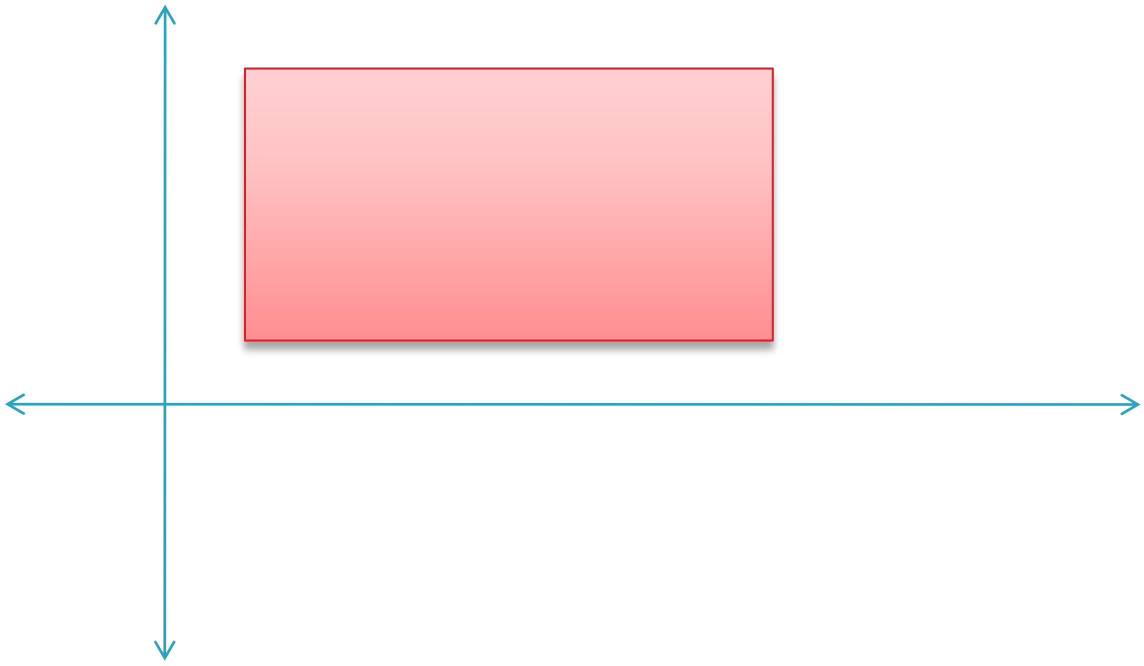
PAC Learning and The VC Dimension

Rectangle Game

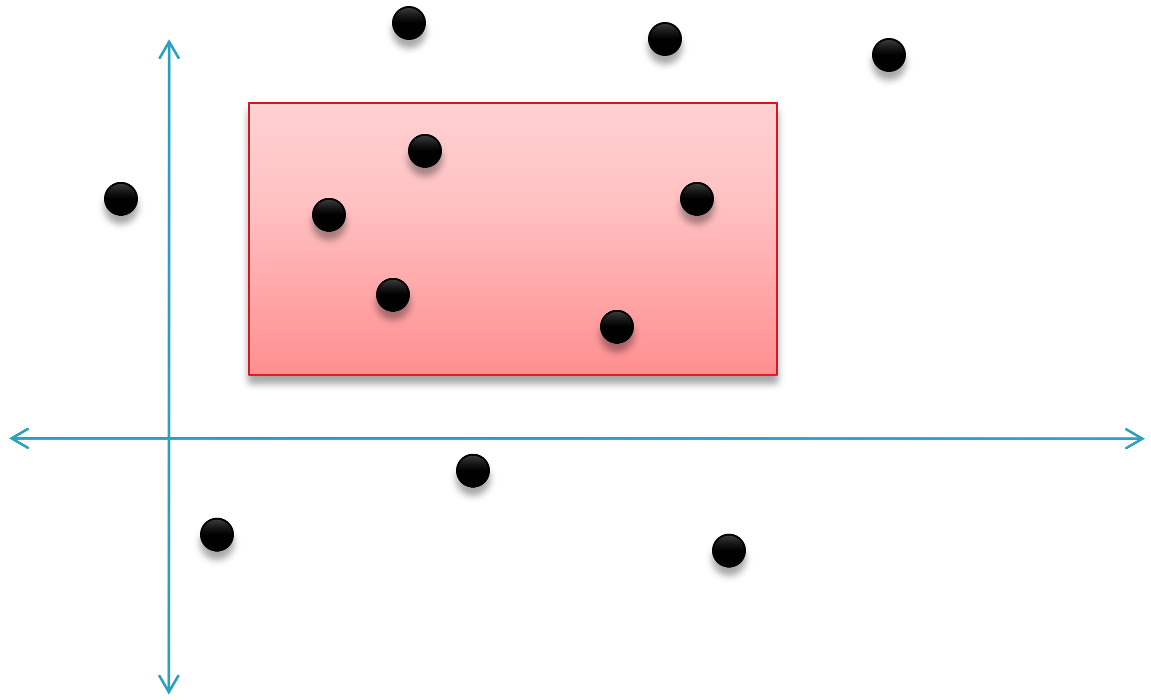
- Fix a rectangle (unknown to you):



