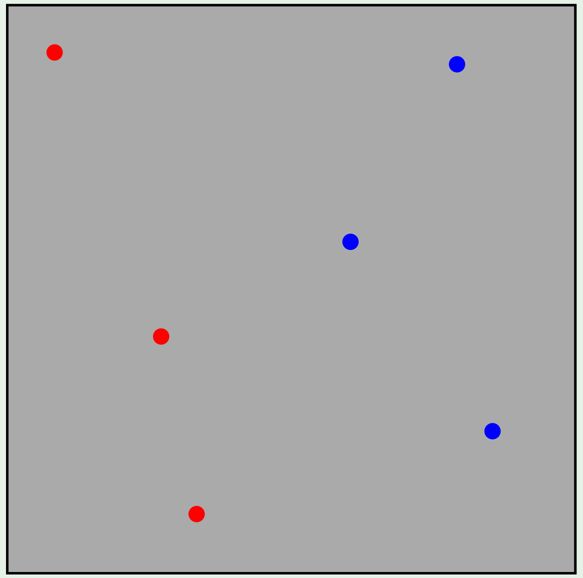
Review of Lecture 5

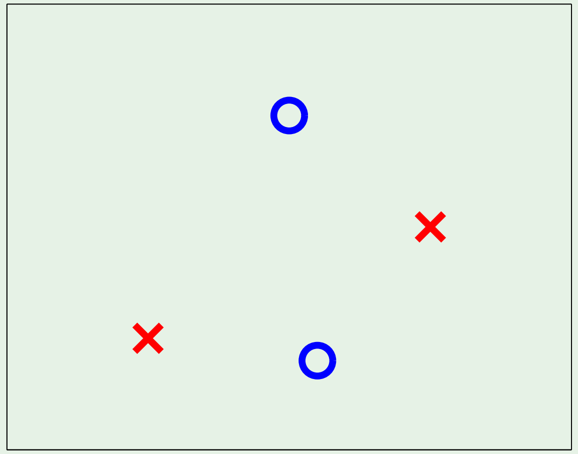
- Dichotomies



- Growth function

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathcal{X}} |\mathcal{H}(\mathbf{x}_1, \cdots, \mathbf{x}_N)|$$

- Break point



- Maximum # of dichotomies

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

Learning From Data

Yaser S. Abu-Mostafa
California Institute of Technology

Lecture 6: Theory of Generalization

Sponsored by Caltech's Provost Office, E&AS Division, and IST • Thursday, April 19, 2012

Outline

- Proof that $m_{\mathcal{H}}(N)$ is polynomial

- Proof that $m_{\mathcal{H}}(N)$ can replace M

Bounding $m_{\mathcal{H}}(N)$

To show: $m_{\mathcal{H}}(N)$ is polynomial

We show: $m_{\mathcal{H}}(N) \ \leq \ \cdots \ \leq \ \cdots \ \leq$ a polynomial

Key quantity:

$B(N, k)$: Maximum number of dichotomies on N points, with break point k

Recursive bound on $B(N,k)$

Consider the following table:

$$B(N,k) = \ \alpha + 2\beta$$

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
			$+1$	$+1$	$\ldots$	$+1$	$+1$
			-1	$+1$	$\ldots$	$+1$	-1
	S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	-1	-1
			-1	$+1$	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	$+1$
			-1	-1	$\ldots$	$+1$	$+1$
	S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	$+1$
S_2			-1	-1	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	$+1$	-1
	S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	-1	-1

Estimating α and β

Focus on $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N-1}$ columns:

$$\alpha + \beta \ \leq \ B(N-1, k)$$

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\dots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
			$+1$	$+1$	$\dots$	$+1$	$+1$
			-1	$+1$	$\dots$	$+1$	-1
	S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\dots$	-1	-1
			-1	$+1$	$\dots$	-1	$+1$
			$+1$	-1	$\dots$	$+1$	$+1$
			-1	-1	$\dots$	$+1$	$+1$
	S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\dots$	$+1$	$+1$
S_2			-1	-1	$\dots$	-1	$+1$
			$+1$	-1	$\dots$	$+1$	-1
			-1	-1	$\dots$	$+1$	-1
	S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\dots$	$+1$	-1
			-1	-1	$\dots$	-1	-1

Estimating β by itself

Now, focus on the $S_2 = S_2^+ \cup S_2^-$ rows:

$$\beta \;\leq\; B(N-1, k-1)$$

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
			$+1$	$+1$	$\ldots$	$+1$	$+1$
			-1	$+1$	$\ldots$	$+1$	-1
	S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	-1	-1
			-1	$+1$	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	$+1$
			-1	-1	$\ldots$	$+1$	$+1$
	S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	$+1$
S_2			-1	-1	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	$+1$	-1
	S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	-1	-1

