10-301/601: Introduction to Machine Learning Lecture 15 – Learning Theory (Infinite Case)

Henry Chai & Matt Gormley & Hoda Heidari

10/23/23

Front Matter

- Announcements
  - HW5 released 10/9, due 10/27 (Friday) at 11:59 PM
  - Exam 3 scheduled
    - Tuesday, December 12th from 5:30 PM to 8:30 PM
  - Sign up for peer tutoring! See Piazza for more details

Recall - Theorem 1: Finite, Realizable Case

- For a *finite* hypothesis set $\mathcal{H}$ such that $c^* \in \mathcal{H}$ (*realizable*) and arbitrary distribution p^*, if the number of labelled training data points satisfies

$$N \geq \frac{1}{\epsilon}\left(\ln(|\mathcal{H}|) + \ln\left(\frac{1}{\delta}\right)\right)$$

then with probability at least $1 - \delta$, all $h \in \mathcal{H}$ with $\hat{R}(h) = 0$ have $R(h) \leq \epsilon$

Recall - Theorem 1: Finite, Realizable Case

- For a *finite* hypothesis set $\mathcal{H}$ such that $c^* \in \mathcal{H}$ (*realizable*) and arbitrary distribution p^*, if the number of labelled training data points satisfies

$$N = \frac{1}{\epsilon}\left(\ln(|\mathcal{H}|) + \ln\left(\frac{1}{\delta}\right)\right)$$

then with probability at least $1 - \delta$, all $h \in \mathcal{H}$ with $\hat{R}(h) = 0$ have $R(h) \leq \epsilon$

- Making the bound tight and solving for ϵ gives…

Statistical Learning Theory Corollary

- For a *finite* hypothesis set $\mathcal{H}$ such that $c^* \in \mathcal{H}$ (*realizable*) and arbitrary distribution p^*, given a training dataset S where $|S| = N$, all $h \in \mathcal{H}$ with $\hat{R}(h) = 0$ have

$$R(h) \leq \frac{1}{N}\left(\ln(|\mathcal{H}|) + \ln\left(\frac{1}{\delta}\right)\right)$$

with probability at least $1 - \delta$.

Recall - Theorem 2: Finite, Agnostic Case

Agnostic: $\begin{cases} c^* \notin \mathcal{H} \\ c^* \in \mathcal{H} \end{cases}$

- For a *finite* hypothesis set $\mathcal{H}$ and arbitrary distribution p^*, if the number of labelled training data points satisfies

$$N \geq \frac{1}{2\epsilon^2}\left(\ln(|\mathcal{H}|) + \ln\left(\frac{2}{\delta}\right)\right)$$

then with probability at least $1 - \delta$, all $h \in \mathcal{H}$ satisfy $|R(h) - \hat{R}(h)| \leq \epsilon$

- Bound is inversely quadratic in ϵ, e.g., halving ϵ means we need four times as many labelled training data points

Statistical Learning Theory Corollary

- For a *finite* hypothesis set $\mathcal{H}$ and arbitrary distribution p^*, given a training dataset S where $|S| = N$, all $h \in \mathcal{H}$ have

$$R(h) \leq \hat{R}(h) + \sqrt{\frac{1}{2N}\left(\ln(|\mathcal{H}|) + \ln\left(\frac{2}{\delta}\right)\right)}$$

ϵ

with probability at least $1 - \delta$.

What happens when $|\mathcal{H}| = \infty$?

- For a *finite* hypothesis set $\mathcal{H}$ and arbitrary distribution p^*, given a training data set S where $|S| = N$, all $h \in \mathcal{H}$ have

$$R(h) \leq \hat{R}(h) + \sqrt{\frac{1}{2N}\left(\ln(|\mathcal{H}|) + \ln\left(\frac{2}{\delta}\right)\right)}$$

with probability at least $1 - \delta$.

Labellings

- Given some finite set of data points $S = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$ and some hypothesis $h \in \mathcal{H}$, applying h to each point in S results in a **labelling**
  - $\left[h\left(\boldsymbol{x}^{(1)}\right), \ldots, h\left(\boldsymbol{x}^{(N)}\right)\right]$ is a vector of N +1's and -1's (recall: our discussion of PAC learning assumes binary classification)
- Given $S = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$, each hypothesis in $\mathcal{H}$ induces a labelling but not necessarily a unique labelling
  - The set of labellings induced by $\mathcal{H}$ on S is

$$\mathcal{H}(S) = \left\{\left[h\left(\boldsymbol{x}^{(1)}\right), \ldots, h\left(\boldsymbol{x}^{(N)}\right)\right] \mid h \in \mathcal{H}\right\}$$

(binary classification)

