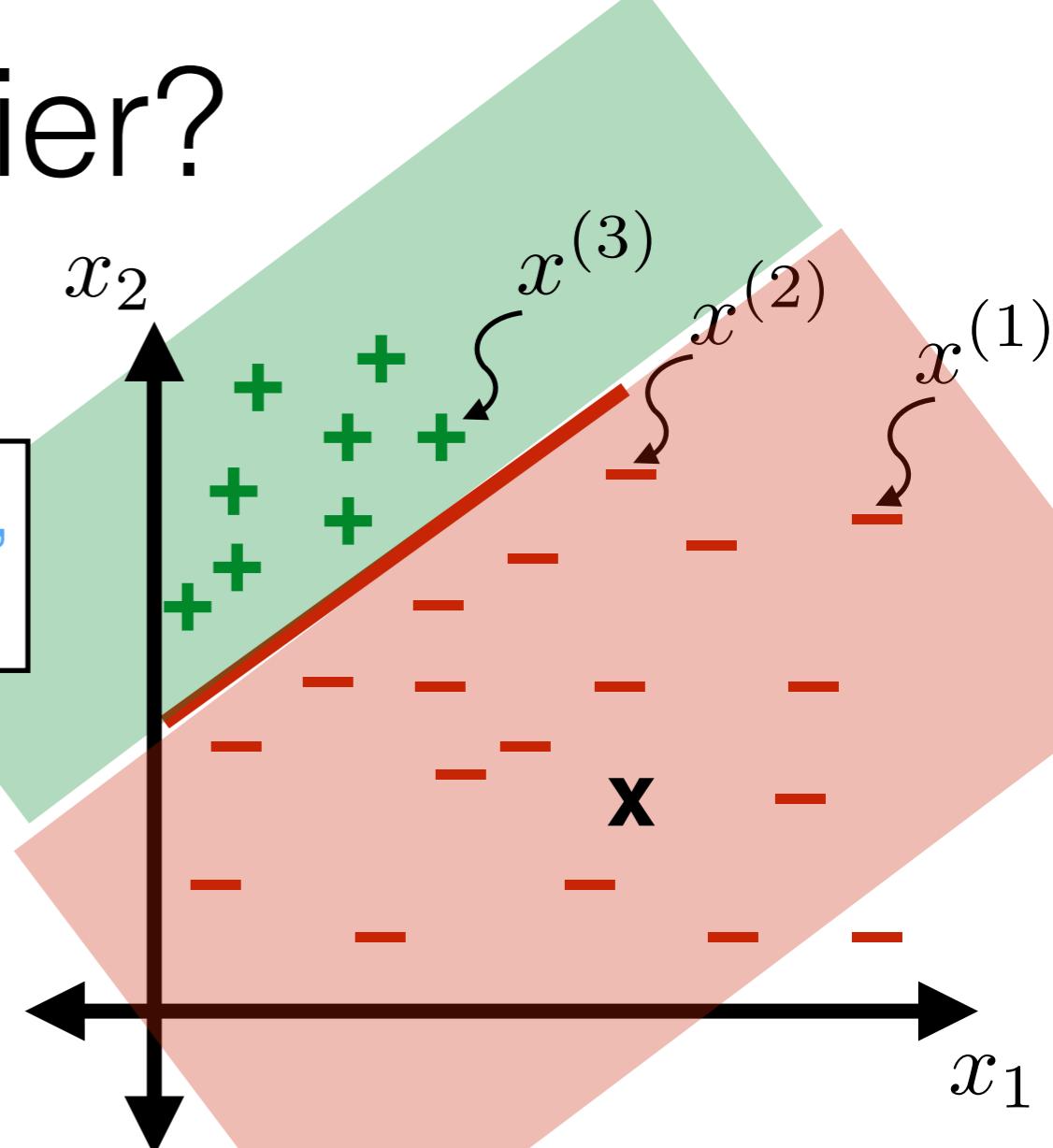
How good is a classifier?

- Should predict well on future data
- How good is a classifier at a single point? Loss $L(g, a)$
 - Example: 0-1 loss

$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

- Example: asymmetric loss

**g: guess,
a: actual**



How good is a classifier?

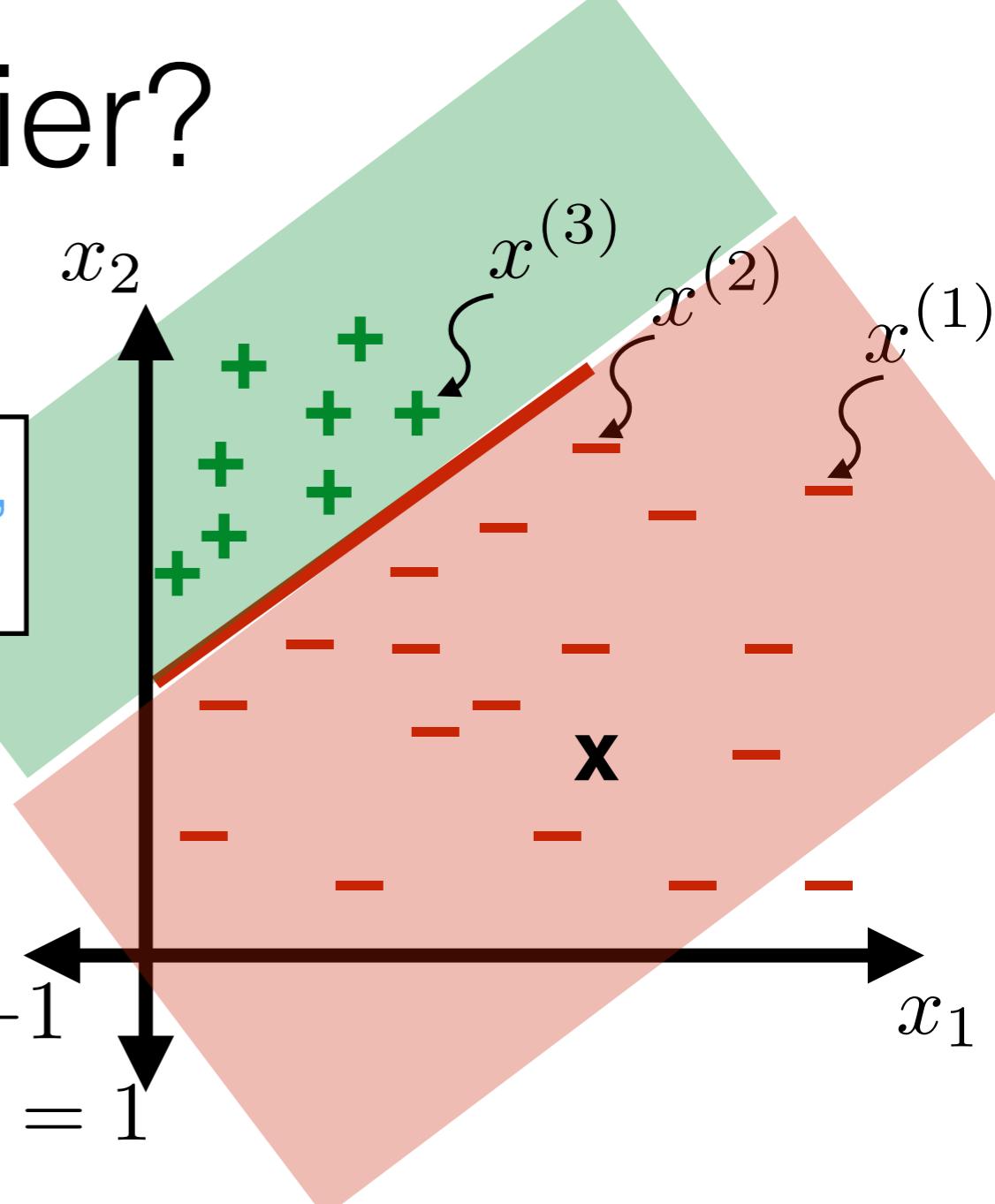
- Should predict well on future data
- How good is a classifier at a single point? Loss $L(g, a)$
 - Example: 0-1 loss

$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

- Example: asymmetric loss

$$L(g, a) = \begin{cases} 1 & \text{if } g = 1, a = -1 \\ 100 & \text{if } g = -1, a = 1 \\ 0 & \text{else} \end{cases}$$

**g: guess,
a: actual**



How good is a classifier?

- Should predict well on future data

- How good is a classifier at a single point? Loss $L(g, a)$

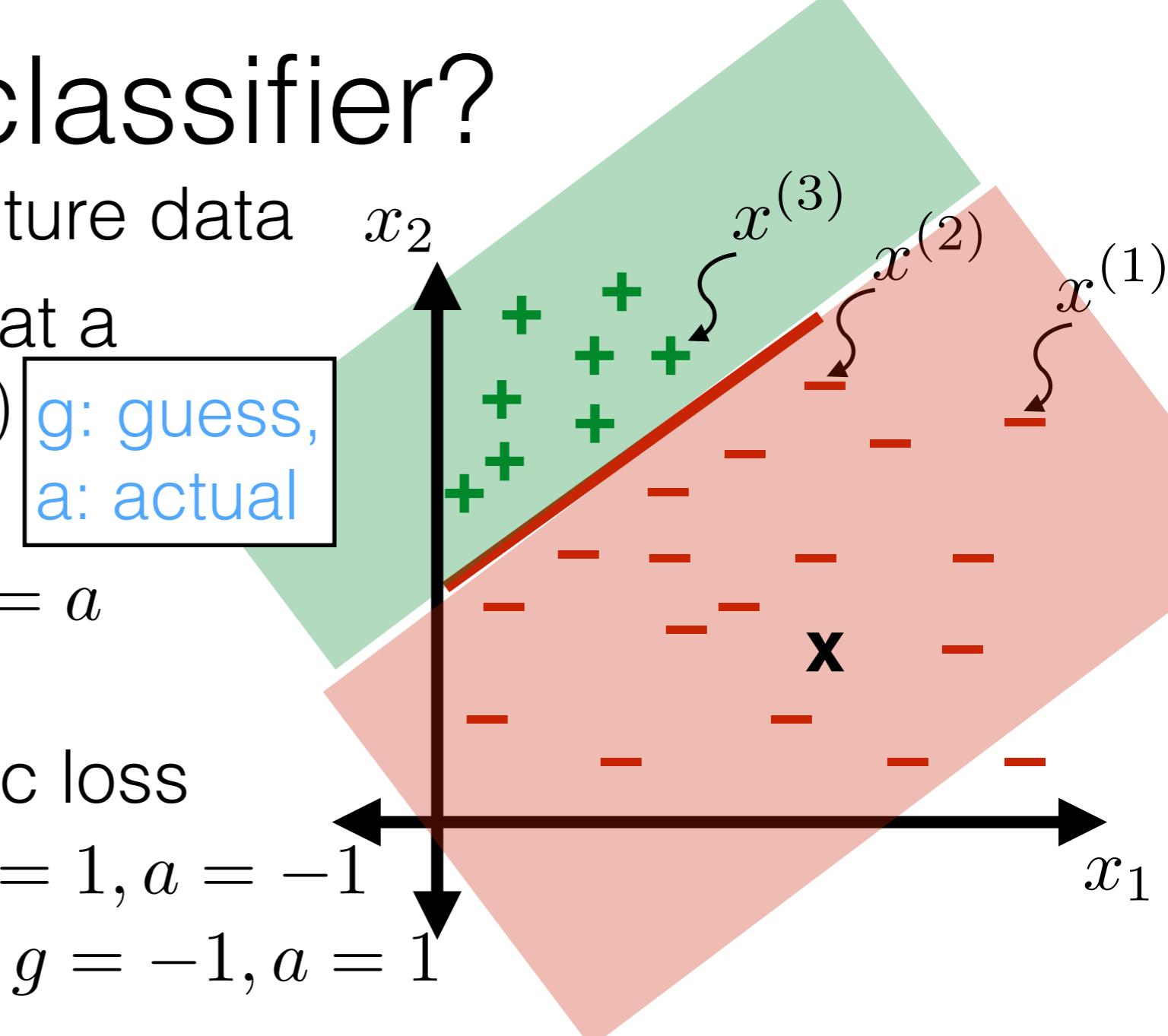
- Example: 0-1 loss

$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

- Example: asymmetric loss

$$L(g, a) = \begin{cases} 1 & \text{if } g = 1, a = -1 \\ 100 & \text{if } g = -1, a = 1 \\ 0 & \text{else} \end{cases}$$

- Test error (n' new points):



How good is a classifier?

- Should predict well on future data

- How good is a classifier at a single point? Loss $L(g, a)$

- Example: 0-1 loss

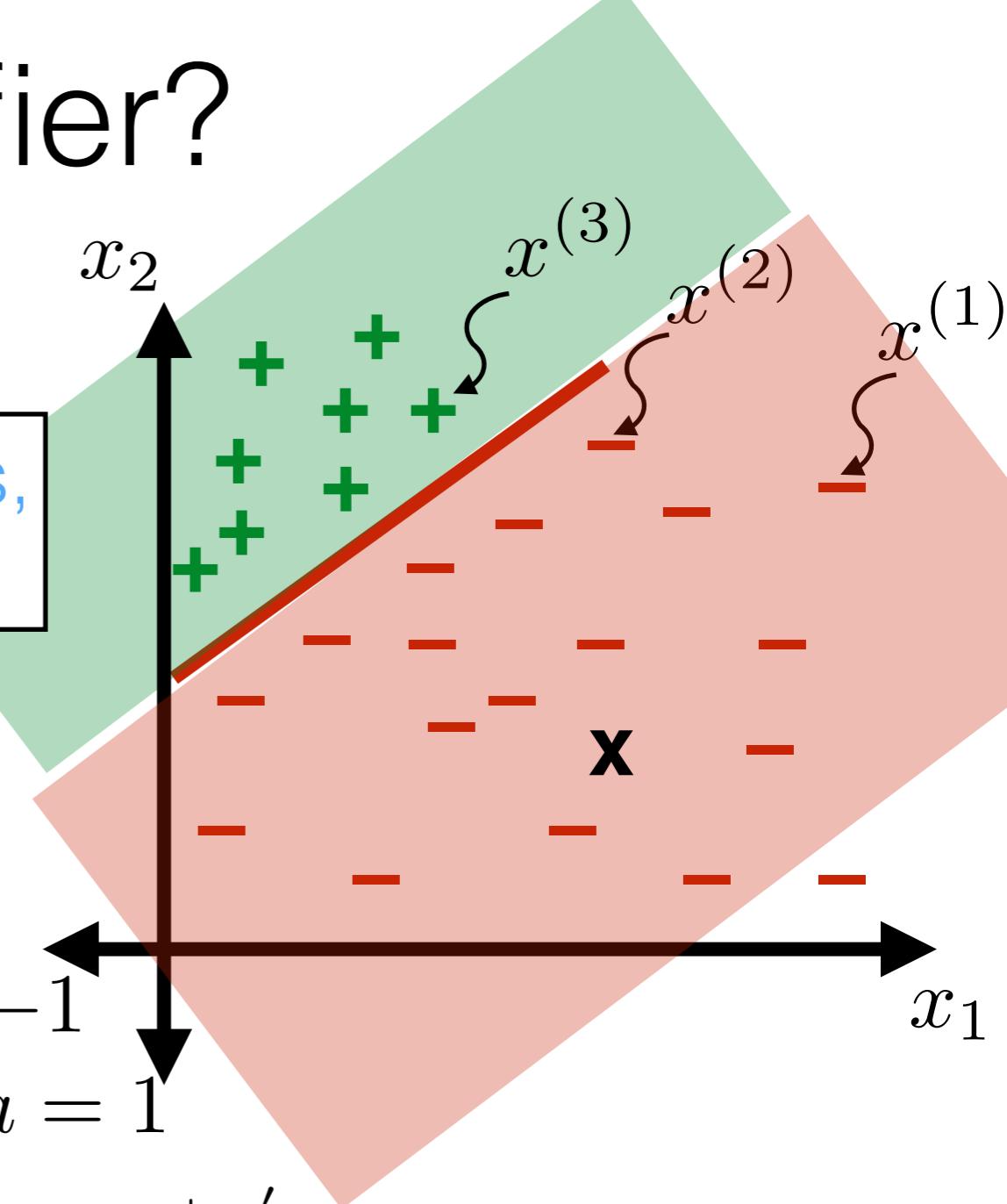
$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

- Example: asymmetric loss

$$L(g, a) = \begin{cases} 1 & \text{if } g = 1, a = -1 \\ 100 & \text{if } g = -1, a = 1 \\ 0 & \text{else} \end{cases}$$

- Test error (n' new points): $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} L(h(x^{(i)}), y^{(i)})$

**g: guess,
a: actual**



How good is a classifier?

