Algorithms and Hardness for Learning Linear Thresholds from Label Proportions

Rishi Saket

Google Research India
Bangalore, India

NeurIPS 2022

Learning from Label Proportions (LLP)

- Feature-vector space $\mathscr{X} = \mathbb{R}^d$, $f : \mathscr{X} \to \{0,1\}$.
- Define *label proportion* $\sigma(B,f) \coloneqq \text{Avg}\{f(\mathbf{x}) : \mathbf{x} \in B\}$ for *bag* $B \subseteq \mathscr{X}$
- Training examples $(B, \sigma(B,f))$, goal is to train h consistent with f.
- $h : \mathscr{X} \to \{0,1\}$ *satisfies* B if $\sigma(B,h) = \sigma(B,f)$

Learning from Label Proportions (LLP)

- Feature-vector space $\mathscr{X} = \mathbb{R}^d$, $f : \mathscr{X} \to \{0,1\}$.
- Define *label proportion* $\sigma(B,f) \coloneqq \text{Avg}\{f(\mathbf{x}) : \mathbf{x} \in B\}$ for *bag* $B \subseteq \mathscr{X}$
- Training examples $(B, \sigma(B,f))$, goal is to train h consistent with f.
- $h : \mathscr{X} \to \{0,1\}$ *satisfies* B if $\sigma(B,h) = \sigma(B,f)$

Goal: Given $(B_k, \sigma(B_k,f))$ sampled from some distribution, $(k=1,...,m)$ find hypothesis
$h : \mathscr{X} \to \{0,1\}$ maximizing # satisfied bags B_k.

Learning from Label Proportions (LLP)

- Feature-vector space $\mathscr{X} = \mathbb{R}^d$, $f : \mathscr{X} \rightarrow \{0,1\}$.
- Define *label proportion* $\sigma(B,f) \coloneqq \text{Avg}\{f(\mathbf{x}) : \mathbf{x} \in B\}$ for *bag* $B \subseteq \mathscr{X}$
- Training examples $(B, \sigma(B,f))$, goal is to train h consistent with f.
- $h : \mathscr{X} \rightarrow \{0,1\}$ *satisfies* B if $\sigma(B,h) = \sigma(B,f)$

Goal: Given $(B_k, \sigma(B_k,f))$ sampled from some distribution, $(k=1,...,m)$ find hypothesis
$h : \mathscr{X} \rightarrow \{0,1\}$ maximizing # satisfied bags B_k.

Our focus: When the target concept f is a linear threshold function (LTF) or halfspace.

- $f = \text{pos}(\langle \mathbf{r}, \mathbf{x} \rangle + c)$ where $\text{pos}(a) = 1$ if $a > 0$, 0 otherwise.

Previous Work

[Saket, NeurIPS'21]: Given $(\{(B_k, \sigma(B_k,f))\} : k = 1,...,m)$ s.t. $|B_k| \leq 2$, f is unknown LTF:

- Efficient algorithm that finds an LTF satisfying ⅖ fraction of all the bags.
- NP-hard to find any fn. of constantly many LTFs satisfying $(\frac{1}{2} + \delta)$-frac. of the bags.

Previous Work

[Saket, NeurIPS'21]: Given $(\{(B_k, \sigma(B_k, f))\} : k = 1,...,m)$ s.t. $|B_k| \leq 2$, f is unknown LTF:

- Efficient algorithm that finds an LTF satisfying ⅖ fraction of all the bags.
- NP-hard to find any fn. of constantly many LTFs satisfying $(\frac{1}{2} + \delta)$-frac. of the bags.

Is there algorithm satisfying $\Omega(1)$-fraction of bags of size > 2 ?

Our Contributions

Given $(\{(B_k, \sigma(B_k,f))\} : k = 1,...,m)$ s.t. $|B_k| \leq q$, f is unknown LTF:

Our Contributions

Given $(\{(B_k, \sigma(B_k,f))\} : k = 1,...,m)$ s.t. $|B_k| \leq q$, f is unknown LTF:

Algorithms:

- If $q = 3$, an efficient algorithm to satisfy at least (1/12)-fraction of the bags.

Our Contributions

Given $(\{(B_k, \sigma(B_k, f))\} : k = 1,...,m)$ s.t. $|B_k| \leq q$, f is unknown LTF:

Algorithms:

- If $q = 3$, an efficient algorithm to satisfy at least (1/12)-fraction of the bags.
- For $q > 3$, an efficient algorithm to *weakly-satisfy* $\mathbf{\Omega}(1/q)$-fraction of bags.