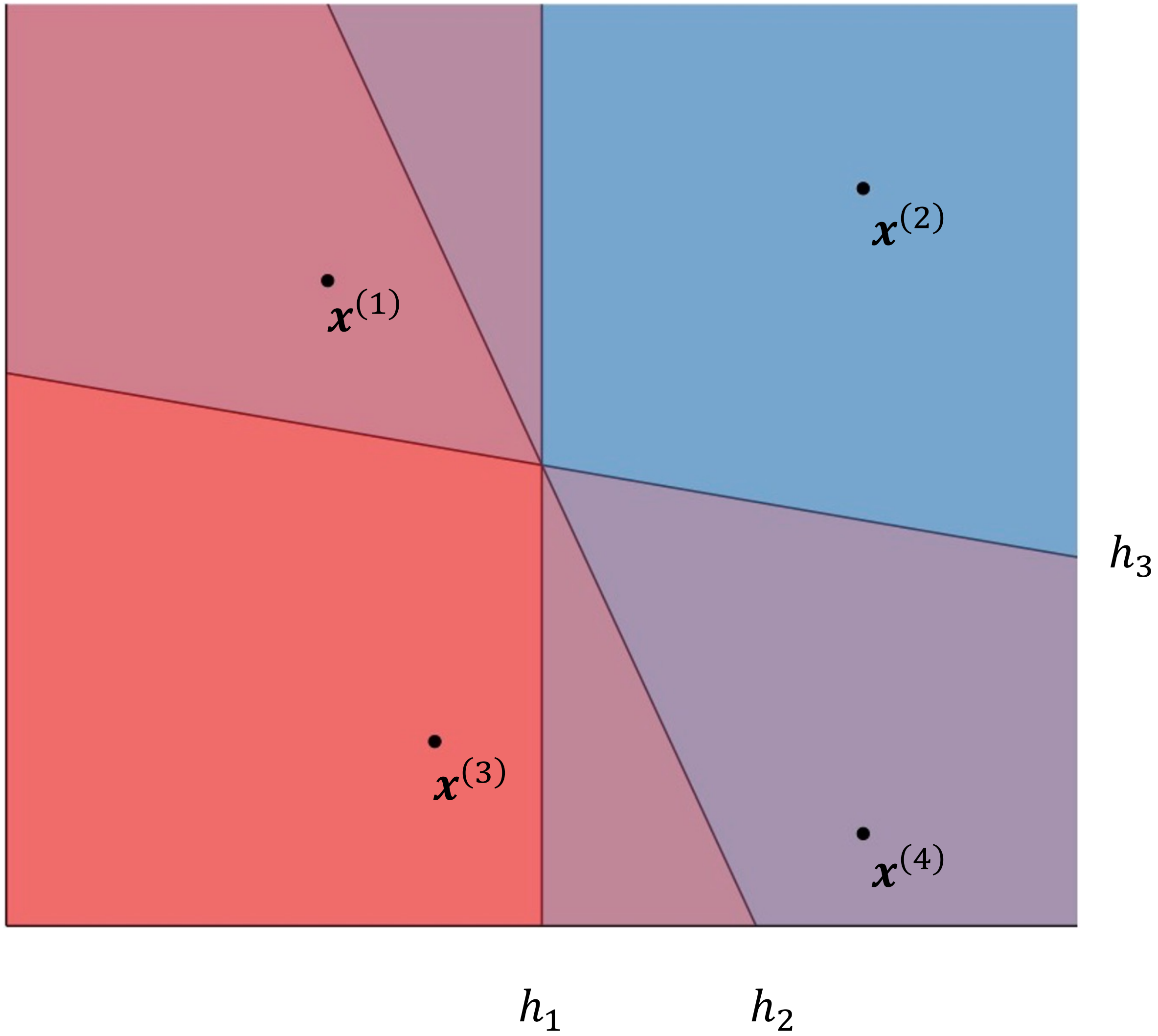
$|\mathcal{H}(S)| \leq 2^{|S|}$

Example: Labellings

$\mathcal{H} = \{h_1, h_2, h_3\}$

$\mathcal{H}(S) = ?$

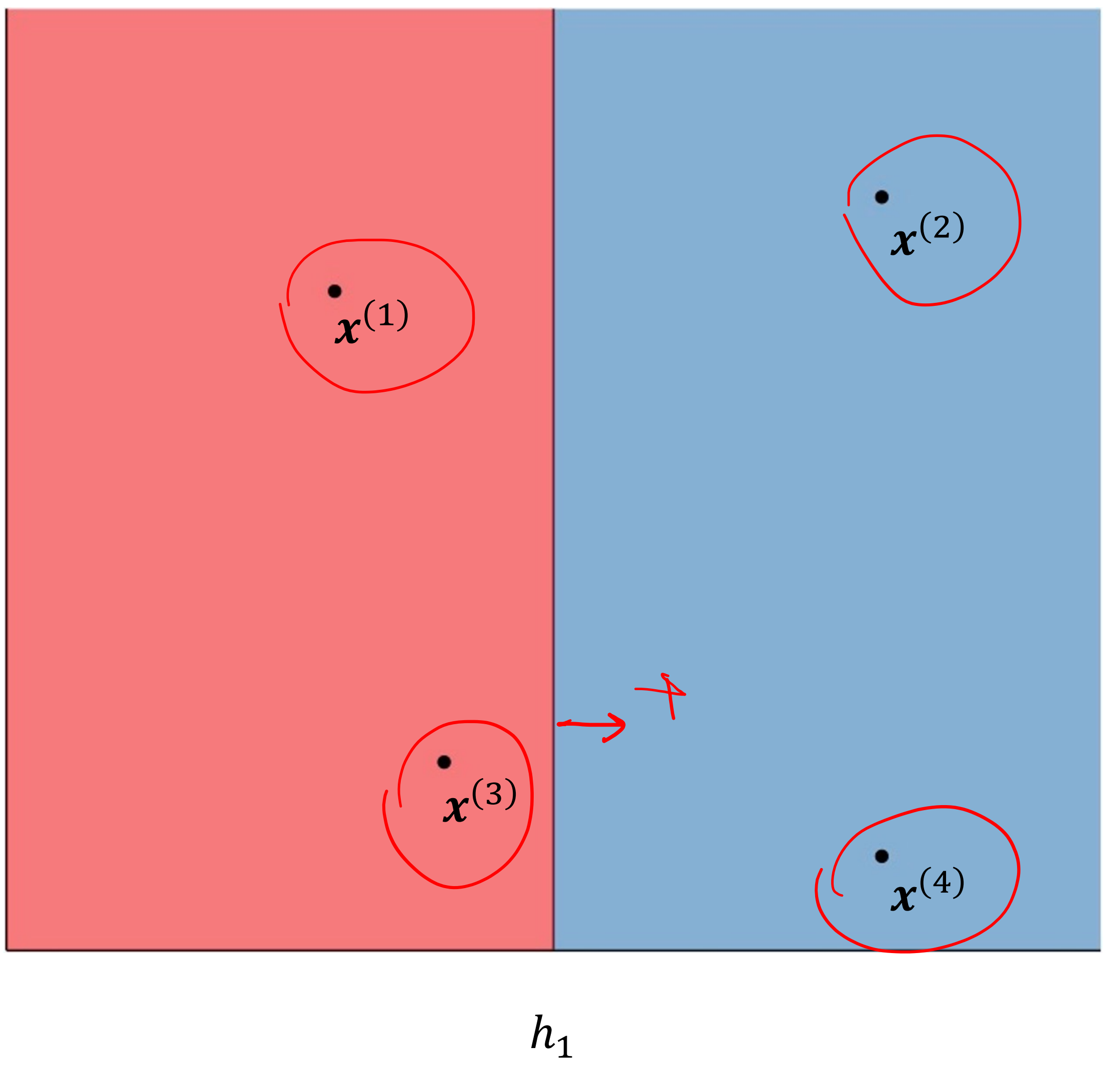
$S = \{x^{(1)}, \ldots, x^{(4)}\}$



Example: Labellings

$\mathcal{H} = \{h_1, h_2, h_3\}$

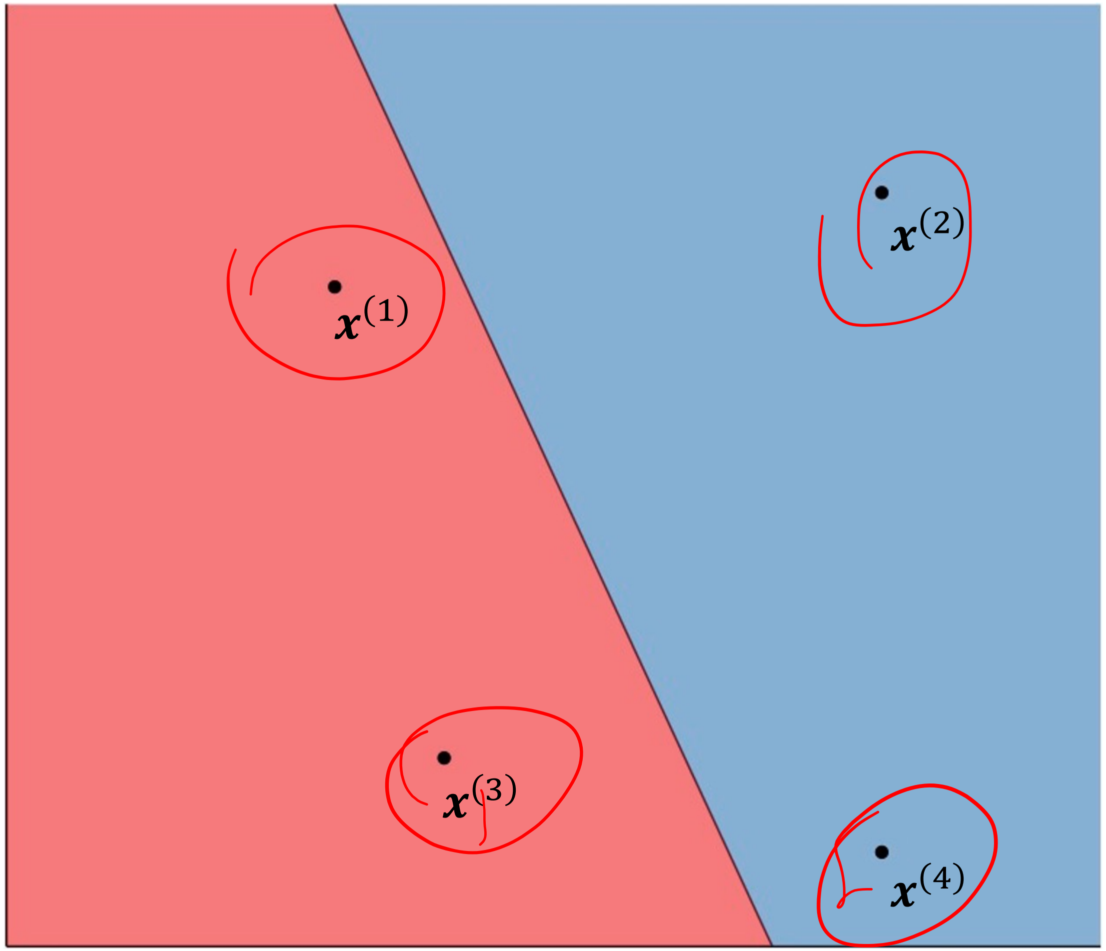
$[h_1(\boldsymbol{x}^{(1)}), h_1(\boldsymbol{x}^{(2)}), h_1(\boldsymbol{x}^{(3)}), h_1(\boldsymbol{x}^{(4)})]$
$= (-1, +1, -1, +1)$



Example: Labellings

$\mathcal{H} = \{h_1, h_2, h_3\}$

$[h_2(\boldsymbol{x}^{(1)}), h_2(\boldsymbol{x}^{(2)}), h_2(\boldsymbol{x}^{(3)}), h_2(\boldsymbol{x}^{(4)})]$
$= (-1, +1, -1, +1)$



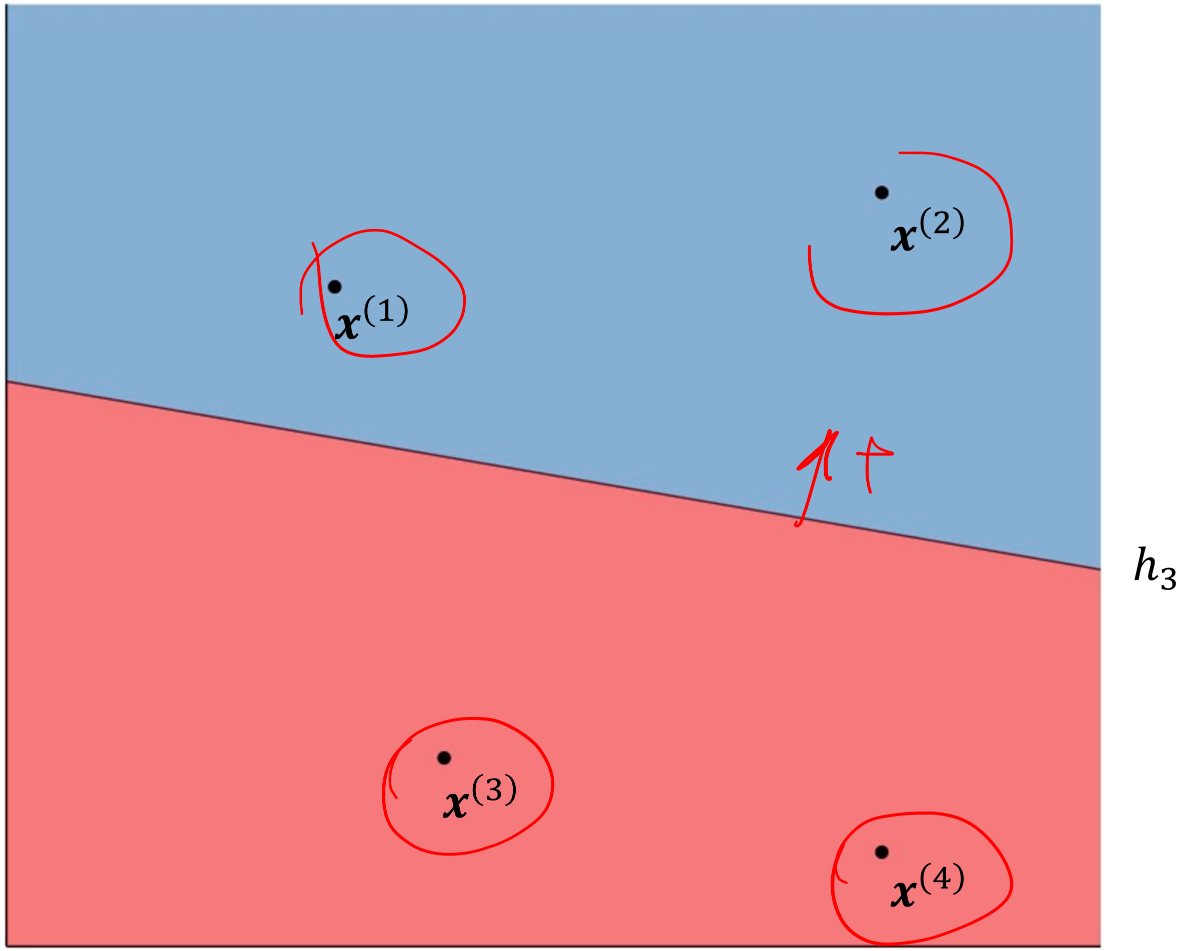
h_2

Example: Labellings

$\mathcal{H} = \{h_1, h_2, h_3\}$

$[h_1(\boldsymbol{x}^{(1)}), h_1(\boldsymbol{x}^{(2)}), h_1(\boldsymbol{x}^{(3)}), h_1(\boldsymbol{x}^{(4)})]$

