Lecture 6: Theory of Generalization

Hsuan-Tien Lin (林軒田)
htlin@csie.ntu.edu.tw

Department of Computer Science & Information Engineering
National Taiwan University
(國立台灣大學資訊工程系)
Roadmap

1. When Can Machines Learn?
2. **Why** Can Machines Learn?

Lecture 5: Training versus Testing
- **effective** price of choice in training: *(wishfully)*
- growth function \( m_H(N) \) with a break point

Lecture 6: Theory of Generalization
- Restriction of Break Point
- Bounding Function: Basic Cases
- Bounding Function: Inductive Cases
- A Pictorial Proof

3. How Can Machines Learn?
4. How Can Machines Learn Better?
The Four Break Points

growth function $m_{\mathcal{H}}(N)$: max number of dichotomies

- positive rays: $m_{\mathcal{H}}(2) = 3 < 2^2$: break point at 2
  $m_{\mathcal{H}}(N) = N + 1$

- positive intervals: $m_{\mathcal{H}}(3) = 7 < 2^3$: break point at 3
  $m_{\mathcal{H}}(N) = \frac{1}{2} N^2 + \frac{1}{2} N + 1$

- convex sets: $m_{\mathcal{H}}(N) = 2^N$ always: no break point

- 2D perceptrons: $m_{\mathcal{H}}(4) = 14 < 2^4$: break point at 4
  $m_{\mathcal{H}}(N) < 2^N$ in some cases

break point $k$, break point $k + 1$, ... what else?
Restriction of Break Point (1/2)

What ‘must be true’ when minimum break point \( k = 2 \)

- \( N = 1 \): every \( m_{\mathcal{H}}(N) = 2 \) by definition
- \( N = 2 \): every \( m_{\mathcal{H}}(N) < 4 \) by definition
  (so maximum possible = 3)

Maximum possible \( m_{\mathcal{H}}(N) \) when \( N = 3 \) and \( k = 2 \)?

1 dichotomy, shatter any two points? No

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bigcirc )</td>
<td>( \bigcirc )</td>
<td>( \bigcirc )</td>
</tr>
</tbody>
</table>
Restriction of Break Point (1/2)

what ‘must be true’ when \textbf{minimum break point} \( k = 2 \)

- \( N = 1 \): every \( m_{\mathcal{H}}(N) = 2 \) by definition
- \( N = 2 \): every \( m_{\mathcal{H}}(N) < 4 \) by definition
  (so \textbf{maximum possible} = 3)

maximum possible \( m_{\mathcal{H}}(N) \) when \( N = 3 \) and \( k = 2 \)?

2 dichotomies , shatter any two points? \textbf{no}

\[
\begin{array}{ccc}
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\
\circ & \circ & \circ \\
\circ & \circ & \times \\
\end{array}
\]
Theory of Generalization

Restriction of Break Point

Restriction of Break Point (1/2)

what ‘must be true’ when minimum break point $k = 2$

- $N = 1$: every $m_{\mathcal{H}}(N) = 2$ by definition
- $N = 2$: every $m_{\mathcal{H}}(N) < 4$ by definition
  (so maximum possible = 3)

maximum possible $m_{\mathcal{H}}(N)$ when $N = 3$ and $k = 2$?

3 dichotomies, shatter any two points? no

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>o</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>x</td>
<td>o</td>
</tr>
</tbody>
</table>
Theory of Generalization

Restriction of Break Point

**Restriction of Break Point (1/2)**

what ‘must be true’ when **minimum break point** $k = 2$

- $N = 1$: every $m_H(N) = 2$ by definition
- $N = 2$: every $m_H(N) < 4$ by definition
  (so **maximum possible** = 3)

**maximum possible $m_H(N)$ when $N = 3$ and $k = 2$?**

4 dichotomies, shatter any two points? **yes**

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>○</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>○</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>○</td>
<td>×</td>
<td>○</td>
</tr>
<tr>
<td>○</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>
Restriction of Break Point (1/2)

what ‘must be true’ when \textbf{minimum break point} $k = 2$

- $N = 1$: every $m_H(N) = 2$ by definition
- $N = 2$: every $m_H(N) < 4$ by definition
  (so \textbf{maximum possible} = 3)

maximum possible $m_H(N)$ when $N = 3$ and $k = 2$?