- Should predict well on future data

- How good is a classifier at a single point? Loss $L(g, a)$

- Example: 0-1 loss

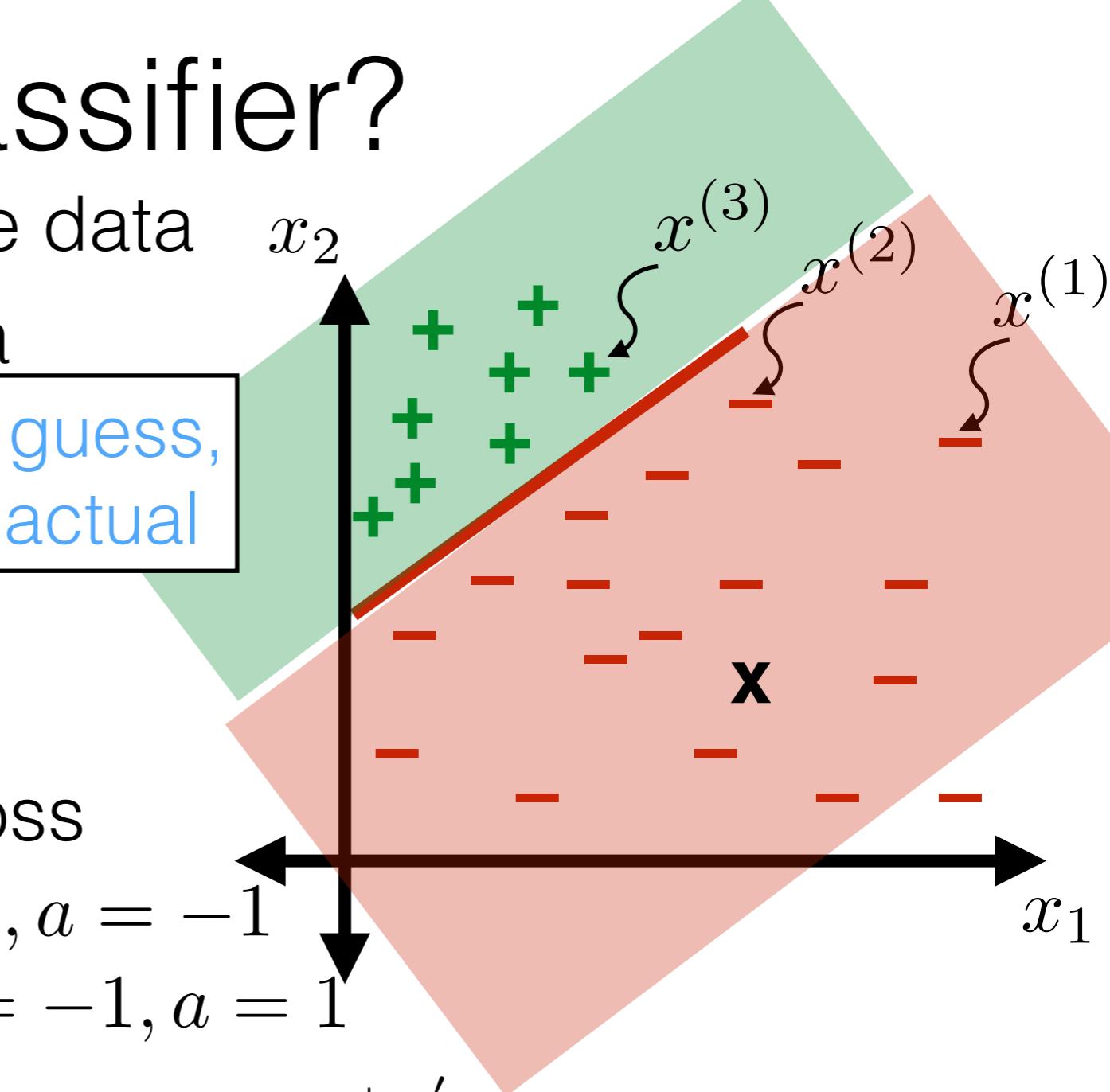
$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

- Example: asymmetric loss

$$L(g, a) = \begin{cases} 1 & \text{if } g = 1, a = -1 \\ 100 & \text{if } g = -1, a = 1 \\ 0 & \text{else} \end{cases}$$

- Test error (n' new points): $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} L(h(x^{(i)}), y^{(i)})$
- Training error:

**g: guess,
a: actual**



How good is a classifier?

- Should predict well on future data

- How good is a classifier at a single point? Loss $L(g, a)$

- Example: 0-1 loss

$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

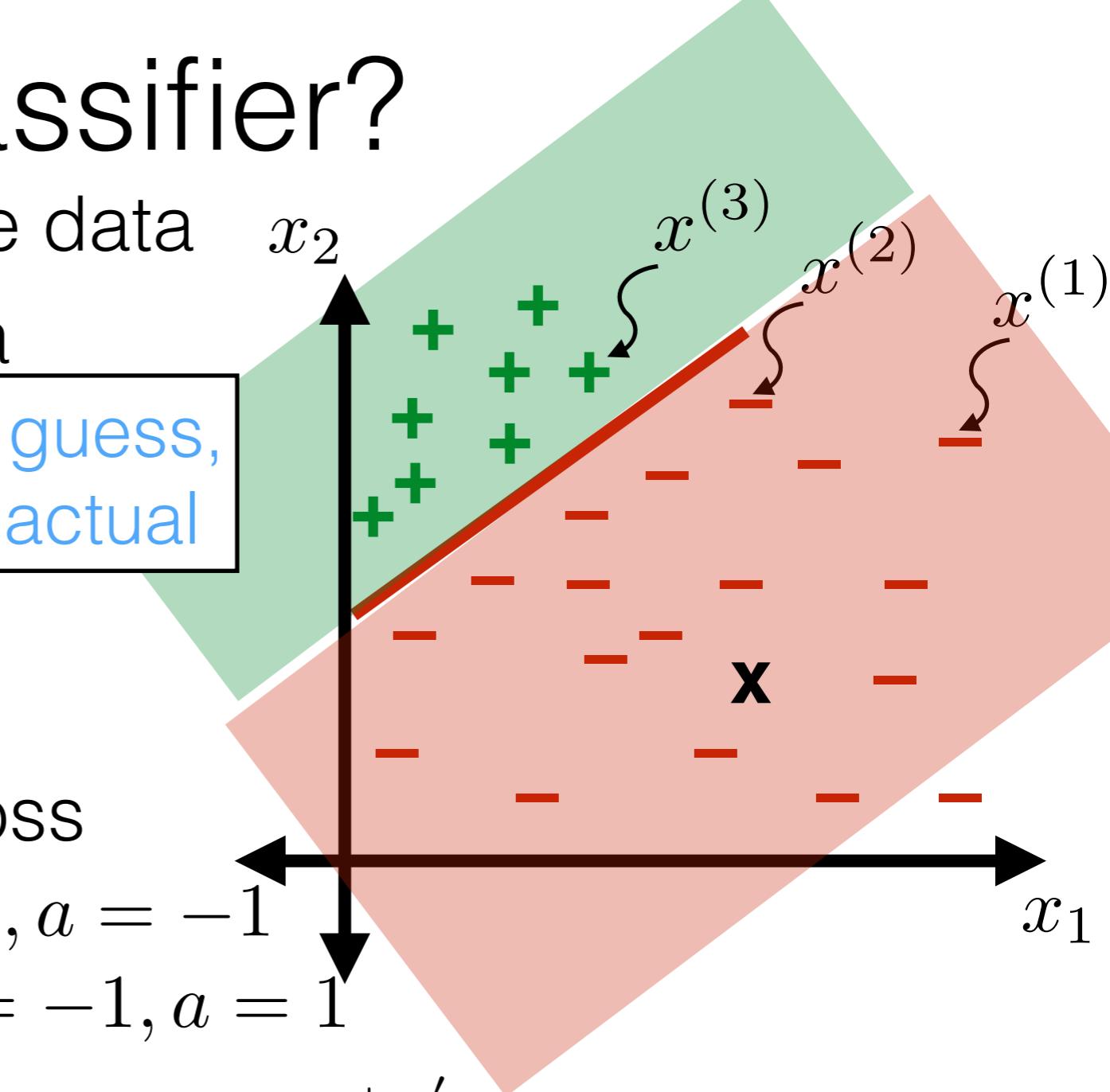
- Example: asymmetric loss

$$L(g, a) = \begin{cases} 1 & \text{if } g = 1, a = -1 \\ 100 & \text{if } g = -1, a = 1 \\ 0 & \text{else} \end{cases}$$

- Test error (n' new points): $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} L(h(x^{(i)}), y^{(i)})$

- Training error: $\mathcal{E}_n(h) = \frac{1}{n} \sum_{i=1}^n L(h(x^{(i)}), y^{(i)})$

**g: guess,
a: actual**



How good is a classifier?

- Should predict well on future data

- How good is a classifier at a single point? Loss $L(g, a)$

- Example: 0-1 loss

$$L(g, a) = \begin{cases} 0 & \text{if } g = a \\ 1 & \text{else} \end{cases}$$

- Example: asymmetric loss

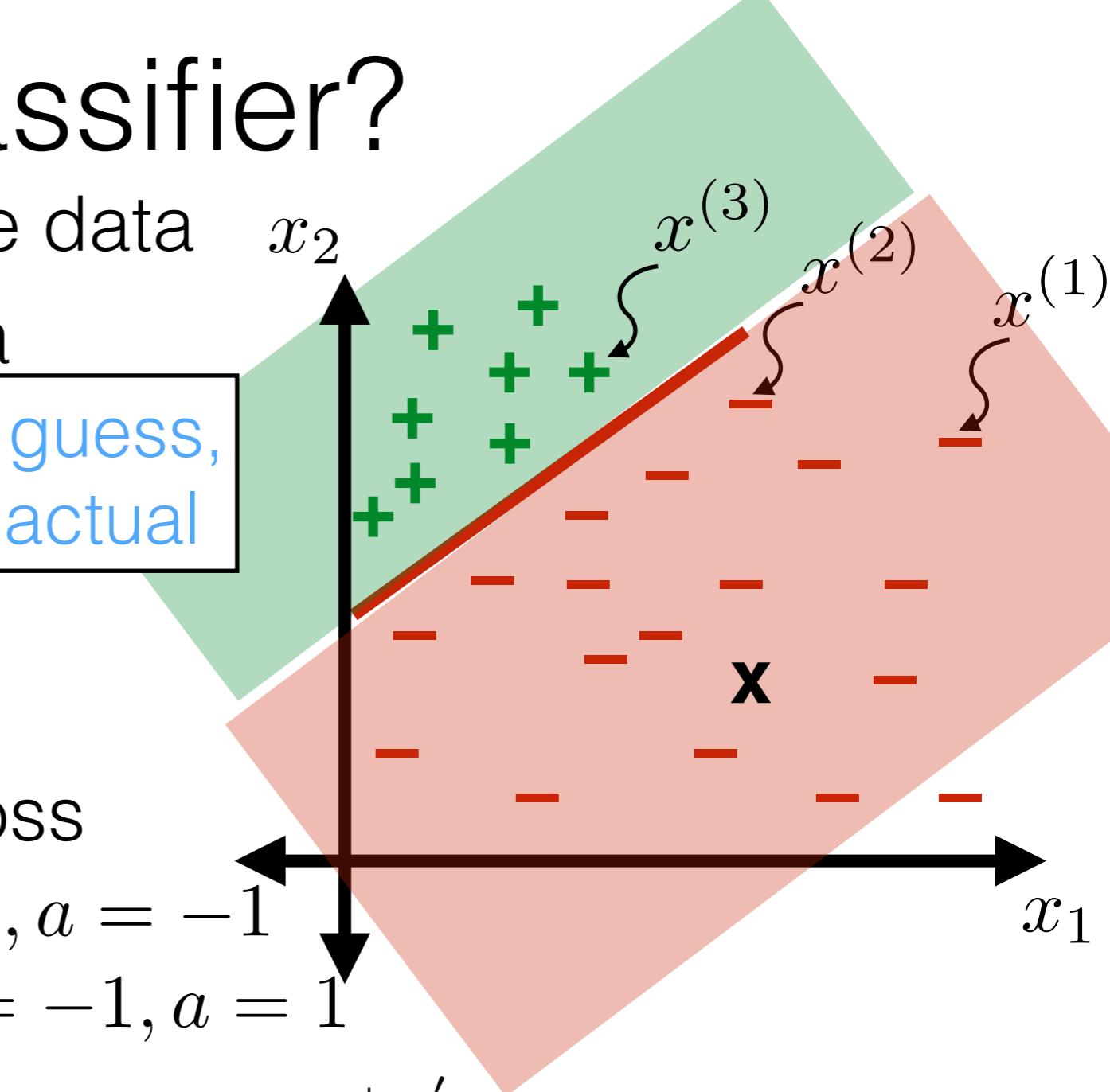
$$L(g, a) = \begin{cases} 1 & \text{if } g = 1, a = -1 \\ 100 & \text{if } g = -1, a = 1 \\ 0 & \text{else} \end{cases}$$

- Test error (n' new points): $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} L(h(x^{(i)}), y^{(i)})$

- Training error: $\mathcal{E}_n(h) = \frac{1}{n} \sum_{i=1}^n L(h(x^{(i)}), y^{(i)})$