Putting it together

$$B(N,k) = \ \alpha \ + \ 2\beta$$

$$\alpha + \beta \ \le \ B(N-1,k)$$

$$\beta \ \le \ B(N-1,k-1)$$

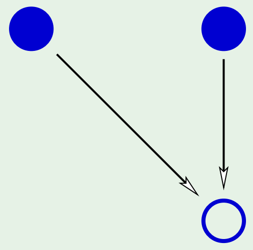
$$B(N,k) \le$$

$$B(N-1,k) + B(N-1,k-1)$$

	# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\dots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
		$+1$	$+1$	$\dots$	$+1$	$+1$
		-1	$+1$	$\dots$	$+1$	-1
S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		$+1$	-1	$\dots$	-1	-1
		-1	$+1$	$\dots$	-1	$+1$
		$+1$	-1	$\dots$	$+1$	$+1$
		-1	-1	$\dots$	$+1$	$+1$
S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		$+1$	-1	$\dots$	$+1$	$+1$
S_2		-1	-1	$\dots$	-1	$+1$
		$+1$	-1	$\dots$	$+1$	-1
		-1	-1	$\dots$	$+1$	-1
S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		$+1$	-1	$\dots$	$+1$	-1
		-1	-1	$\dots$	-1	-1

Numerical computation of $B(N,k)$ bound

$$B(N,k) \leq B(N-1,k) + B(N-1,k-1)$$



N \ k	1	2	3	4	5	6	..
1	1	2	2	2	2	2	..
2	1	3	4	4	4	4	..
3	1	4	7	8	8	8	..
4	1	5	11	..	..	..	..
5	1	6	:	.			
6	1	7	:		.		
:	:	:	:			.	

Analytic solution for $B(N, k)$ bound

$$B(N,k) \leq B(N-1,k) + B(N-1,k-1)$$

Theorem:

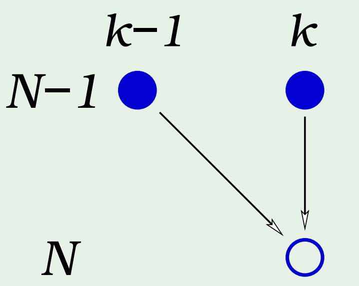
$$B(N,k) \;\leq\; \sum_{i=0}^{k-1} \binom{N}{i}$$

1. Boundary conditions: easy

		k						
		1	2	3	4	5	6	..
	1	1	2	2	2	2	2	..
	2	1						
	3	1						
N	4	1						
	5	1						
	6	1						
	$\vdots$	$\vdots$						

2. The induction step

$$
\begin{aligned}
\sum_{i=0}^{k-1} \binom{N}{i} &= \sum_{i=0}^{k-1} \binom{N-1}{i} + \sum_{i=0}^{k-2} \binom{N-1}{i} \; {\color{red}?} \\
&= 1 + \sum_{i=1}^{k-1} \binom{N-1}{i} + \sum_{i=1}^{k-1} \binom{N-1}{i-1} \\
&= 1 + \sum_{i=1}^{k-1} \left[\binom{N-1}{i} + \binom{N-1}{i-1} \right] \\
&= 1 + \sum_{i=1}^{k-1} \binom{N}{i} = \sum_{i=0}^{k-1} \binom{N}{i} \; {\color{green}\checkmark}
\end{aligned}
$$



It is polynomial!

For a given $\mathcal{H}$, the break point k is fixed

$$m_{\mathcal{H}}(N) \quad \leq \quad \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{maximum power is } N^{k-1}}$$

Three examples

$$\sum_{i=0}^{k-1} \binom{N}{i}$$

- $\mathcal{H}$ is **positive rays:** (break point $k = 2$)

$$m_{\mathcal{H}}(N) = N + 1 \ \leq \ N + 1$$

- $\mathcal{H}$ is **positive intervals:** (break point $k = 3$)

$$m_{\mathcal{H}}(N) = \tfrac{1}{2}N^2 + \tfrac{1}{2}N + 1 \ \leq \ \tfrac{1}{2}N^2 + \tfrac{1}{2}N + 1$$