From **An Introduction to Computational Learning Theory** by Keanrs and Vazirani

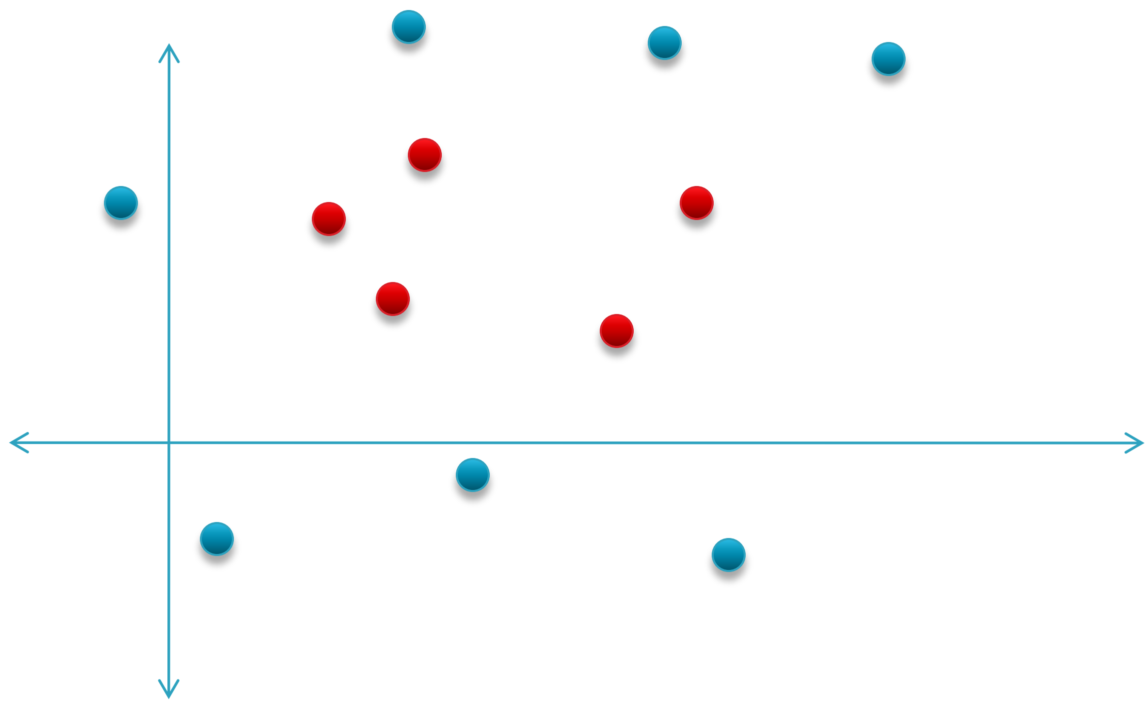
Rectangle Game

- Draw points from some fixed unknown distribution:



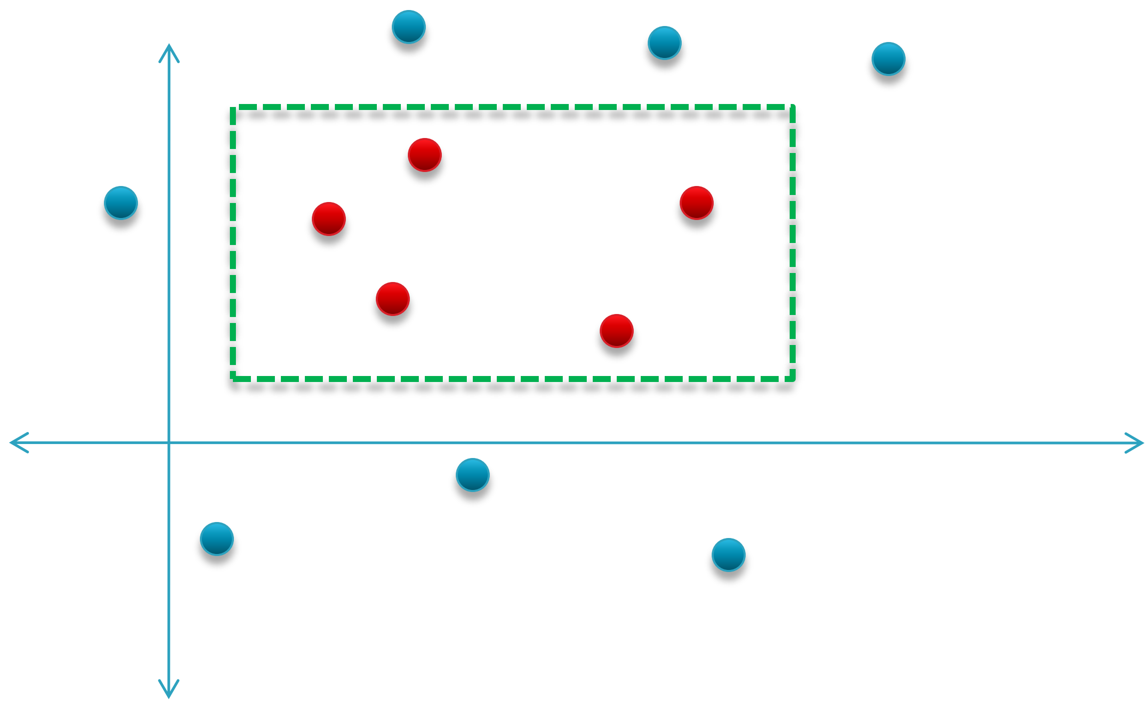
Rectangle Game

- You are told the points and whether they are in or out:



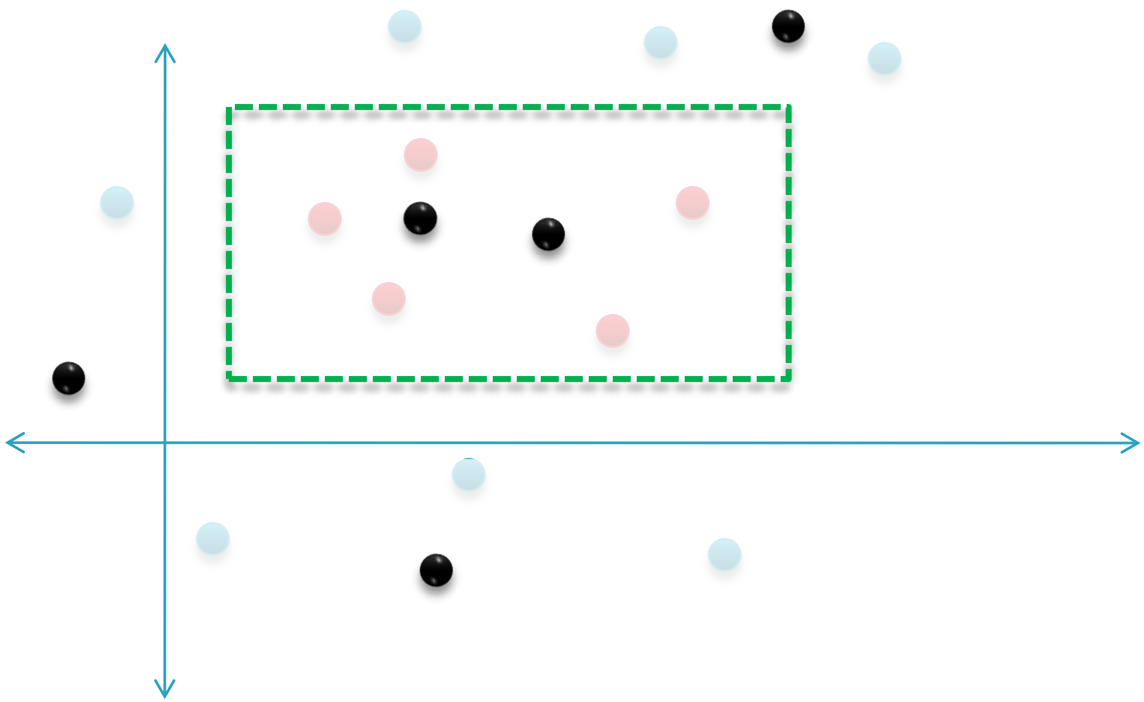
Rectangle Game

- You propose a hypothesis:



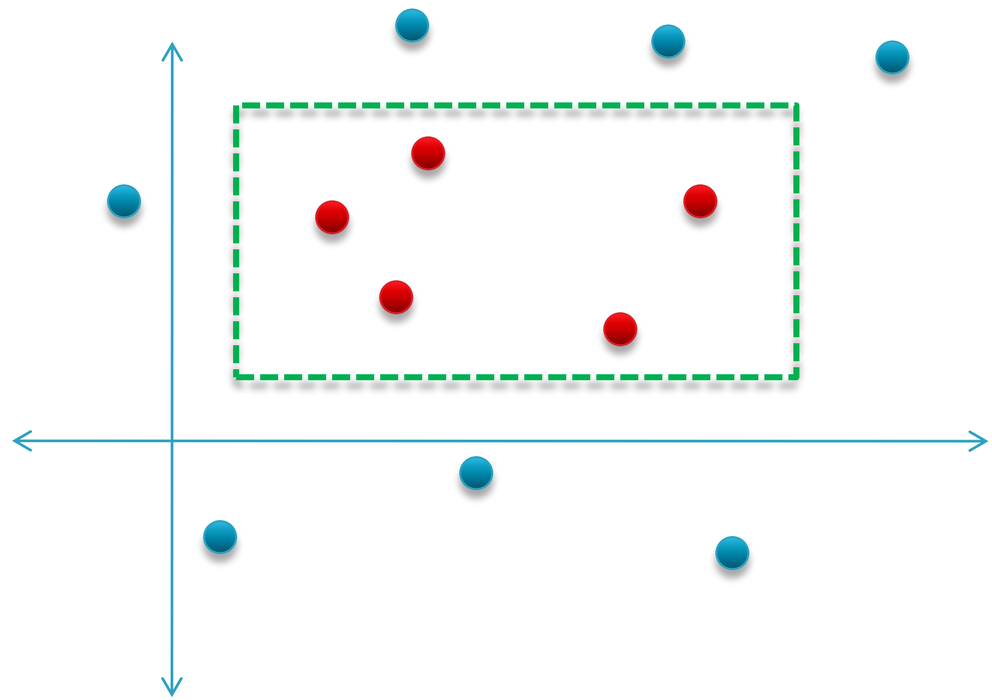
Rectangle Game

- Your hypothesis is tested on points drawn from the same distribution:



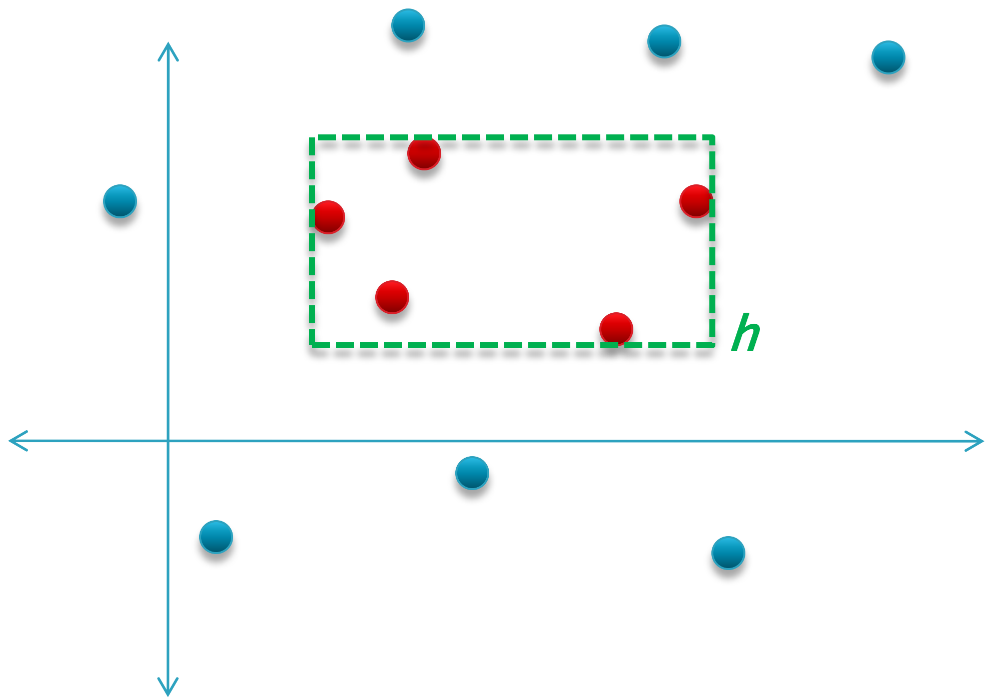
Goal

- We want an algorithm that:
  - With high probability will choose a hypothesis that is approximately correct.



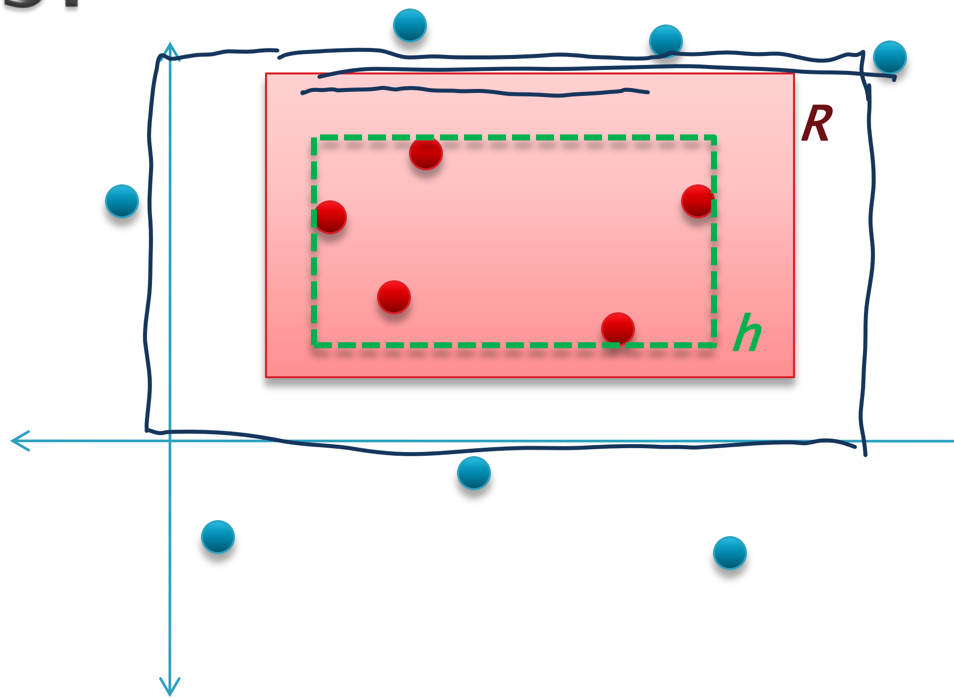
Minimum Rectangle Learner:

- Choose the minimum area rectangle containing all the positive points:



How Good is this?

