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 ± 1 labeled examples
- At iteration $t = 1, \ldots, T$:
  - Receive "weak" hypothesis h_t from oracle
  - Update $\mathbf{d}_{t-1}$ to $\mathbf{d}_t$: put more weights on "hard" examples
- Output convex combination of the T weak hypotheses

Protocol of Boosting

- Maintain distribution on N ± 1 labeled examples
- At iteration $t = 1, \ldots, T$:
  - Receive "weak" hypothesis h_t from oracle
  - Update $\mathbf{d}_{t-1}$ to $\mathbf{d}_t$: 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 [FS97]

- Game matrix $\underset{N\times t}{\mathbf{U}_t}$ 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 $\mathrm{val}(\mathbf{U}_t) = \min_{\mathbf{d}\in\mathcal{S}^n} \max_{\mathbf{w}\in\mathcal{S}^t} \mathbf{d}^\top \mathbf{U}_t\, \mathbf{w}$

		w_1	w_2	w_3
		.33	.33	.33
d_1	.33	0	1	-1
d_2	.33	-1	0	1
d_3	.33	1	-1	0

hypotheses (↓)

examples (→)

Game value increases along iterations

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

iteration 1

$\mathbf{u}_1$
0
1
-1

$\text{val}(\mathbf{U}_1) = -1$

iteration 2

$\mathbf{u}_1$	$\mathbf{u}_2$
0	1
1	0
-1	1

$\text{val}(\mathbf{U}_2) = -1/3$

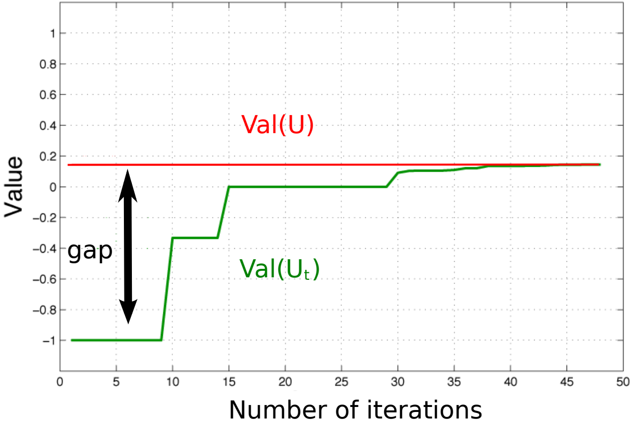
iteration 3

$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$
0	-1	1
1	0	-1
-1	1	0

$\text{val}(\mathbf{U}_3) = 0$

Gap of the game value

$\mathrm{val}(\mathbf{U}_t)$ converges to $\mathrm{val}(\mathbf{U})$ where $\mathbf{U}$ 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 $\mathbf{U}_t$ that

$$\mathrm{val}(\mathbf{U}) - \mathrm{val}(\mathbf{U}_t) \le O\left(\sqrt{\frac{\log N}{t}}\right)$$

[S92,F95,RW05]

- Matching lower bound against a **weak** oracle which [F95]
  - Iteratively returns a hypothesis of error ϵ w.r.t. $\mathbf{d}_t$
  - Makes this go on for $\Omega(\frac{\log N}{\epsilon^2})$ iterations.

Lower bound against best choice of hypotheses

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

$$\mathrm{val}(\mathbf{U}) - \max_{\substack{\text{over all submat.} \\ \mathbf{U}_t \text{ of } t \text{ cols}}} \mathrm{val}(\mathbf{U}_t)$$

Lower bound against best choice of hypotheses

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

$$\mathrm{val}(\mathbf{U}) - \max_{\substack{\text{over all submat.} \\ \mathbf{U}_t \text{ of } t \text{ cols}}} \mathrm{val}(\mathbf{U}_t)$$

- To prove a lower bound, find a ± 1 matrix $\mathbf{U}$ 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

$$\mathrm{val}(\mathbf{U}) - \max_{\substack{\text{over all submat.}\\ \mathbf{U}_t \text{ of } t \text{ cols}}} \mathrm{val}(\mathbf{U}_t)$$

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

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

Hadamard matrices

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

$$\mathbf{H}^{(2)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \mathbf{H}^{(2n)} = \begin{bmatrix} \mathbf{H}^{(n)} & \mathbf{H}^{(n)} \\ \mathbf{H}^{(n)} & -\mathbf{H}^{(n)} \end{bmatrix}$$

Game matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & -1 & \dots & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & \dots & -1 & 1 \end{bmatrix}$$

$$\hat{\mathbf{H}} = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & -1 & \dots & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & \dots & -1 & 1 \end{bmatrix}$$

Game matrix

$$\hat{\mathbf{H}} = \begin{bmatrix} \cancel{1} & \cancel{1} & \cancel{\ldots} & \cancel{1} & \cancel{1} \\ 1 & -1 & \ldots & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & \ldots & -1 & 1 \end{bmatrix}$$

$$-\hat{\mathbf{H}} = \begin{bmatrix} \cancel{-1} & \cancel{-1} & \cancel{\ldots} & \cancel{-1} & \cancel{-1} \\ -1 & 1 & \ldots & 1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 1 & \ldots & 1 & -1 \end{bmatrix}$$

Game matrix

$$
\mathbf{U} = \begin{bmatrix}
\cancel{1} & \cancel{1} & \cancel{\dots} & \cancel{1} & \cancel{1} \\
1 & -1 & \dots & -1 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & -1 & \dots & -1 & 1 \\
 & & & & \\
\cancel{-1} & \cancel{-1} & \cancel{\dots} & \cancel{-1} & \cancel{-1} \\
-1 & 1 & \dots & 1 & -1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
-1 & 1 & \dots & 1 & -1
\end{bmatrix}
$$

Lower bound with Hadamard matrices

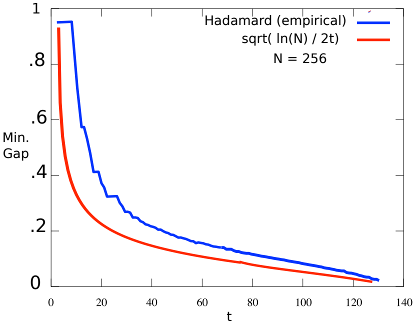
- We can easily to show that for $1 \leq t \leq N/2$,

$$\mathrm{val}(\mathbf{U}) - \max_{\mathbf{U}_t} \mathrm{val}(\mathbf{U}_t) \;\geq\; \sqrt{\frac{1}{2t}}$$

Conjecture 1

$$\mathrm{val}(\mathbf{U}) - \max_{\mathbf{U}_t} \mathrm{val}(\mathbf{U}_t) \;\geq\; \sqrt{\frac{\log N}{2t}}, \qquad \text{for } \log N \leq t \leq \text{const } N$$

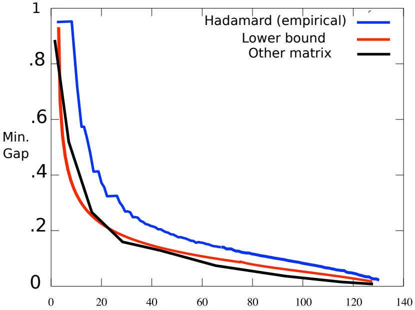
Experimental evidence



Conjecture 2

For all the ± 1 valued matrices,

the **minimum gap** of game matrix $\mathbf{U} = \begin{bmatrix} \hat{\mathbf{H}} \\ -\hat{\mathbf{H}} \end{bmatrix}$ decreases the slowest



Evidence:

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