Boost-by-Majority

COMS 4995-1 Spring 2020 (Daniel Hsu)
Boosting game

- Boosting game takes place over multiple rounds
- Goal of Booster is to get hypotheses $h_1, \ldots, h_T$ from WLer s.t.

$$
\sum_{t=1}^{T} y_i h_t(x_i) > 0 \quad \forall i = 1, \ldots, n.
$$

- Let $z_{t,i} := y_i h_t(x_i) \in \{-1, 1\}$ for all $i$, and

$$
s_t := z_1 + \cdots + z_t.
$$

- So goal of Booster is to get $s_{T,i} > 0$ for all $i$. 
Initialize: $s_0 := 0 \in \mathbb{R}^n$

For $t = 1, 2, \ldots, T$:

- Booster choose $p_t \in \Delta([n])$
- WLer “picks” $z_t \in \{\pm 1\}^n$ such that $p_t^Tz_t \geq \theta$.
- $s_t := s_{t-1} + z_t$

Booster wins if $s_{T,i} > 0$ for all $i$
- I.e., if $\Lambda_T(s_T) := \sum_{i=1}^{n} \mathbb{1}\{s_{T,i} \leq 0\} < 1$

How large should $T$ be, and how should Booster choose $p_t$’s?
Last round of the game

- Suppose $s$ is the state after first $T - 1$ rounds. How should Booster choose $p_T$ in round $T$?
  - Min-max strategy: assume WLer will choose worst $z_T$, so state becomes $s + z_T$, and final loss is $\Lambda_T(s + z_T)$
  - Booster should choose $p_T$ to achieve $\min$ below:

  $$\Lambda_{T-1}(s) := \min_{p \in \Delta([n])} \max_{z \in \{-1, 1\}^n} \Lambda_T(s + z).$$

  - Semantics of $\Lambda_{T-1}(s)$: the loss of the Booster if state after round $T - 1$ is $s$ and both Booster and WLer play optimally in round $T$

- General case: $\Lambda_t(s)$ is the loss of the Booster if state after round $t$ is $s$ and both Booster and WLer play optimally in rounds $t + 1, \ldots, T$
  - By same argument:

  $$\Lambda_{t-1}(s) = \min_{p \in \Delta([n])} \max_{z \in \{-1, 1\}^n} \Lambda_t(s + z).$$
Sequential min-max optimal boosting

- Game starts in state $s_0 = 0$
- Loss achievable by Booster:

\[
\min_{p_1} \max_{z_1 \in \{\pm 1\}^n: p_1 z_1 \geq \theta} \min_{p_2} \max_{z_2 \in \{\pm 1\}^n: p_2 z_2 \geq \theta} \cdots \min_{p_T} \max_{z_T \in \{\pm 1\}^n: p_T z_T \geq \theta} \Lambda_T \left( \sum_{t=1}^T z_t \right).
\]

- But this seems intractable to compute and reason about
- Instead, we will develop a tractable upper bound on $\Lambda_t$ that (amazingly) decomposes over $i = 1, \ldots, n$.
  - This will also suggest an efficient algorithm, “Boost-by-Majority” (BBM) due to Freund.
  - Our analysis will follow the “Drifting Games” analysis of Schapire.
Define $\phi_T(s_i) := \mathbb{1}\{s_i \leq 0\}$, so for round $T$, we have

$$\Lambda_T(s) = \sum_{i=1}^{n} \phi_T(s_i).$$

Assuming $\phi_t$ has been defined for some $t \geq 1$, let

$$\phi_{t-1}(s_i) := \min_{q_i \geq 0} \max_{z_i \in \{\pm 1\}} \phi_t(s_i + z_i) + q_i(z_i - \theta).$$

Claim: For all $t = 0, \ldots, T$,

$$\Lambda_t(s) \leq \sum_{i=1}^{n} \phi_t(s_i), \quad \text{for all } s \in \mathbb{R}^n.$$
Proof of upper bound

\[
\Lambda_{t-1}(s) = \min_{p \in \Delta([n])} \max_{z \in \{\pm 1\}^n} \Lambda_t(s + z)
\]

\[
= \min_{p \in \Delta([n])} \max_{z \in \{\pm 1\}^n} \min_{\lambda \geq 0} \Lambda_t(s + z) + \lambda(p^T z - \theta)
\]

\[
\leq \min_{p \in \Delta([n])} \min_{\lambda \geq 0} \max_{z \in \{\pm 1\}^n} \Lambda_t(s + z) + \lambda(p^T z - \theta)
\]

\[
= \min_{q \in \mathbb{R}_+^n} \max_{z \in \{\pm 1\}^n} \Lambda_t(s + z) + \sum_{i=1}^{n} q_i(z_i - \theta)
\]

\[
\leq \min_{q \in \mathbb{R}_+^n} \max_{z \in \{\pm 1\}^n} \sum_{i=1}^{n} \phi_t(s_i) + q_i(z_i - \theta)
\]

\[
= \sum_{i=1}^{n} \min_{q_i \geq 0} \max_{z_i \in \{\pm 1\}} \phi_t(s_i + z_i) + q_i(z_i - \theta)
\]

\[
= \sum_{i=1}^{n} \phi_{t-1}(s_i)
\]
Achieving the bound

▶ Since the upper bound on $\Lambda_t$ decomposes over $i$, it is easy to achieve the bound (i.e., find the minimizing $q_i$)

