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

A new measure for the diversity of a hypothesis set.

$$\mathcal{H}(x_1, \ldots, x_N) = \{(h(x_1), \ldots, h(x_N))\}$$

The dichotomies ($N$-tuples) $\mathcal{H}$ implements on $x_1, \ldots, x_N$.

The growth function $m_{\mathcal{H}}(N)$ considers the worst possible $x_1, \ldots, x_N$.

$$m_{\mathcal{H}}(N) = \max_{x_1, \ldots, x_N} |\mathcal{H}(x_1, \ldots, x_N)|.$$  

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

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D perceptron</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1-D pos. ray</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2-D pos. rectangles</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>$&lt; 2^5$</td>
<td>...</td>
</tr>
</tbody>
</table>

- $m_{\mathcal{H}}(N)$ drops below $2^N$ – **there is hope**.
- A **break point** is any $k$ for which $m_{\mathcal{H}}(k) < 2^k$. 
I give you a set of \( k^* \) points \( x_1, \ldots, x_{k^*} \) 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.
I give you a set of $k^*$ points $x_1, \ldots, x_{k^*}$ 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.
For every set of $k^*$ points $x_1, \ldots, x_{k^*}$, $\mathcal{H}$ 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 $x_1, \ldots, x_{k^*}$, $\mathcal{H}$ 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.
To show that $k$ is not a break point for $\mathcal{H}$:

(a) Show a set of $k$ points $x_1, \ldots, x_k$ which $\mathcal{H}$ can shatter.

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

(c) Show a set of $k$ points $x_1, \ldots, x_k$ 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 \textit{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 $x_1, \ldots, x_k$ which $\mathcal{H}$ can shatter.

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

(c) Show a set of $k$ points $x_1, \ldots, x_k$ which $\mathcal{H}$ cannot shatter.

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

(e) Show $m_{\mathcal{H}}(k) > 2^k$. 
To show that $k$ is a break point for $\mathcal{H}$:

(a) Show a set of $k$ points $x_1, \ldots, x_k$ which $\mathcal{H}$ can shatter.

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

(c) Show a set of $k$ points $x_1, \ldots, x_k$ which $\mathcal{H}$ cannot shatter.

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

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

Bounding the Growth Function: 11 / 31 

Combinatorial puzzle again →
Back to Our Combinatorial Puzzle

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

Can we add a 6th dichotomy?

<table>
<thead>
<tr>
<th></th>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>X₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Can’t add a 6th dichotomy
Can’t Add A 6th Dichotomy

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>
The Combinatorial Quantity $B(N, k)$

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

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

\[
\begin{array}{ccc}
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\
\circ & \circ & \circ \\
\circ & \circ & \bullet \\
\circ & \bullet & \circ \\
\bullet & \circ & \circ \\
\end{array}
\quad
\begin{array}{cccc}
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \\
\circ & \circ & \circ & \circ \\
\circ & \circ & \circ & \bullet \\
\circ & \bullet & \circ & \circ \\
\bullet & \circ & \circ & \circ \\
\end{array}
\]

$B(3, 2) = 4$ \hspace{1cm} $B(4, 2) = 5$
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.

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>●</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
<td>●</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>●</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>●</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>O</td>
<td>●</td>
<td>O</td>
<td>●</td>
</tr>
<tr>
<td>●</td>
<td>O</td>
<td>O</td>
<td>●</td>
</tr>
<tr>
<td>O</td>
<td>●</td>
<td>●</td>
<td>O</td>
</tr>
<tr>
<td>●</td>
<td>O</td>
<td>●</td>
<td>O</td>
</tr>
<tr>
<td>●</td>
<td>●</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>●</td>
<td>●</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Bounding the Growth Function: 15/31
Two Kinds of Dichotomys

Prefix appears once or prefix appears twice.

\[
\begin{array}{cccc|c}
X_1 & X_2 & X_3 & X_4 \\
\circ & \circ & \circ & \circ \\
\circ & \circ & \circ & \bullet \\
\circ & \circ & \bullet & \circ \\
\circ & \bullet & \circ & \circ \\
\bullet & \circ & \circ & \circ \\
\circ & \circ & \bullet & \bullet \\
\circ & \bullet & \circ & \bullet \\
\bullet & \circ & \circ & \bullet \\
\bullet & \bullet & \bullet & \circ \\
\bullet & \circ & \bullet & \circ \\
\bullet & \bullet & \circ & \circ \\
\bullet & \bullet & \bullet & \circ \\
\end{array}
\]
Reorder the Dichotomys

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\alpha$: prefix appears once

$\beta$: prefix appears twice

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

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>$\beta$</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>$\beta$</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
</tbody>
</table>

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

A list on 3 points, with no 3 shattered (why?)
Second, Bound $\beta$

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

$\beta \leq B(3, 2)$
Combining to Bound $\alpha + 2\beta$

The argument generalizes to $(N, k)$

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

$$B(N, k) \leq B(N - 1, k) + B(N - 1, k - 1)$$
## Boundary Cases: $B(N, 1)$ and $B(N, N)$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1   3</td>
</tr>
<tr>
<td>3</td>
<td>1   7</td>
</tr>
<tr>
<td>4</td>
<td>1   15</td>
</tr>
<tr>
<td>5</td>
<td>1   31</td>
</tr>
<tr>
<td>6</td>
<td>1   63</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- $B(N, 1) = 1$ (why?)
- $B(N, N) = 2^N - 1$ (why?)
Recursion Gives $B(N, k)$ Bound

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

<table>
<thead>
<tr>
<th>$N$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>\ldots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>\ldots</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>\ldots</td>
</tr>
</tbody>
</table>
Recursion Gives $B(N, k)$ Bound

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

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6</td>
<td>16</td>
<td>26</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>7</td>
<td>22</td>
<td>42</td>
<td>57</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Theorem.

\[ B(N, k) \leq \sum_{i=0}^{k-1} \binom{N}{i}. \]

Proof: (Induction on \( N \).)
1. Verify for \( N = 1 \): \( B(1, 1) \leq \binom{1}{0} = 1 \quad \checkmark \)
2. Suppose \( B(N, k) \leq \sum_{i=0}^{k-1} \binom{N}{i} \).

\[ \text{Lemma} \cdot \binom{N}{k} + \binom{N}{k-1} = \binom{N+1}{k}. \]

\[ B(N + 1, k) \leq B(N, k) + B(N, k - 1) \]
\[ \leq \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i} \]
\[ = \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=1}^{k-1} \binom{N}{i-1} \]
\[ = 1 + \sum_{i=1}^{k-1} \left( \binom{N}{i} + \binom{N}{i-1} \right) \]
\[ = 1 + \sum_{i=1}^{k-1} \binom{N+1}{i} \quad (\text{lemma}) \]
\[ = \sum_{i=0}^{k-1} \binom{N+1}{i} \]
$m_{\mathcal{H}}(N)$ is bounded by $B(N, k)$!

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

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

Consider any $k$ points.

They cannot be shattered (otherwise $k$ would 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):

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

This is huge: if we can replace $|\mathcal{H}|$ with $m_\mathcal{H}(N)$ in the bound, then learning is feasible.
Bounding the Growth Function: 27 / 31
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 $E_{\text{out}}$ (Sketch)

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

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

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

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

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

We can analyze deviations between two in-sample errors.
(ii) Real Plus Ghost Data Set = $2N$ points

Number of dichotomys is at most $m_H(2N)$.

Up to technical details, analyze a “hypothesis set” of size at most $m_H(2N)$. 
The Vapnik-Chervonenkis Bound (VC Bound)

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

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

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