Learning From Data
Lecture 6
Bounding The Growth Function

Bounding the Growth Function
Models are either Good or Bad
The VC Bound - replacing $|\mathcal{H}|$ with $m_{\mathcal{H}}(N)$

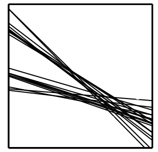
M. Magdon-Ismail
CSCI 4100/6100

RECAP: The Growth Function $m_{\mathcal{H}}(N)$

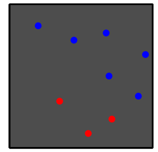
A new measure for the diversity of a hypothesis set.

$$\mathcal{H}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \{(h(\mathbf{x}_1), \ldots, h(\mathbf{x}_N))\}$$

The dichotomies (N-tuples) $\mathcal{H}$ implements on $\mathbf{x}_1, \ldots, \mathbf{x}_N$.



$\mathcal{H}$



$\mathcal{H}$ viewed through $\mathcal{D}$

The *growth function* $m_{\mathcal{H}}(N)$ considers the worst possible $\mathbf{x}_1, \ldots, \mathbf{x}_N$.

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1,\ldots,\mathbf{x}_N} |\mathcal{H}(\mathbf{x}_1, \ldots, \mathbf{x}_N)|.$$

This lecture: **Can we bound $m_{\mathcal{H}}(N)$ by a polynomial in N?**
Can we replace $|\mathcal{H}|$ by $m_{\mathcal{H}}(N)$ in the generalization bound?

Example Growth Functions

	N: 1	2	3	4	5	…
2-D perceptron	2	4	8	14	…	
1-D pos. ray	2	3	4	5	…	
2-D pos. rectangles	2	4	8	16	$< 2^5$	…

- $m_{\mathcal{H}}(N)$ drops below 2^N – **there is hope**.
- A break point is any k for which $m_{\mathcal{H}}(k) < 2^k$.

Pop Quiz I

I give you a set of k^* points $\mathbf{x}_1, \ldots, \mathbf{x}_{k^*}$ on which $\mathcal{H}$ implements $< 2^{k^*}$ dichotomys.

(a) k^* is a break point.

(b) k^* is not a break point.

(c) all break points are $> k^*$.

(d) all break points are $\leq k^*$.

(e) we don't know anything about break points.

Pop Quiz I

I give you a set of k^* points $\mathbf{x}_1, \ldots, \mathbf{x}_{k^*}$ on which $\mathcal{H}$ implements $< 2^{k^*}$ dichotomys.

(a) k^* is a break point.

(b) k^* is not a break point.

(c) all break points are $> k^*$.

(d) all break points are $\leq k^*$.

✓ (e) we don't know anything about break points.

Pop Quiz II

For every set of k^* points $\mathbf{x}_1, \ldots, \mathbf{x}_{k^*}$, $\mathcal{H}$ implements $< 2^{k^*}$ dichotomys.

(a) k^* is a break point.

(b) k^* is not a break point.

(c) all $k \geq k^*$ are break points.

(d) all $k < k^*$ are break points.

(e) we don't know anything about break points.

Pop Quiz II

For every set of k^* points $\mathbf{x}_1, \ldots, \mathbf{x}_{k^*}$, $\mathcal{H}$ implements $< 2^{k^*}$ dichotomys.

✓ (a) k^* is a break point.

(b) k^* is not a break point.

✓ (c) all $k \geq k^*$ are break points.

(d) all $k < k^*$ are break points.

(e) we don't know anything about break points.

Pop Quiz III

To show that k is *not* a break point for $\mathcal{H}$:

(a) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ can shatter.

(b) Show $\mathcal{H}$ can shatter any set of k points.

(c) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ cannot shatter.

(d) Show $\mathcal{H}$ cannot shatter any set of k points.

(e) Show $m_{\mathcal{H}}(k) = 2^k$.

Pop Quiz III

To show that k is *not* a break point for $\mathcal{H}$:

✓ **(a)** Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ can shatter.

overkill **(b)** Show $\mathcal{H}$ can shatter any set of k points.

(c) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ cannot shatter.

(d) Show $\mathcal{H}$ cannot shatter any set of k points.

✓ **(e)** Show $m_{\mathcal{H}}(k) = 2^k$.

Pop Quiz IV

To show that k *is* a break point for $\mathcal{H}$:

(a) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ can shatter.

(b) Show $\mathcal{H}$ can shatter any set of k points.

(c) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ cannot shatter.