$= (+1, +1, -1, -1)$



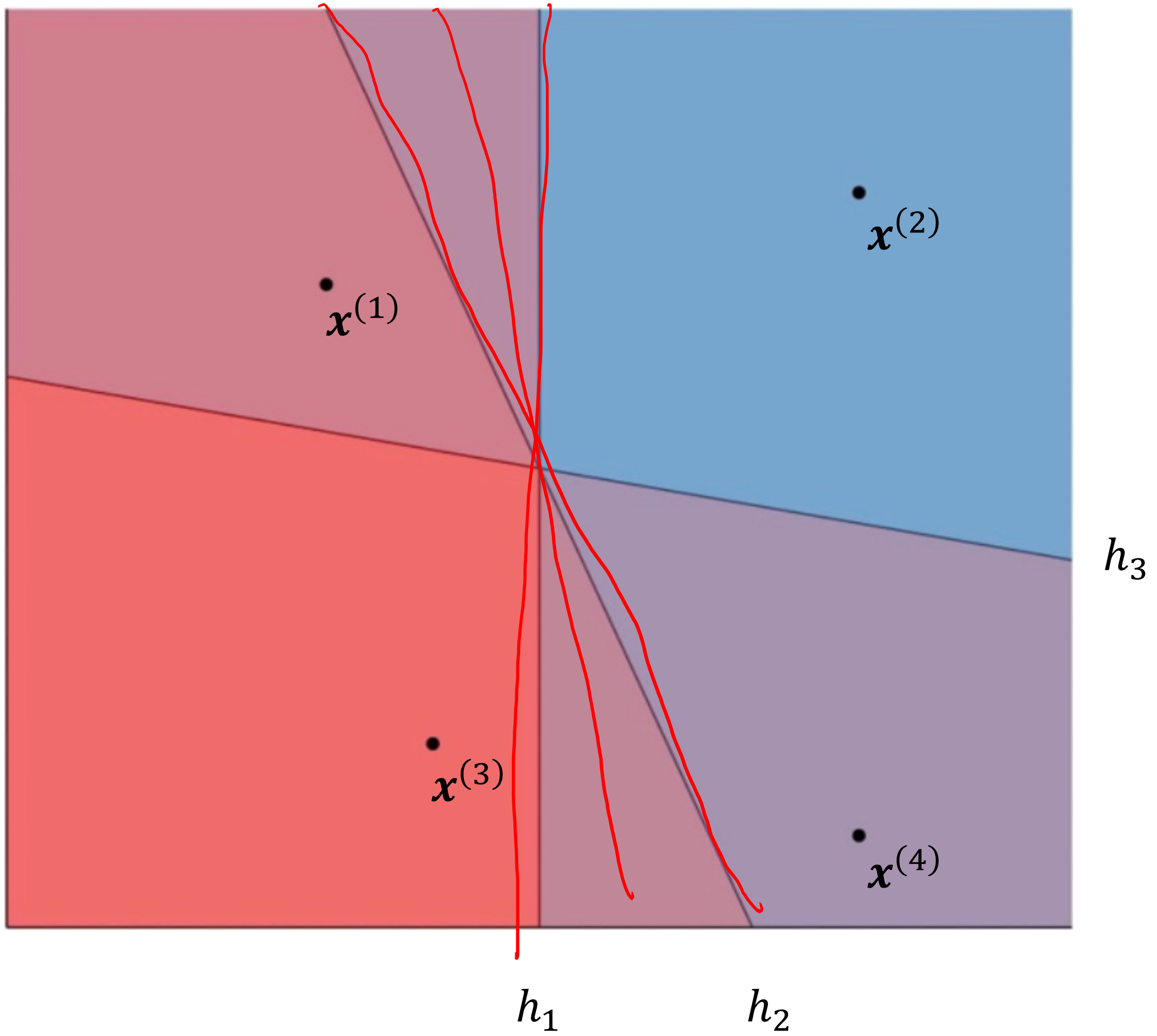
Example: Labellings

$\mathcal{H} = \{h_1, h_2, h_3\}$

$\mathcal{H}(S)$

$= \{[+1, +1, -1, -1], [-1, +1, -1, +1]\}$

$|\mathcal{H}(S)| = 2$



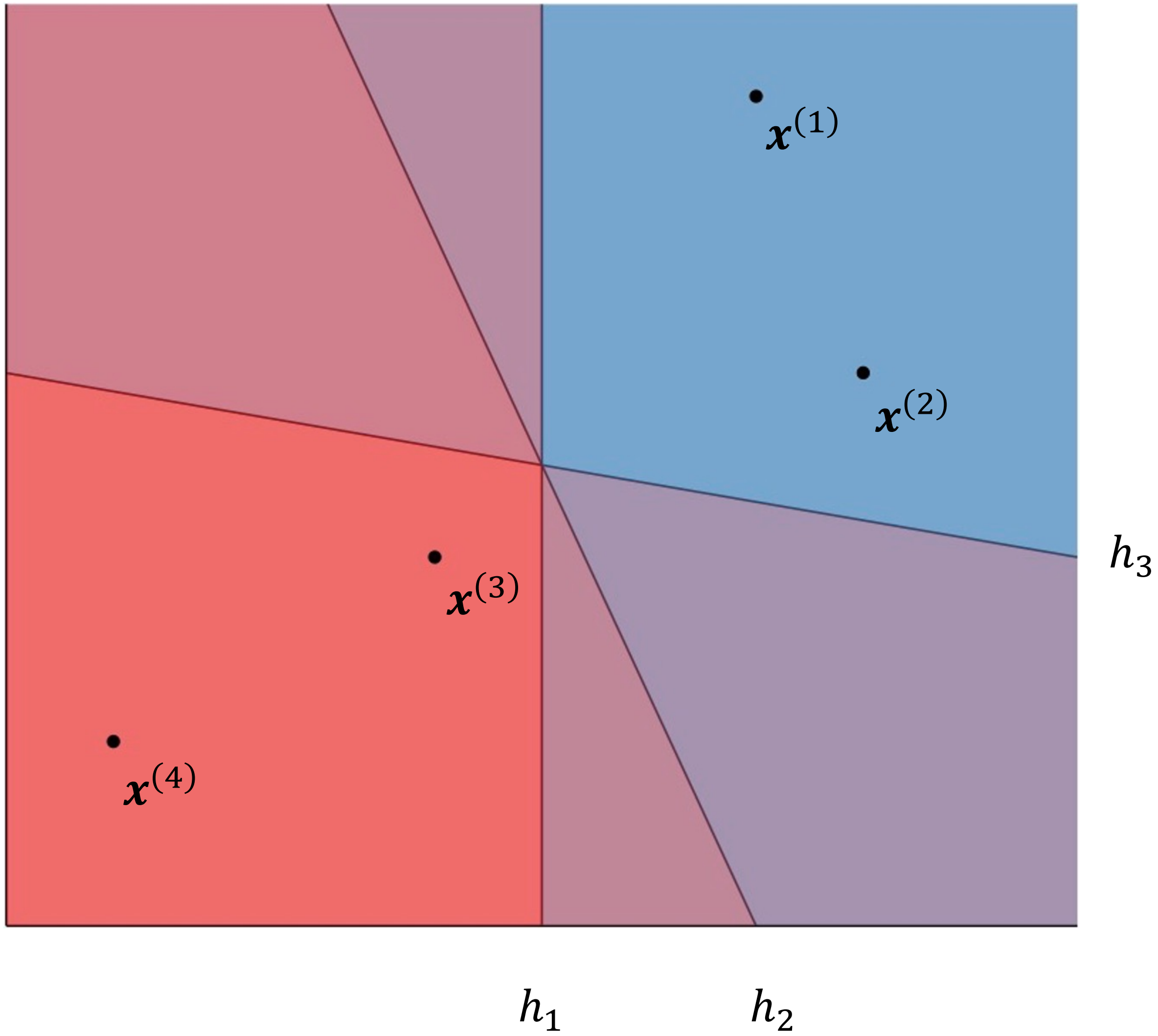
Example: Labellings

$\mathcal{H} = \{h_1, h_2, h_3\}$

$\mathcal{H}(S)$

$= \{[+1, +1, -1, -1]\}$

$|\mathcal{H}(S)| = 1$



VC-Dimension

- $\mathcal{H}(S)$ is the set of all labellings induced by $\mathcal{H}$ on S
  - If $|S| = N$, then $|\mathcal{H}(S)| \leq 2^N$
  - $\mathcal{H}$ **shatters** S if $|\mathcal{H}(S)| = 2^N$
- The **VC-dimension** of $\mathcal{H}$, $VC(\mathcal{H})$, is the size of the largest set S that can be shattered by $\mathcal{H}$.
  - If $\mathcal{H}$ can shatter arbitrarily large finite sets, then $VC(\mathcal{H}) = \infty$
- To prove that $VC(\mathcal{H}) = d$, you need to show

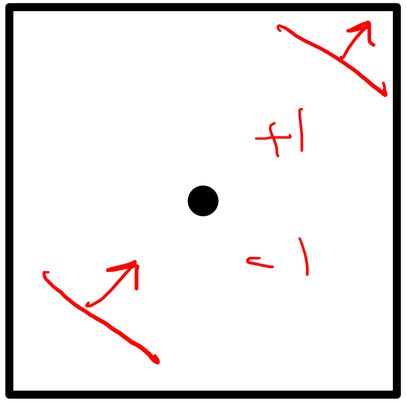
$VC(H) \geq d$ 1. $\exists$ some set of d data points that $\mathcal{H}$ can shatter and

$VC(H) \ngtr d$ 2. $\nexists$ a set of $d + 1$ data points that $\mathcal{H}$ can shatter

VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?

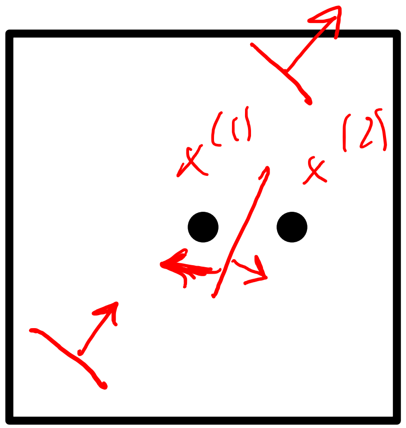
VC(H) ≥ 1



S

VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points? $VC(H) \geq 2$

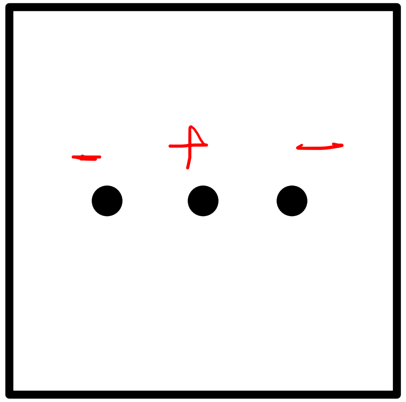


S

$x^{(1)}$	+	+	−	−
$x^{(2)}$	+	−	+	−
	✓	✓	✓	✓

VC-Dimension: Example

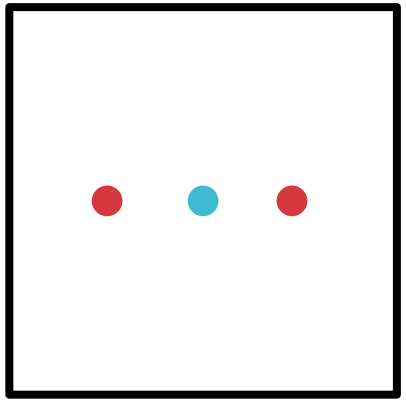
- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - VC(H) ≥ 3 Can $\mathcal{H}$ shatter some set of 3 points?



