Toy Example

$D_1$

weak classifiers = vertical or horizontal half-planes
Round 1

$h_1$

$\varepsilon_1 = 0.30$

$\alpha_1 = 0.42$

$D_2$
Round 2

$\alpha_2 = 0.65$

$\varepsilon_2 = 0.21$

$D_3$
Round 3

$\varepsilon_3 = 0.14$

$\alpha_3 = 0.92$
Final Classifier

\[ H_{\text{final}} = \text{sign} \left( \begin{array}{c} 0.42 \\ + 0.65 \\ + 0.92 \end{array} \right) \]
Actual Typical Run

- Test error does **not** increase, even after 1000 rounds
  - (total size > 2,000,000 nodes)
- Test error continues to drop even after training error is zero!

<table>
<thead>
<tr>
<th># rounds</th>
<th>5</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>train error</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>test error</td>
<td>8.4</td>
<td>3.3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

- Occam’s razor **wrongly** predicts “simpler” rule is better
Empirical Evidence: The Margin Distribution

- margin distribution
  = cumulative distribution of margins of training examples

![Graph showing margin distribution](image)

<table>
<thead>
<tr>
<th># rounds</th>
<th>5</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>train error</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>test error</td>
<td>8.4</td>
<td>3.3</td>
<td>3.1</td>
</tr>
<tr>
<td>% margins ≤ 0.5</td>
<td>7.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>minimum margin</td>
<td>0.14</td>
<td>0.52</td>
<td>0.55</td>
</tr>
</tbody>
</table>