- $\mathcal{H}$ is **2D perceptrons:** (break point $k = 4$)

$$m_{\mathcal{H}}(N) = \ ? \ \leq \ \tfrac{1}{6}N^3 + \tfrac{5}{6}N + 1$$

Outline

- Proof that $m_{\mathcal{H}}(N)$ is polynomial

- Proof that $m_{\mathcal{H}}(N)$ can replace M

What we want

Instead of:

$$\mathbb{P}[\ |E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\ ] \ \leq\ 2 \quad M \quad e^{-2\epsilon^2 N}$$

We want:

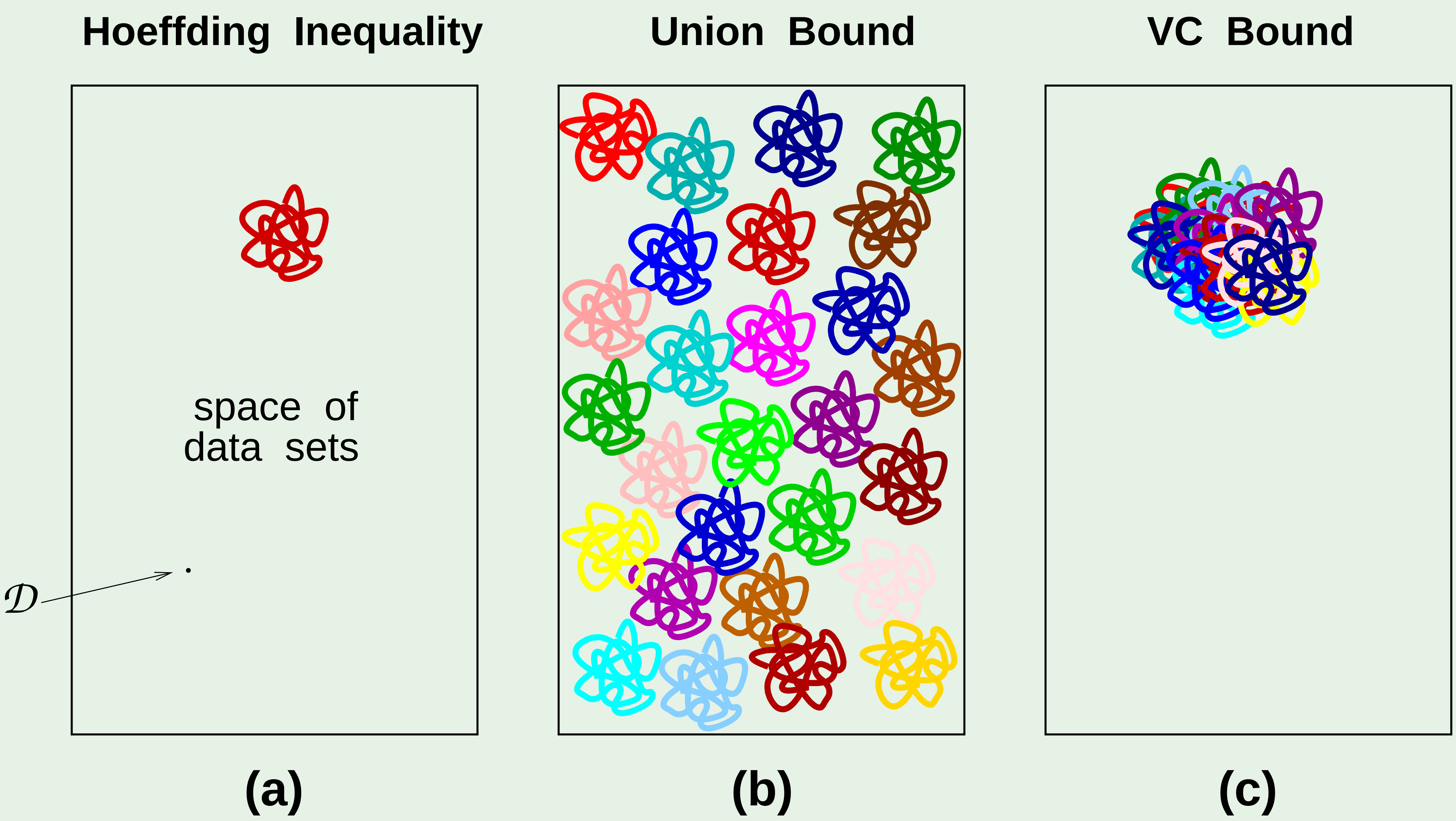
$$\mathbb{P}[\ |E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\ ] \ \leq\ 2\ m_{\mathcal{H}}(N)\ e^{-2\epsilon^2 N}$$

