Machine Learning from Data

Lecture 6: Spring 2021

Today's Lecture

- Bounding the Growth Function
- Models are either Good or Bad
- The VC Bound

Putting Everything Together

- The growth function:

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

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

Quiz Q1

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.

Quiz Q2

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.

Quiz Q3

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

Quiz Q4

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 the puzzle

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

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

The combinatorial relationship

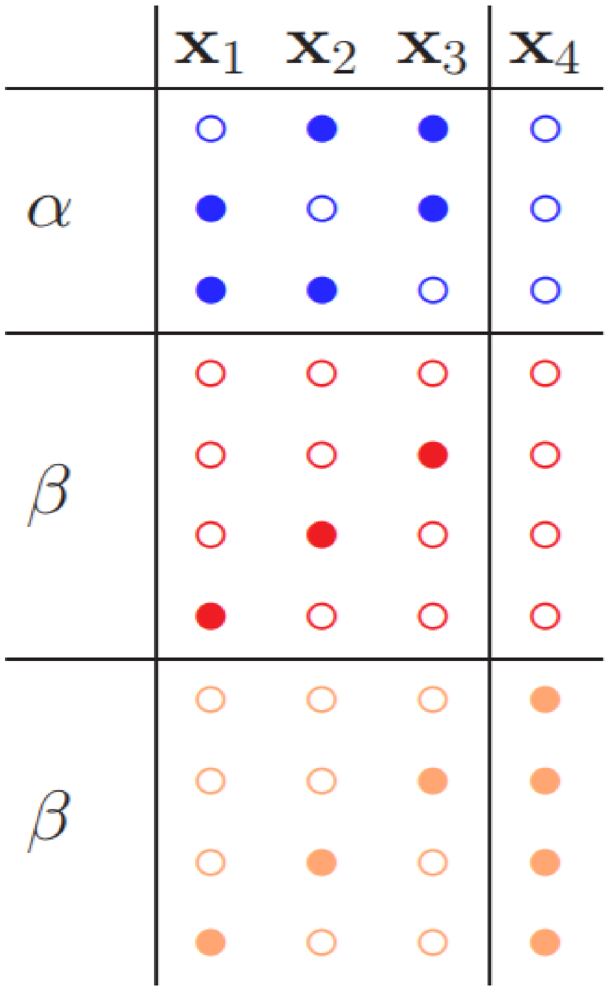
$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

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

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

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

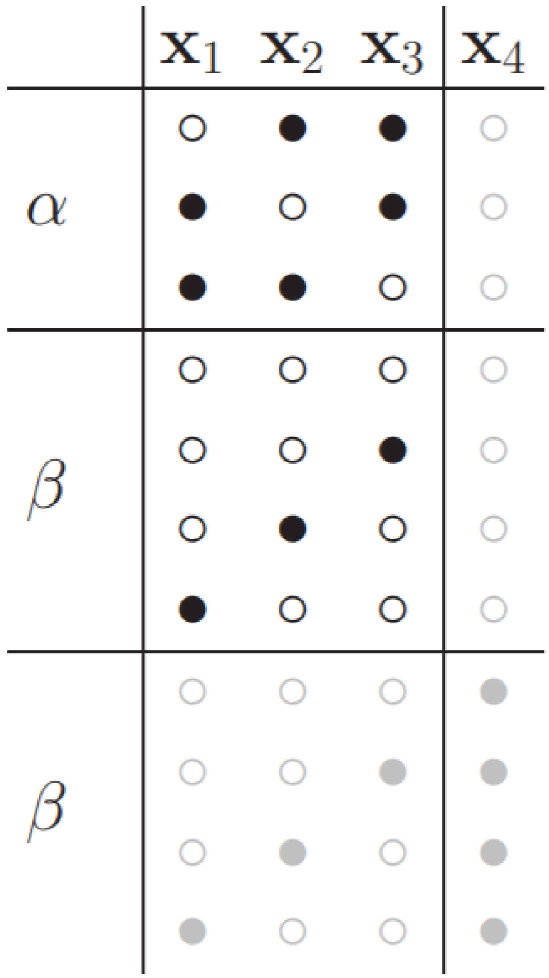
$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○