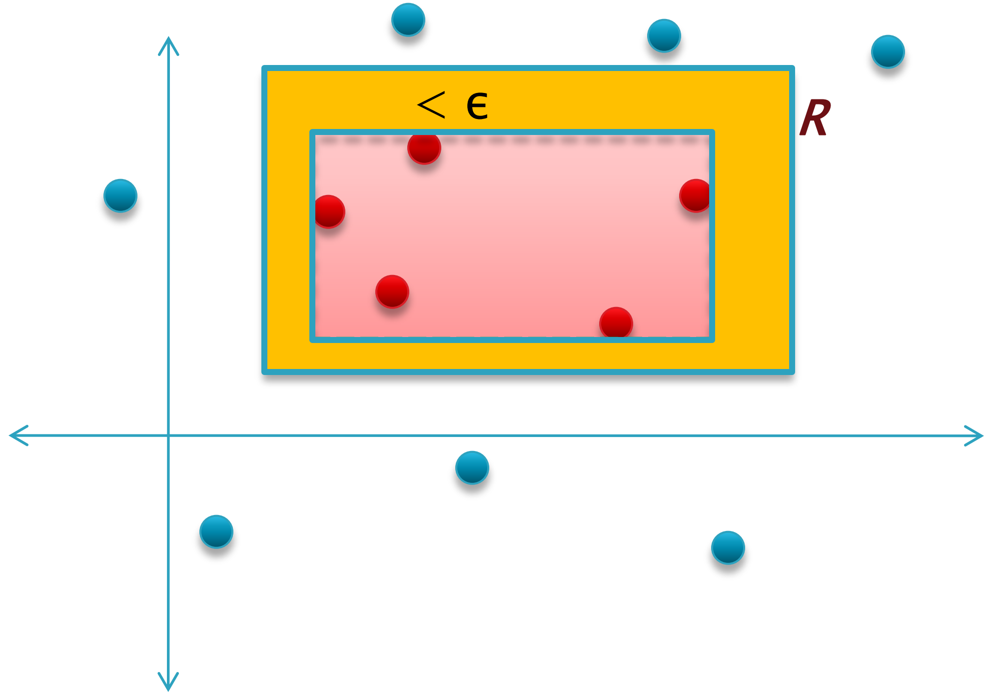
- Derive a PAC bound:
- For fixed:
  - R : Rectangle
  - D : Data Distribution
  - ϵ : Test Error
  - δ : Probability of failing
  - m : Number of Samples



$$\mathbf{P}\left(\text{error}_{\text{test}}(h) \leq \epsilon\right) \leq 1 - \delta$$

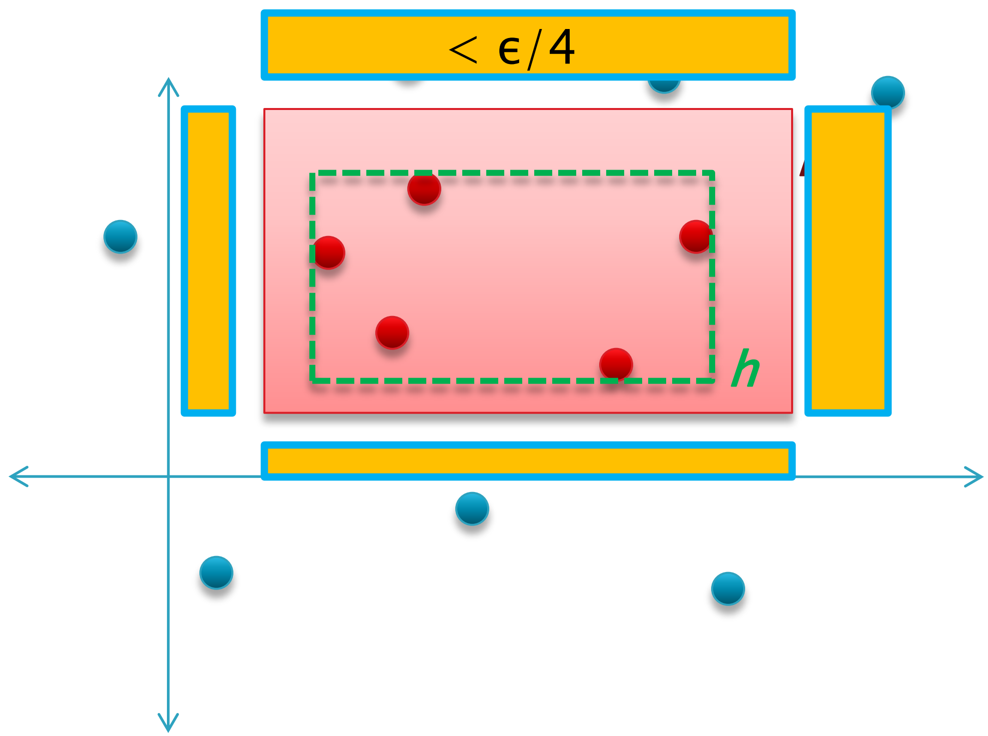
Proof:

- We want to show that with high probability the area below measured with respect to *D* is bounded by ϵ :



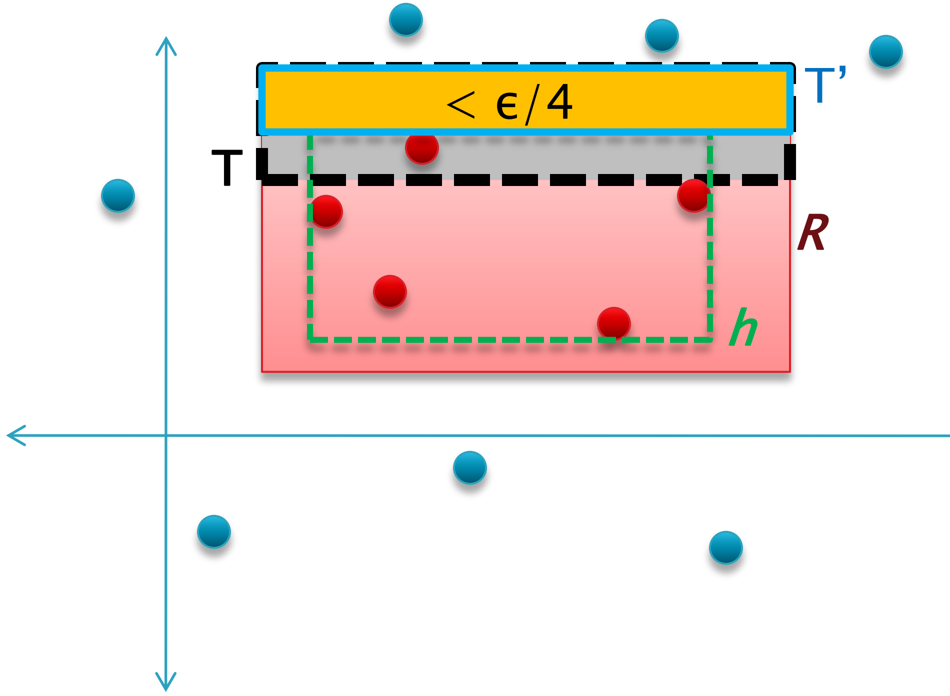
Proof:

- We want to show that with high probability the area below measured with respect to *D* is bounded by ϵ :



Proof:

- Define T to be the region that contains exactly $\epsilon/4$ of the mass in D sweeping down from the top of R.
- $p(T') > \epsilon/4 = p(T)$ IFF T' contains T
- T' contains T IFF none of our *m* samples are from T
- What is the probability that all samples miss T

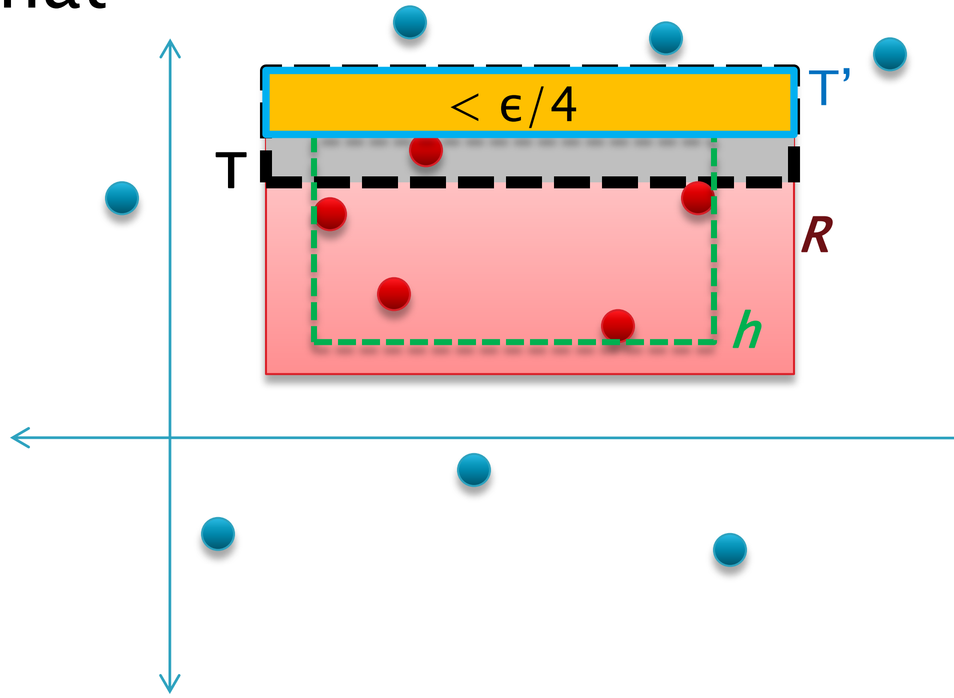


Proof:

- What is the probability that all *m* samples miss T:

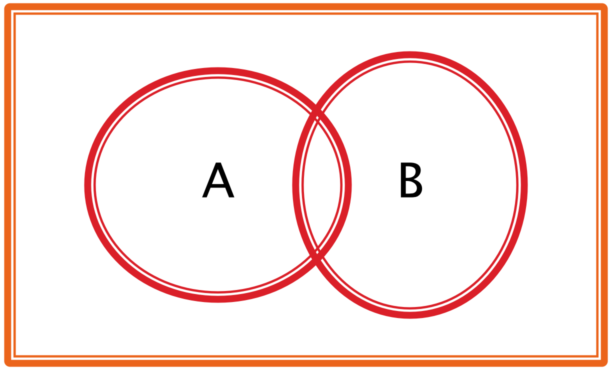
$$\mathbf{P}\left(m \text{ samples miss } T\right) = \left(1 - \frac{\epsilon}{4}\right)^m$$

- What is the probability that we miss any of the rectangles?
  - Union Bound



Union Bound

$$\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$$

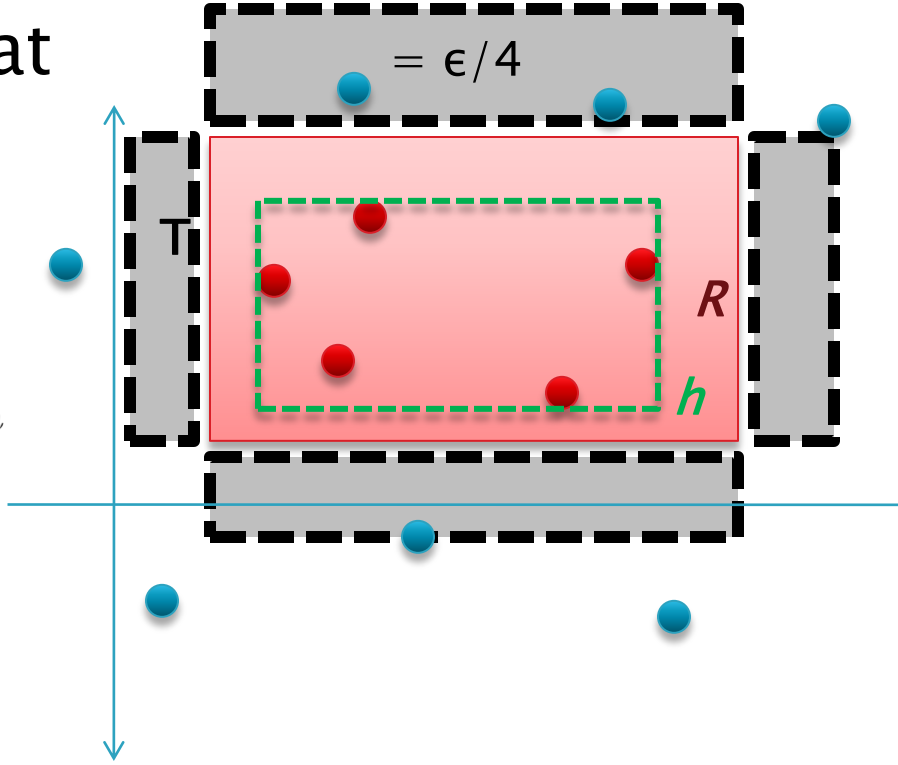


Proof:

- What is the probability that all *m* samples miss T:

$$\mathbf{P}\left(m \text{ samples miss } T\right) = \left(1 - \frac{\epsilon}{4}\right)^m$$