(d) Show $\mathcal{H}$ cannot shatter any set of k points.

(e) Show $m_{\mathcal{H}}(k) > 2^k$.

Pop Quiz IV

To show that k *is* a break point for $\mathcal{H}$:

(a) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ can shatter.

(b) Show $\mathcal{H}$ can shatter any set of k points.

(c) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ cannot shatter.

✓ **(d)** Show $\mathcal{H}$ cannot shatter any set of k points.

(e) Show $m_{\mathcal{H}}(k) > 2^k$.

Back to Our Combinatorial Puzzle

How many dichotomies can you list on 4 points so that no 2 is shattered.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○

Can we add a 6th dichotomy?

Can't Add A 6th Dichotomy

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	●	●	○

The Combinatorial Quantity $B(N,k)$

How many dichotomies can you list on $\underset{N}{\underset{\uparrow}{4}}$ points so that no $\underset{k}{\underset{\uparrow}{2}}$ are shattered.

$B(N,k)$: Max. number of dichotomys on N points so that no k are shattered.

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

$B(3,2)=4$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○

$B(4,2)=5$

Let's Try To Bound $B(4,3)$

How many dichotomies can you list on 4 points so that no subset of 3 is shattered.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○

Two Kinds of Dichotomys

Prefix appears once or prefix appears twice.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○

Reorder the Dichotomys

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

α: prefix appears once

β: prefix appears twice

$$B(4,3) = \alpha + 2\beta$$

First, Bound $\alpha + \beta$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

$$\alpha + \beta \leq B(3,3)$$

$\uparrow$

A list on 3 points, with no 3 shattered (why?)

Second, Bound β

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

$$\beta \leq B(3,2)$$

$\uparrow$

If 2 points are shattered, then using the mirror dichotomies you shatter 3 points (why?)

Combining to Bound $\alpha + 2\beta$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

$$\begin{aligned} B(4,3) &= \alpha + \beta \ + \beta \\ &\leq B(3,3) + B(3,2) \end{aligned}$$

The argument generalizes to (N, k)

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

Boundary Cases: $B(N, 1)$ and $B(N, N)$

N \ k	1	2	3	4	5	6	$\cdots$
1	1						
2	1	3					
3	1		7				
4	1			15			
5	1				31		
6	1					63	
$\vdots$	$\vdots$						$\ddots$

$$B(N, 1) = 1 \qquad \text{(why?)}$$

$$B(N, N) = 2^N - 1 \qquad \text{(why?)}$$

Recursion Gives $B(N, k)$ Bound

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

N \ k	1	2	3	4	5	6	$\cdots$
1	1						
2	1	3					
3	1	**4**	7				
4	1			15			
5	1				31		
6	1					63	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Recursion Gives $B(N, k)$ Bound

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

N \ k	1	2	3	4	5	6	$\cdots$
1	1						
2	1	3					
3	1	**4**	7				
4	1	**5**	$\boxed{\mathbf{11}}$	15			
5	1	**6**	**16**	**26**	31		
6	1	**7**	**22**	**42**	**57**	63	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Analytic Bound for $B(N, k)$

Theorem.

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

Proof: (Induction on N.)

1. Verify for $N = 1$: $B(1, 1) \le \binom{1}{0} = 1$ ✓
2. Suppose $B(N, k) \le \sum_{i=0}^{k-1} \binom{N}{i}$.

Lemma. $\binom{N}{k} + \binom{N}{k-1} = \binom{N+1}{k}$.

$$\begin{aligned} B(N+1, k) &\le B(N, k) + B(N, k-1) \\ &\le \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i} \\ &= \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=1}^{k-1} \binom{N}{i-1} \\ &= 1 + \sum_{i=1}^{k-1} \left(\binom{N}{i} + \binom{N}{i-1} \right) \\ &= 1 + \sum_{i=1}^{k-1} \binom{N+1}{i} \qquad \text{(lemma)} \\ &= \sum_{i=0}^{k-1} \binom{N+1}{i} \end{aligned}$$

$m_{\mathcal{H}}(N)$ is bounded by $B(N, k)$!

Theorem. Suppose that $\mathcal{H}$ has a break point at k. Then,

$$m_{\mathcal{H}}(N) \leq B(N, k).$$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	...	$\mathbf{x}_N$
○	○	○	○	...	●
○	○	○	●	...	○
○	○	●	○	...	○
○	●	○	○	...	○
●	○	○	○	...	●
○	○	●	●	...	●
○	●	○	●	...	○
⋮	⋮	⋮	⋮	...	⋮

Consider any k points.