S

VC-Dimension: Example

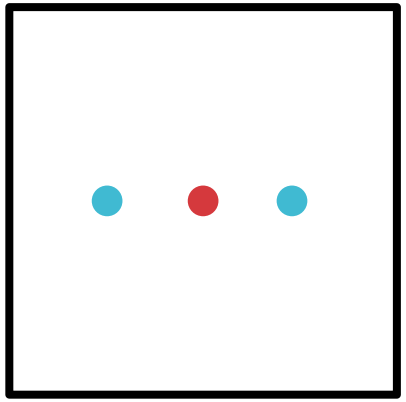
- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?



S

VC-Dimension: Example

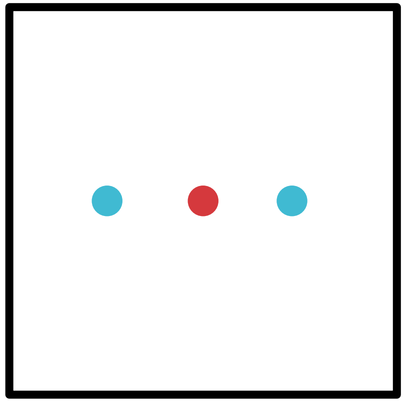
- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H} =$ all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?



S

VC-Dimension: Example

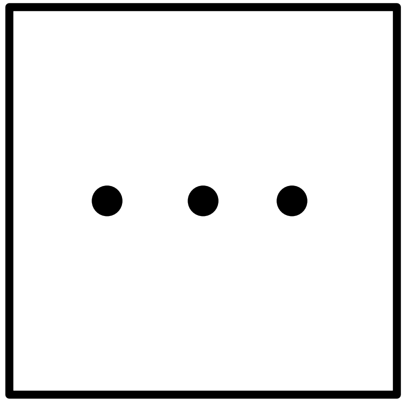
- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H} =$ all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter ***some*** set of 3 points?



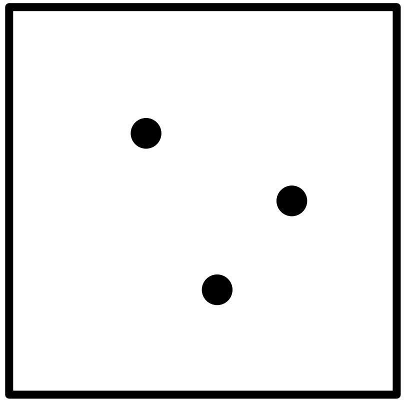
S

VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?



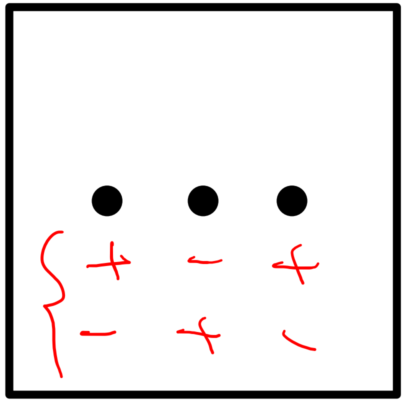
S_1



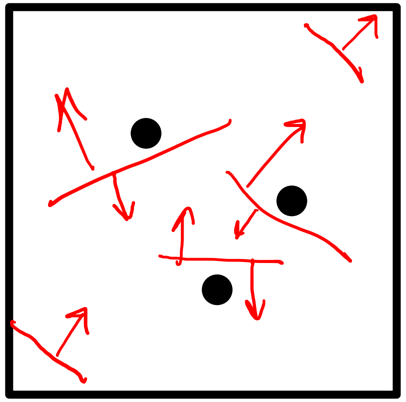
S_2

VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?



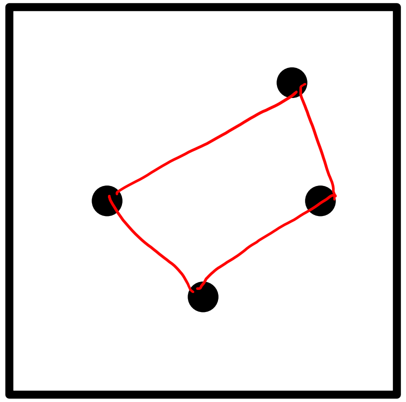
$|\mathcal{H}(S_1)| = 6$



$|\mathcal{H}(S_2)| = 8$

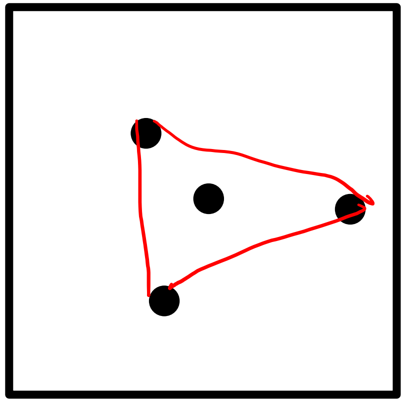
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



S_1

All points on the convex hull

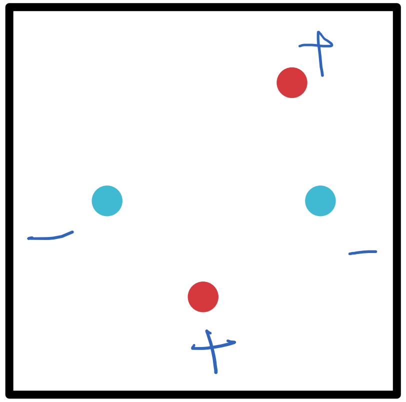


S_2

At least one point inside the convex hull

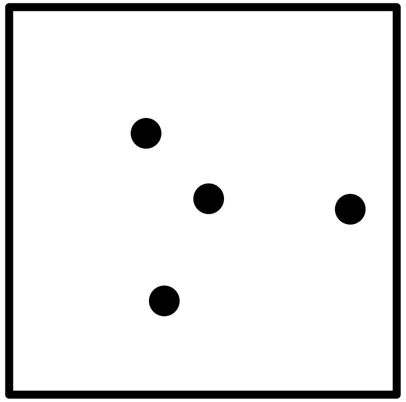
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



S_1

All points on the convex hull

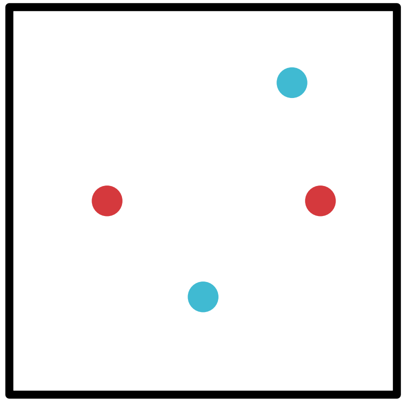


S_2

At least one point inside the convex hull

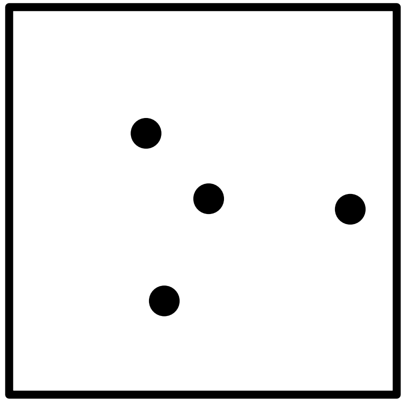
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



S_1

All points on the convex hull

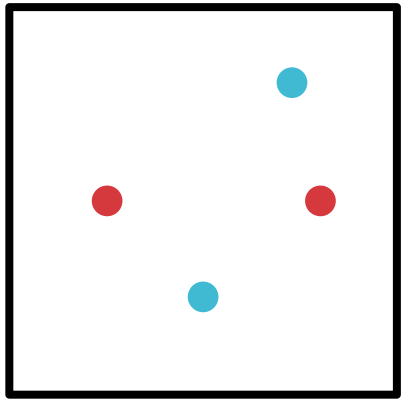


S_2

At least one point inside the convex hull

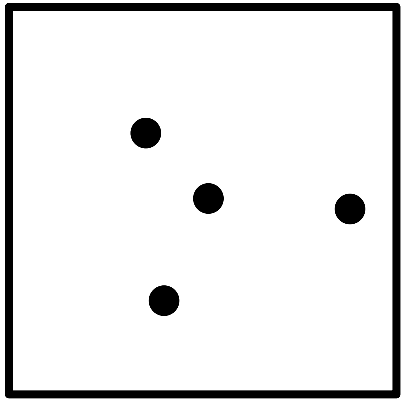
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



$|\mathcal{H}(S_1)| = 14$

All points on the convex hull

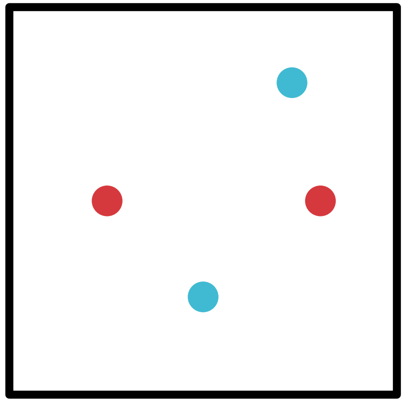


S_2

At least one point inside the convex hull

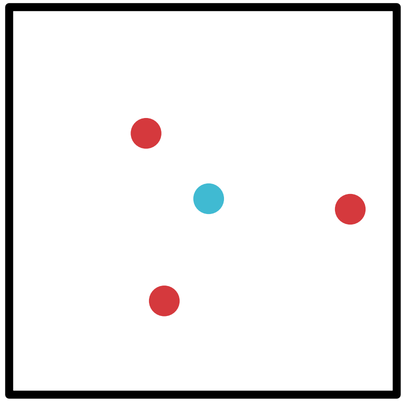
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