4 dichotomies, shatter any two points? \textbf{no}

\begin{tabular}{ccc}
  \hline
  $x_1$ & $x_2$ & $x_3$ \\
  \hline
  \circ & \circ & \circ \\
  \circ & \circ & \times \\
  \circ & \times & \circ \\
  \times & \circ & \circ \\
  \hline
\end{tabular}
Restriction of Break Point (1/2)

what ‘must be true’ when minimum break point \( k = 2 \)

- \( N = 1 \): every \( m_\mathcal{H}(N) = 2 \) by definition
- \( N = 2 \): every \( m_\mathcal{H}(N) < 4 \) by definition (so maximum possible = 3)

maximum possible \( m_\mathcal{H}(N) \) when \( N = 3 \) and \( k = 2 \)?

5 dichotomies, shatter any two points? yes

<table>
<thead>
<tr>
<th>( \mathbf{x}_1 )</th>
<th>( \mathbf{x}_2 )</th>
<th>( \mathbf{x}_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>x</td>
</tr>
<tr>
<td>o</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>x</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>x</td>
<td>o</td>
<td>x</td>
</tr>
</tbody>
</table>
Restriction of Break Point (1/2)

what ‘must be true’ when minimum break point $k = 2$

- $N = 1$: every $m_{\mathcal{H}}(N) = 2$ by definition
- $N = 2$: every $m_{\mathcal{H}}(N) < 4$ by definition
  (so maximum possible = 3)

maximum possible $m_{\mathcal{H}}(N)$ when $N = 3$ and $k = 2$?

5 dichotomies, shatter any two points? yes

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>○</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>○</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>×</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>×</td>
<td>×</td>
<td>○</td>
</tr>
</tbody>
</table>

Hsuan-Tien Lin (NTU CSIE)
Restriction of Break Point (1/2)

what ‘must be true’ when minimum break point \( k = 2 \)

- \( N = 1 \): every \( m_{\mathcal{H}}(N) = 2 \) by definition
- \( N = 2 \): every \( m_{\mathcal{H}}(N) < 4 \) by definition
  (so maximum possible = 3)

maximum possible \( m_{\mathcal{H}}(N) \) when \( N = 3 \) and \( k = 2 \)?

5 dichotomies, shatter any two points? yes

<table>
<thead>
<tr>
<th></th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</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>
</tbody>
</table>
Restriction of Break Point (1/2)

what ‘must be true’ when minimum break point \( k = 2 \)

- \( N = 1 \): every \( m_H(N) = 2 \) by definition
- \( N = 2 \): every \( m_H(N) < 4 \) by definition (so maximum possible = 3)

maximum possible \( m_H(N) \) when \( N = 3 \) and \( k = 2 \)?

maximum possible so far: 4 dichotomies

<table>
<thead>
<tr>
<th></th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>2</td>
<td>o</td>
<td>o</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>o</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

:-)  :-)  :-)
Restriction of Break Point (2/2)

what ‘must be true’ when minimum break point $k = 2$

- $N = 1$: every $m_H(N) = 2$ by definition
- $N = 2$: every $m_H(N) < 4$ by definition
  (so maximum possible = 3)
- $N = 3$: maximum possible = 4 $\ll 2^3$

—break point $k$ restricts maximum possible $m_H(N)$ a lot for $N > k$

idea: $m_H(N)$

$\leq$ maximum possible $m_H(N)$ given $k$

$\leq poly(N)$
When minimum break point $k = 1$, what is the maximum possible $m_{\mathcal{H}}(N)$ when $N = 3$?

Reference Answer: 1

Because $k = 1$, the hypothesis set cannot even shatter one point. Thus, every ‘column’ of the table cannot contain both $\circ$ and $\times$. Then, after including the first dichotomy, it is not possible to include any other different dichotomy. Thus, the maximum possible $m_{\mathcal{H}}(N)$ is 1.
Bounding Function

**Bounding function** $B(N, k)$:
maximum possible $m_{\mathcal{H}}(N)$ when break point $= k$

- combinatorial quantity:
  maximum number of length-$N$ vectors with $(\circ, \times)$
  while *no shatter* any length-$k$ subvectors

- irrelevant of the details of $\mathcal{H}$
  e.g. $B(N, 3)$ bounds both
  - positive intervals ($k = 3$)
  - 1D perceptrons ($k = 3$)

**new goal:** $B(N, k) \leq \text{poly}(N)$?
### Table of Bounding Function (1/4)

<table>
<thead>
<tr>
<th>( B(N, k) )</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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Known*

- \( B(2, 2) = 3 \) (maximum < 4)
- \( B(3, 2) = 4 \) (‘pictorial’ proof previously)
### Table of Bounding Function (2/4)