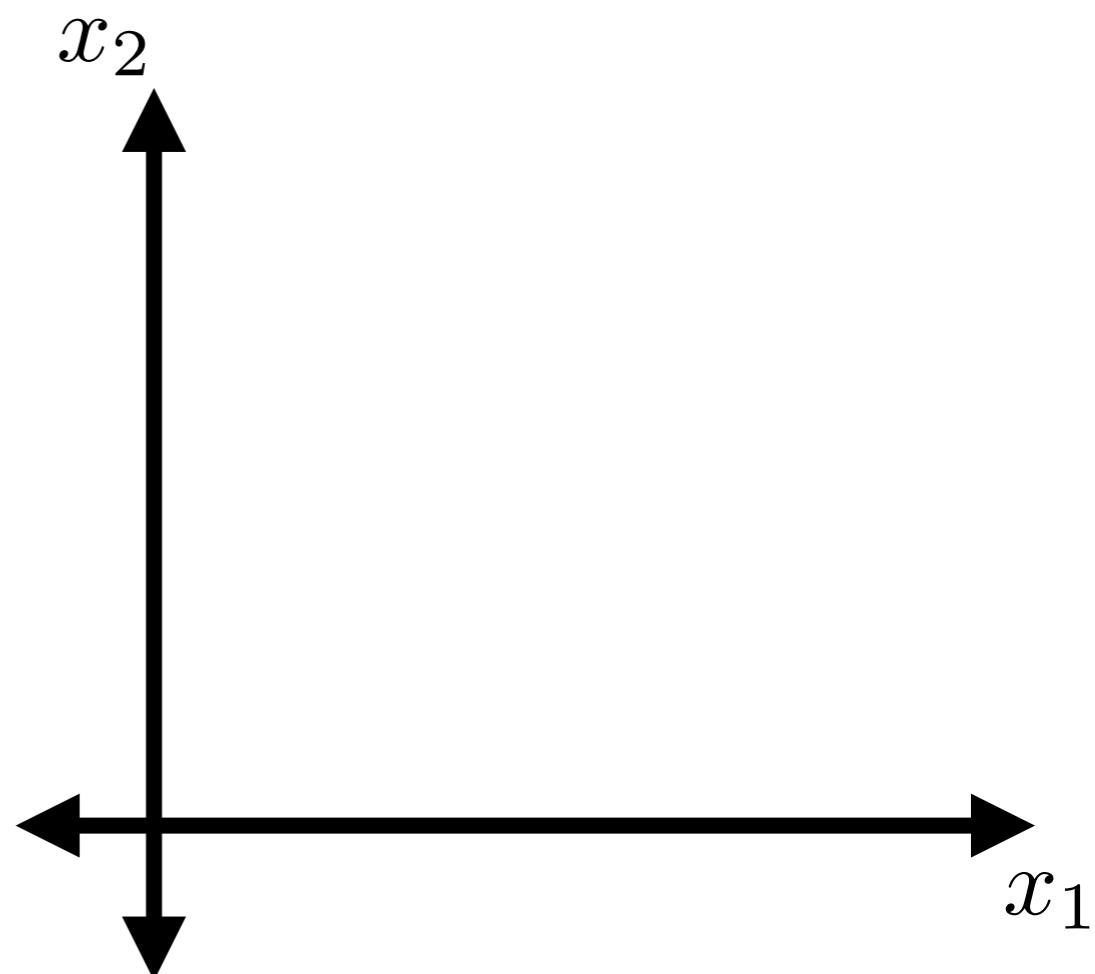
- Prefer h to \tilde{h} if $\mathcal{E}_n(h) < \mathcal{E}_n(\tilde{h})$

g: guess,
a: actual



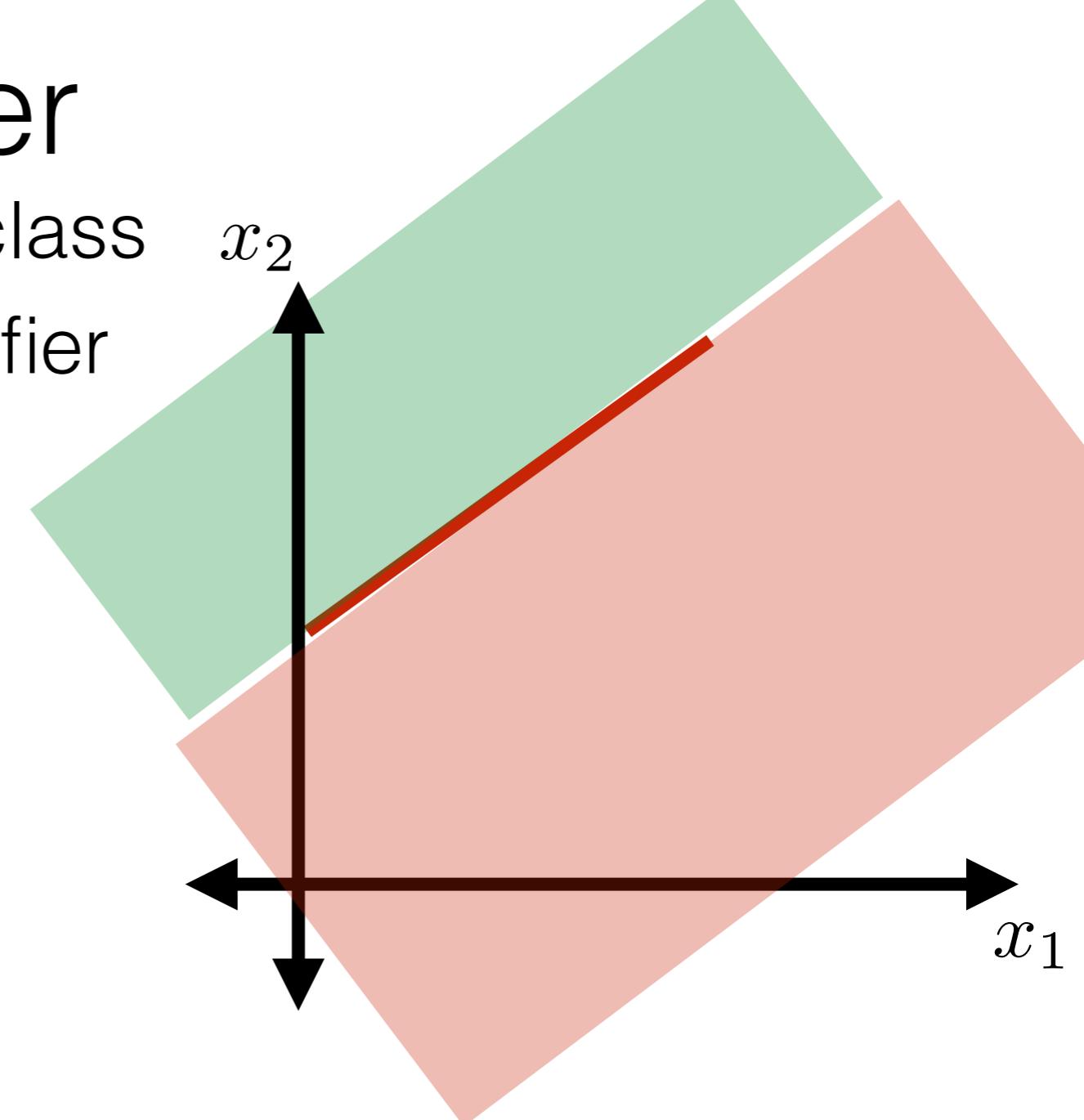
Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$



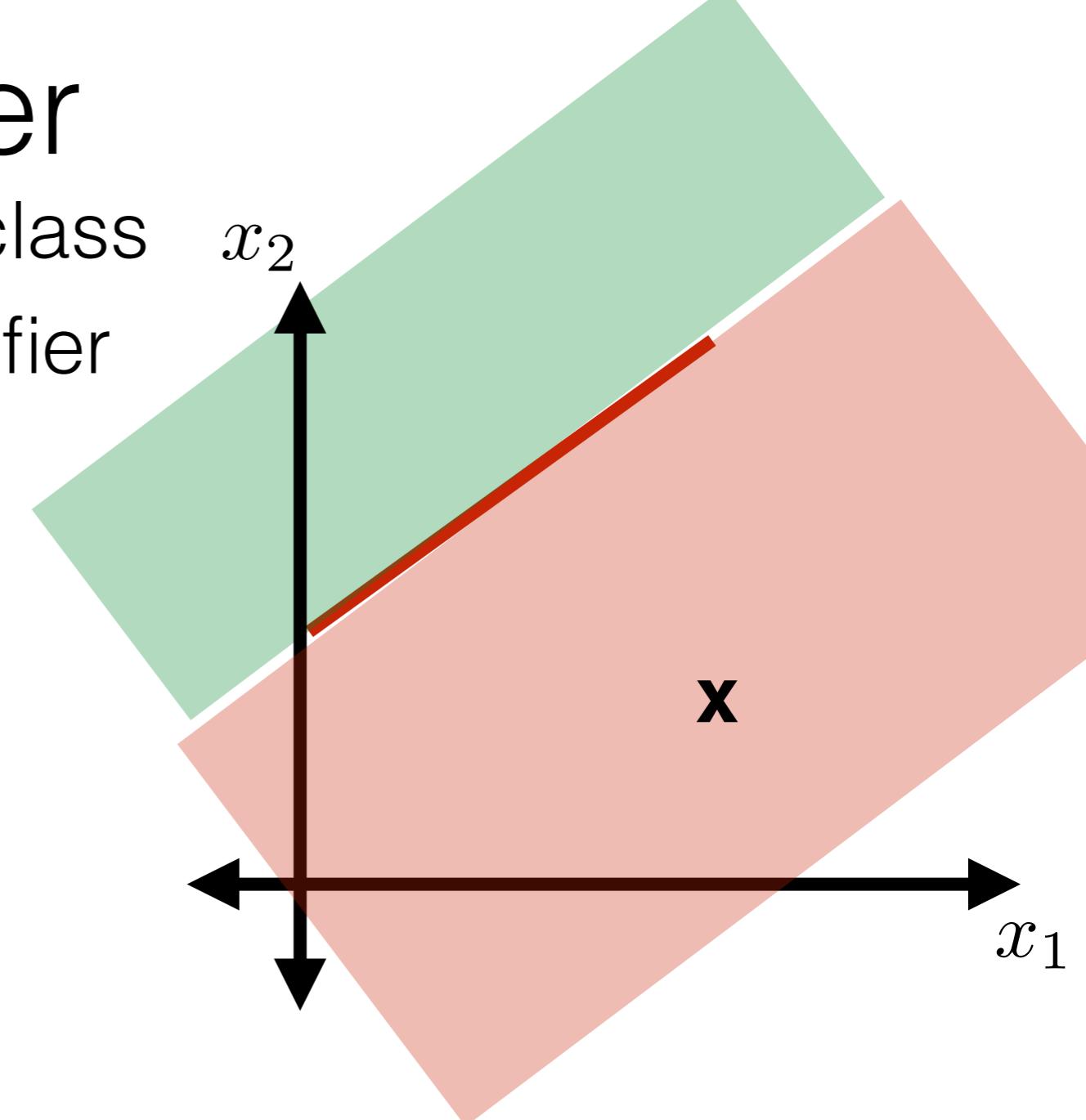
Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$



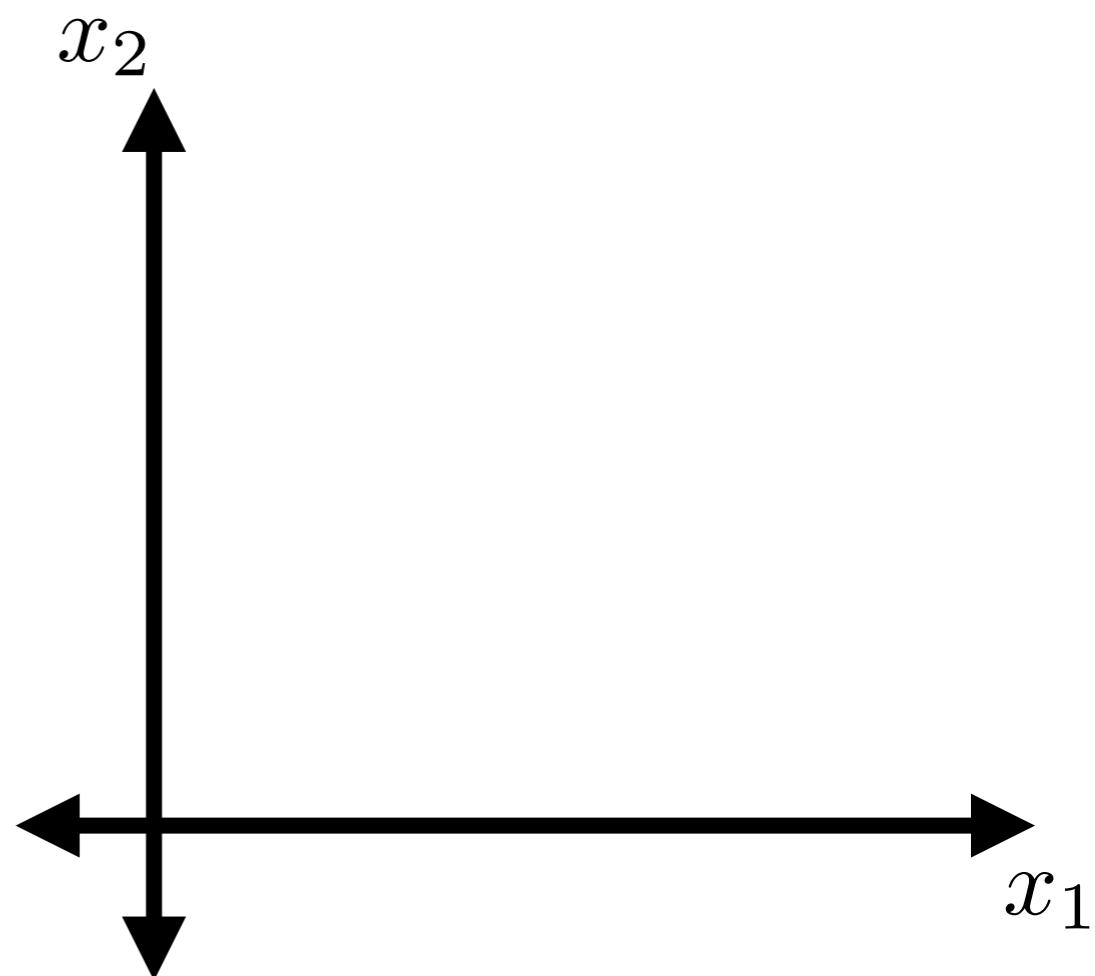
Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$



Learning a classifier

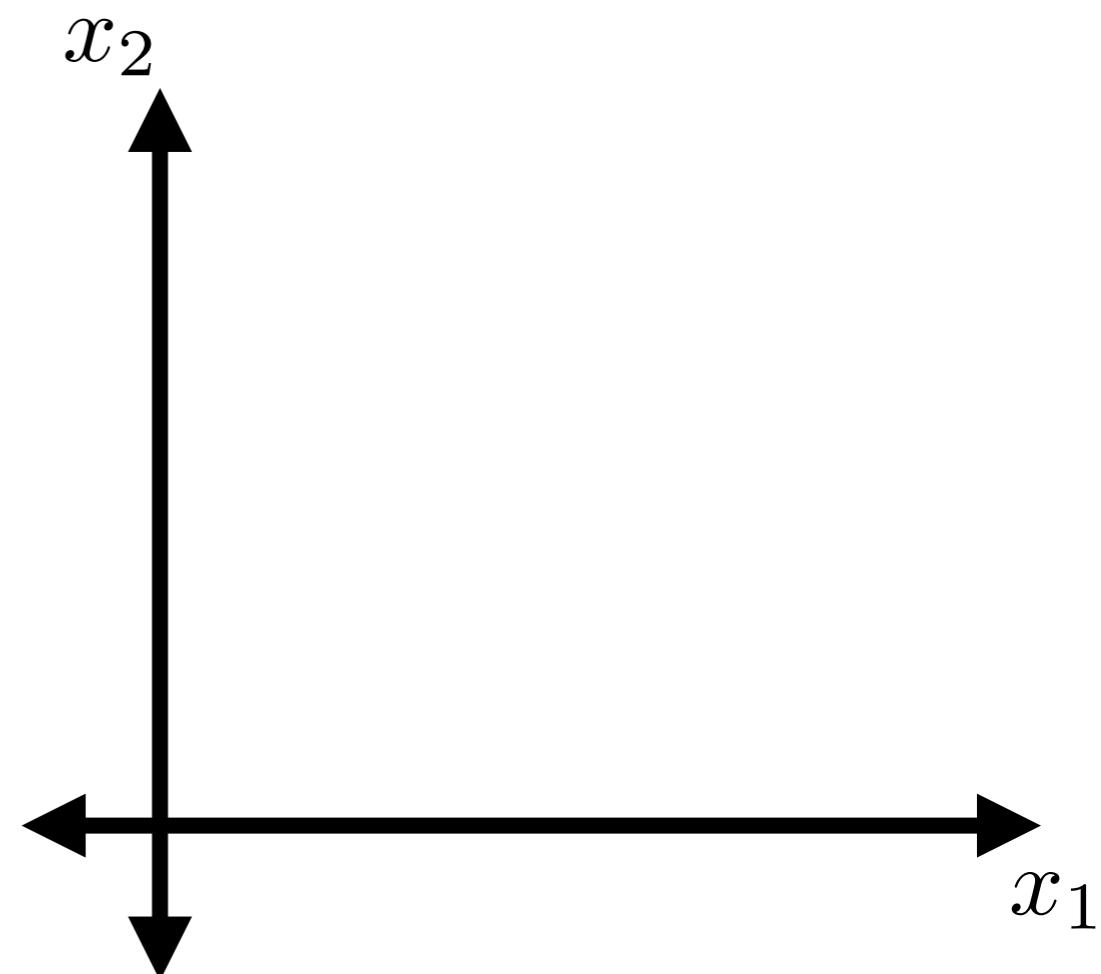
- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$