Our Contributions

Given $(\{(B_k, \sigma(B_k, f))\} : k = 1,...,m)$ s.t. $|B_k| \le q$, f is unknown LTF:

Algorithms:

- If $q = 3$, an efficient algorithm to satisfy at least (1/12)-fraction of the bags.
- For $q > 3$, an efficient algorithm to *weakly-satisfy* $\mathbf{\Omega}(1/q)$-fraction of bags.

Hardness: NP-hard to find any function of constantly many LTFs that

- satisfies $(1/q + \delta)$-fraction of bags for any constant $q \in \mathbb{Z}^+$,
- satisfies $(4/9 + \delta)$-fraction of bags for $q = 2$.

for any constant $\delta > 0$.

SDP of [Saket, NeurIPS'21] for $q = 2$:

We can assume that the satisfying LTF is $\text{pos}(\langle \mathbf{r}_*, \mathbf{x} \rangle)$ with non-zero margin.

For bag $B = \{\mathbf{x}_1, \mathbf{x}_2\}$: $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_2 \rangle \leq 0$ if B is non-monochromatic.

SDP of [Saket, NeurIPS'21] for $q = 2$:

We can assume that the satisfying LTF is $\text{pos}(\langle \mathbf{r}_*, \mathbf{x} \rangle)$ with non-zero margin.

For bag $B = \{\mathbf{x}_1, \mathbf{x}_2\}$: $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_2 \rangle \leq 0$ if B is non-monochromatic.

With $\mathbf{r}_*(\mathbf{r}_*)^T$ as a soln. write the feasible SDP for symmetric psd $\mathbf{R}$:

$$(\mathbf{x}_1)^T\mathbf{R}\,\mathbf{x}_2 \leq 0 \text{ for all non-mon. bags B} \quad \& \quad (\mathbf{x}_i)^T\mathbf{R}\,\mathbf{x}_i > 0 \text{ for all } \mathbf{x}_i.$$

SDP of [Saket, NeurIPS'21] for q = 2:

We can assume that the satisfying LTF is pos($\langle \mathbf{r}_*, \mathbf{x} \rangle$) with non-zero margin.

For bag B = $\{\mathbf{x}_1, \mathbf{x}_2\}$: $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_2 \rangle \leq 0$ if B is non-monochromatic.

With $\mathbf{r}_*(\mathbf{r}_*)^T$ as a soln. write the feasible SDP for symmetric psd $\mathbf{R}$:

$$(\mathbf{x}_1)^T\mathbf{R}\,\mathbf{x}_2 \leq 0 \text{ for all non-mon. bags B} \quad \& \quad (\mathbf{x}_i)^T\mathbf{R}\,\mathbf{x}_i > 0 \text{ for all } \mathbf{x}_i.$$

Factor $\mathbf{R} = \mathbf{L}^T\mathbf{L}$. Rounding based on sign of $\langle \mathbf{L}\mathbf{x}_i, \mathbf{g} \rangle$ for random gaussian vector $\mathbf{g}$.

SDP of [Saket, NeurIPS'21] for $q = 2$:

We can assume that the satisfying LTF is $\text{pos}(\langle \mathbf{r}^*, \mathbf{x} \rangle)$ with non-zero margin.

For bag $B = \{\mathbf{x}_1, \mathbf{x}_2\}$: $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_2 \rangle \leq 0$ if B is non-monochromatic.

With $\mathbf{r}_*(\mathbf{r}_*)^T$ as a soln. write the feasible SDP for symmetric psd $\mathbf{R}$:

$$(\mathbf{x}_1)^T \mathbf{R}\, \mathbf{x}_2 \leq 0 \text{ for all non-mon. bags B} \quad \& \quad (\mathbf{x}_i)^T \mathbf{R}\, \mathbf{x}_i > 0 \text{ for all } \mathbf{x}_i.$$

Factor $\mathbf{R} = \mathbf{L}^T\mathbf{L}$. Rounding based on sign of $\langle \mathbf{L}\mathbf{x}_i, \mathbf{g} \rangle$ for random gaussian vector $\mathbf{g}$.

Problem: For $q = 3$: the sign of $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_2 \rangle$ not determined by the label proportion for non-monochromatic bags.

Our novel SDP for q = 3:

Key idea: Use relaxations of **R** using $\succcurlyeq$ partial ordering (*Loewner* order).

Our novel SDP for q = 3:

Key idea: Use relaxations of **R** using $\succcurlyeq$ partial ordering (*Loewner* order).

Observation: For a non-monochromatic bag B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

at least one of $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_2\rangle$ or $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_3\rangle$ is negative.