<table>
<thead>
<tr>
<th>$B(N, k)$</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></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
</tbody>
</table>

**Known**

- $B(N, 1) = 1$ (see previous quiz)
## Table of Bounding Function (3/4)

<table>
<thead>
<tr>
<th>$B(N, k)$</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>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>...</td>
</tr>
<tr>
<td>$N$</td>
<td>4</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td></td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

**Known**

- $B(N, k) = 2^N$ for $N < k$
  —including all dichotomies not violating ‘breaking condition’
## Theory of Generalization

### Bounding Function: Basic Cases

#### Table of Bounding Function (4/4)

<table>
<thead>
<tr>
<th>(B(N, k))</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>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>(\ldots)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>(\ldots)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>(\ldots)</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>(\ldots)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>31</td>
<td>32</td>
<td>(\ldots)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>63</td>
<td>(\ldots)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\vdots)</td>
<td>(\vdots)</td>
<td>(\vdots)</td>
<td>(\vdots)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Known

- \(B(N, k) = 2^N - 1\) for \(N = k\)
  - removing a single dichotomy satisfies ‘breaking condition’

---

more than halfway done! :-)
For the 2D perceptrons, which of the following claim is true?

1. minimum break point \( k = 2 \)
2. \( m_{\mathcal{H}}(4) = 15 \)
3. \( m_{\mathcal{H}}(N) < B(N, k) \) when \( N = k = \) minimum break point
4. \( m_{\mathcal{H}}(N) > B(N, k) \) when \( N = k = \) minimum break point

Reference Answer: 3

As discussed previously, minimum break point for 2D perceptrons is 4, with \( m_{\mathcal{H}}(4) = 14 \). Also, note that \( B(4, 4) = 15 \). So bounding function \( B(N, k) \) can be ‘loose’ in bounding \( m_{\mathcal{H}}(N) \).
### Estimating $B(4, 3)$

<table>
<thead>
<tr>
<th>$B(N, k)$</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>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td>?</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>...</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td>31</td>
<td>32</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>63</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

**Motivation**

- $B(4, 3)$ shall be related to $B(3, ?)$
  —‘adding’ one point from $B(3, ?)$

next: reduce $B(4, 3)$ to $B(3, ?)$
‘Achieving’ Dichotomies of $B(4,3)$

after checking all $2^{24}$ sets of dichotomies, the winner is ...

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>02</td>
<td>x</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>03</td>
<td>o</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>04</td>
<td>o</td>
<td>o</td>
<td>x</td>
</tr>
<tr>
<td>05</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>06</td>
<td>x</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>07</td>
<td>x</td>
<td>o</td>
<td>x</td>
</tr>
<tr>
<td>08</td>
<td>x</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>09</td>
<td>o</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10</td>
<td>o</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>11</td>
<td>o</td>
<td>o</td>
<td>x</td>
</tr>
</tbody>
</table>

how to reduce $B(4,3)$ to $B(3,?)$ cases?
Reorganized Dichotomies of $B(4, 3)$

after checking all $2^{24}$ sets of dichotomies, **the winner is** . . .

<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>01</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>02</td>
<td>×</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>03</td>
<td>○</td>
<td>×</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>04</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>○</td>
</tr>
<tr>
<td>05</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>06</td>
<td>×</td>
<td>×</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>07</td>
<td>×</td>
<td>○</td>
<td>×</td>
<td>○</td>
</tr>
<tr>
<td>08</td>
<td>×</td>
<td>○</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>09</td>
<td>○</td>
<td>×</td>
<td>×</td>
<td>○</td>
</tr>
<tr>
<td>10</td>
<td>○</td>
<td>×</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>11</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

**orange**: pair; **purple**: single
Estimating Part of $B(4, 3)$ (1/2)

$$B(4, 3) = 11 = 2\alpha + \beta$$

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</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>

<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>$\beta$</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>×</td>
<td>○</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>○</td>
<td>×</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>○</td>
</tr>
</tbody>
</table>

- $\alpha + \beta$: dichotomies on $(x_1, x_2, x_3)$
- $B(4, 3)$ ‘no shatter’ any 3 inputs
  $\implies \alpha + \beta$ ‘no shatter’ any 3

$$\alpha + \beta \leq B(3, 3)$$
Estimating Part of $B(4, 3)$ (2/2)

$B(4, 3) = 11 = 2\alpha + \beta$

- $\alpha$: dichotomies on $(x_1, x_2, x_3)$ with $x_4$ paired
- $B(4, 3)$ ‘no shatter’ any 3 inputs $\implies \alpha$ ‘no shatter’ any 2