$|\mathcal{H}(S_1)| = 14$

All points on the convex hull

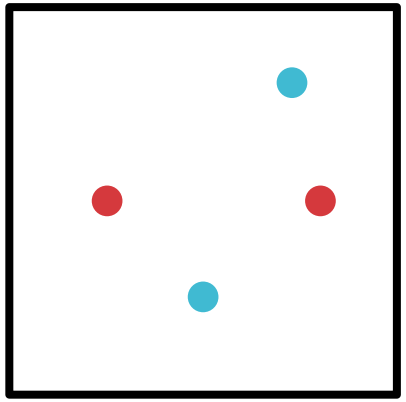


S_2

At least one point inside the convex hull

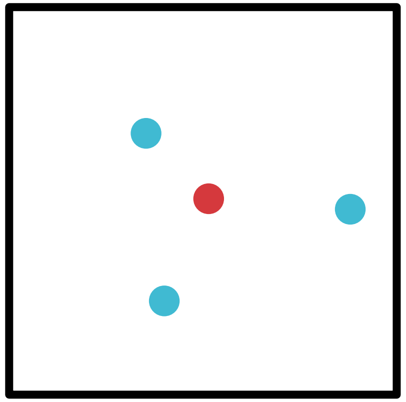
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



$|\mathcal{H}(S_1)| = 14$

All points on the convex hull

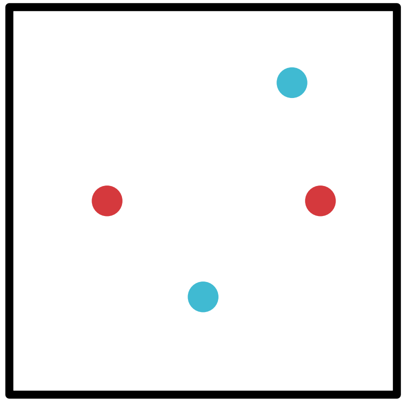


S_2

At least one point inside the convex hull

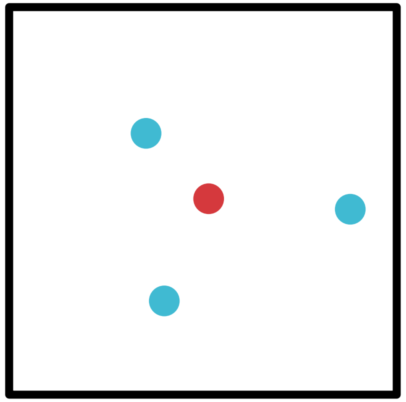
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- What is $VC(\mathcal{H})$?
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



$|\mathcal{H}(S_1)| = 14$

All points on the convex hull

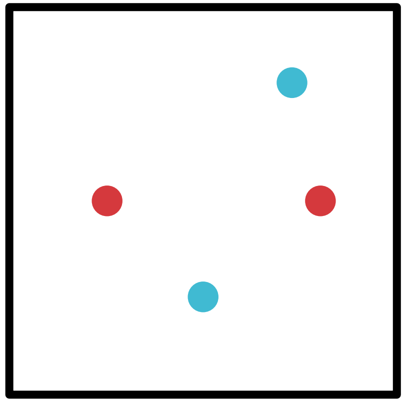


$|\mathcal{H}(S_2)| = 14$

At least one point inside the convex hull

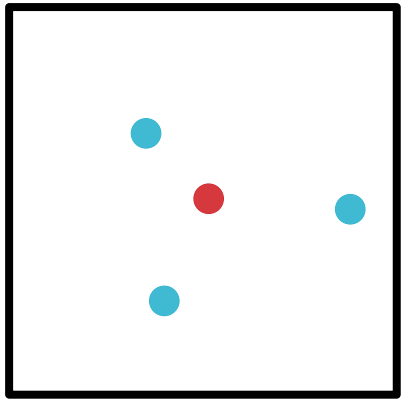
VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{H}$ = all 2-dimensional linear separators
- $VC(\mathcal{H}) = 3$
  - Can $\mathcal{H}$ shatter some set of 1 point?
  - Can $\mathcal{H}$ shatter some set of 2 points?
  - Can $\mathcal{H}$ shatter some set of 3 points?
  - Can $\mathcal{H}$ shatter some set of 4 points?



$|\mathcal{H}(S_1)| = 14$

All points on the convex hull



$|\mathcal{H}(S_2)| = 14$

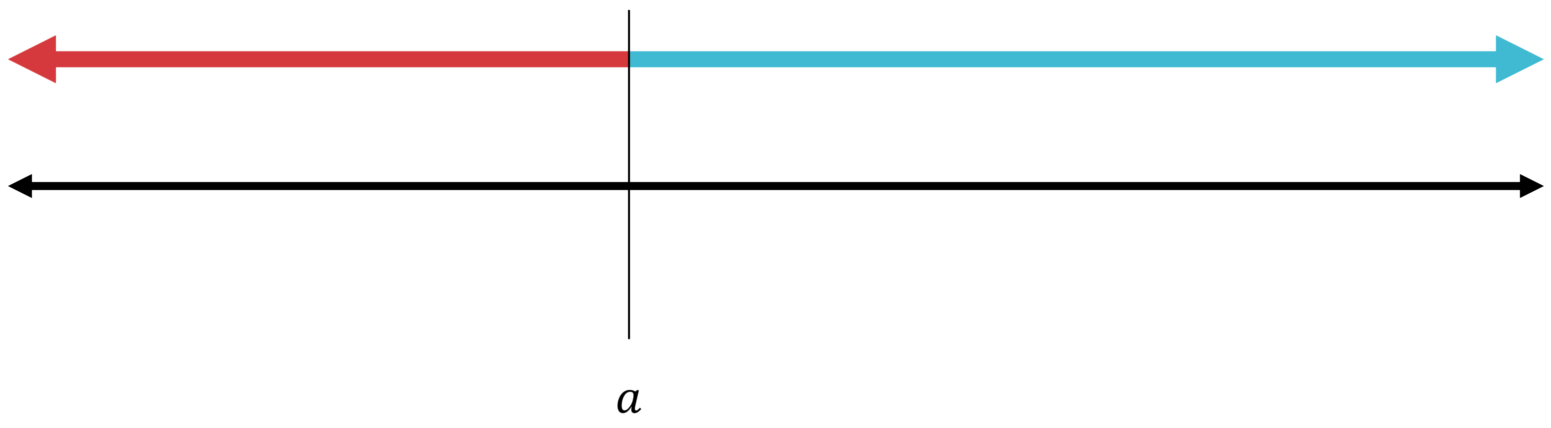
At least one point inside the convex hull

VC-Dimension: Example

- $\boldsymbol{x} \in \mathbb{R}^d$ and $\mathcal{H}$ = all d-dimensional linear separators
- $VC(\mathcal{H}) = d + 1$

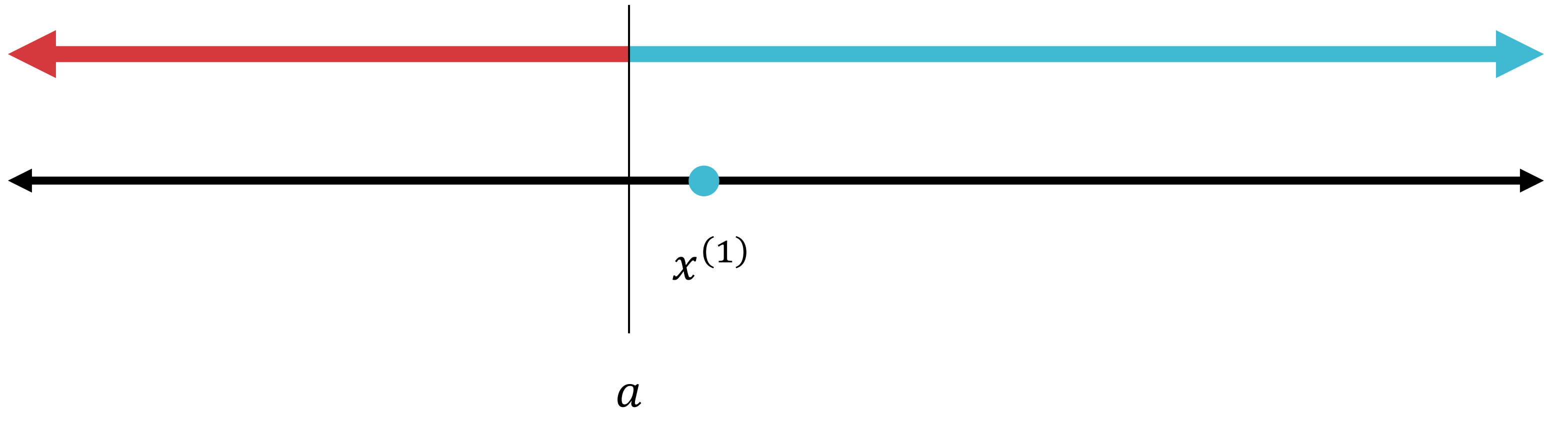
VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



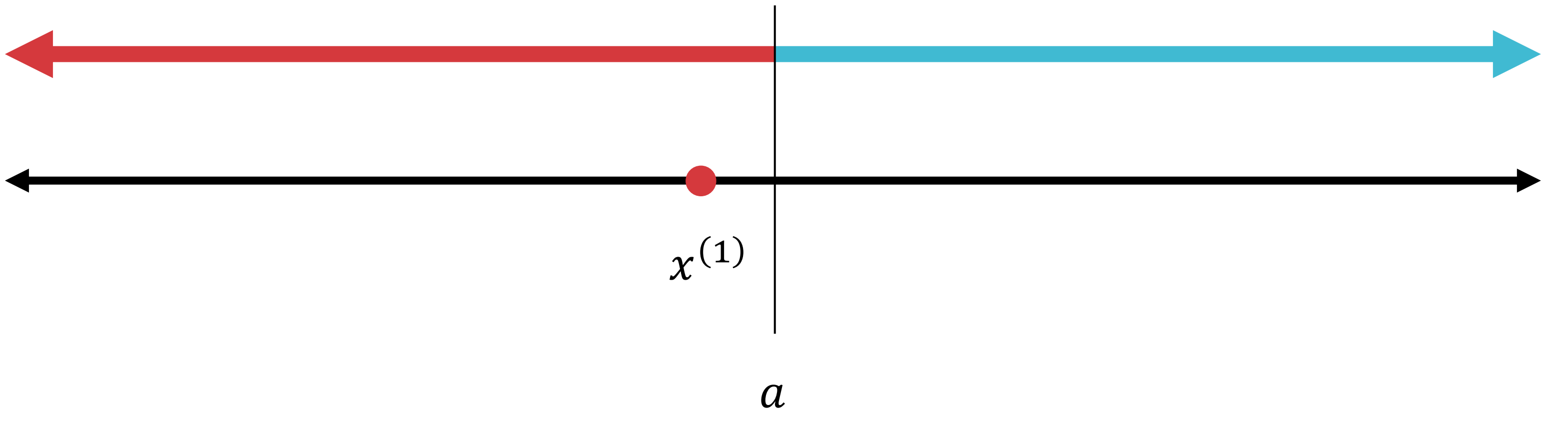
VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