Learning a classifier

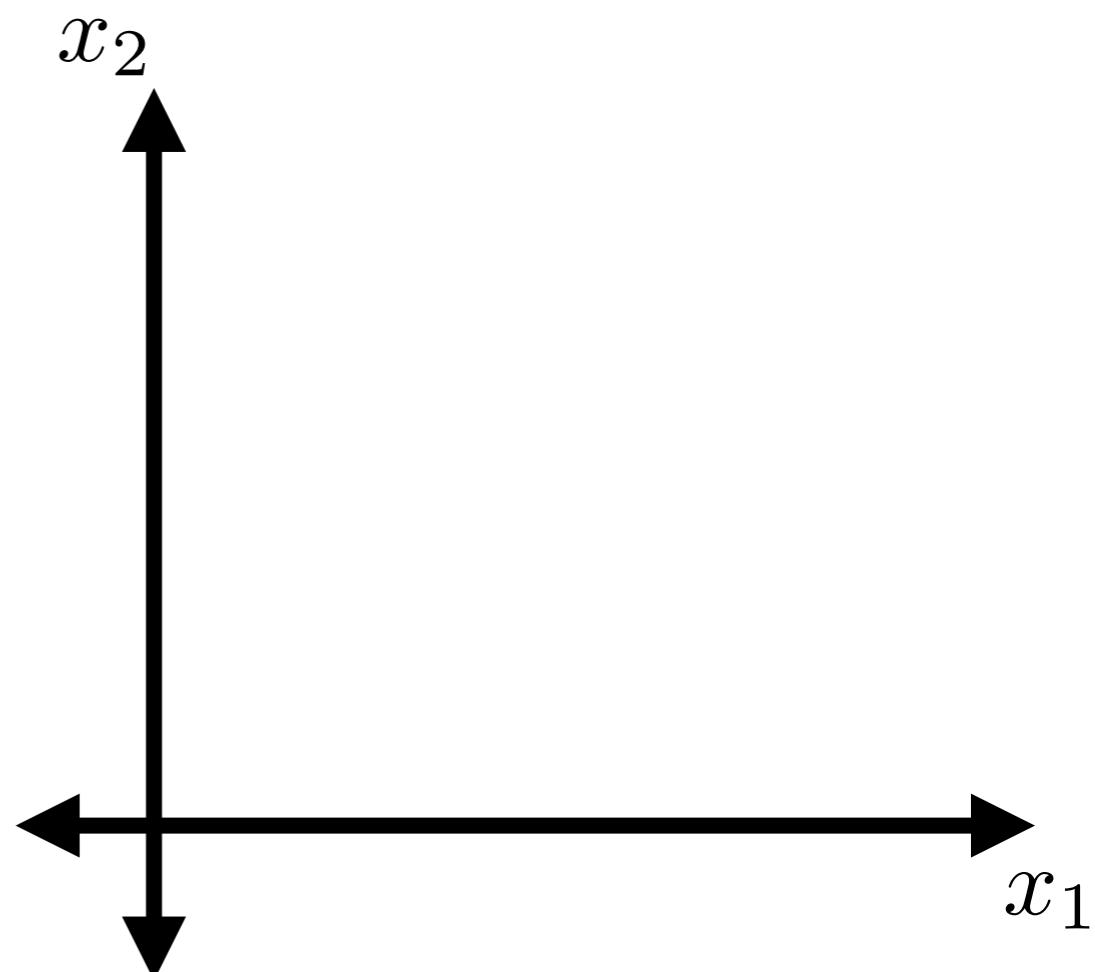
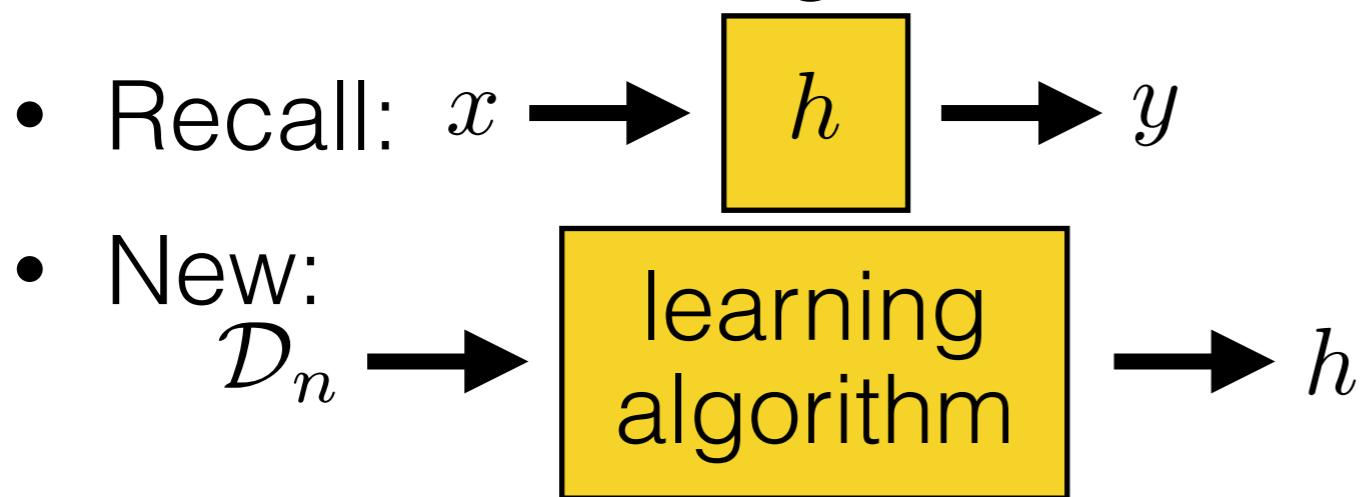
- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$
- New:



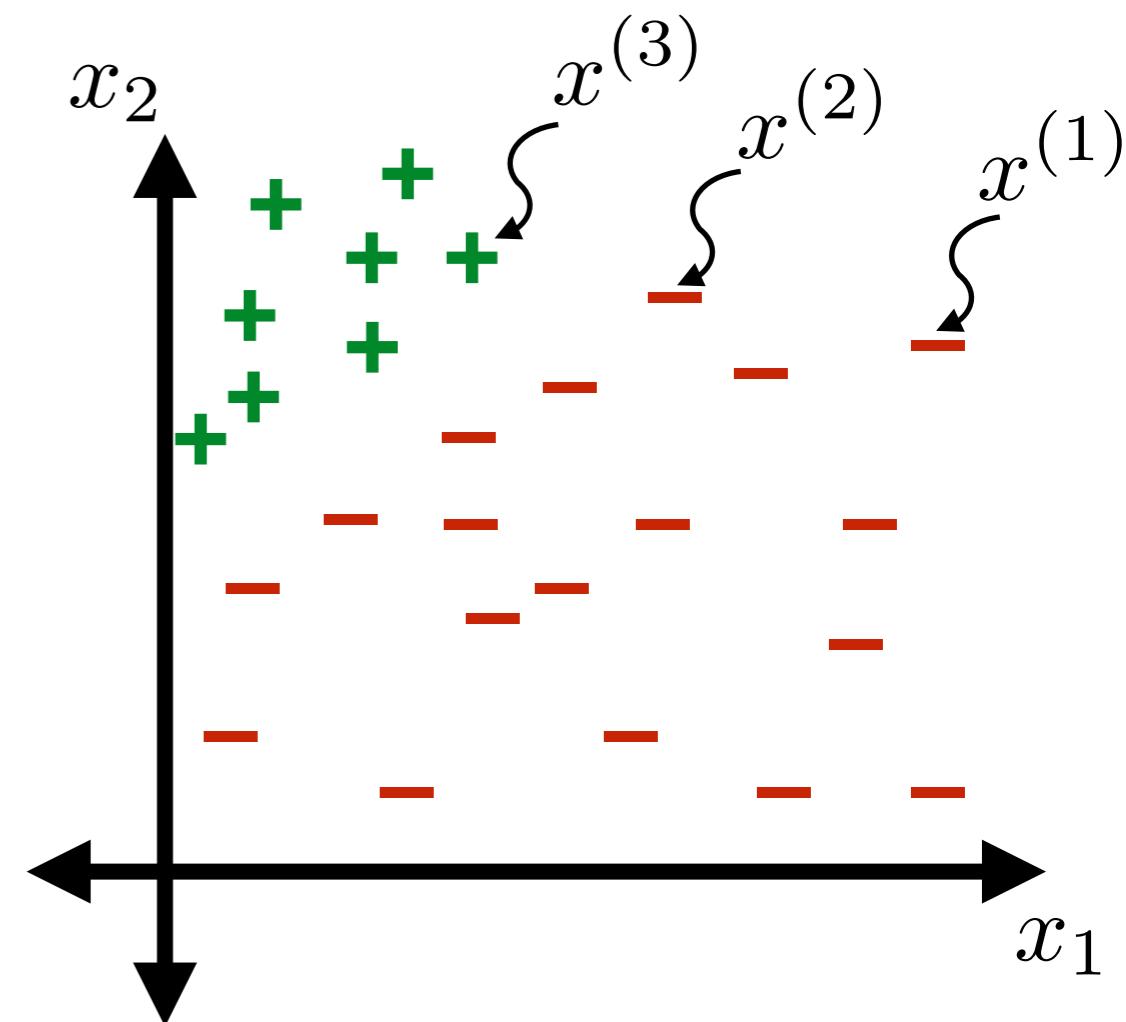
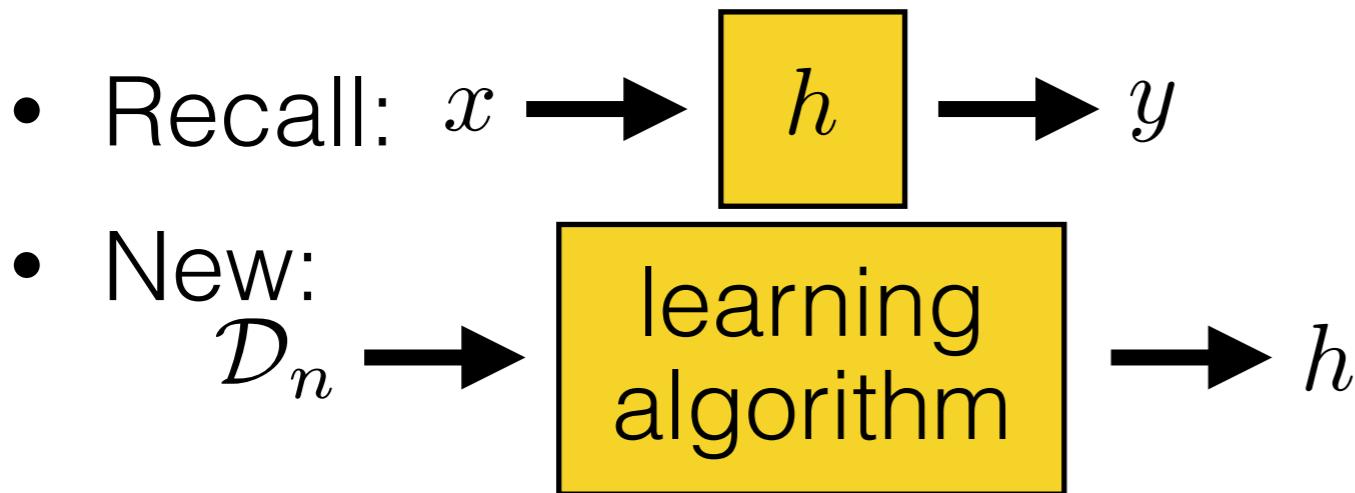
Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