Our novel SDP for q = 3:

Key idea: Use relaxations of **R** using $\succcurlyeq$ partial ordering (*Loewner* order).

Observation: For a non-monochromatic bag B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

at least one of $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_2\rangle$ or $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_3\rangle$ is negative.

$\mathbf{R}^{\{1,j\}} =$ $\mathbf{R} := \mathbf{r}_*(\mathbf{r}_*)^T$ if $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_j\rangle < 0$ and $\mathbf{0}$ o/w., is a feasible soln to:

$(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\mathbf{x}_2 \leq 0$, $(\mathbf{x}_1)^T\mathbf{R}^{\{1,3\}}\mathbf{x}_3 \leq 0$, $(\mathbf{x}_1)^T(\mathbf{R}^{\{1,2\}} + \mathbf{R}^{\{1,3\}})\mathbf{x}_1 \geq (\mathbf{x}_1)^T\mathbf{R}\mathbf{x}_1$, $\mathbf{R} \succcurlyeq \mathbf{R}^{\{1,j\}}$ for j=2,3

$\forall$ non-monochromatic bags B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$.

Our novel SDP for q = 3:

Key idea: Use relaxations of **R** using $\succcurlyeq$ partial ordering (*Loewner* order).

Observation: For a non-monochromatic bag B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

at least one of $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_2\rangle$ or $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_3\rangle$ is negative.

$\mathbf{R}^{\{1,j\}} = \mathbf{R} := \mathbf{r}_*(\mathbf{r}_*)^T$ if $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle\langle \mathbf{r}_*, \mathbf{x}_j\rangle < 0$ and $\mathbf{0}$ o/w., is a feasible soln to:

$(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\mathbf{x}_2 \leq 0$, $(\mathbf{x}_1)^T\mathbf{R}^{\{1,3\}}\mathbf{x}_3 \leq 0$, $(\mathbf{x}_1)^T(\mathbf{R}^{\{1,2\}} + \mathbf{R}^{\{1,3\}})\mathbf{x}_1 \geq (\mathbf{x}_1)^T\mathbf{R}\mathbf{x}_1$, $\mathbf{R} \succcurlyeq \mathbf{R}^{\{1,j\}}$ for j=2,3

$\forall$ non-monochromatic bags B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$.

Rounding: WLOG $(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\mathbf{x}_1 \geq (\mathbf{x}_1)^T\mathbf{R}\mathbf{x}_1 / 2$. Factor $\mathbf{R} = \mathbf{L}^T\mathbf{L}$.

Our novel SDP for q = 3:

Key idea: Use relaxations of **R** using $\succcurlyeq$ partial ordering (*Loewner* order).

Observation: For a non-monochromatic bag B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

at least one of $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_2 \rangle$ or $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_3 \rangle$ is negative.

$\mathbf{R}^{\{1,j\}} =$ $\mathbf{R} := \mathbf{r}_*(\mathbf{r}_*)^T$ if $\langle \mathbf{r}_*, \mathbf{x}_1 \rangle \langle \mathbf{r}_*, \mathbf{x}_j \rangle < 0$ and **0** o/w., is a feasible soln to:

$(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}} \mathbf{x}_2 \leq 0$, $(\mathbf{x}_1)^T\mathbf{R}^{\{1,3\}} \mathbf{x}_3 \leq 0$, $(\mathbf{x}_1)^T(\mathbf{R}^{\{1,2\}} + \mathbf{R}^{\{1,3\}})\mathbf{x}_1 \geq (\mathbf{x}_1)^T\mathbf{R}\mathbf{x}_1$, $\mathbf{R} \succcurlyeq \mathbf{R}^{\{1,j\}}$ for j=2,3

$\forall$ non-monochromatic bags B = $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$.

Rounding: WLOG $(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\mathbf{x}_1 \geq (\mathbf{x}_1)^T\mathbf{R}\mathbf{x}_1 / 2$. Factor $\mathbf{R} = \mathbf{L}^T\mathbf{L}$.

Can we show that $\angle\, \mathbf{L}\mathbf{x}_1, \mathbf{L}\mathbf{x}_2$ is at least some constant $\boldsymbol{\theta}_0 > 0$?

Using our main technical Lemma

A novel characterization of $\mathbf{A} \succcurlyeq \mathbf{B}$ (‡) for symmetric psd matrices.

Lemma: For sym. psd $\mathbf{A}$ we can efficiently factor $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ s.t. for all sym. psd $\mathbf{B}$,

(‡) ⇔ there exists $\mathbf{C}$ s.t. $\mathbf{B} = \mathbf{L}^T\mathbf{C}$ and $\mathbf{A} \succcurlyeq \mathbf{C}^T\mathbf{C}$.