- What is the probability that we miss any of the rectangles:
  - Union Bound



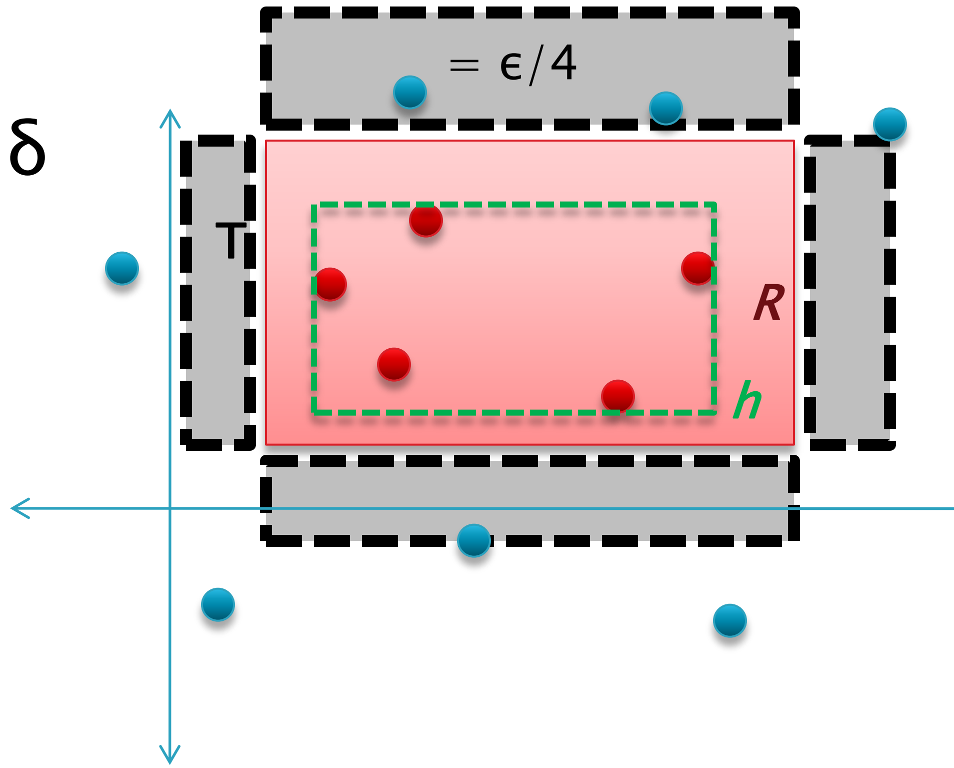
$$\mathbf{P}\left(m \text{ samples miss any } T\right) \leq 4\left(1 - \frac{\epsilon}{4}\right)^m$$

Proof:

- Probability that any region has weight greater than ϵ/4 after m samples is at most: $4\left(1-\frac{\epsilon}{4}\right)^m$
- If we fix *m* such that:

$$4\left(1-\frac{\epsilon}{4}\right)^m \leq \delta$$

- Than with probability 1– δ we achieve an error rate of at most ϵ



Extra Inequality

- Common Inequality:

$$1 - x \leq e^{-x}$$

- We can show:

$$4\left(1 - \frac{\epsilon}{4}\right)^m \quad \leq \quad 4e^{-m\epsilon/4}$$

- Obtain a lower bound on the samples:

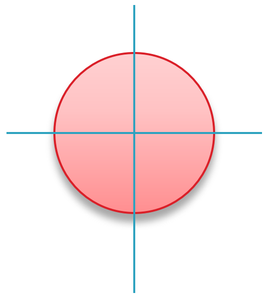
$$m \quad \geq \quad \frac{4}{\epsilon}\ln\left(\frac{4}{\delta}\right)$$

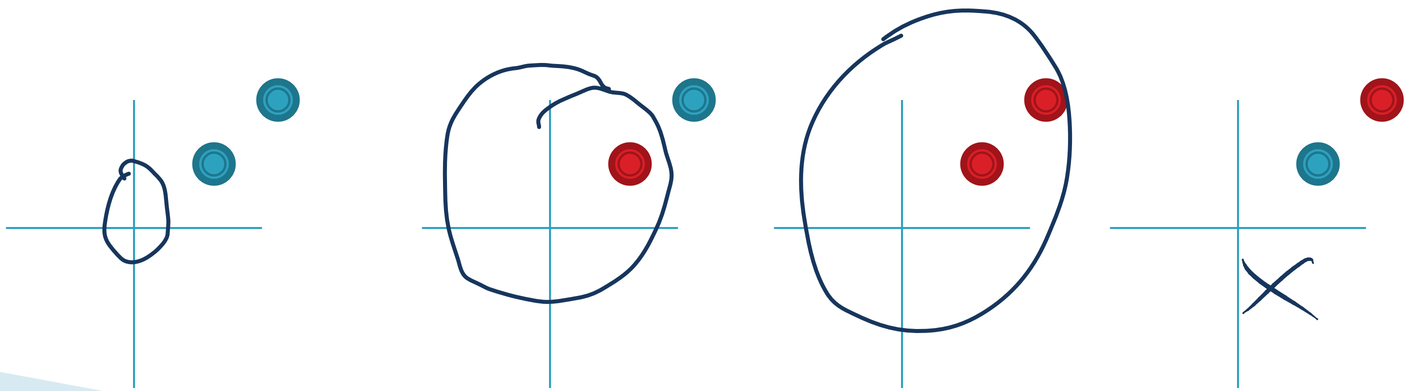
VC – Dimension

- Provides a measure of the complexity of a "hypothesis space" or the "power" of "learning machine"
- Higher VC dimension implies the ability to represent more complex functions
- The VC dimension is the maximum number of points that can be arranged so that **f** shatters them.
- What does it mean to shatter?

Define: Shattering

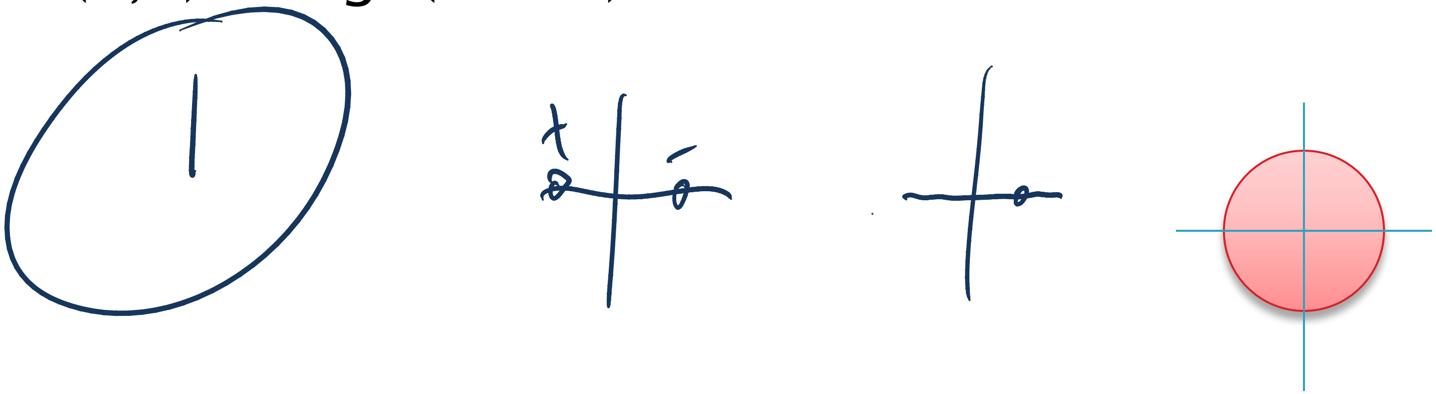
- A classifier **f** can shatter a set of points if and only if for all truth assignments to those points **f** gets zero training error
- Example: f(x,b) = sign(x.x−b)