They cannot be shattered (otherwise k woud not be a break point).

$B(N, k)$ is largest such list.

$m_{\mathcal{H}}(N) \leq B(N, k)$

Once bitten, twice shy ... Once Broken, Forever Polynomial

Theorem. If k is any break point for $\mathcal{H}$, so $m_{\mathcal{H}}(k) < 2^k$, then

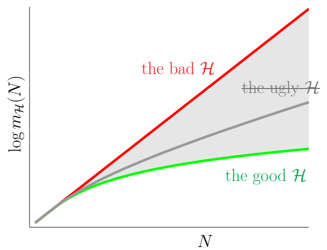
$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}.$$

Facts (Problems 2.5 and 2.6):

$$\sum_{i=0}^{k-1} \binom{N}{i} \leq \begin{cases} N^{k-1} + 1 \\ \left(\dfrac{eN}{k-1}\right)^{k-1} \end{cases} \qquad \text{(polynomial in } N\text{)}$$

This is **huge**: if we can replace $|\mathcal{H}|$ with $m_{\mathcal{H}}(N)$ in the bound, then learning is feasible.

A Hypothesis Set is either Good and Bad



	N 1	2	3	4	5	$\cdots$	$m_{\mathcal{H}}(N)$
2-D perceptron	2	4	8	14	$\cdots$	$\cdots$	$\leq N^3 + 1$
1-D pos. ray	2	3	4	5	$\cdots$	$\cdots$	$\leq N^1 + 1$
2-D pos. rectangles	2	4	8	16	$< 2^5$	$\cdots$	$\leq N^4 + 1$

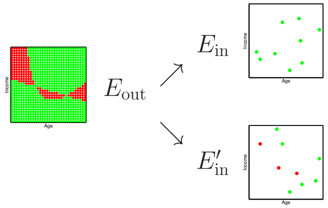
We have One Step in the Puzzle

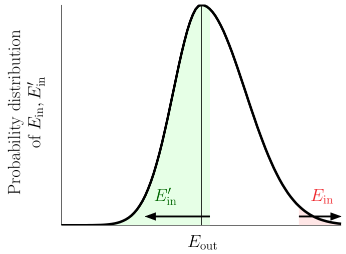
✓ Can we get a polynomial bound on $m_{\mathcal{H}}(N)$ even for infinite $\mathcal{H}$?

Can we replace $|\mathcal{H}|$ with $m_{\mathcal{H}}(N)$ in the generalization bound?

(i) How to Deal With E_{out} (Sketch)

The *ghost data set*: a 'fictitious' data set $\mathcal{D}'$:





E'_{in} is like a test error on N new points.

E_{in} deviates from E_{out} implies E_{in} deviates from E'_{in}.

E_{in} and E'_{in} have the same distribution.

$\mathbb{P}[(E'_{\text{in}}(g), E_{\text{in}}(g))$ "deviate"$] \geq \frac{1}{2}\,\mathbb{P}\,[(E_{\text{out}}(g), E_{\text{in}}(g))$ "deviate"$]$

We can analyze deviations between two in-sample errors.

(ii) Real Plus Ghost Data Set $= 2N$ points

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\ldots$	$\mathbf{x}_N$	$\mathbf{x}_{N+1}$	$\mathbf{x}_{N+2}$	$\mathbf{x}_{N+3}$	$\ldots$	$\mathbf{x}_{2N}$
○	○	●	$\ldots$	○	●	●	○	$\ldots$	○

Number of dichotomys is at most $m_{\mathcal{H}}(2N)$.

Up to technical details, analyze a "hypothesis set" of size at most $m_{\mathcal{H}}(2N)$.

The Vapnik-Chervonenkis Bound (VC Bound)

$$\mathbb{P}\left[|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\right] \leq 4m_{\mathcal{H}}(2N)e^{-\epsilon^2 N/8}, \qquad \text{for any } \epsilon > 0.$$

$$\mathbb{P}\left[|E_{\text{in}}(g) - E_{\text{out}}(g)| \leq \epsilon\right] \geq 1 - 4m_{\mathcal{H}}(2N)e^{-\epsilon^2 N/8}, \qquad \text{for any } \epsilon > 0.$$

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{8}{N}\log\frac{4m_{\mathcal{H}}(2N)}{\delta}}, \qquad \text{w.p. at least } 1-\delta.$$