Using our main technical Lemma

A novel characterization of $\mathbf{A} \succcurlyeq \mathbf{B}$ (‡) for symmetric psd matrices.

<u>Lemma</u>: For sym. psd $\mathbf{A}$ we can efficiently factor $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ s.t. for all sym. psd **B,**

(‡) ⇔ there exists $\mathbf{C}$ s.t. $\mathbf{B} = \mathbf{L}^T\mathbf{C}$ and $\mathbf{A} \succcurlyeq \mathbf{C}^T\mathbf{C}$.

Apply to $\mathbf{R} \succcurlyeq \mathbf{R}^{\{1,2\}}$ to get $\mathbf{R}^{\{1,2\}} = \mathbf{L}^T\mathbf{C}$ s.t. $\mathbf{R} \succcurlyeq \mathbf{C}^T\mathbf{C}$

Using our main technical Lemma

A novel characterization of $\mathbf{A} \succcurlyeq \mathbf{B}$ (‡) for symmetric psd matrices.

<u>Lemma</u>: For sym. psd $\mathbf{A}$ we can efficiently factor $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ s.t. for all sym. psd $\mathbf{B}$,

$$(\ddagger) \Leftrightarrow \text{there exists } \mathbf{C} \text{ s.t. } \mathbf{B} = \mathbf{L}^T\mathbf{C} \text{ and } \mathbf{A} \succcurlyeq \mathbf{C}^T\mathbf{C}.$$

Apply to $\mathbf{R} \succcurlyeq \mathbf{R}^{\{1,2\}}$ to get $\mathbf{R}^{\{1,2\}} = \mathbf{L}^T\mathbf{C}$ s.t. $\mathbf{R} \succcurlyeq \mathbf{C}^T\mathbf{C}$

Now, $(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\,\mathbf{x}_2 \le 0$ means that $\angle\, \mathbf{Lx}_2, \mathbf{Cx}_1 \ge \pi/2$

OTOH $\mathbf{R} \succcurlyeq \mathbf{C}^T\mathbf{C} \Rightarrow \mathbf{L}^T\mathbf{L} \succcurlyeq \mathbf{C}^T\mathbf{C} \Rightarrow ||\mathbf{Lx}_1|| \ge ||\mathbf{Cx}_1||$ - (1).

(1) along with $(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\mathbf{x}_1 \ge (\mathbf{x}_1)^T\mathbf{R}\mathbf{x}_1 / 2$ imply that $\angle\, \mathbf{Lx}_1, \mathbf{Cx}_1 \le \pi/3$. Thus, $\angle\, \mathbf{Lx}_1, \mathbf{Lx}_2 \ge \pi/6$.

Using our main technical Lemma

A novel characterization of $\mathbf{A} \succcurlyeq \mathbf{B}$ (‡) for symmetric psd matrices.

Lemma: For sym. psd $\mathbf{A}$ we can efficiently factor $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ s.t. for all sym. psd $\mathbf{B}$,

(‡) ⇔ there exists $\mathbf{C}$ s.t. $\mathbf{B} = \mathbf{L}^T\mathbf{C}$ and $\mathbf{A} \succcurlyeq \mathbf{C}^T\mathbf{C}$.

Apply to $\mathbf{R} \succcurlyeq \mathbf{R}^{\{1,2\}}$ to get $\mathbf{R}^{\{1,2\}} = \mathbf{L}^T\mathbf{C}$ s.t. $\mathbf{R} \succcurlyeq \mathbf{C}^T\mathbf{C}$

Now, $(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\,\mathbf{x}_2 \leq 0$ means that $\angle\, \mathbf{Lx}_2, \mathbf{Cx}_1 \geq \pi/2$

OTOH $\mathbf{R} \succcurlyeq \mathbf{C}^T\mathbf{C} \Rightarrow \mathbf{L}^T\mathbf{L} \succcurlyeq \mathbf{C}^T\mathbf{C} \Rightarrow ||\mathbf{Lx}_1|| \geq ||\mathbf{Cx}_1||$ - (1).

(1) along with $(\mathbf{x}_1)^T\mathbf{R}^{\{1,2\}}\mathbf{x}_1 \geq (\mathbf{x}_1)^T\mathbf{Rx}_1 / 2$ imply that $\angle\, \mathbf{Lx}_1, \mathbf{Cx}_1 \leq \pi/3$. Thus, $\angle\, \mathbf{Lx}_1, \mathbf{Lx}_2 \geq \pi/6$.

Future Work: Algorithm for satisfying bags of size > 3.

LLP-learning other classifiers, deviation-based objectives.