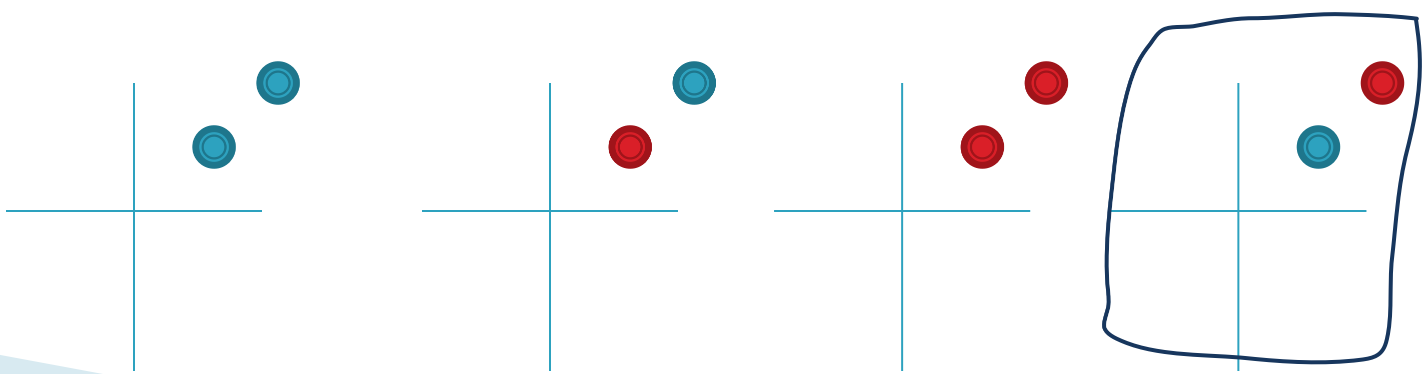




Example Continued:

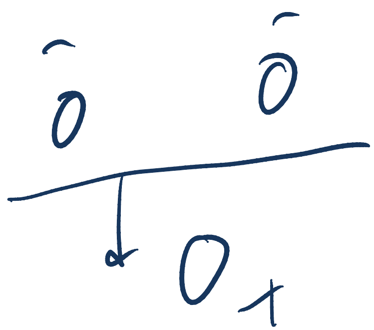
- What is the VC Dimension of the classifier:
  - f(x,b) = sign(x.x−b)

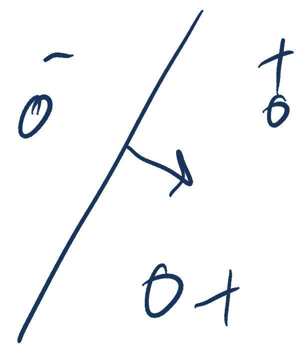


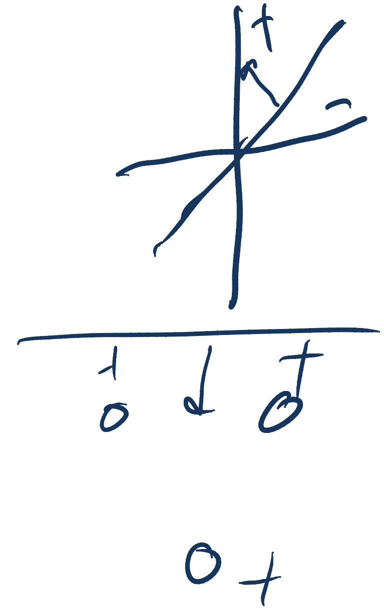


VC Dimension of 2D Half-Space:

- Conjecture: 3
- Easy Proof (lower Bound):

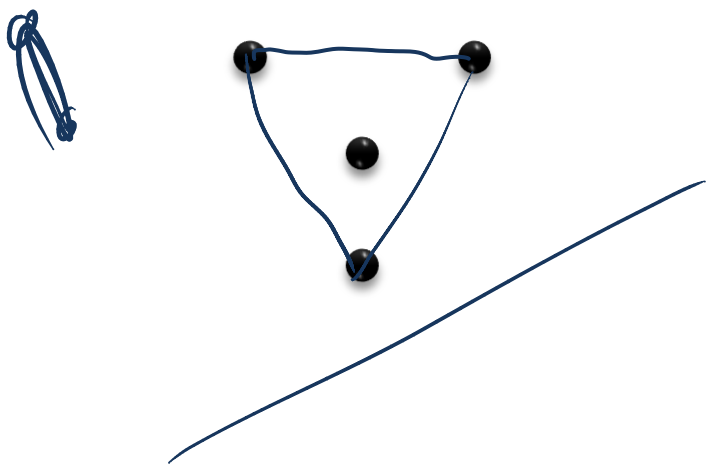


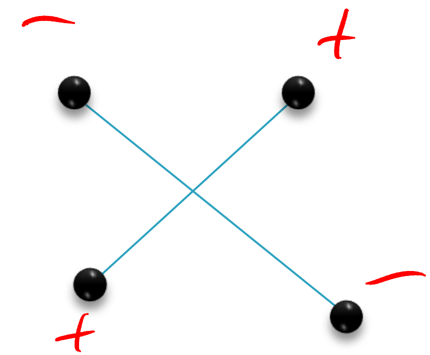




VC Dimension of 2D Half-Space:

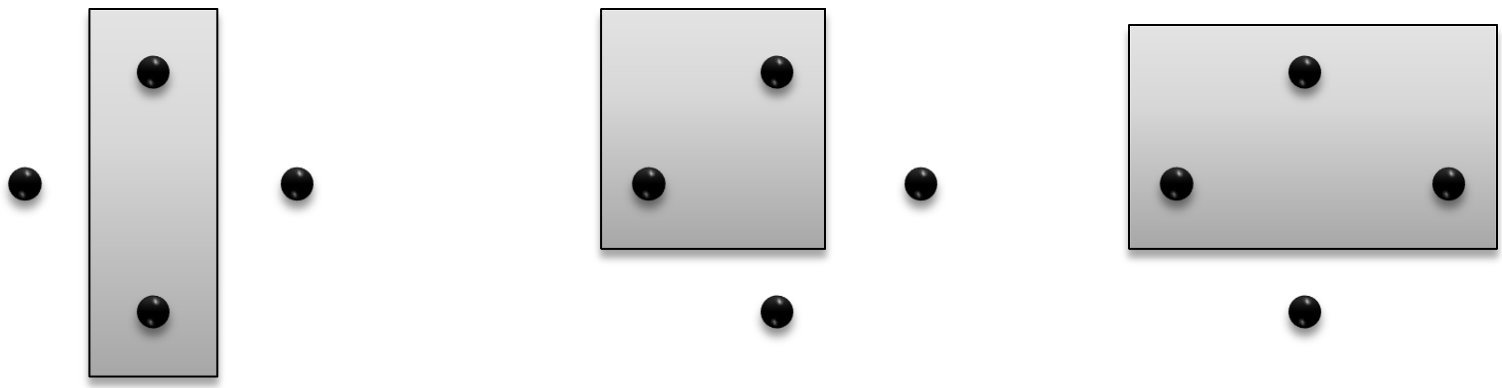
- Harder Proof (Upper Bound):