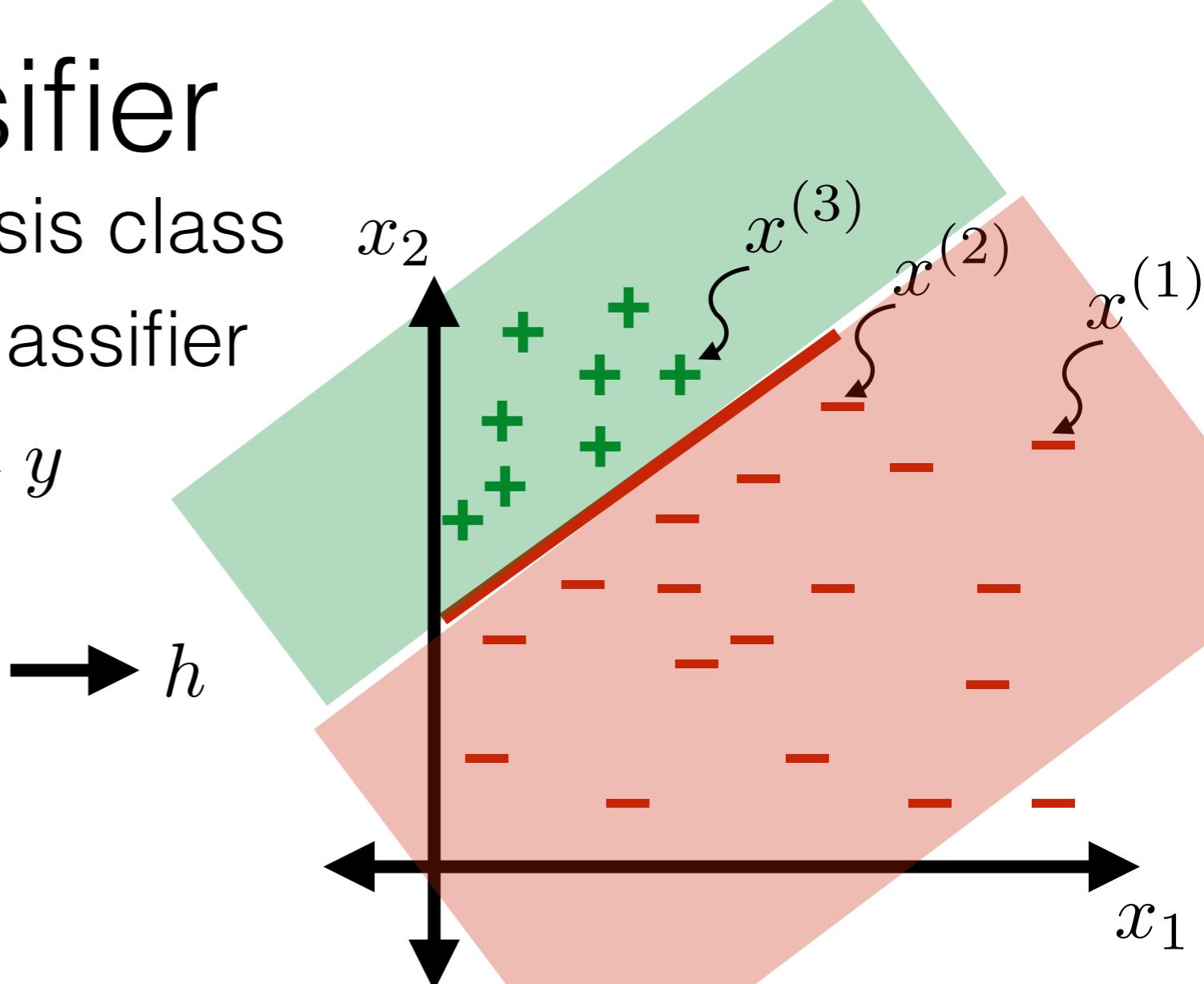


Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

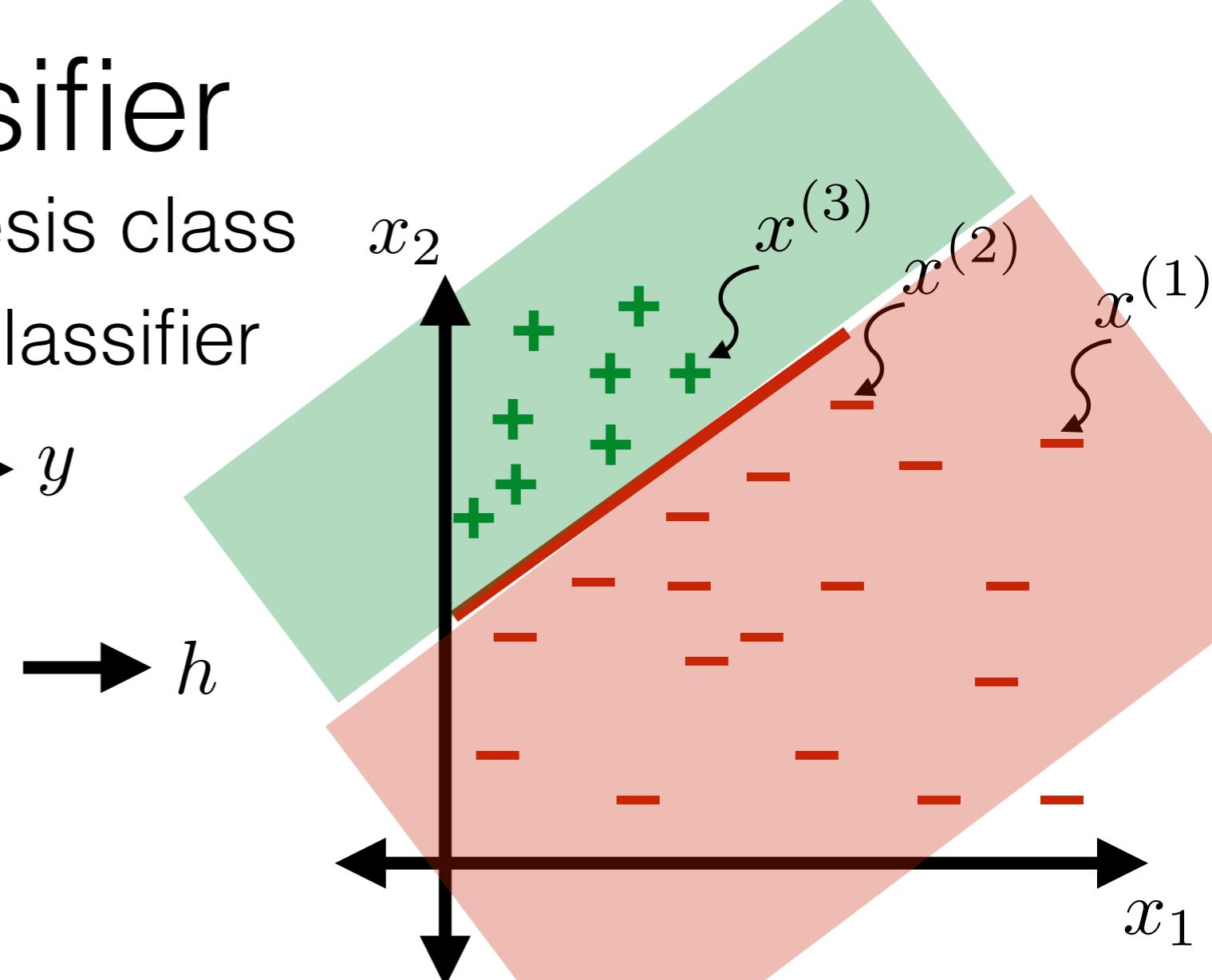
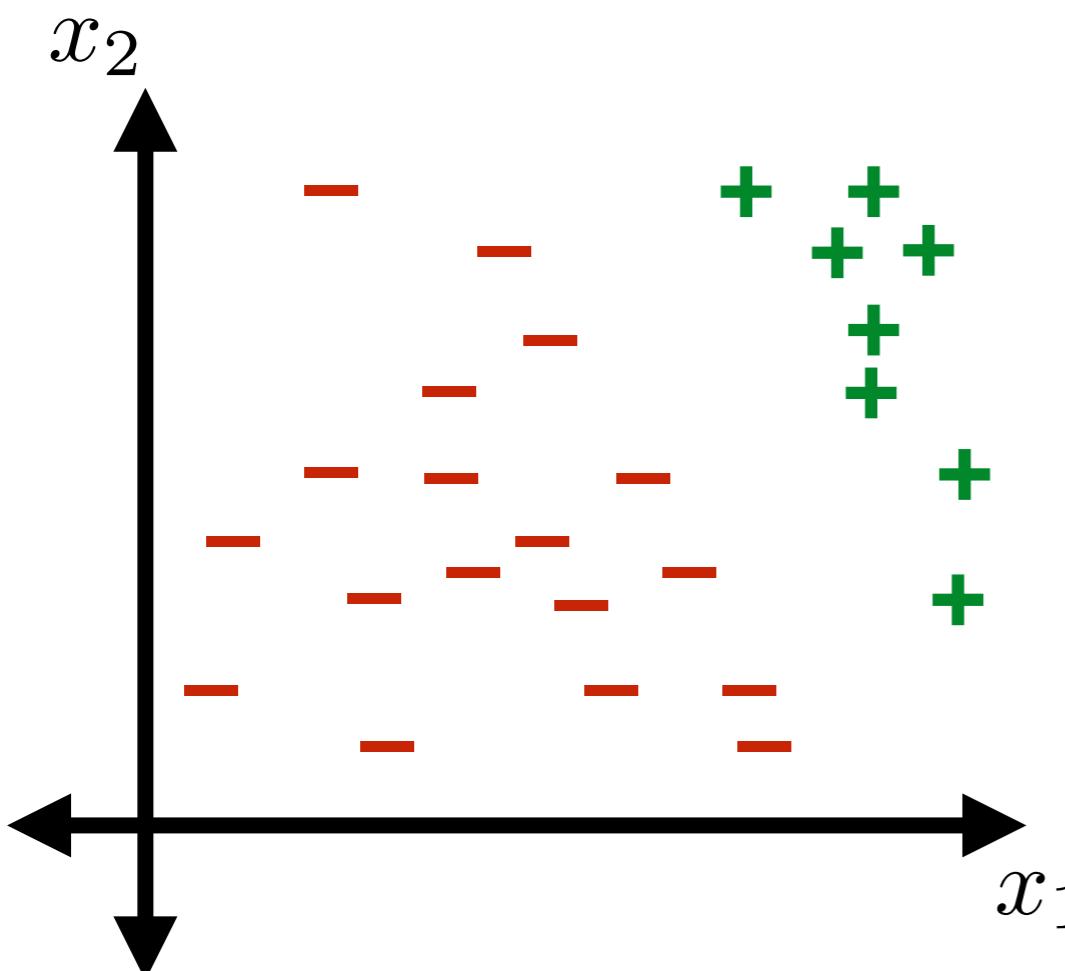


Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

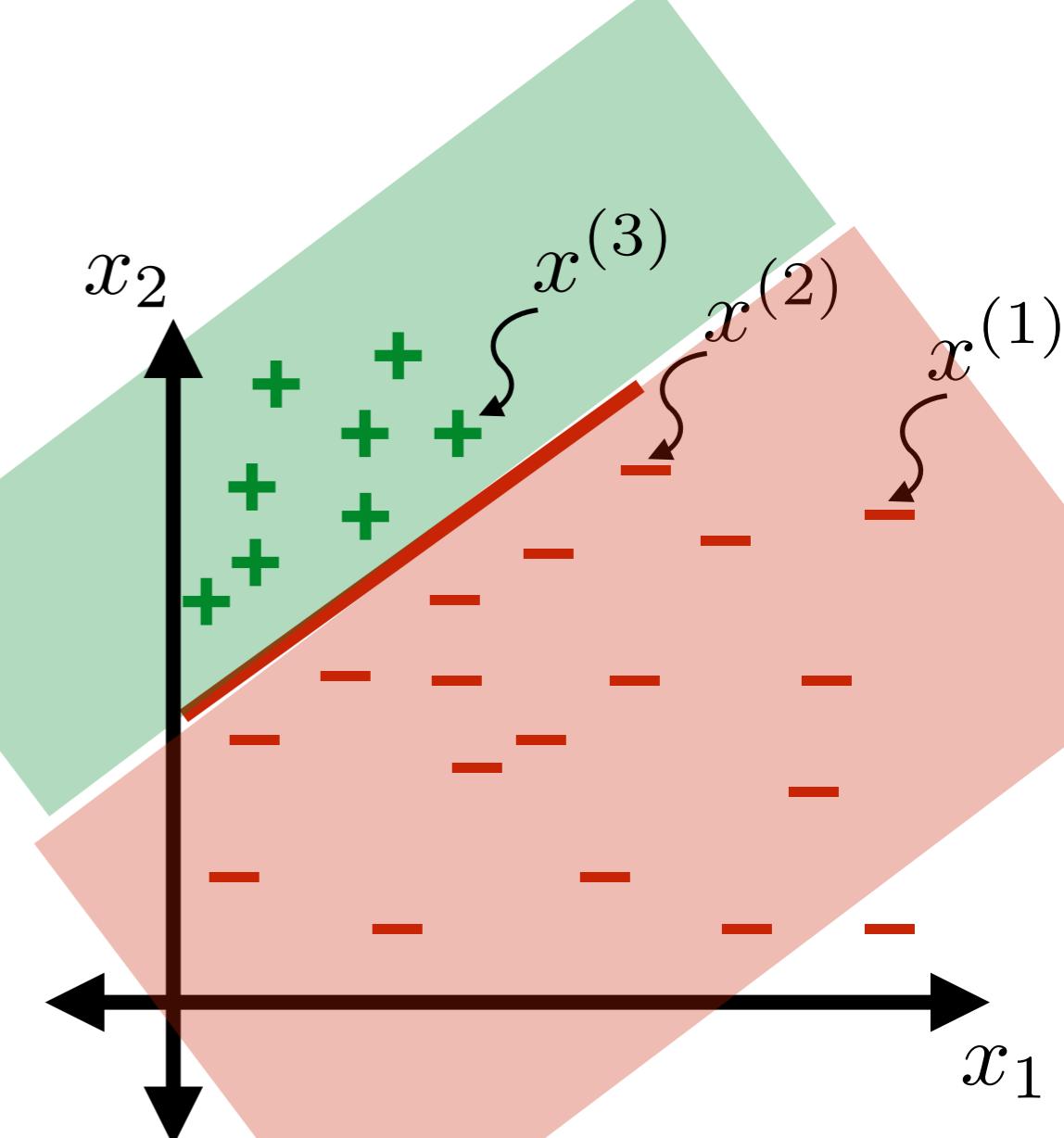
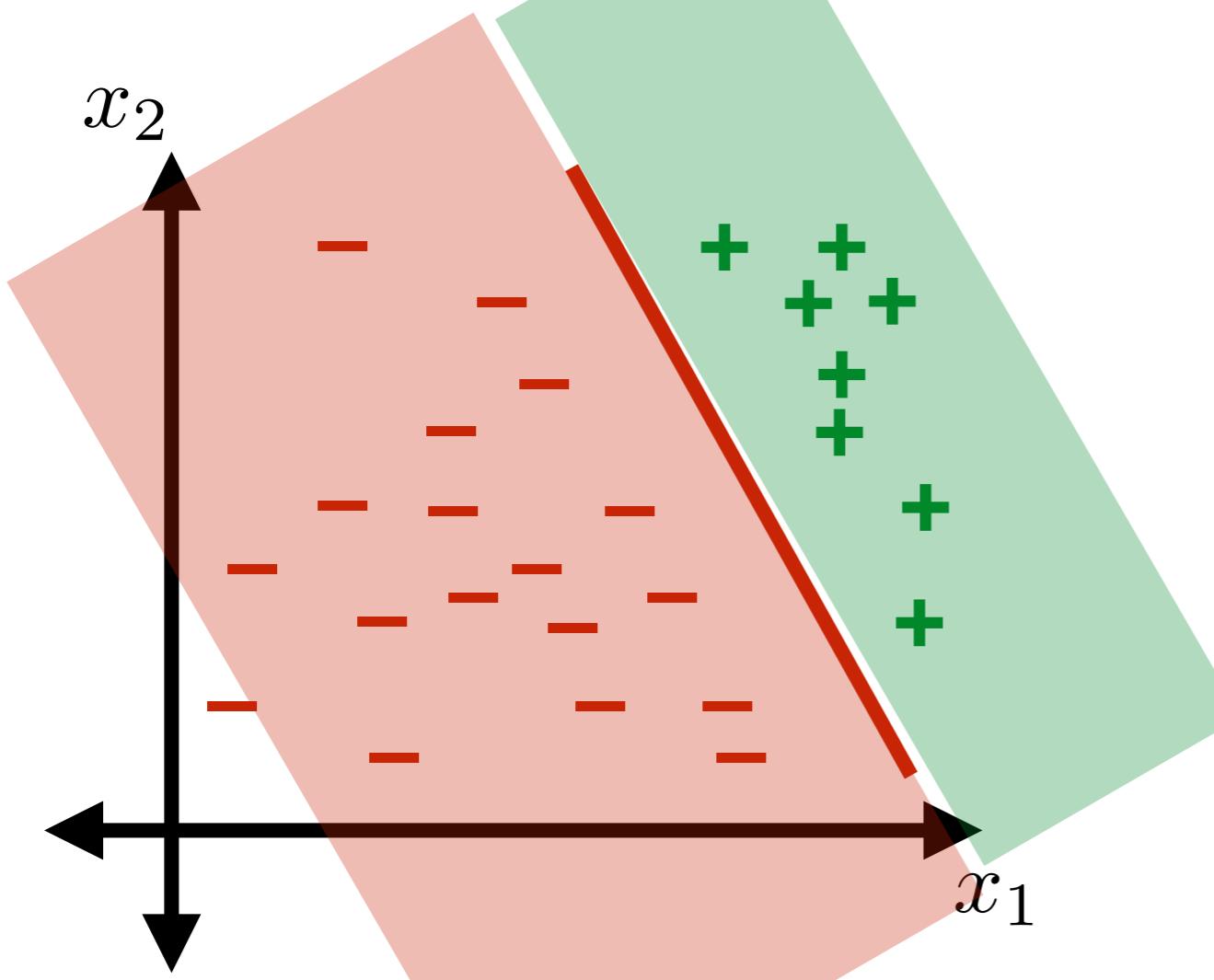


Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm $\rightarrow h$

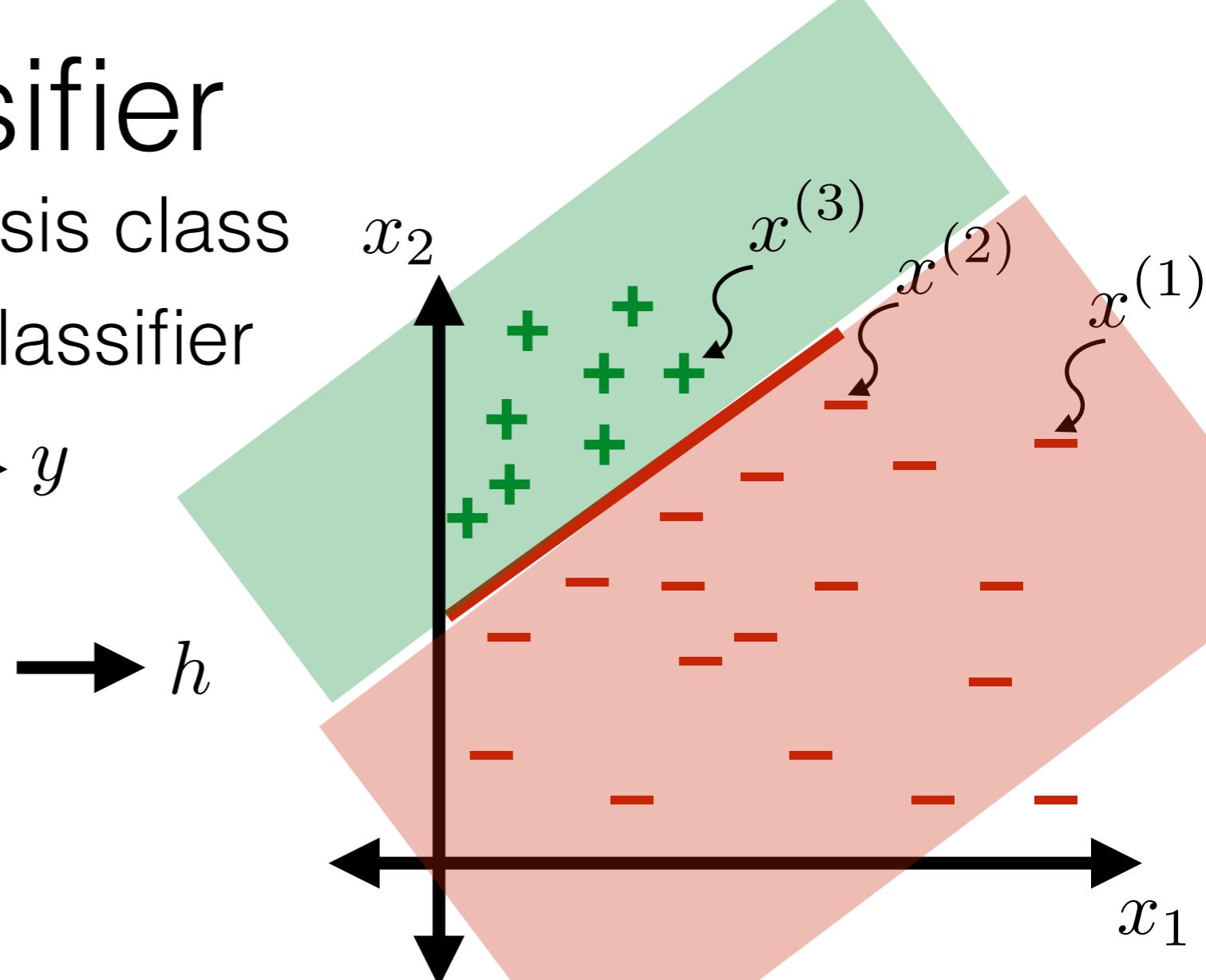


Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

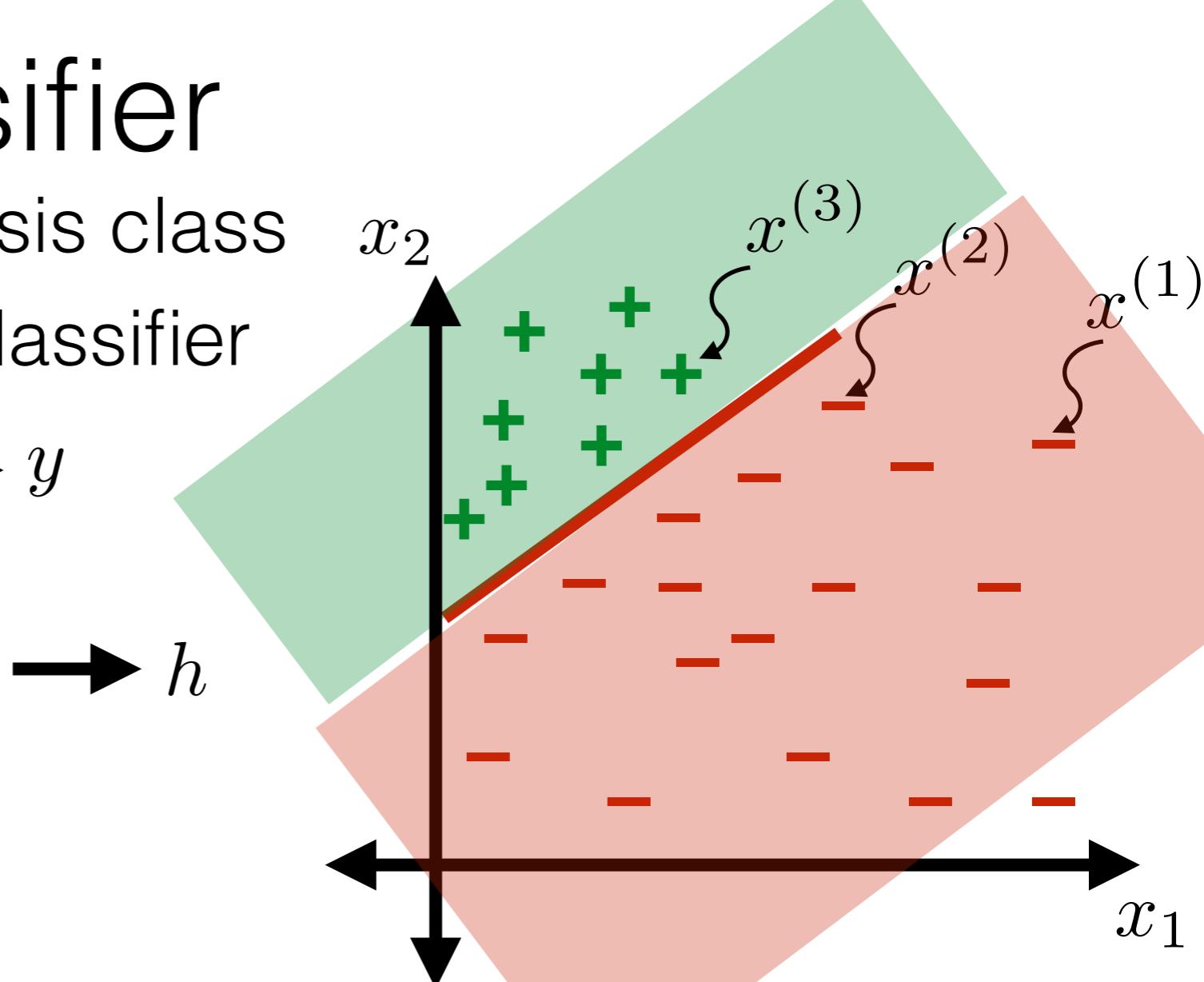
- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm



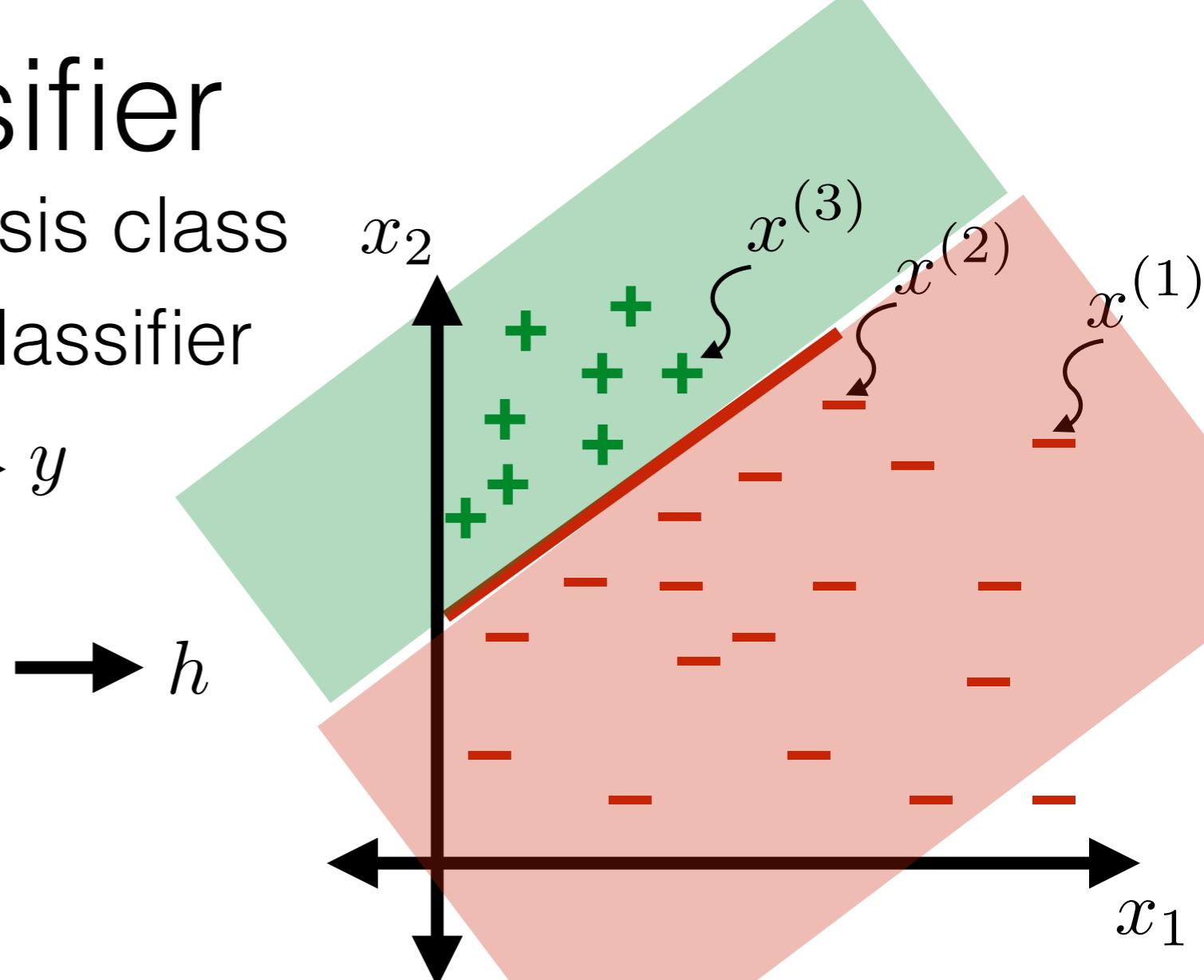
Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$
 - New:
 $\mathcal{D}_n \rightarrow$ learning algorithm
- Example:



Learning a classifier

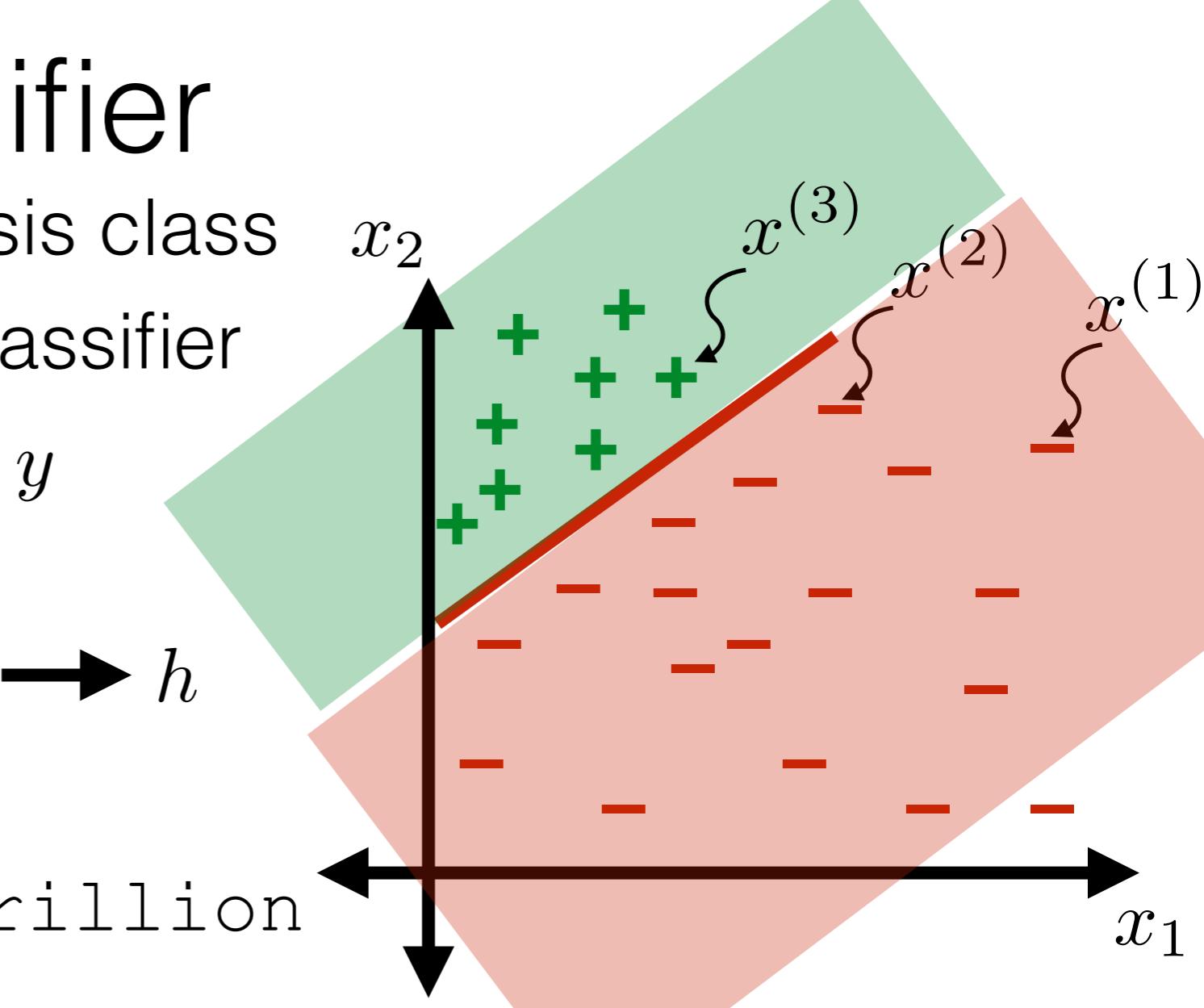
- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$
 - New:
 $\mathcal{D}_n \rightarrow$ learning algorithm
- Example:



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier
 - Recall: $x \rightarrow h \rightarrow y$
 - New:
 $\mathcal{D}_n \rightarrow$ learning algorithm
- Example:

for $j = 1, \dots, 1 \text{ trillion}$



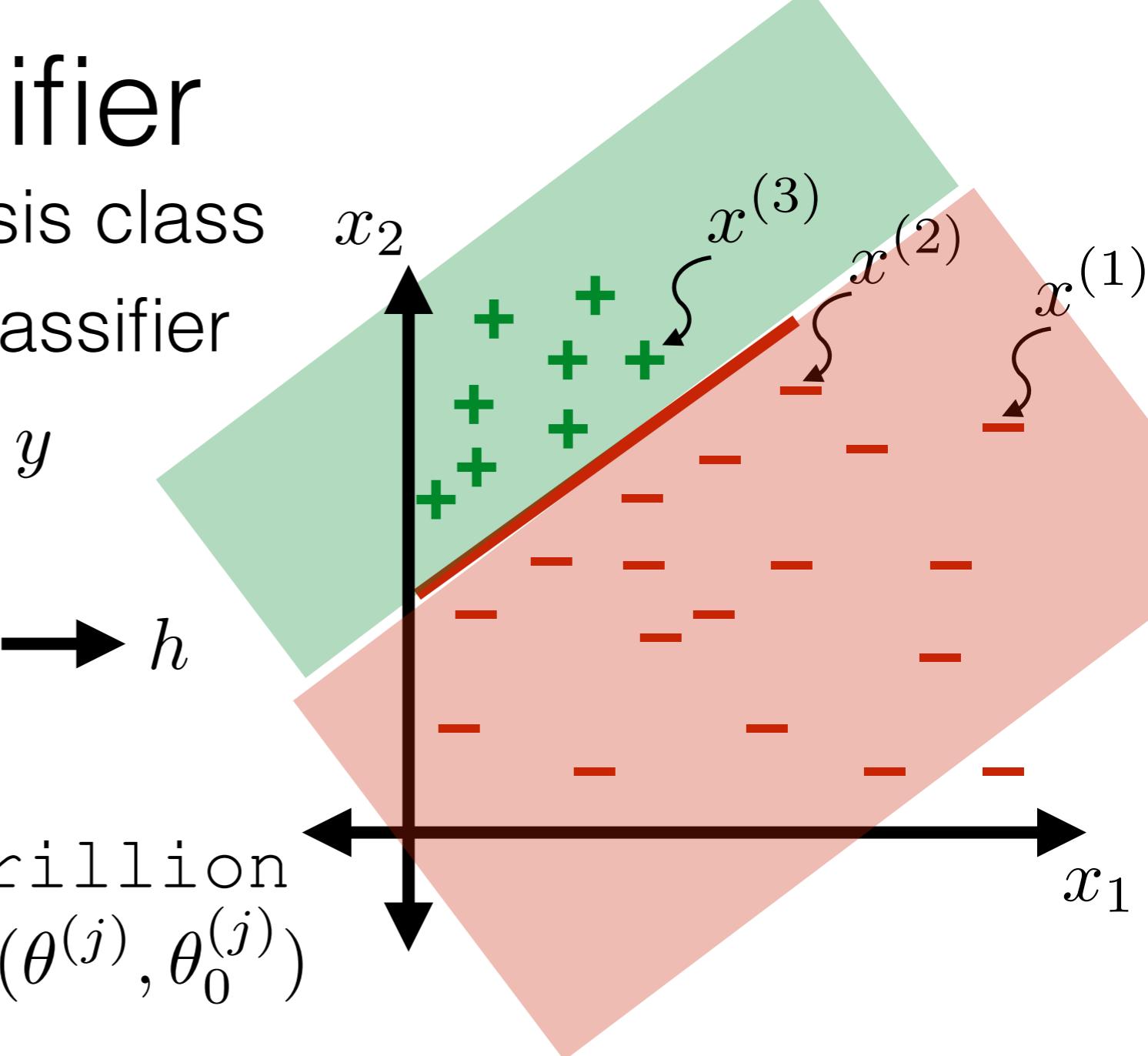
Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