$\alpha \leq B(3, 2)$
Putting It All Together

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

\[
\alpha + \beta \leq B(3, 3)
\]

\[
\alpha \leq B(3, 2)
\]

\[
\Rightarrow B(4, 3) \leq B(3, 3) + B(3, 2)
\]

<table>
<thead>
<tr>
<th>(B(N, k))</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>(\leq 5)</td>
<td>11</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>(\leq 6)</td>
<td>(\leq 16)</td>
<td>(\leq 26)</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>(\leq 7)</td>
<td>(\leq 22)</td>
<td>(\leq 42)</td>
<td>(\leq 57)</td>
<td>63</td>
</tr>
</tbody>
</table>

now have upper bound of bounding function
Putting It All Together

\[ B(N, k) = 2\alpha + \beta \]
\[ \alpha + \beta \leq B(N - 1, k) \]
\[ \alpha \leq B(N - 1, k - 1) \]
\[ \Rightarrow B(N, k) \leq B(N - 1, k) + B(N - 1, k - 1) \]

\[
\begin{array}{cccccc}
B(N, k) & 1 & 2 & 3 & 4 & 5 & 6 \\
1 & 1 & 2 & 2 & 2 & 2 & 2 \\
2 & 1 & 3 & 4 & 4 & 4 & 4 \\
3 & 1 & \text{4} & 7 & \text{8} & 8 & 8 \\
N & 4 & 1 & \leq 5 & 11 & 15 & 16 & 16 \\
5 & 1 & \leq 6 & \leq 16 & \leq 26 & 31 & 32 \\
6 & 1 & \leq 7 & \leq 22 & \leq 42 & \leq 57 & 63 \\
\end{array}
\]

now have upper bound of bounding function
### Bounding Function: The Theorem

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

- highest term \(N^{k-1}\)

- simple induction using **boundary and inductive formula**
- for fixed \(k\), \(B(N, k)\) upper bounded by \(\text{poly}(N)\)
  \(\implies m_H(N)\) is \(\text{poly}(N)\) if break point exists

\[\leq\] can be \(=\) actually,

**go play and prove it if math lover!** :-)

---

Hsuan-Tien Lin (NTU CSIE)
The Three Break Points

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

highest term \( N^{k-1} \)

- **positive rays:**
  \[ m_\mathcal{H}(2) = 3 < 2^2: \text{break point at 2} \]
  \[ m_\mathcal{H}(3) = 7 < 2^3: \text{break point at 3} \]

- **positive intervals:**
  \[ m_\mathcal{H}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 \leq \frac{1}{2}N^2 + \frac{1}{2}N + 1 \]

- **2D perceptrons:**
  \[ m_\mathcal{H}(N) = ? \leq \frac{1}{6}N^3 + \frac{5}{6}N + 1 \]

\[ m_\mathcal{H}(4) = 14 < 2^4: \text{break point at 4} \]

\( \text{can bound } m_\mathcal{H}(N) \text{ by only one break point} \)
For 1D perceptrons (positive and negative rays), we know that $m_{\mathcal{H}}(N) = 2N$. Let $k$ be the minimum break point. Which of the following is not true?

1. $k = 3$
2. for some integers $N > 0$, $m_{\mathcal{H}}(N) = \sum_{i=0}^{k-1} \binom{N}{i}$
3. for all integers $N > 0$, $m_{\mathcal{H}}(N) = \sum_{i=0}^{k-1} \binom{N}{i}$
4. for all integers $N > 2$, $m_{\mathcal{H}}(N) < \sum_{i=0}^{k-1} \binom{N}{i}$

Reference Answer: 3

The proof is generally trivial by listing the definitions. For (2), $N = 1$ or 2 gives the equality. One thing to notice is (4): the upper bound can be ‘loose’.
BAD Bound for General $\mathcal{H}$

want:

$$\mathbb{P}\left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon \right] \leq 2 \cdot m_{\mathcal{H}}(N) \cdot \exp \left( -2 \cdot \epsilon^2 N \right)$$

actually, when $N$ large enough,

$$\mathbb{P}\left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon \right] \leq 2 \cdot 2^{m_{\mathcal{H}}(2N)} \cdot \exp \left( -2 \cdot \frac{1}{16} \epsilon^2 N \right)$$

next: **sketch** of proof
Step 1: Replace $E_{out}$ by $E_{in}'$

\[
\frac{1}{2} \mathbb{P} \left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{in}(h) - E_{out}(h)| > \epsilon \right] \\
\leq \mathbb{P} \left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{in}(h) - E_{in}'(h)| > \frac{\epsilon}{2} \right]
\]

