**Review of Lecture 4**

- **Error measures**
  - User-specified \( e(h(x), f(x)) \)
  
  ![Fingerprint diagram](image)

  \[ E_{\text{in}}(h) = \frac{1}{N} \sum_{n=1}^{N} e(h(x_n), f(x_n)) \]

  - In-sample:
  - Out-of-sample

  \[ E_{\text{out}}(h) = \mathbb{E}_x[e(h(x), f(x))] \]

- **Noisy targets**

  \[ y = f(x) \quad \rightarrow \quad y \sim P(y | x) \]

  ![Diagram showing training examples and distributions](image)

  - Unknown target distribution
  - Target function \( f: X \rightarrow Y \) plus noise

  \[ P(x, y) = P(x)P(y|x) \]

  - \((x_1, y_1), \ldots, (x_N, y_N)\) generated by

  \[ E_{\text{out}}(h) \text{ is now } \mathbb{E}_{x,y}[e(h(x), y)] \]
Learning From Data

Yaser S. Abu-Mostafa
California Institute of Technology

Lecture 5: Training versus Testing
Outline

• From training to testing

• Illustrative examples

• Key notion: break point

• Puzzle
The final exam

Testing:

\[ \mathbb{P} \left[ |E_{\text{in}} - E_{\text{out}}| > \epsilon \right] \leq 2 e^{-2\epsilon^2 N} \]

Training:

\[ \mathbb{P} \left[ |E_{\text{in}} - E_{\text{out}}| > \epsilon \right] \leq 2M e^{-2\epsilon^2 N} \]
Where did the $M$ come from?

The $\mathcal{B}$ad events $\mathcal{B}_m$ are

$"|E_{\text{in}}(h_m) - E_{\text{out}}(h_m)| > \epsilon"$

The union bound:

$\mathbb{P}[^{\mathcal{B}_1 \text{ or } \mathcal{B}_2 \text{ or } \ldots \text{ or } \mathcal{B}_M}]$

$\leq \mathbb{P}[\mathcal{B}_1] + \mathbb{P}[\mathcal{B}_2] + \cdots + \mathbb{P}[\mathcal{B}_M]$

no overlaps: $M$ terms
Can we improve on \( M \)?

Yes, bad events are very overlapping!

\[ \Delta E_{\text{out}}: \text{change in } +1 \text{ and } -1 \text{ areas} \]

\[ \Delta E_{\text{in}}: \text{change in labels of data points} \]

\[ |E_{\text{in}}(h_1) - E_{\text{out}}(h_1)| \approx |E_{\text{in}}(h_2) - E_{\text{out}}(h_2)| \]
What can we replace $M$ with?

Instead of the whole input space,

we consider a finite set of input points,

and count the number of *dichotomies*.
Dichotomies: mini-hypotheses

A hypothesis \( h : \mathcal{X} \rightarrow \{-1, +1\} \)

A dichotomy \( h : \{x_1, x_2, \cdots, x_N\} \rightarrow \{-1, +1\} \)

Number of hypotheses \( |\mathcal{H}| \) can be infinite

Number of dichotomies \( |\mathcal{H}(x_1, x_2, \cdots, x_N)| \) is at most \( 2^N \)

Candidate for replacing \( M \)
The growth function

The growth function counts the most dichotomies on any $N$ points

$$m_H(N) = \max_{x_1, \ldots, x_N \in \mathcal{X}} |\mathcal{H}(x_1, \ldots, x_N)|$$

The growth function satisfies:

$$m_H(N) \leq 2^N$$

Let's apply the definition.
Applying $m_H(N)$ definition - perceptrons

$N = 3$

$m_H(3) = 8$

$N = 3$

$\begin{align*}
N = 4 \quad & m_H(4) = 14
\end{align*}$
Outline

- From training to testing
- Illustrative examples
- Key notion: break point
- Puzzle
Example 1: positive rays

\[ h(x) = \begin{cases} -1 & \text{if } x < a \\ +1 & \text{if } x \geq a \end{cases} \]

\[ m_\mathcal{H}(N) = N + 1 \]

\( \mathcal{H} \) is set of \( h : \mathbb{R} \rightarrow \{-1, +1\} \)

\[ h(x) = \text{sign}(x - a) \]
Example 2: positive intervals

\[ h(x) = -1 \quad h(x) = +1 \quad h(x) = -1 \]

\[ x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_N \]

\[ \mathcal{H} \text{ is set of } h: \mathbb{R} \rightarrow \{-1, +1\} \]

Place interval ends in two of \( N + 1 \) spots

\[ m_{\mathcal{H}}(N) = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1 \]
Example 3: convex sets

$\mathcal{H}$ is set of $h : \mathbb{R}^2 \rightarrow \{-1, +1\}$

$h(x) = +1$ is convex

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

The $N$ points are 'shattered' by convex sets
The 3 growth functions

- $\mathcal{H}$ is positive rays:
  
  $$m_{\mathcal{H}}(N) = N + 1$$

- $\mathcal{H}$ is positive intervals:
  
  $$m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$$

- $\mathcal{H}$ is convex sets:
  
  $$m_{\mathcal{H}}(N) = 2^N$$
Back to the big picture

Remember this inequality?

\[
P \left[ |E_{\text{in}} - E_{\text{out}}| > \epsilon \right] \leq 2M e^{-2\epsilon^2 N}
\]

What happens if \( m_\mathcal{H}(N) \) replaces \( M \)?

\( m_\mathcal{H}(N) \) polynomial \implies \text{Good!}

Just prove that \( m_\mathcal{H}(N) \) is polynomial?
Outline

• From training to testing

• Illustrative examples

• Key notion: break point

• Puzzle
Break point of $\mathcal{H}$

Definition:

If no data set of size $k$ can be shattered by $\mathcal{H}$, then $k$ is a break point for $\mathcal{H}$

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

For 2D perceptrons, $k = 4$

A bigger data set cannot be shattered either
Break point - the 3 examples

- Positive rays $m_{\mathcal{H}}(N) = N + 1$
  - break point $k = 2$

- Positive intervals $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$
  - break point $k = 3$

- Convex sets $m_{\mathcal{H}}(N) = 2^N$
  - break point $k = \infty$
Main result

No break point \( \implies m_H(N) = 2^N \)

Any break point \( \implies m_H(N) \) is polynomial in \( N \)
Puzzle

<table>
<thead>
<tr>
<th>X₁</th>
<th>X₂</th>
<th>X₃</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>