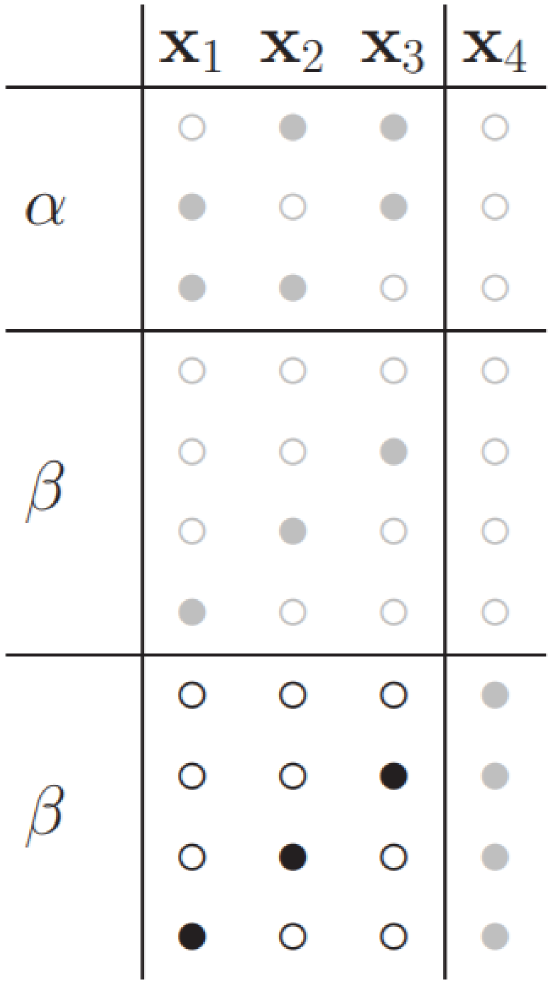
α: prefix appears once

β: prefix appears twice

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



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



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

Fill the table values

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

$$B(N, 1) = 1$$

$$B(N, N) = 2^N - 1$$

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

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

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

		k						
		1	2	3	4	5	6	$\cdots$
	1	1						
	2	1	3					
N	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

-

Proof: (Induction on N.)

1. Verify for $N = 1$: $B(1,1) \le \binom{1}{0} = 1$ ✓

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

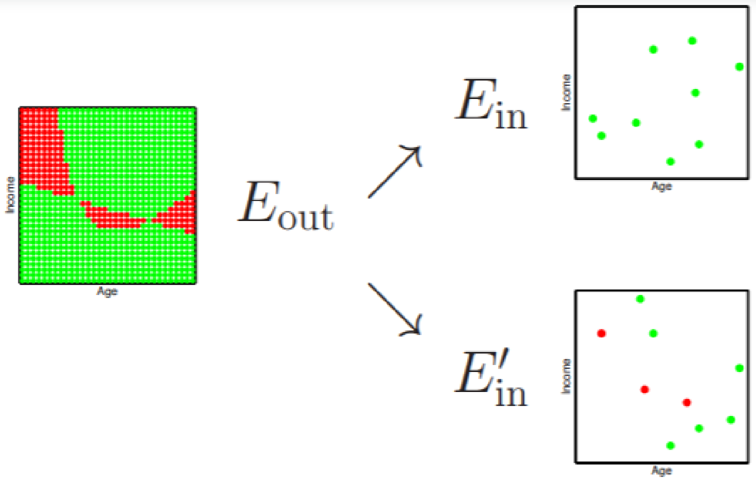
Lemma. $\binom{N}{k} + \binom{N}{k-1} = \binom{N+1}{k}$.

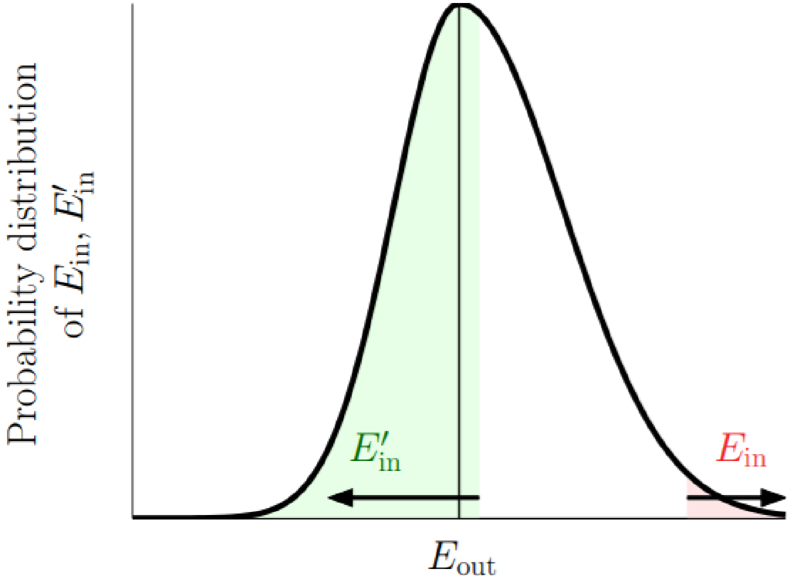
$$\begin{aligned}
B(N+1,k) &\le B(N,k) + B(N,k-1) \\
&\le \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i} \\
&= \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=1}^{k-1} \binom{N}{i-1} \\
&= 1 + \sum_{i=1}^{k-1} \left(\binom{N}{i} + \binom{N}{i-1} \right) \\
&= 1 + \sum_{i=1}^{k-1} \binom{N+1}{i}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

✓ 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?

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”$] \geq \frac{1}{2}\,\mathbb{P}\left[(E_{\text{out}}(g), E_{\text{in}}(g))\right.$ “deviate”$\left.\right]$

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

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)

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

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

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

Thanks!