- Example:
for $j = 1, \dots, 1 \text{ trillion}$
Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

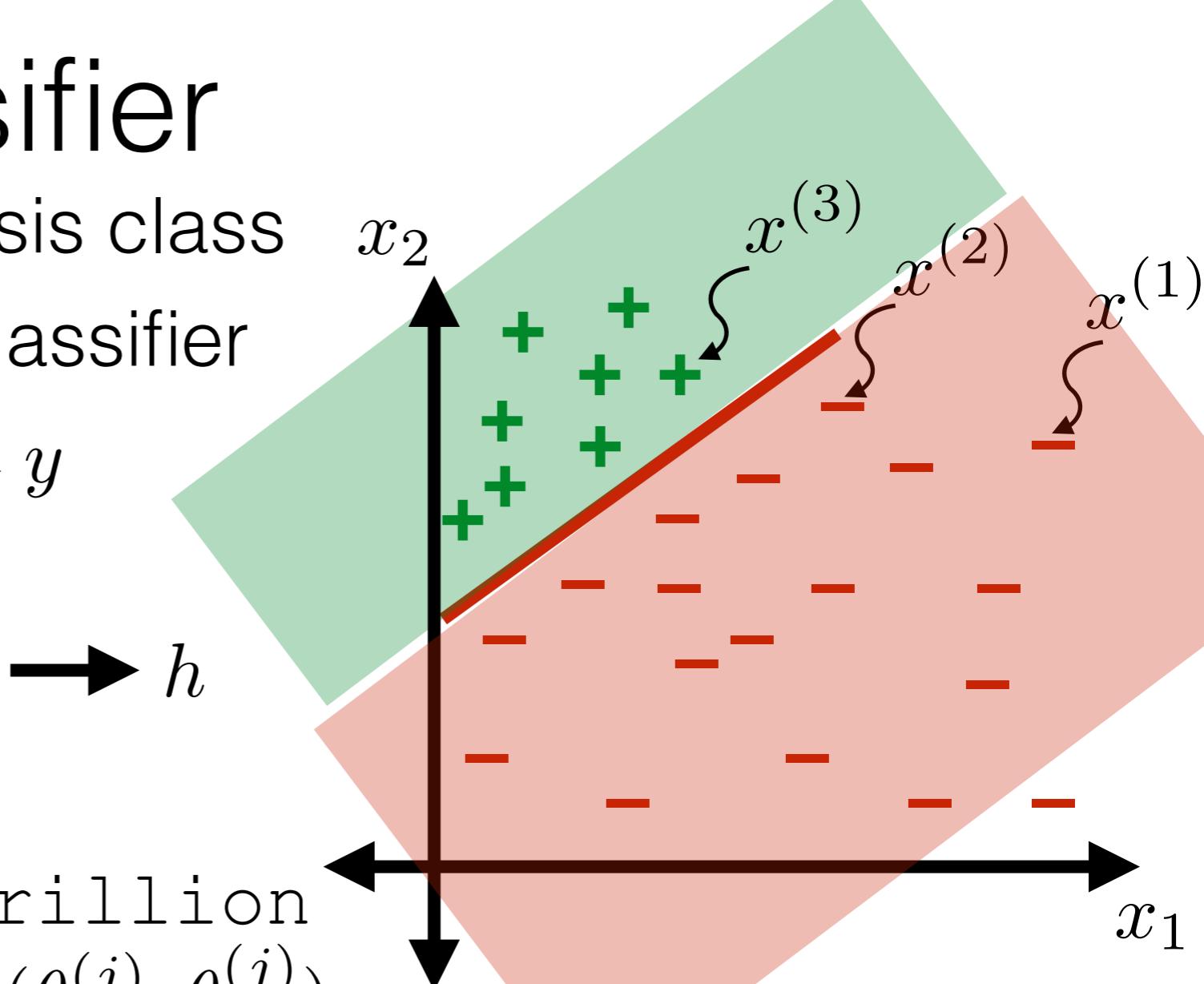
- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

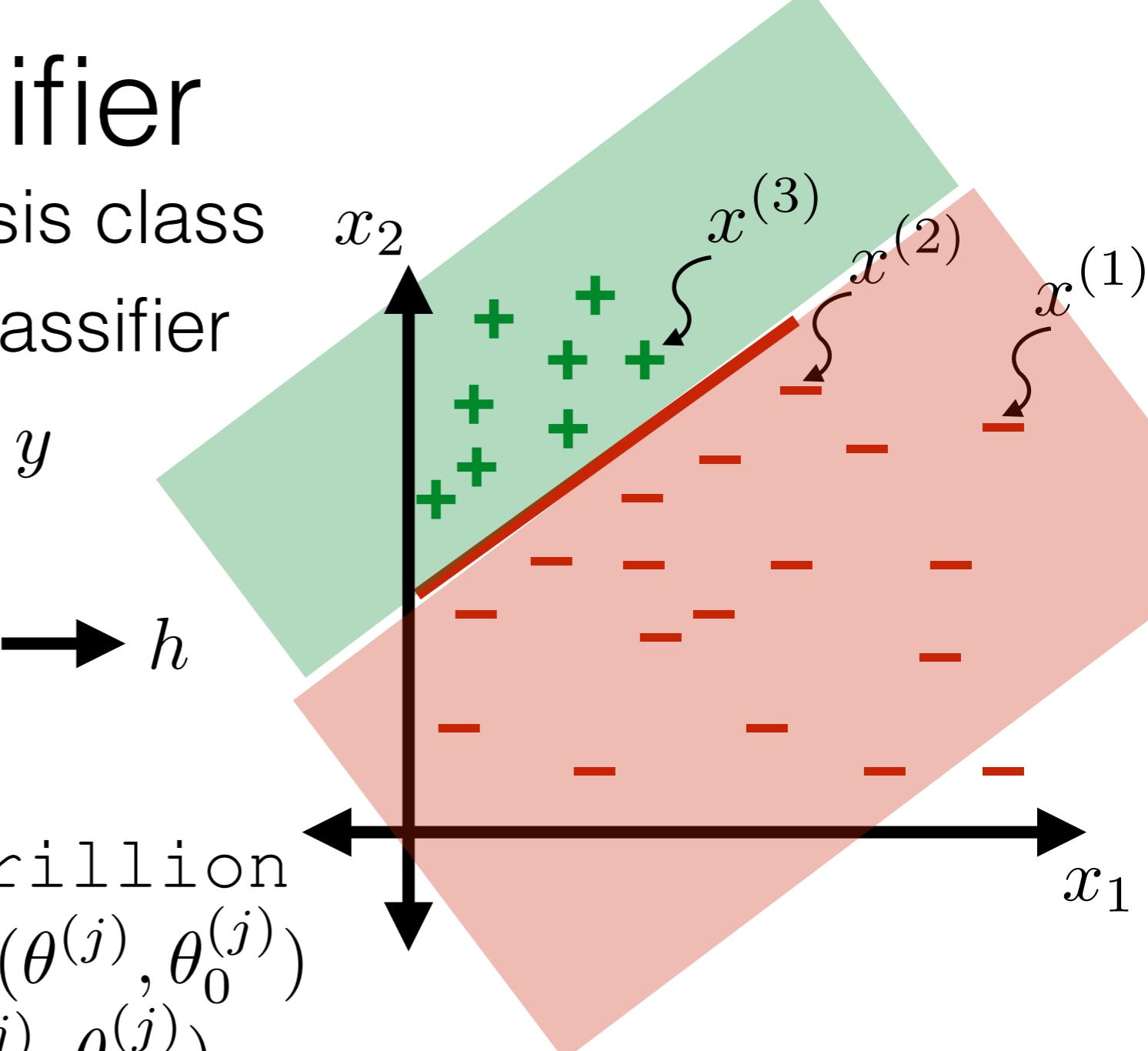
- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

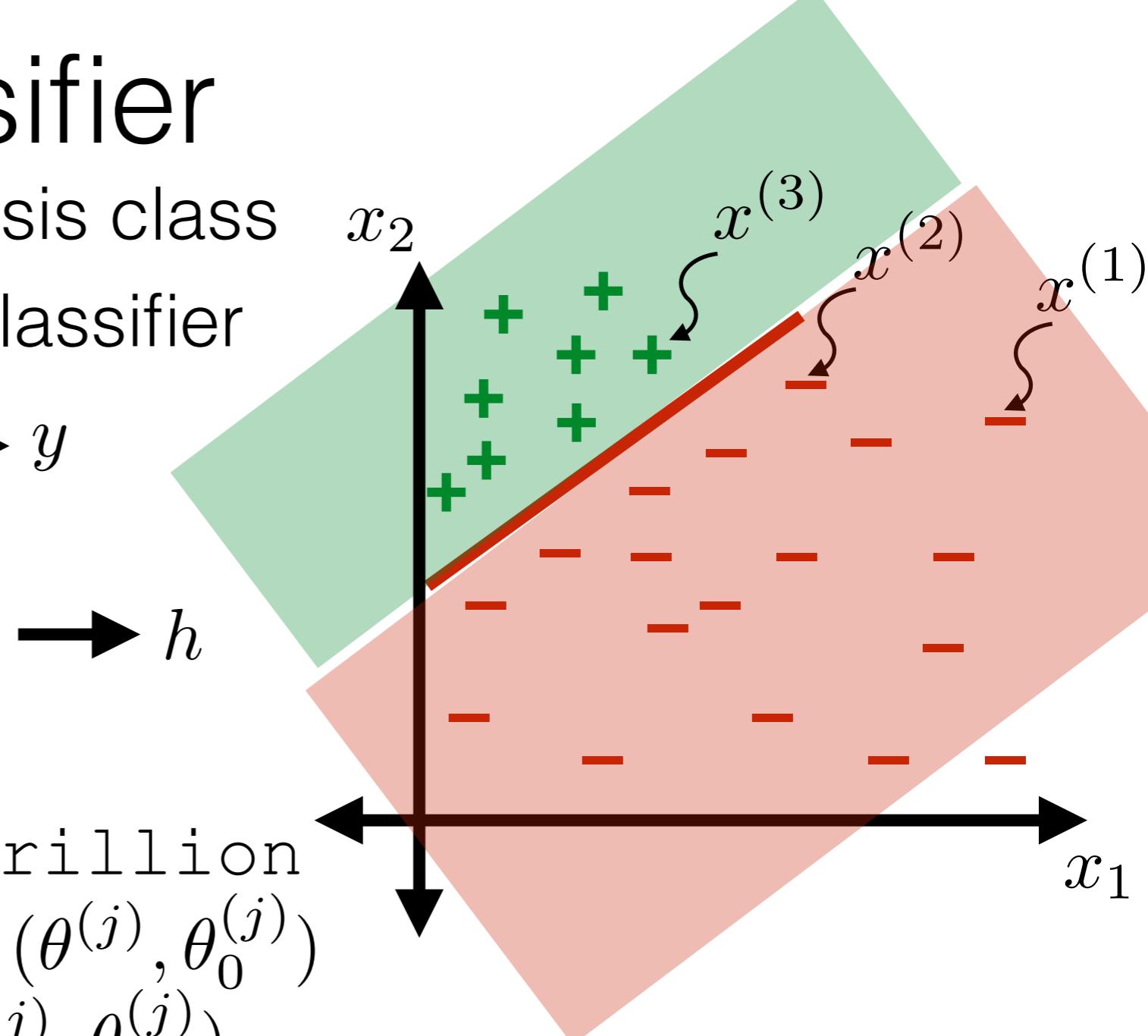
- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

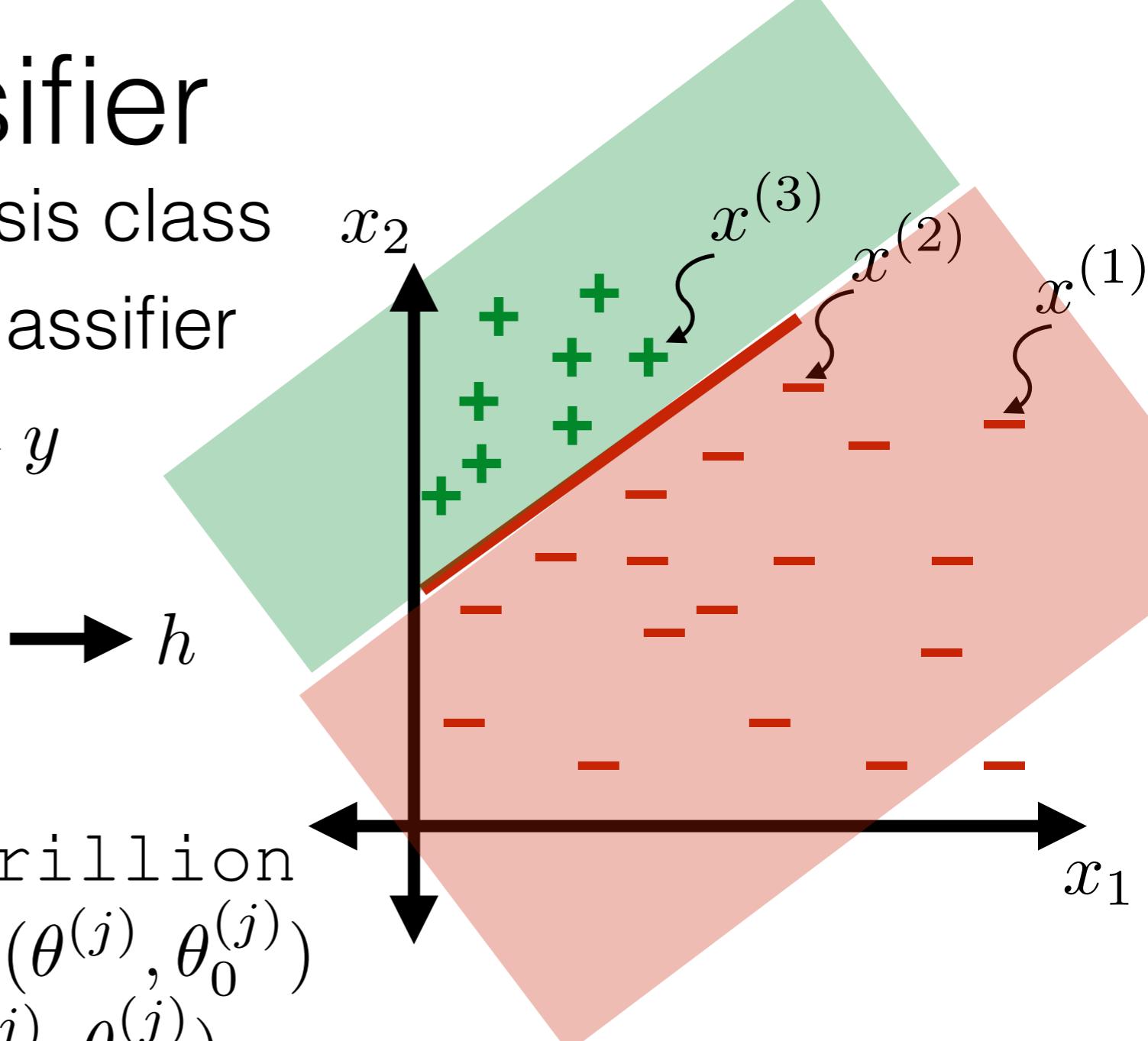
- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg(\mathcal{D}_n)