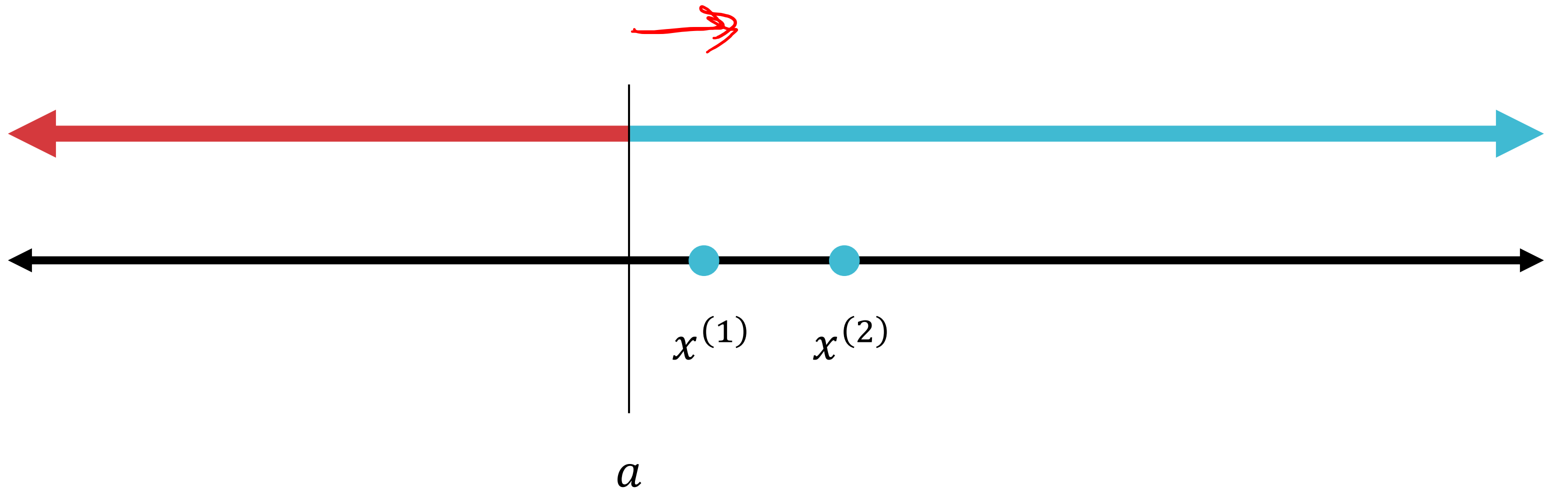
VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



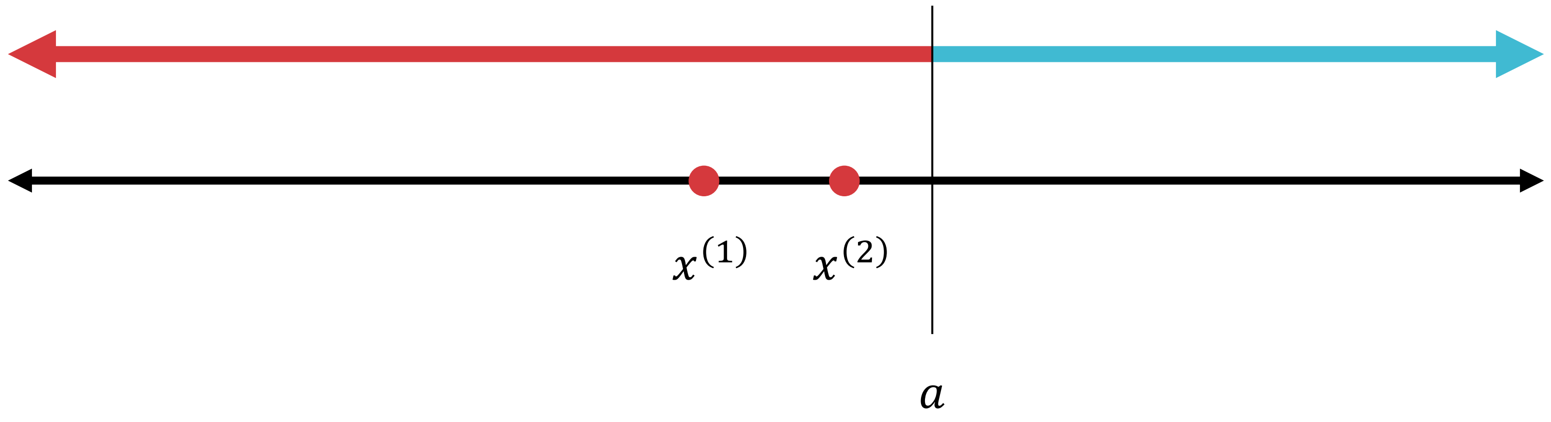
VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



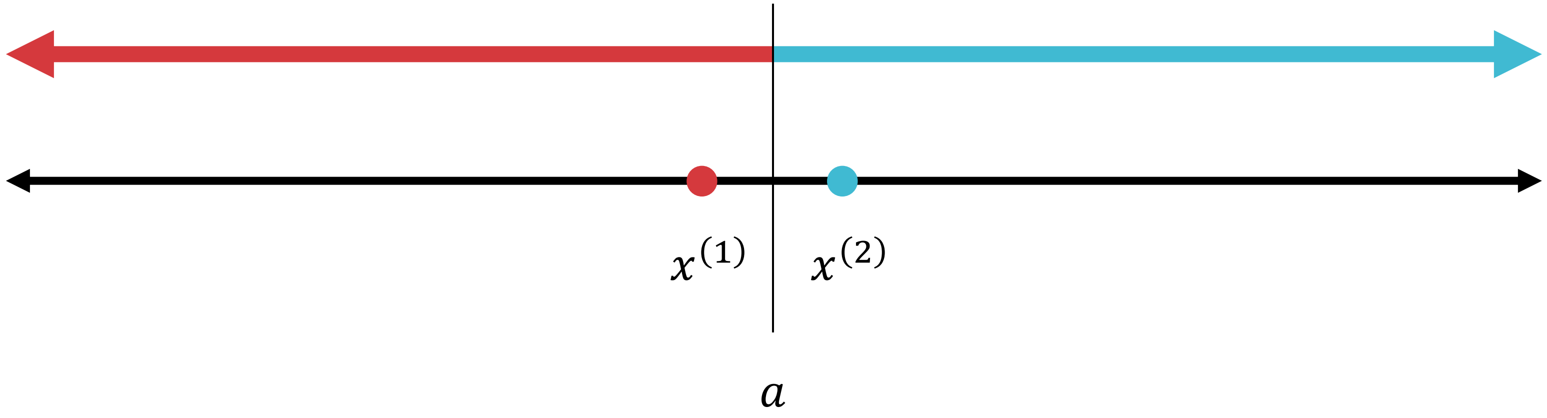
VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



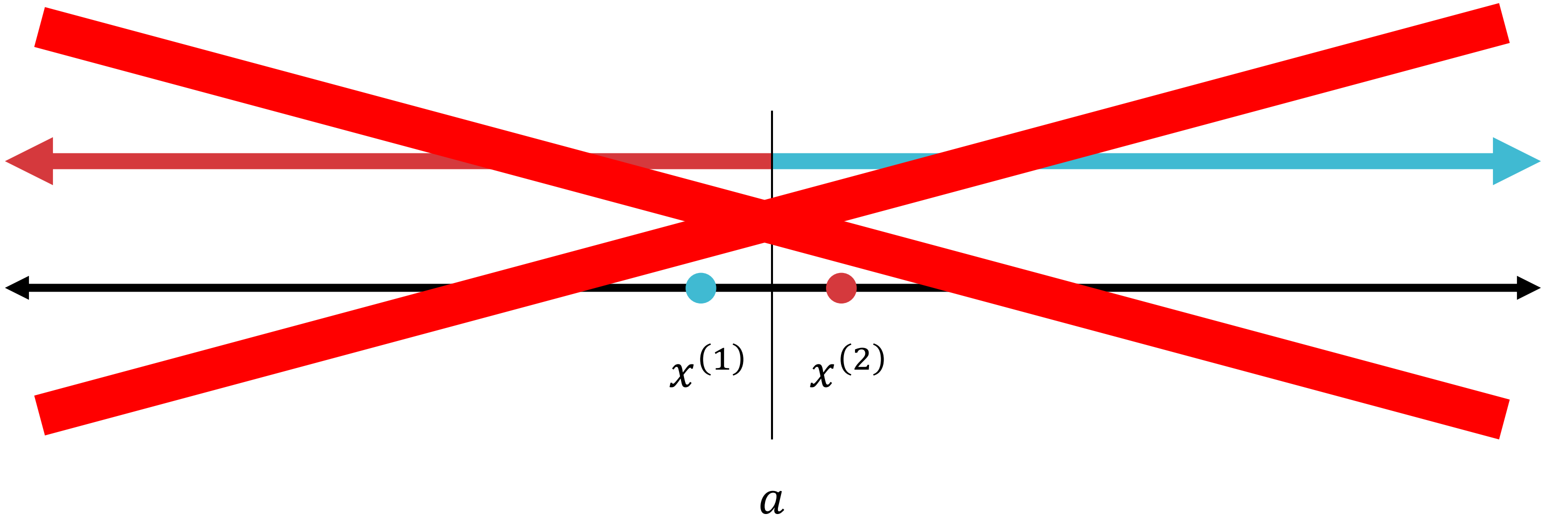
VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