Pictorial proof ☺

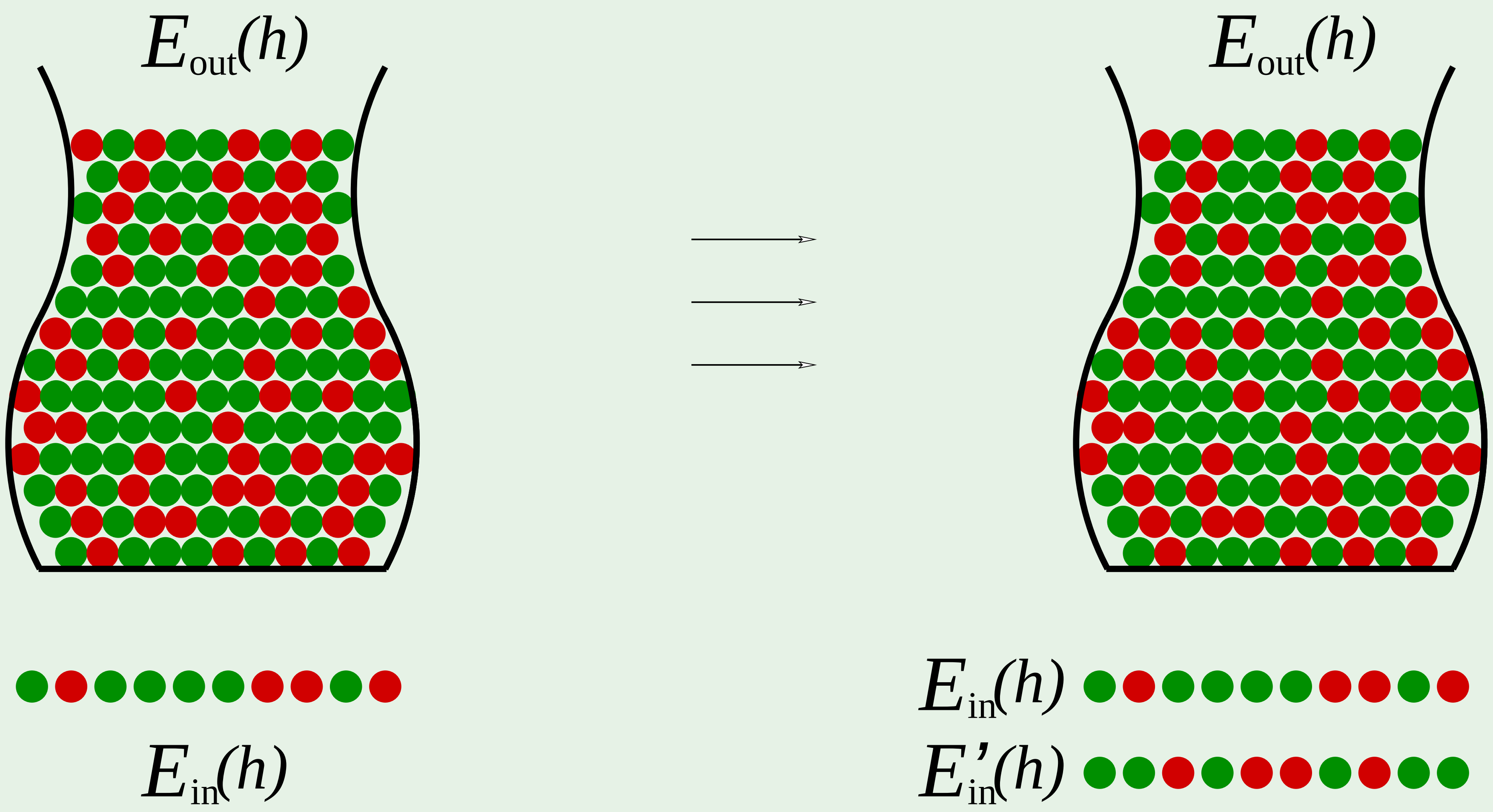
- How does $m_{\mathcal{H}}(N)$ relate to overlaps?

- What to do about E_{out}?

- Putting it together


Hoeffding Inequality
Union Bound
VC Bound
space of
data sets
$\mathcal{D}$
(a)
(b)
(c)

What to do about E_{out}



Putting it together

Not quite:

$$\mathbb{P}[\ |E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\ ] \ \leq\ 2\ \ m_{\mathcal{H}}(\ N\ )\ \ e^{-\ 2\ \epsilon^2 N}$$

but rather:

$$\mathbb{P}[\ |E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\ ] \ \leq\ 4\ \ m_{\mathcal{H}}(2N)\ \ e^{-\ \frac{1}{8}\ \epsilon^2 N}$$

The Vapnik-Chervonenkis Inequality