Boosting as Frank-Wolfe

Ryotaro Mitsuboshi[1,2] Kohei Hatano[1,2] Eiji Takimoto[1]

[1]Kyushu University

[2]Riken AIP

OPT 2022

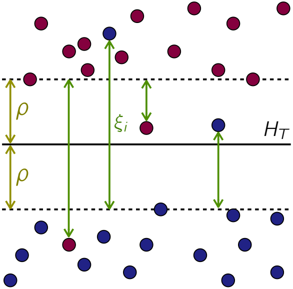
Soft margin optimization

- **Input:** $S = ((\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_m, y_m)) \in (\mathcal{X} \times \{\pm 1\})^m$.
- **Output:** A combined hypothesis $H_T = \sum_{h \in \mathcal{H}} \bar{w}_h h$, where $\bar{\boldsymbol{w}}$ is an optimal solution of:

$$\max_{\rho, \boldsymbol{w}, \boldsymbol{\xi}} \quad \rho - \frac{1}{\nu} \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad y_i \sum_{h \in \mathcal{H}} w_h h(\boldsymbol{x}_i) \geq \rho - \xi_i, \quad i \in [m],$$

$$\sum_{h \in \mathcal{H}} w_h = 1, \boldsymbol{w} \geq \mathbf{0}, \boldsymbol{\xi} \geq \mathbf{0}.$$



- A linear program.
- Hard for off-the-shelf solvers when $\mathcal{H}$ is a huge set.
- Boosting is a standard approach for such a situation.

Boosting

- Solves the dual problem of soft margin optimization:

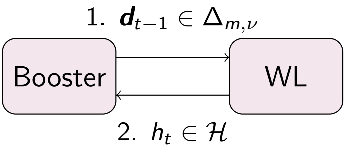
$$\min_{\boldsymbol{d} \in \Delta_{m,\nu}} \max_{h \in \mathcal{H}} (\boldsymbol{d}^\top A)_h, \quad \Delta_{m,\nu} = \{\boldsymbol{d} \in [0, 1/\nu]^m \mid \|\boldsymbol{d}\|_1 = 1\}.$$

- Boosting is a protocol between *Booster* and *Weak Learner (WL)*.

In each step $t = 1, 2, \ldots, T$,

1. Booster sends $\boldsymbol{d}_{t-1}$ to WL.
2. Booster obtains a hypothesis $h_t \in \mathcal{H}$ from WL.
3. Booster updates the distribution $\boldsymbol{d}_t \in \Delta_{m,\nu}$ over training instances.

Output $H_T = \sum_{t=1}^{T} w_t h_t$.



Related works

	LPBoost	ERLPBoost	C-ERLPBoost
Rounds	$\Omega(m)$	$O(\frac{1}{\epsilon^2} \ln \frac{m}{\nu})$	$O(\frac{1}{\epsilon^2} \ln \frac{m}{\nu})$
Sub-problem	LP	CP	LP

- LPBoost is practical but takes $\Omega(m)$ rounds for the worst case.
- ERLPBoost has a favorable bound
  but involves a convex program per round. $\Longrightarrow$ Slower than LPBoost.
- C-ERLPBoost has the same bound and solves LP, but it takes many rounds, so it is slower than ERLPBoost.

Our objective

Find a practical boosting algorithm that has a theoretical guarantee.

Contributions

Two contributions.

1. A unified view of the boosting algorithms.
   - LPBoost, ERLPBoost, and C-ERLPBoost are instances of the Frank-Wolfe algorithm.
2. Propose a "boosting scheme".
   - ERLPBoost and C-ERLPBoost are the instances of this scheme.
   - Terminates in $O(\frac{1}{\epsilon^2} \ln \frac{m}{\nu})$ rounds.
   - One can incorporate any heuristic algorithm into this scheme.

Experiments