VC-Dimension: Example

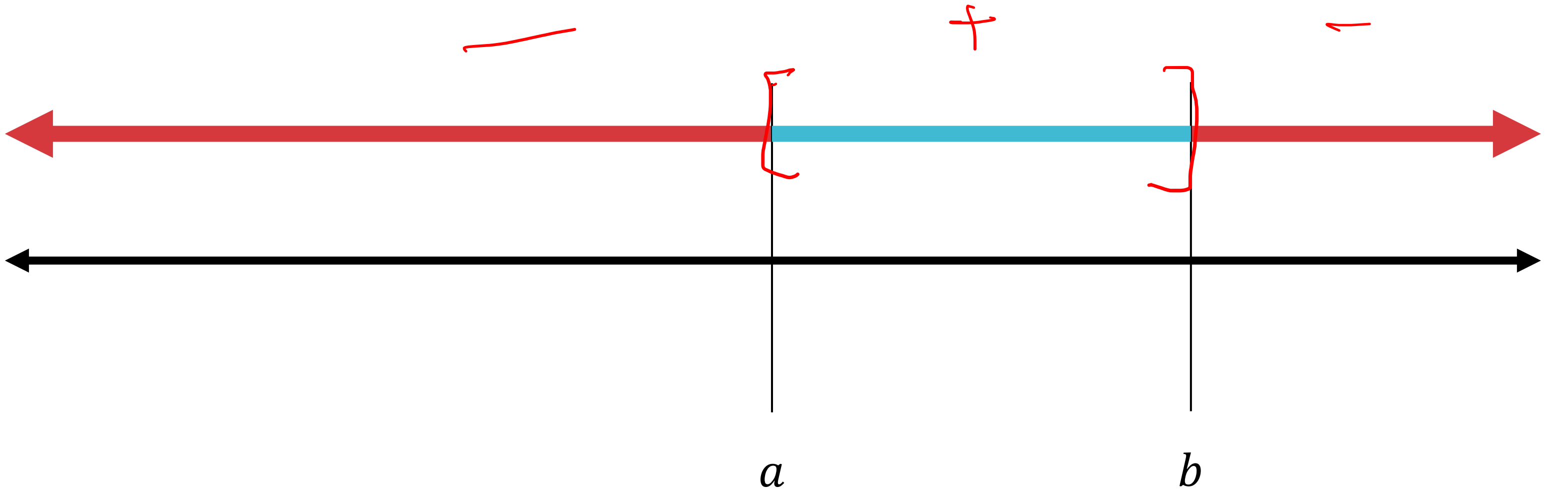
- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive rays, i.e., all hypotheses of the form $h(x; a) = \text{sign}(x - a)$



- $VC(\mathcal{H}) = 1$

VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive intervals

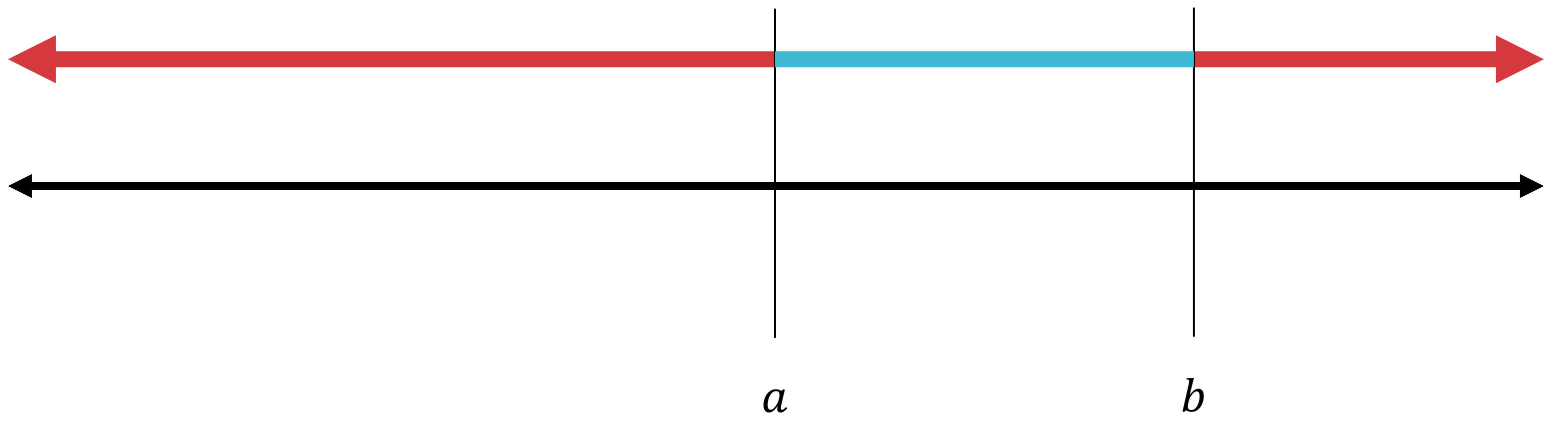


Poll Question 1:

What is $VC(\mathcal{H})$?

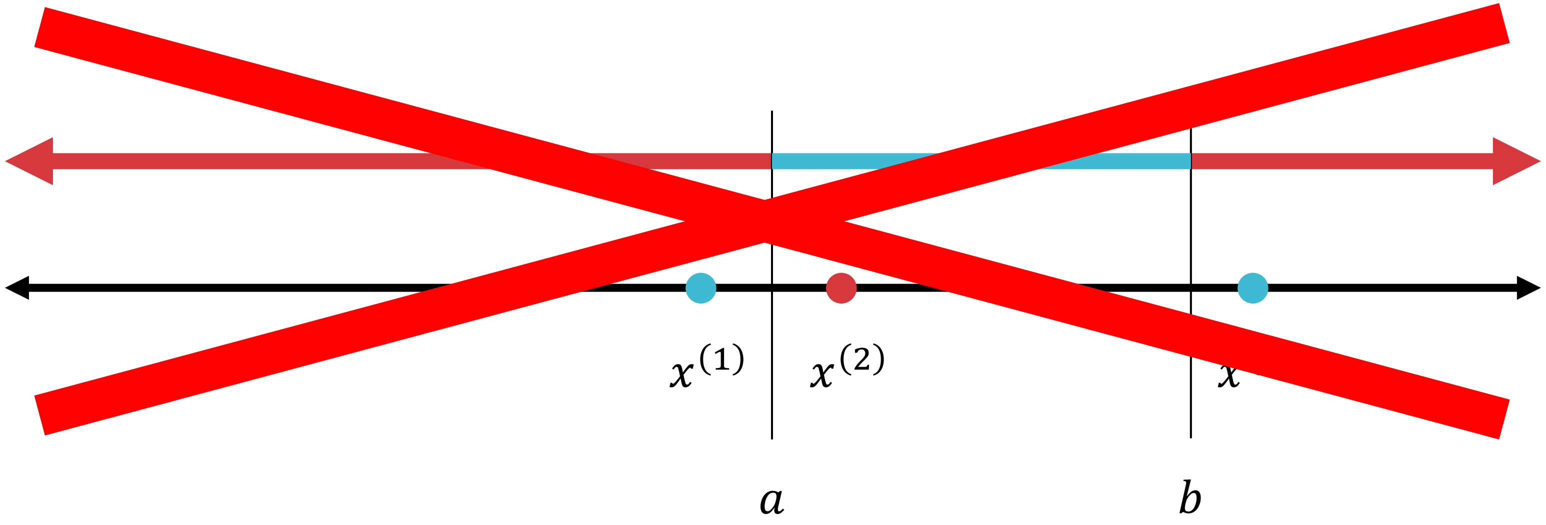
A. 0
B. 1
C. 1.5 **(TOXIC)**
D. 2
E. 3

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive intervals



VC-Dimension: Example

- $x \in \mathbb{R}$ and $\mathcal{H}$ = all 1-dimensional positive intervals



- $VC(\mathcal{H}) = 2$

Theorem 3: Vapnik-Chervonenkis (VC)-Bound

$|\mathcal{H}| = \infty$ but $VC(\mathcal{H}) < \infty$

- Infinite, realizable case: for any hypothesis set $\mathcal{H}$ such that $c^* \in \mathcal{H}$ and arbitrary distribution p^*, if the number of labelled training data points satisfies

$\ln|\mathcal{H}|$

$$M = O\left(\frac{1}{\epsilon}\left(VC(\mathcal{H})\log\left(\frac{1}{\epsilon}\right) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

then with probability at least $1 - \delta$, all $h \in \mathcal{H}$ with $\hat{R}(h) = 0$ have $R(h) \leq \epsilon$

consistent

Statistical Learning Theory Corollary 3

- Infinite, realizable case: for any hypothesis set $\mathcal{H}$ such that $c^* \in \mathcal{H}$ and arbitrary distribution p^*, given a training dataset S where $|S| = N$, all $h \in \mathcal{H}$ with $\hat{R}(h) = 0$ have

$$R(h) \leq O\left(\frac{1}{N}\left(VC(\mathcal{H}) \log\left(\frac{N}{VC(\mathcal{H})}\right) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

with probability at least $1 - \delta$.

Theorem 4: Vapnik-Chervonenkis (VC)-Bound

$|\mathcal{H}| = \infty$ but $VC(\mathcal{H}) < \infty$

- Infinite, agnostic case: for any hypothesis set $\mathcal{H}$ and arbitrary distribution p^*, if the number of labelled training data points satisfies

$$N = O\left(\frac{1}{\epsilon^2}\left(VC(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

$1/\epsilon$ $\ln(|\mathcal{H}|)$

then with probability at least $1 - \delta$, all $h \in \mathcal{H}$ have $|R(h) - \hat{R}(h)| \leq \epsilon$

Statistical Learning Theory Corollary 4

- Infinite, agnostic case: for any hypothesis set $\mathcal{H}$ and arbitrary distribution p^*, given a training dataset S where $|S| = N$, all $h \in \mathcal{H}$ have

$$R(h) \leq \hat{R}(h) + O\left(\sqrt{\frac{1}{N}\left(VC(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right)}\right)$$

with probability at least $1 - \delta$.

Approximation Generalization Tradeoff

How well does h generalize?

true error

empirical

$$R(h) \leq \hat{R}(h) + O\left(\sqrt{\frac{1}{N}\left(VC(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right)}\right)$$

How well does h approximate c^*?