▶ Recall

$$\phi_{t-1}(s_i) = \min_{q_i \geq 0} \max_{z_i \in \{\pm 1\}} \phi_t(s_i + z_i) + q_i(z_i - \theta).$$

▶ Objective to minimize is the maximum of two linear functions, one with positive slope and the other with negative slope

▶ Claim: $\phi_t(s_i + 1) \leq \phi_t(s_i - 1)$ (proof is by backward induction)

▶ Given this claim, we find that the minimizing $q_i$ is

$$q_{t,i} := \frac{\phi_t(s_i - 1) - \phi_t(s_i + 1)}{2},$$

and the value achieved is

$$\phi_{t-1}(s_i) = \frac{1 + \theta}{2} \phi_t(s_i + 1) + \frac{1 - \theta}{2} \phi_t(s_i - 1).$$

▶ So Booster should choose $p_{t,i} \propto q_{t,i}$ defined above (at $s = s_{t-1}$). (This is the “Boost-by-Majority” algorithm.)
Explicit form of the upper bound

- Solving recurrence $\phi_{t-1}(s_i) = \frac{1+\theta}{2} \phi_t(s_i + 1) + \frac{1-\theta}{2} \phi_t(s_i - 1)$
with $\phi_T(s_i) = 1\{s_i \leq 0\}$ yields

$$\phi_t(s_i) = \text{BinomialCDF} \left( \frac{T - t - s_i}{2} \middle| T - t, \frac{1 + \theta}{2} \right)$$

$$= \sum_{k=0}^{\left\lfloor \frac{T-t-s_i}{2} \right\rfloor} \binom{T-t}{k} \left( \frac{1 + \theta}{2} \right)^k \left( \frac{1 - \theta}{2} \right)^{T-t-k}.$$
Monotonicity of the upper bound

Claim: Pick any \( t = 1, \ldots, T \). Let \( s \in \mathbb{R}^n \) be the state after \( t - 1 \) rounds. Assume the Booster chooses \( p_t \) (through \( q_t \)) to achieve the the minimum value in the definition of \( \phi_{t-1}(s_i) \). Then for any \( z_t \in \{\pm 1\}^n \) such that \( p_t^T z_t \geq \theta \),

\[
\sum_{i=1}^n \phi_t(s_i + z_t,i) \leq \sum_{i=1}^n \phi_{t-1}(s_i).
\]

Proof: Recall that

\[
\phi_{t-1}(s_i) = \min_{q_i \geq 0} \max_{z_i \in \{\pm 1\}} \phi_t(s_i + z_i) + q_i(z_i - \theta).
\]

Since \( p_t \) is chosen (via choice of \( q_t \)) to achieve the min,

\[
\sum_i \phi_{t-1}(s_i) = \sum_i \max_{z_i \in \{\pm 1\}} \phi_t(s_i + z_i) + q_{t,i}(z_i - \theta)
\]

\[
\geq \sum_i \phi_t(s_i + z_{t,i}) + q_{t,i}(z_{t,i} - \theta)
\]

\[
\geq \sum_i \phi_t(s_i + z_{t,i}),
\]

where the last inequality uses the constraint \( p_t^T z_t \geq \theta \).
By monotonicity property, in the sequence of states $s_0, s_1, \ldots, s_T$ actually encountered by running BBM, we have

$$\sum_{i=1}^{n} \phi_T(s_T, i) \leq \sum_{i=1}^{n} \phi_{T-1}(s_{T-1}, i) \leq \cdots \leq \sum_{i=1}^{n} \phi_0(s_0, i) = n \cdot \phi_0(0).$$

So,

$$\sum_{i=1}^{n} 1\{s_T, i \leq 0\} \leq n \cdot \text{BinomialCDF} \left( \frac{T}{2} \left| T, \frac{1 + \theta}{2} \right. \right) \leq ne^{-\theta^2T/2}.$$

Conclusion: BBM requires $T = \frac{2\ln n}{\theta^2}$ rounds to have a majority-vote classifier that correctly classifies all training examples.
Exponential weights variant of BBM

- Use $\phi_T(s_i) = e^{-\eta s_i}$ for some $\eta > 0$.
  - This is an upper bound on $\mathbb{1}\{s_i \leq 0\}$.
  - Analysis is exactly the same except for the final conclusion.
  - Solution to backwards recurrence defining $\phi_t$ is very simple:

$$
\phi_t(s_i) = \left( \frac{1 + \theta}{2} e^{-\eta} + \frac{1 - \theta}{2} e^{\eta} \right)^{T-t} e^{-\eta s_i}.
$$

- Booster’s distribution over examples:

$$
\rho_{t,i} \propto e^{-\eta s_{t-1,i}}.
$$

- Final conclusion:

$$
\sum_{i=1}^{n} \mathbb{1}\{s_{T,i} \leq 0\} \leq \sum_{i=1}^{n} e^{-\eta s_{T,i}} \leq \cdots \leq n \left( \frac{1 + \theta}{2} e^{-\eta} + \frac{1 - \theta}{2} e^{\eta} \right)^T.
$$

- Choose $\eta = \frac{1}{2} \ln \frac{1+\theta}{1-\theta}$ to minimize the bound.
- Just need $T = \frac{2 \ln(n)}{\theta^2}$ rounds.