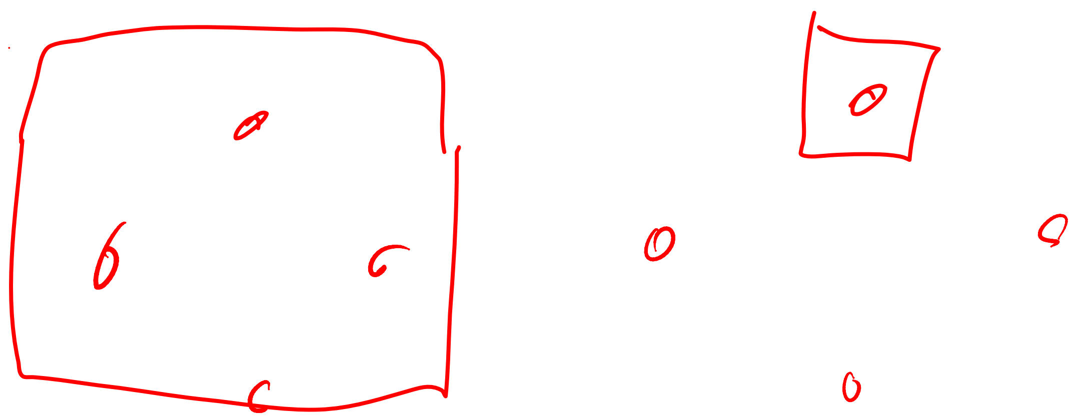




VC-Dim: Axis Aligned Rectangles

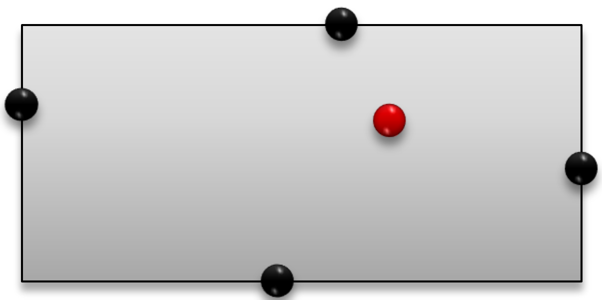
- VC Dimension Conjecture: 4

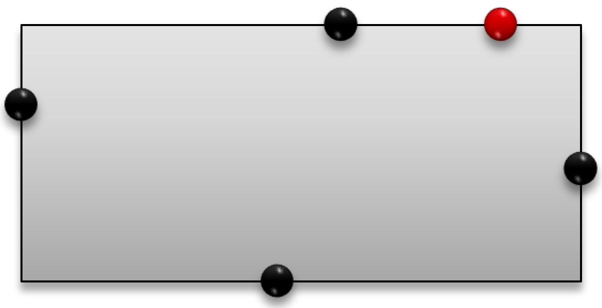


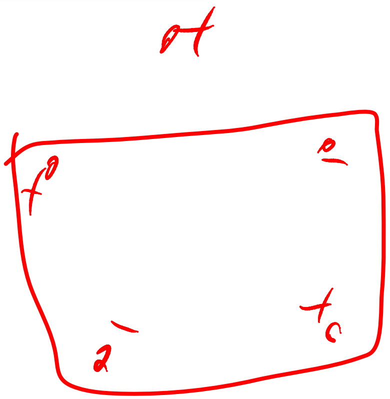


VC-Dim: Axis Aligned Rectangles

- VC Dimension Conjecture: 4
- Upper bound (more Difficult):

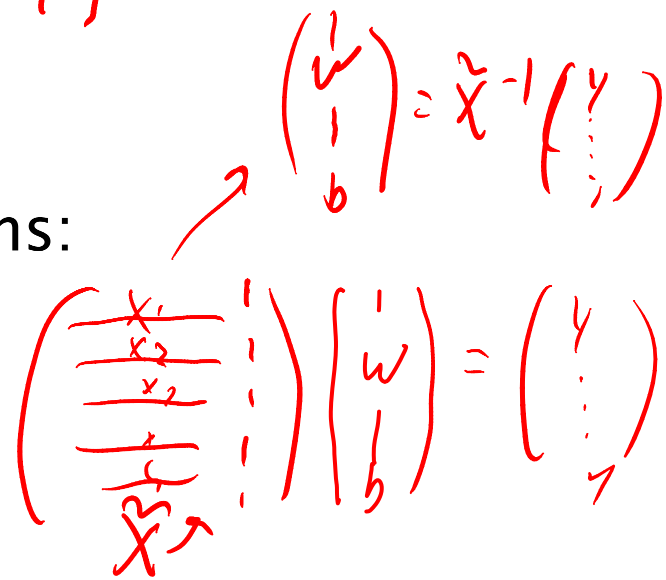






General Half-Spaces in (d – dim)

- What is the VC Dimension of:
  - f(x,{w,b})=sign(w . x + b)
  - X in R^d
- Proof (lower bound):
  - Pick {x_1, …, x_n} (point) locations:
  - Adversary gives assignments {y_1, …, y_n} and you choose {w_1, …, w_n} and b:



Extra Space:

General Half-Spaces

- Proof (upper bound): VC-Dim = d+1
  - Observe that the last d+1 points can always be expressed as:

$$x_{d+2} = \sum_{i=1}^{d+1} \alpha_i x_i$$

$$1 = \sum_{i=1}^{d+1} \alpha_i$$

$$\tilde{x}_{d+2} = \begin{pmatrix} x_{d+2} \\ 1 \end{pmatrix} = \sum \alpha_i \begin{pmatrix} x_i \\ 1 \end{pmatrix} = \begin{pmatrix} \sum \alpha_i x_i \\ \sum \alpha_i \end{pmatrix}$$

- Proof (upper bound): VC-Dim = d+1
  - Observe that the last d+1 points can always be expressed as:

$$x_{d+2} = \sum_{i=1}^{d+1} \alpha_i x_i$$

$$1 = \sum_{i=1}^{d+1} \alpha_i$$

$\hat{w} = \begin{pmatrix} w \\ b \end{pmatrix}$, $\hat{x}_{d+2} = \begin{pmatrix} x_{d+2} \\ 1 \end{pmatrix}$

$y_{d+2} = \text{sign}(\hat{w} \cdot \hat{x}_{d+2})$

$= \text{sign}(\hat{w} \cdot \sum \alpha \hat{x}_i)$

$= \text{sign}(\sum \alpha \hat{w} \cdot \hat{x}_i)$

$y_i = \begin{cases} \alpha < 0 \Rightarrow -1 \\ \text{otherwise} \Rightarrow 1 \end{cases}$

$\sum \alpha w \cdot \hat{x}_i \geq 0$

$y_{d+2} = -1$ ⟷ Contradiction ⟷ $\Rightarrow y_{d+2} = +1$

Extra Space