Outline

• What is overfitting?

• The role of noise

• Deterministic noise

• Dealing with overfitting
Illustration of overfitting

Simple target function

5 data points- noisy

4th-order polynomial fit

\[ E_{\text{in}} = 0, \quad E_{\text{out}} \text{ is huge} \]
Overfitting versus bad generalization

Neural network fitting noisy data

Overfitting: \( E_{\text{in}} \downarrow \quad E_{\text{out}} \uparrow \)
The culprit

**Overfitting**: “fitting the data more than is warranted”

**Culprit**: fitting the noise - **harmful**
Case study

10th-order target + noise

50th-order target
Two fits for each target

Noisy low-order target

<table>
<thead>
<tr>
<th></th>
<th>2nd Order</th>
<th>10th Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{\text{in}}$</td>
<td>0.050</td>
<td>0.034</td>
</tr>
<tr>
<td>$E_{\text{out}}$</td>
<td>0.127</td>
<td>9.00</td>
</tr>
</tbody>
</table>

Noiseless high-order target

<table>
<thead>
<tr>
<th></th>
<th>2nd Order</th>
<th>10th Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{\text{in}}$</td>
<td>0.029</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>$E_{\text{out}}$</td>
<td>0.120</td>
<td>7680</td>
</tr>
</tbody>
</table>
An irony of two learners

Two learners $O$ and $R$

They know the target is 10th order

$O$ chooses $H_{10}$ \hspace{1cm} $R$ chooses $H_{2}$

Learning a 10th-order target
We have seen this case

Remember learning curves?

$$\mathcal{H}_2$$

$$\mathcal{H}_{10}$$

Expected Error

Expected Error

$E_{\text{out}}$

$E_{\text{in}}$

Number of Data Points, $N$

Number of Data Points, $N$
Even without noise

The two learners $\mathcal{H}_{10}$ and $\mathcal{H}_2$

They know there is no noise.

Is there really no noise?

Learning a 50th-order target