Lower bounds for Boosting with Hadamard Matrices

Jiazhong Nie
Manfred K. Warmuth
S.V. N. Vishwanathan
Xinhua Zhang

June 13, 2013
Protocol of Boosting

- Maintain distribution on $N \pm 1$ labeled examples
- At iteration $t = 1, \ldots, T$:
  - Receive “weak” hypothesis $h_t$ from oracle
  - Update $d_{t-1}$ to $d_t$: put more weights on “hard” examples
- Output convex combination of the $T$ weak hypotheses
Protocol of Boosting

- Maintain distribution on $N \pm 1$ labeled examples
- At iteration $t = 1, \ldots, T$:
  - Receive “weak” hypothesis $h_t$ from oracle
  - Update $d_{t-1}$ to $d_t$: put more weights on “hard” examples
- Output convex combination of the $T$ weak hypotheses

Goals:
- Final hypothesis highly accurate
- Small number of iterations
View Boosting as a zero-sum game

- Game matrix $U_t$ at iteration $t$:
  - columns are the $t$ hypotheses returned by the oracle

- A game between Row player $\mathbf{d}$ and Column player $\mathbf{w}$

- Value of the game $\text{val}(U_t) = \min_{\mathbf{d} \in S^n} \max_{\mathbf{w} \in S^t} \mathbf{d}^\top U_t \mathbf{w}$

<table>
<thead>
<tr>
<th>$\mathbf{w}_1$</th>
<th>$\mathbf{w}_2$</th>
<th>$\mathbf{w}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mathbf{d}_1$</th>
<th>$\mathbf{d}_2$</th>
<th>$\mathbf{d}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Examples

Hypotheses
Game value increases along iterations

A each iteration, a new hypothesis is added to the game matrix and the game value increases

<table>
<thead>
<tr>
<th>iteration 1</th>
<th>iteration 2</th>
<th>iteration 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>$u_1$ $u_2$</td>
<td>$u_1$ $u_2$ $u_3$</td>
</tr>
<tr>
<td>0</td>
<td>0 1</td>
<td>0 -1 1</td>
</tr>
<tr>
<td>1</td>
<td>1 0</td>
<td>1 0 -1</td>
</tr>
<tr>
<td>-1</td>
<td>-1 1</td>
<td>-1 1 0</td>
</tr>
</tbody>
</table>

$\text{val}(U_1) = -1$  
$\text{val}(U_2) = -1/3$  
$\text{val}(U_3) = 0$
\( \text{val}(U_t) \) converges to \( \text{val}(U) \) where \( U \) has all the hypothesis as columns

Gap measures the distance from the optimal learning accuracy
Bounds on the gap

- Upper bound: after $t$ iterations, Boosting algorithm chooses a $t$ columns/hypotheses submatrix $U_t$ that

$$\text{val}(U) - \text{val}(U_t) \leq O\left(\sqrt{\frac{\log N}{t}}\right)$$

[S92,F95,RW05]

- Matching lower bound against a weak oracle which
  - Iteratively returns a hypothesis of error $\epsilon$ w.r.t. $d_t$
  - Makes this go on for $\Omega\left(\frac{\log N}{\epsilon^2}\right)$ iterations.

[F95]
The **minimum gap**: the gap against the best choice of hypothesis

\[ \text{val}(U) - \max_{\text{val}(U_t)} \text{val}(U_t) \]

over all submat.

\( U_t \) of \( t \) cols
Lower bound against best choice of hypotheses

- **The minimum gap**: the gap against the best choice of hypothesis

  \[ \text{val}(U) - \max_{\text{over all submat.}} \text{val}(U_t) \]

  \[ \text{U}_t \text{ of } t \text{ cols} \]

- To prove a lower bound, find a ±1 matrix \( U \) whose minimum gap is \( \Omega(\sqrt{\frac{\log N}{t}}) \)
Lower bound against best choice of hypotheses

- **The minimum gap**: the gap against the best choice of hypothesis
  \[ \text{val}(U) - \max_{\text{over all submat.}} \text{val}(U_t) \]
  where \( U_t \) of \( t \) cols

- To prove a lower bound, find a \( \pm 1 \) matrix \( U \) whose **minimum gap** is \( \Omega(\sqrt{\frac{\log N}{t}}) \)

- **Previous result** [KY99]
  For random bit matrices \( U \), matching lower bound is proved with high probability when \( \log n \leq t \leq \sqrt{N} \).
Hadamard matrices

- $\pm 1$ orthogonal matrices of size $n = 2^k$

$$H^{(2)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H^{(2n)} = \begin{bmatrix} H^{(n)} & H^{(n)} \\ H^{(n)} & -H^{(n)} \end{bmatrix}$$
Game matrix

\[ H = \begin{bmatrix}
  1 & 1 & \ldots & 1 & 1 \\
  1 & -1 & \ldots & -1 & 1 \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  1 & -1 & \ldots & -1 & 1 \\
\end{bmatrix} \]
Game matrix

$$\hat{H} = \begin{bmatrix}
1 & 1 & \ldots & 1 & 1 \\
1 & -1 & \ldots & -1 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & -1 & \ldots & -1 & 1
\end{bmatrix}$$
\[
\hat{H} = \begin{bmatrix}
1 & 1 & \ldots & 1 & 1 \\
1 & -1 & \ldots & -1 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & -1 & \ldots & -1 & 1
\end{bmatrix}
\]

\[
-\hat{H} = \begin{bmatrix}
-1 & -1 & \ldots & -1 & -1 \\
-1 & 1 & \ldots & 1 & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-1 & 1 & \ldots & 1 & -1
\end{bmatrix}
\]
Game matrix

\[ U = \begin{pmatrix}
1 & 1 & \ldots & 1 & 1 \\
1 & -1 & \ldots & -1 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & -1 & \ldots & -1 & 1 \\
-1 & -1 & \ldots & -1 & -1 \\
-1 & 1 & \ldots & 1 & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-1 & 1 & \ldots & 1 & -1 \\
\end{pmatrix} \]
We can easily show that for $1 \leq t \leq N/2$,

$$\text{val}(U) - \max_{U_t} \text{val}(U_t) \geq \sqrt{\frac{1}{2t}}$$
Conjecture 1

\[ \text{val}(U) - \max_{U_t} \text{val}(U_t) \geq \sqrt{\frac{\log N}{2t}}, \]

for \( \log N \leq t \leq \text{const } N \)

Experimental evidence
Conjecture 2

For all the ±1 valued matrices, the \textbf{minimum gap} of game matrix $U = \begin{bmatrix} \hat{H} \\ -\hat{H} \end{bmatrix}$ decreases the slowest.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{chart.png}
\caption{Comparison of minimum gaps for different matrices.}
\end{figure}

\textbf{Evidence:}

- Verified by tedious combinatorial arguments for $N = 2, 4, 8$ and $t \leq N$ as well as for $N - 2 \leq t \leq N$. 