Approximation Generalization Tradeoff

Increases as
$VC(\mathcal{H})$ increases

$$R(h) \leq \hat{R}(h) + O\left(\sqrt{\frac{1}{N}\left(VC(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right)}\right)$$

Decreases as
$VC(\mathcal{H})$ increases

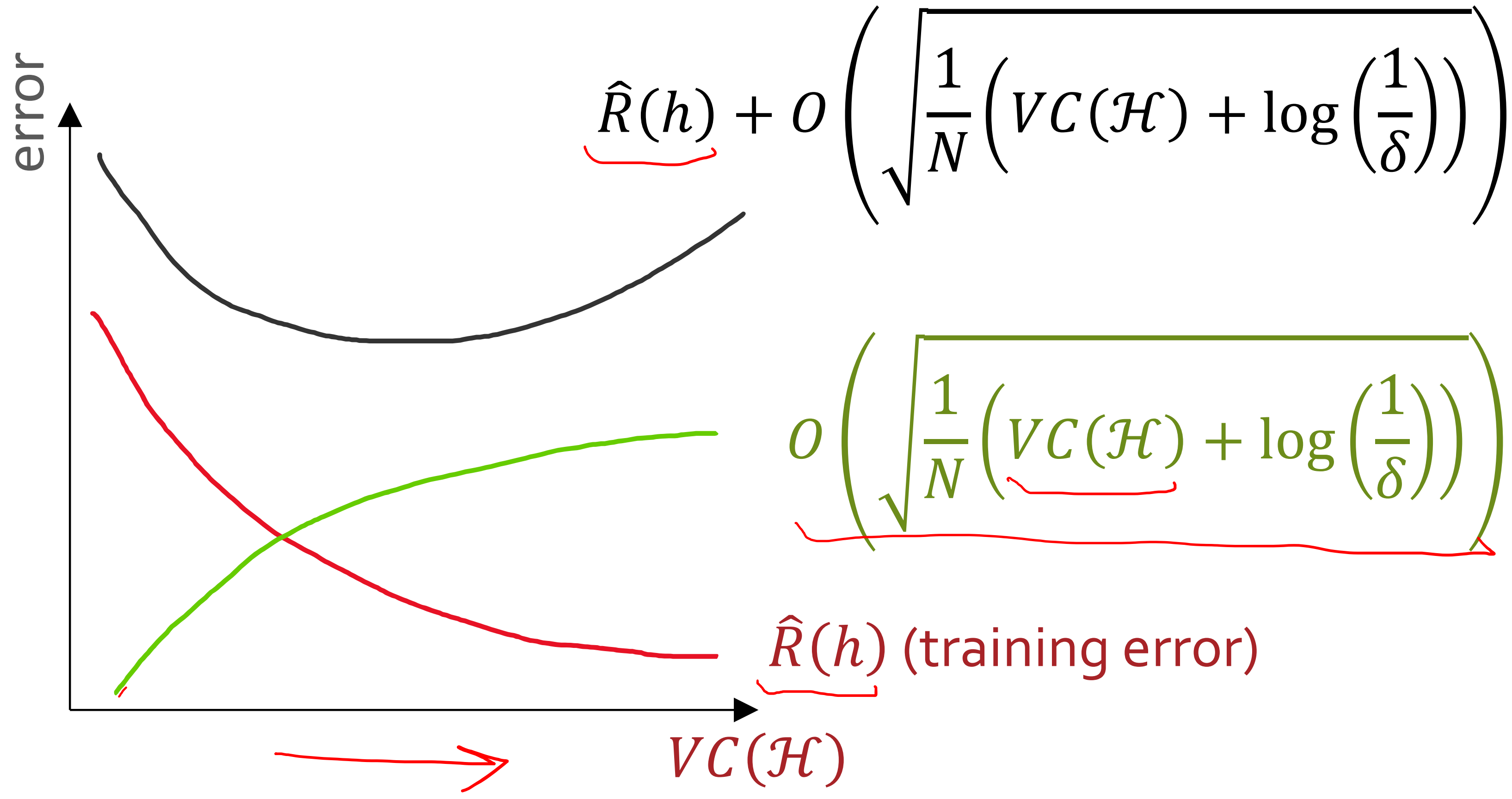
Can we use this corollary to guide model selection?

- Infinite, agnostic case: for any hypothesis set $\mathcal{H}$ and arbitrary distribution p^*, given a training dataset S where $|S| = N$, all $h \in \mathcal{H}$ have

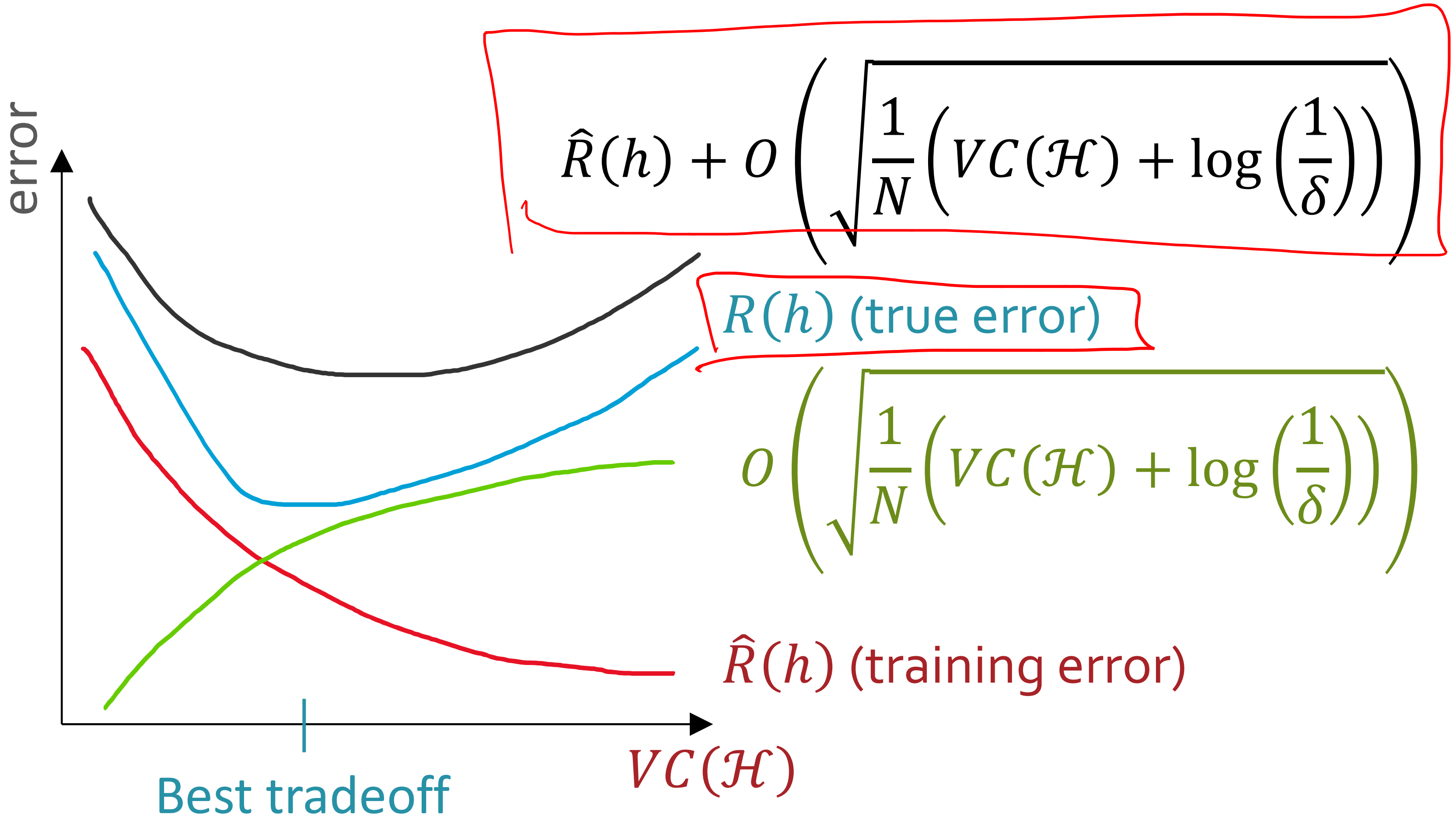
$$R(h) \leq \hat{R}(h) + O\left(\sqrt{\frac{1}{N}\left(VC(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)\right)}\right)$$

with probability at least $1 - \delta$.

Learning Theory and Model Selection



Learning Theory and Model Selection



- How can we find this "best tradeoff" for linear separators?
- Use a regularizer! By (effectively) reducing the number of features our model considers, we reduce its VC-dimension.

Learning Theory Learning Objectives

You should be able to…

- Identify the properties of a learning setting and assumptions required to ensure low generalization error
- Distinguish true error, train error, test error
- Define PAC and explain what it means to be approximately correct and what occurs with high probability
- Apply sample complexity bounds to real-world machine learning examples
- Theoretically motivate regularization

10-301/601: Introduction to Machine Learning Lecture 15 – Societal Impacts of ML

Henry Chai & Matt Gormley & Hoda Heidari

10/23/23

ML in Societal Applications

Deep learning is being used to predict critical COVID-19 cases

8 WAYS MACHINE LEARNING WILL IMPROVE EDUCATION

BY MATTHEW LYNCH / JUNE 12, 2018 / 5

SCI-FI VISIONS

Your Future Doctor May Not be Human. This Is the Rise of AI in Medicine.

From mental health apps to robot surgeons, artificial intelligence is already changing the practice of medicine.

techworld FROM IDG Features Technology Innovation Partner Zone the techies

Home > Features > Emerging tech & innovation Features

Researcher explains how algorithms can create a fairer legal system

TheUpshot

ROBO RECRUITING

Can an Algorithm Hire Better Than a Human?

By Claire Cain Miller

20 JAN 2017 | Insight

Kevin Petrasic | Benjamin Saul

Algorithms and bias: What lenders need to know

The algorithms that power fintech may discriminat… can be difficult to anticipate—and financial institut… accountable even when alleged discrimination is c… unintentional.

HOME > STRATEGY

Artificial intelligence is slated to disrupt 4.5 million jobs for African Americans, who have a 10% greater likelihood of automation-based job loss than other workers

Allana Akhtar Oct 7, 2019, 12:57 PM

Misinformation on coronavirus is proving highly contagious

By DAVID KLEPPER July 29, 2020

ACLU

BECOME A MEMBER / RENEW / TAKE ACT…

ISSUES KNOW YOUR RIGHTS DEFENDING OUR RIGHTS BLOGS ABOUT

SPEAK FREELY

How Facebook Is Giving Sex Discrimination in Employment Ads a New Life

By Galen Sherwin, ACLU Women's Rights Project

SEPTEMBER 18, 2018 | 10:00 AM

The Switch

Wanted: The 'perfect babysitter.' Must pass AI scan for respect and attitude.

The New York Times

I.R.S. Changes Audit Practice That Discriminated Against Black Taxpayers

The agency will overhaul how it scrutinizes returns that claim the earned-income tax credit, which is aimed at alleviating poverty.

MEDICAL MALAISE

If you're not a white male, artificial intelligence's use in healthcare could be dangerous

By Robert David Hart · July 10, 2017

ProPublica

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016