- $E_{in}(h)$ finitely many, $E_{out}(h)$ infinitely many — replace the evil $E_{out}$ first
- how? sample verification set $\mathcal{D}'$ of size $N$ to calculate $E_{in}'$
- BAD $h$ of $E_{in} - E_{out}$ probably $\implies$ BAD $h$ of $E_{in} - E_{in}'$

evil $E_{out}$ removed by verification with ‘ghost data’
Step 2: Decompose $\mathcal{H}$ by Kind

BAD $\leq 2 \mathbb{P} \left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{in}(h) - E'_{in}(h)| > \frac{\epsilon}{2} \right]$

$\leq 2 m_{\mathcal{H}}(2N) \mathbb{P} \left[ \text{fixed } h \text{ s.t. } |E_{in}(h) - E'_{in}(h)| > \frac{\epsilon}{2} \right]$

- $E_{in}$ with $\mathcal{D}$, $E'_{in}$ with $\mathcal{D'}$
  - now $m_{\mathcal{H}}$ comes to play
- how? infinite $\mathcal{H}$ becomes
  $|\mathcal{H}(x_1, \ldots, x_N, x'_1, \ldots, x'_N)|$
  kinds
- union bound on $m_{\mathcal{H}}(2N)$ kinds

use $m_{\mathcal{H}}(2N)$ to calculate BAD-overlap properly
Step 3: Use Hoeffding without Replacement

\[
\text{BAD} \leq 2m_H(2N) \mathbb{P}\left[ \text{fixed } h \text{ s.t. } |E_{\text{in}}(h) - E'_{\text{in}}(h)| > \frac{\epsilon}{2} \right]
\leq 2m_H(2N) \cdot 2 \exp\left(-2 \left(\frac{\epsilon}{4}\right)^2 N\right)
\]

- consider bin of 2N examples, choose N for \(E_{\text{in}}\), leave others for \(E'_{\text{in}}\)
  \(|E_{\text{in}} - E'_{\text{in}}| > \frac{\epsilon}{2} \iff |E_{\text{in}} - \frac{E_{\text{in}} + E'_{\text{in}}}{2}| > \frac{\epsilon}{4}\)
- so? just ‘smaller bin’, ‘smaller \(\epsilon\)’, and Hoeffding without replacement

use Hoeffding after zooming to fixed \(h\)
Vapnik-Chervonenkis (VC) bound:

$$\mathbb{P}\left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{in}(h) - E_{out}(h)| > \epsilon \right] \leq 4m_{\mathcal{H}}(2N) \exp \left( -\frac{1}{8} \epsilon^2 N \right)$$

- replace $E_{out}$ by $E'_{in}$
- decompose $\mathcal{H}$ by kind
- use Hoeffding without replacement

2D perceptron:
- break point? 4
- $m_{\mathcal{H}}(N)$? $O(N^3)$

learning with 2D perceptrons feasible! :-)

That’s All!
For positive rays, \( m_H(N) = N + 1 \). Plug it into the VC bound for \( \epsilon = 0.1 \) and \( N = 10000 \). What is VC bound of BAD events?

\[
\mathbb{P} \left[ \exists h \in H \text{ s.t. } |E_{in}(h) - E_{out}(h)| > \epsilon \right] \leq 4m_H(2N) \exp \left( -\frac{1}{8} \epsilon^2 N \right)
\]

1. \( 2.77 \times 10^{-87} \)
2. \( 5.54 \times 10^{-83} \)
3. \( 2.98 \times 10^{-1} \)
4. \( 2.29 \times 10^{2} \)

Reference Answer: 3

Simple calculation. Note that the BAD probability bound is not very small even with 10000 examples.
Summary

1. **When Can Machines Learn?**
2. **Why Can Machines Learn?**

**Lecture 5: Training versus Testing**

**Lecture 6: Theory of Generalization**

- Restriction of Break Point
  - break point ‘breaks’ consequent points
- Bounding Function: Basic Cases
  - $B(N, k)$ bounds $m_{\mathcal{H}}(N)$ with break point $k$
- Bounding Function: Inductive Cases
  - $B(N, k)$ is poly($N$)
- A Pictorial Proof
  - $m_{\mathcal{H}}(N)$ can replace $M$ with a few changes

- **next: how to ‘use’ the break point?**

3. **How Can Machines Learn?**
4. **How Can Machines Learn Better?**