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

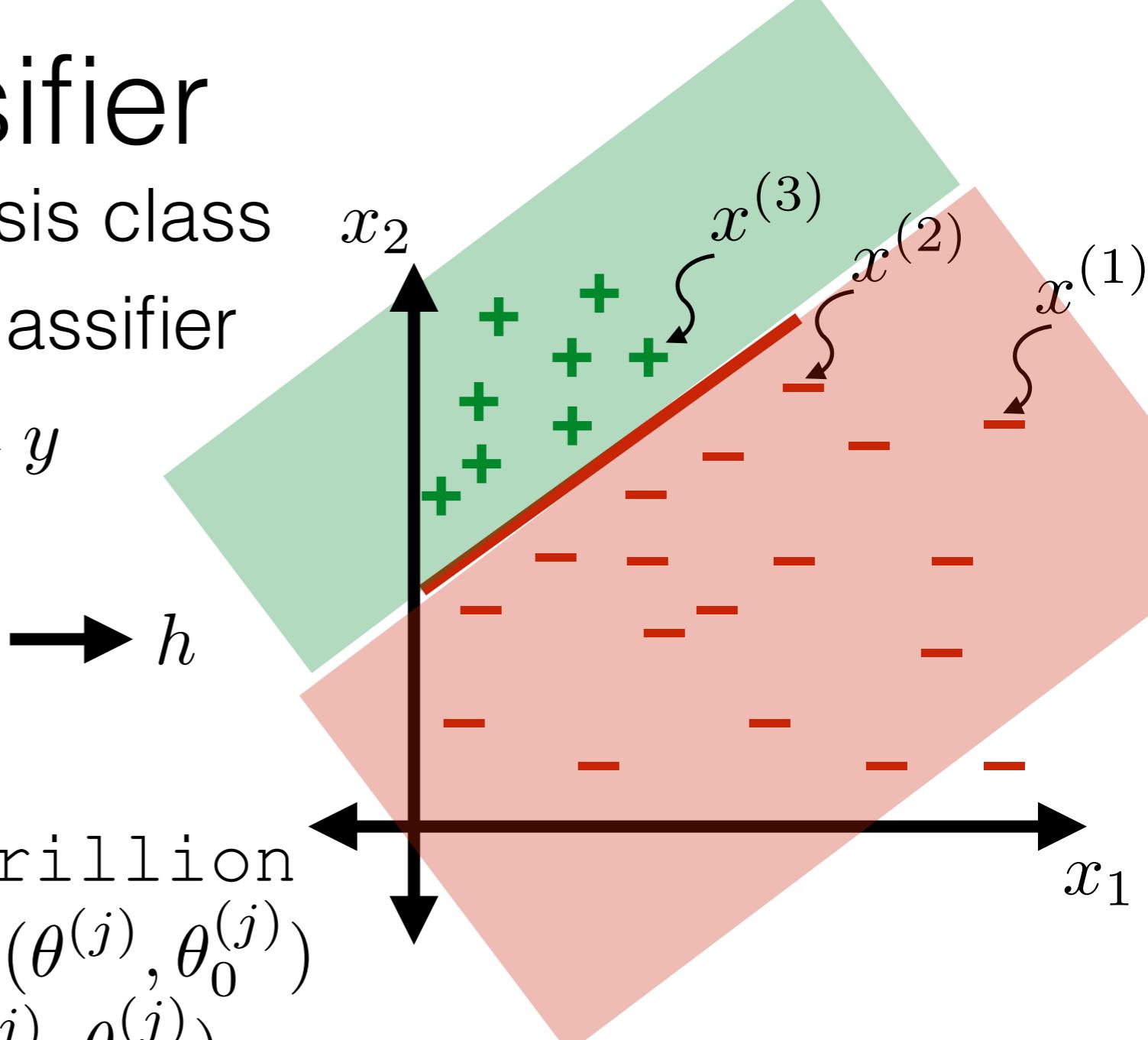
- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg(\mathcal{D}_n ; $k < 1 \text{ trillion}$)



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

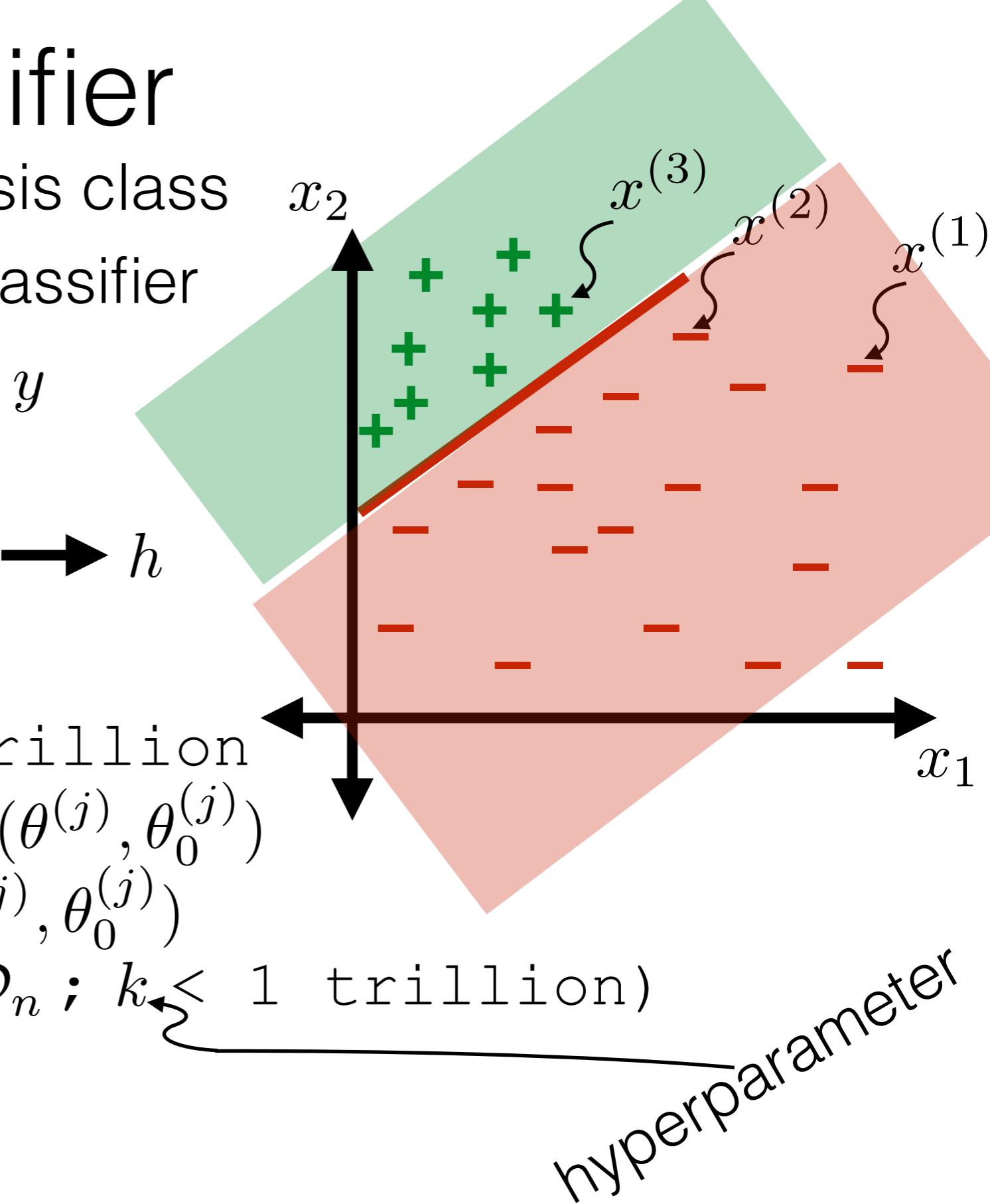
- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg(\mathcal{D}_n ; $k < 1 \text{ trillion}$)



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

- Example:

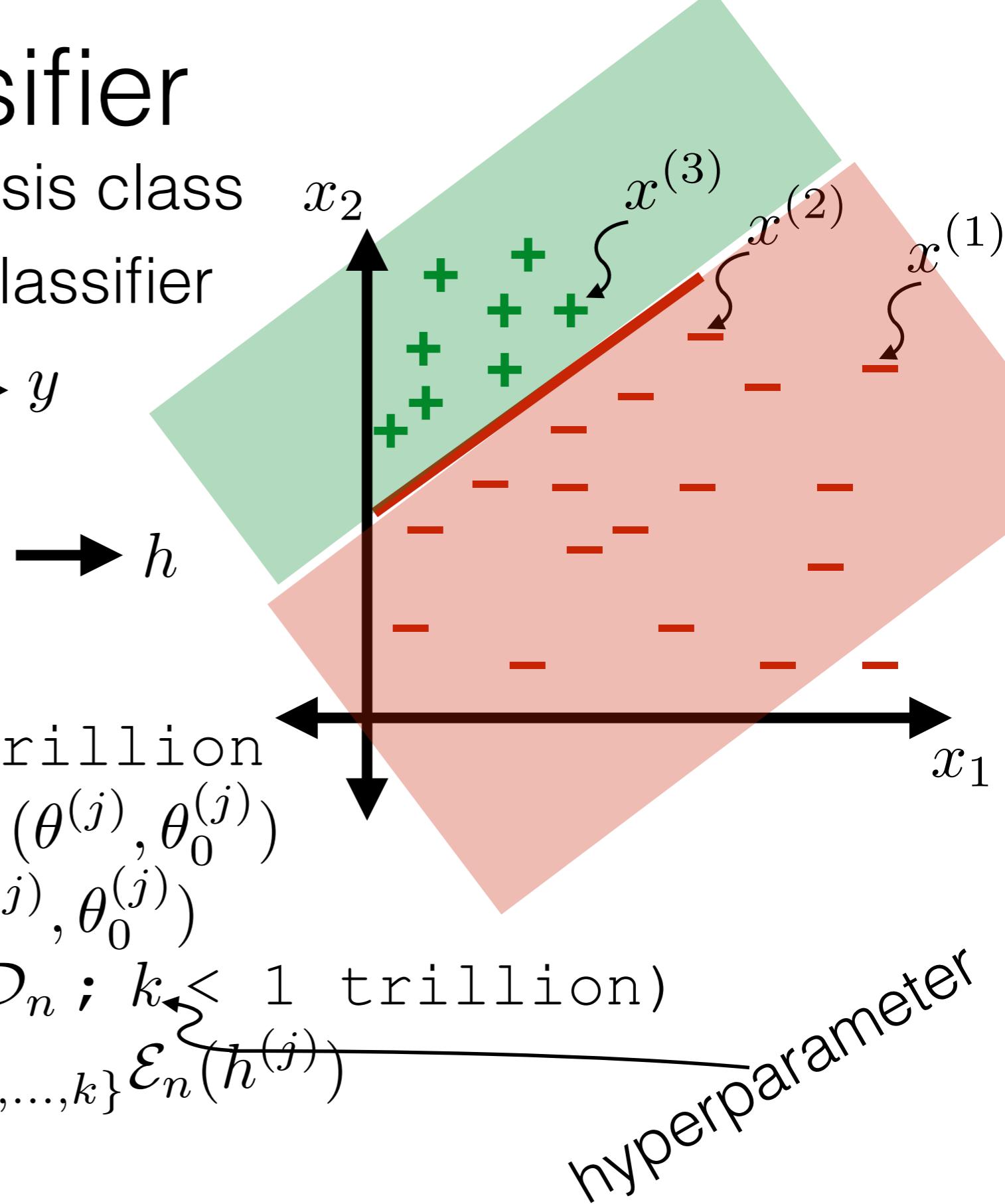
for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg($\mathcal{D}_n ; k < 1 \text{ trillion}$)

Set $j^* = \operatorname{argmin}_{j \in \{1, \dots, k\}} \mathcal{E}_n(h^{(j)})$



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

- Example:

for $j = 1, \dots, 1 \text{ trillion}$

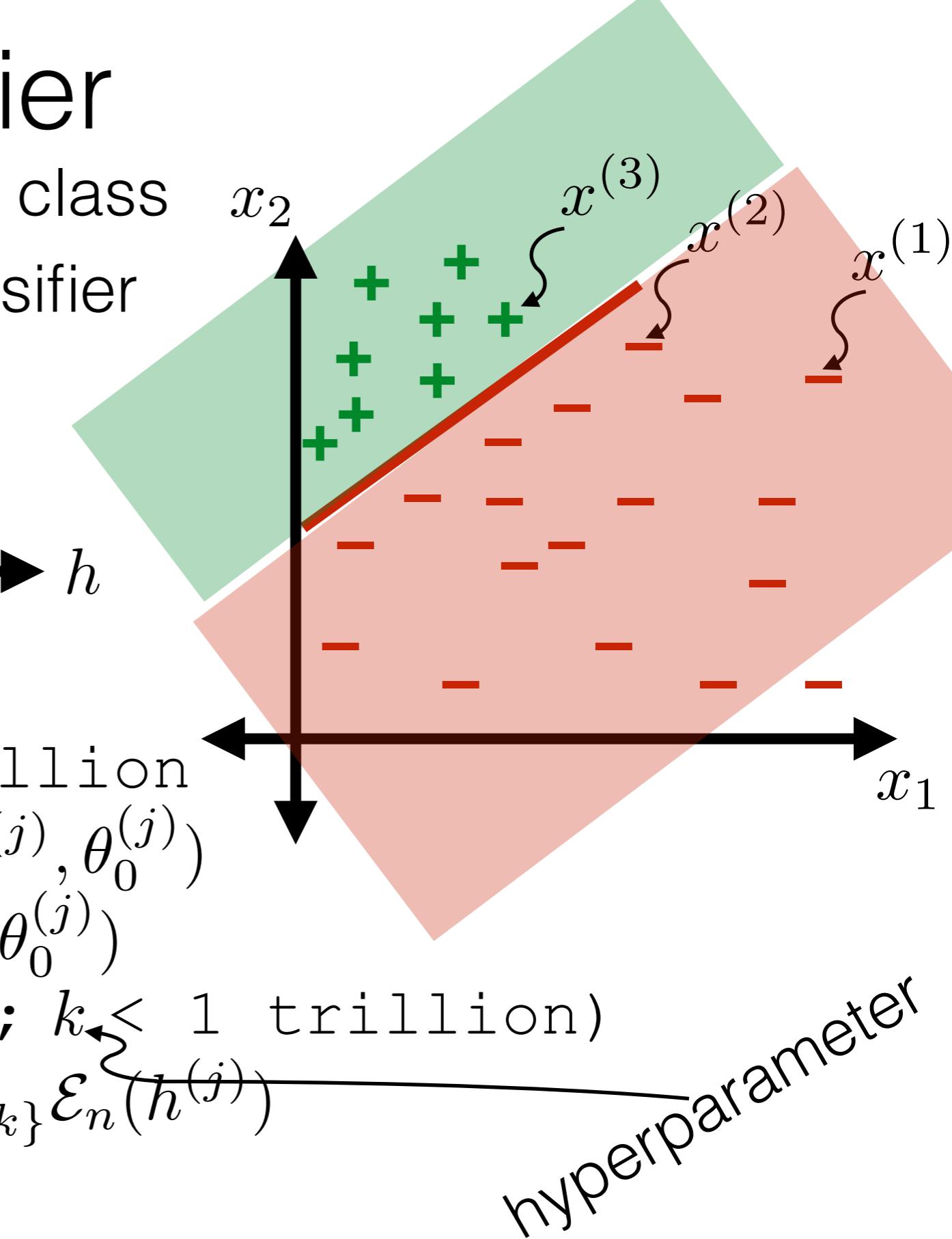
Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg(\mathcal{D}_n ; $k < 1 \text{ trillion}$)

Set $j^* = \operatorname{argmin}_{j \in \{1, \dots, k\}} \mathcal{E}_n(h^{(j)})$

Return $h^{(j^*)}$



Learning a classifier

- Have data; have hypothesis class
- Want to choose a good classifier

- Recall: $x \rightarrow h \rightarrow y$

- New:
 $\mathcal{D}_n \rightarrow$ learning algorithm

- Example:

for $j = 1, \dots, 1 \text{ trillion}$

Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$

Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex_learning_alg(\mathcal{D}_n ; $k < 1 \text{ trillion}$)

Set $j^* = \operatorname{argmin}_{j \in \{1, \dots, k\}} \mathcal{E}_n(h^{(j)})$

Return $h^{(j^*)}$

- How does training error of $\text{Ex_learning_alg}(\mathcal{D}_n; 1)$ compare to the training error of $\text{Ex_learning_alg}(\mathcal{D}_n; 2)$? hyperparameter

