Toy Example

$D_1$

weak classifiers = vertical or horizontal half-planes
Round 1

$\varepsilon_1 = 0.30$
$\alpha_1 = 0.42$

$\theta_1$

$D_2$

$\theta_2$
Round 2

\[ \alpha_2 = 0.65 \]

\[ \varepsilon_2 = 0.21 \]
$h_3$

$\varepsilon_3 = 0.14$

$\alpha_3 = 0.92$
\[ H_{\text{final}} = \text{sign} \left( \begin{array}{c} 0.42 \\ + 0.65 \\ + 0.92 \end{array} \right) \]
Actual Typical Run

(Boosting C4.5 on “letter” dataset)

- 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
The Margin Distribution

- margin distribution
  = distribution of margins of training examples

\[
\begin{array}{cccc}
\text{# rounds} & 5 & 100 & 1000 \\
\text{train error} & 0.0 & 0.0 & 0.0 \\
\text{test error} & 8.4 & 3.3 & 3.1 \\
\% \text{ margins} \leq 0.5 & 7.7 & 0.0 & 0.0 \\
\text{minimum margin} & 0.14 & 0.52 & 0.55 \\
\end{array}
\]
Application: Detecting Faces

- problem: find faces in photograph or movie
- weak classifiers: detect light/dark rectangles in image

• many clever tricks to make extremely fast and accurate