Boosting Algorithms for Maximizing the Soft Margin

Manfred Warmuth¹, Karen Glocer¹, Gunnar Rätsch²

1: UC Santa Cruz
2: Friedrich Miescher Laboratory of the Max Planck Society
manfred,kag@soe.ucsc.edu, gunnar.raetsch@tuebingen.mpg.de

1. Introduction

- Boosting algorithm when there is no consistent convex combination of base hypotheses
- In LPBoost, SoftBoost, iterations produces a convex combination with soft margin within ε of the maximum

Boosting protocol:
- Set of examples $S = ((x_1, y_1), \ldots, (x_N, y_N))$
- Maintains distribution $d$ on examples
- At iteration $t$:
  - Given current distribution $d^{(t)}$, oracle provides hypothesis $h_t$ of edge $\gamma_t = d^{(t)} - d$, where $h'_t = w_A(x_t)$
  - Guarantee $\gamma_t > 0$ known to algorithm
  - Update distribution $d^{(t+1)}$ to $d^{(t)}$

LPBoost computes $d^*$ by solving:

Primal

\begin{align*}
\min \gamma \quad \text{s.t. } d^{(t)} \leq \gamma, & \leq m 0 \leq \gamma \text{ is within } \epsilon \text{ of the optimum}\n\text{max } \sum d(w(x_t)), & \text{maximize minimum soft margin}
\end{align*}

Dual

\begin{align*}
\max \sum \rho + \sum \rho \leq \epsilon \text{ s.t. } & \text{minimize maximum edge}
\end{align*}

Non-standard LPBoost formulation
- Totally correctable
- Capping probabilities in primal $\rightarrow$ soft margin in dual

2. LPBoost does not have O(\ln N) iteration bounds.

LPBoost (Schuurmans et al) works well in practice
- No bounds have been proved for it
- In our counter examples LPBoost takes (\ln N) iterations to achieve margin precision $\delta$ for separable case.
- Forces LPBoost to concentrate its distribution on single example
- Holds regardless of LP optimization algorithm
- Shows need for regularization

\begin{align*}
\text{Algorithm 1: SoftBoost } & \text{ shown in figure } 2
\end{align*}

3. SoftBoost

The counter example suggests that a good algorithm should:
- Cap the weight on any example
- Spread the weight on the examples via a regularization similar to the relative entropy

These two tricks used by the SoftBoost algorithm make it possible to obtain iteration bounds that grow logarithmic in $N$.

- Designed for data that is not necessarily separable by convex combinations of base hypotheses
- Achieves robustness by capping the the weight on any example to be at most $\epsilon$
- Capping the weights on the examples prevents the algorithm from focusing excessively on a few examples that it can't hope to get right
- Produces a convex combination of hypotheses whose soft margin is within $\epsilon$ of the optimum
- SoftBoost terminates after at most $\lceil \ln (N/\epsilon) \rceil$ iterations.
- The algorithm does not need to know the guarantee $\epsilon$ on the base hypotheses

Convergence Speed Comparison

- SoftBoost starts more slowly than LPBoost
- Both converge to within $\epsilon$ of guarantee $\epsilon$ in approximately the same number of iterations
- BrownBoost, which was designed to deal with noisy data but is not a smooth boosting algorithm, does not maximize the soft margin
- SmoothBoost, the best of previously existing smoothing boosting algorithms, converges much more slowly and does not achieve the optimal soft margin

<table>
<thead>
<tr>
<th>UCI Benchmark</th>
<th>AdaBoost</th>
<th>LPBoost</th>
<th>SoftBoost</th>
<th>BrownBoost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult</td>
<td>11.1 ± 0.8</td>
<td>11.1 ± 0.8</td>
<td>12.9 ± 1.0</td>
<td>12.9 ± 0.7</td>
</tr>
<tr>
<td>B.Cancer</td>
<td>32.1 ± 3.8</td>
<td>27.8 ± 4.3</td>
<td>28.0 ± 4.5</td>
<td>30.2 ± 3.9</td>
</tr>
<tr>
<td>Diabetes</td>
<td>27.9 ± 1.5</td>
<td>24.4 ± 1.7</td>
<td>24.4 ± 1.7</td>
<td>27.2 ± 1.6</td>
</tr>
<tr>
<td>German</td>
<td>26.9 ± 1.9</td>
<td>24.6 ± 2.1</td>
<td>24.7 ± 2.1</td>
<td>24.8 ± 1.9</td>
</tr>
<tr>
<td>Heart</td>
<td>20.1 ± 2.7</td>
<td>18.4 ± 3.0</td>
<td>18.2 ± 2.7</td>
<td>20.0 ± 2.8</td>
</tr>
<tr>
<td>Ringnorm</td>
<td>1.9 ± 0.0</td>
<td>1.9 ± 0.2</td>
<td>1.8 ± 0.2</td>
<td>1.9 ± 0.2</td>
</tr>
<tr>
<td>F.Solar</td>
<td>36.1 ± 1.5</td>
<td>35.7 ± 1.6</td>
<td>35.5 ± 1.4</td>
<td>36.1 ± 1.4</td>
</tr>
<tr>
<td>Thyroid</td>
<td>4.4 ± 1.9</td>
<td>4.9 ± 1.9</td>
<td>4.9 ± 1.9</td>
<td>4.6 ± 2.1</td>
</tr>
<tr>
<td>Tictec</td>
<td>22.8 ± 1.0</td>
<td>22.8 ± 1.0</td>
<td>23.0 ± 0.8</td>
<td>22.8 ± 0.8</td>
</tr>
<tr>
<td>Wavform</td>
<td>10.5 ± 0.4</td>
<td>10.1 ± 0.5</td>
<td>9.8 ± 0.5</td>
<td>10.4 ± 0.4</td>
</tr>
</tbody>
</table>

- Generalization error estimates and standard deviations for ten UCI benchmark data sets
- SoftBoost and LPBoost outperform AdaBoost and BrownBoost on most data sets
- SoftBoost and LPBoost perform similarly