Learning From Data
Lecture 5
Training Versus Testing

The Two Questions of Learning Theory of Generalization (Ein ≈ Eout)
An Effective Number of Hypotheses A Combinatorial Puzzle

recap: The Two Questions of Learning
1. Can we make sure that E\text{out}(g) is close enough to E\text{in}(g)?
2. Can we make E\text{in}(g) small enough?

The Hoeffding generalization bound:
\[ E\text{out}(g) \leq E\text{in}(g) + \sqrt{\frac{1}{2N} \ln \frac{2|H|}{\delta}} \]

Ein: training (eg. the practice exam)
Eout: testing (eg. the real exam)

|H| is overkill

\[ E\text{out}(g) \leq E\text{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4mH}{\delta}} \]

Why is |H| an Overkill

How did |H| come in?
Bad events
\[ B_\text{g} = \{|E\text{out}(g) - E\text{in}(g)| > \epsilon\} \]
\[ B_\text{m} = \{|E\text{out}(h_m) - E\text{in}(h_m)| > \epsilon\} \]

We do not know which g, so use a worst case union bound.
\[ P[B_\text{g}] \leq P[\text{any } B_\text{m}] \leq \sum_{m=1}^{m_H} P[B_m]. \]

|H| fails to account for similarity between hypotheses.

The new bound will be applicable to infinite H.
Measuring the Diversity (Size) of $\mathcal{H}$

We need a way to measure the diversity of $\mathcal{H}$.

A simple idea:
Fix any set of $N$ data points.
If $\mathcal{H}$ is diverse it should be able to implement all functions

... on these $N$ points.

A Data Set Reveals the True Colors of an $\mathcal{H}$

From the point of view of $D$, the entire $\mathcal{H}$ is just one dichotomy.
An Effective Number of Hypotheses

If $\mathcal{H}$ is diverse it should be able to implement many dichotomies.

$|\mathcal{H}|$ only captures the maximum possible diversity of $\mathcal{H}$.

Consider an $h \in \mathcal{H}$, and a data set $x_1, \ldots, x_N$.

$h$ gives us an $N$-tuple of $\pm 1$'s:

$$(h(x_1), \ldots, h(x_N)).$$

A dichotomy of the inputs.

If $\mathcal{H}$ is diverse, we get many different dichotomies.

If $\mathcal{H}$ contains similar functions, we only get a few dichotomies.

The growth function quantifies this.

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

Define the restriction of $\mathcal{H}$ to the inputs $x_1, x_2, \ldots, x_N$:

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

The largest set of dichotomies induced by $\mathcal{H}$:

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

$m_{\mathcal{H}}(N) \leq 2^N$.

Can we replace $|\mathcal{H}|$ by $m_{\mathcal{H}}$, an effective number of hypotheses?

- Replacing $|\mathcal{H}|$ with $2^N$ is no help in the bound. (why?)
- We want $m_{\mathcal{H}}(N) \leq \text{poly}(N)$ to get a useful error bar.

Example: 2-D Perceptron Model

Cannot implement

Can implement all 8

Can implement at most 14

$m_{\mathcal{H}}(3) = 8 = 2^3$.

$m_{\mathcal{H}}(4) = 14 < 2^4$.

What is $m_{\mathcal{H}}(5)$?

Example: 1-D Positive Ray Model

$h(x) = \text{sign}(x - w_0)$

Consider $N$ points.

There are $N + 1$ dichotomies depending on where you put $w_0$.

$m_{\mathcal{H}}(N) = N + 1$. 
Example: Positive Rectangles in 2-D

\[ N = 4 \]

\begin{align*}
  & x_1 \quad x_2 \quad x_3 \\
\end{align*}

\[ N = 5 \]

\begin{align*}
  & x_1 \quad x_2 \quad x_3 \\
\end{align*}

\( H \) implements all dichotomies

\[ m_H(4) = 2^4 \]

some point will be inside a rectangle defined by others

\[ m_H(5) < 2^5 \]

We have not computed \( m_H(5) \) – not impossible, but tricky.

---

Example Growth Functions

<table>
<thead>
<tr>
<th>( N )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>\ldots</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D perceptron</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>\ldots</td>
<td></td>
</tr>
<tr>
<td>1-D pos. ray</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>\ldots</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>\ldots</td>
</tr>
</tbody>
</table>

- \( m_H(N) \) drops below \( 2^N \) – there is hope for the generalization bound.
- A break point is any \( n \) for which \( m_H(n) < 2^n \).

---

A Combinatorial Puzzle

\( X_1 \quad X_2 \quad X_3 \)

- ○ ○ ○
- ○ ○ ●
- ● ○ ○
- ● ● ●

A set of dichotomys

---

A Combinatorial Puzzle

\( X_1 \quad X_2 \quad X_3 \)

- ○ ○ ●
- ○ ● ○
- ● ● ○
- ○ ● ●

Two points are \textit{shattered}
A Combinatorial Puzzle

No pair of points is shattered

4 dichotomies is max.

If \( N = 4 \) how many possible dichotomies with no 2 points shattered?