On PAC Learning Halfspaces in Non-interactive Local Privacy Model with Public Unlabeled Data

Jinyan Su

Jinhui Xu

Di Wang

Part of the work was done when Jinyan Su was a research intern at KAUST

Asian Conference on Machine Learning

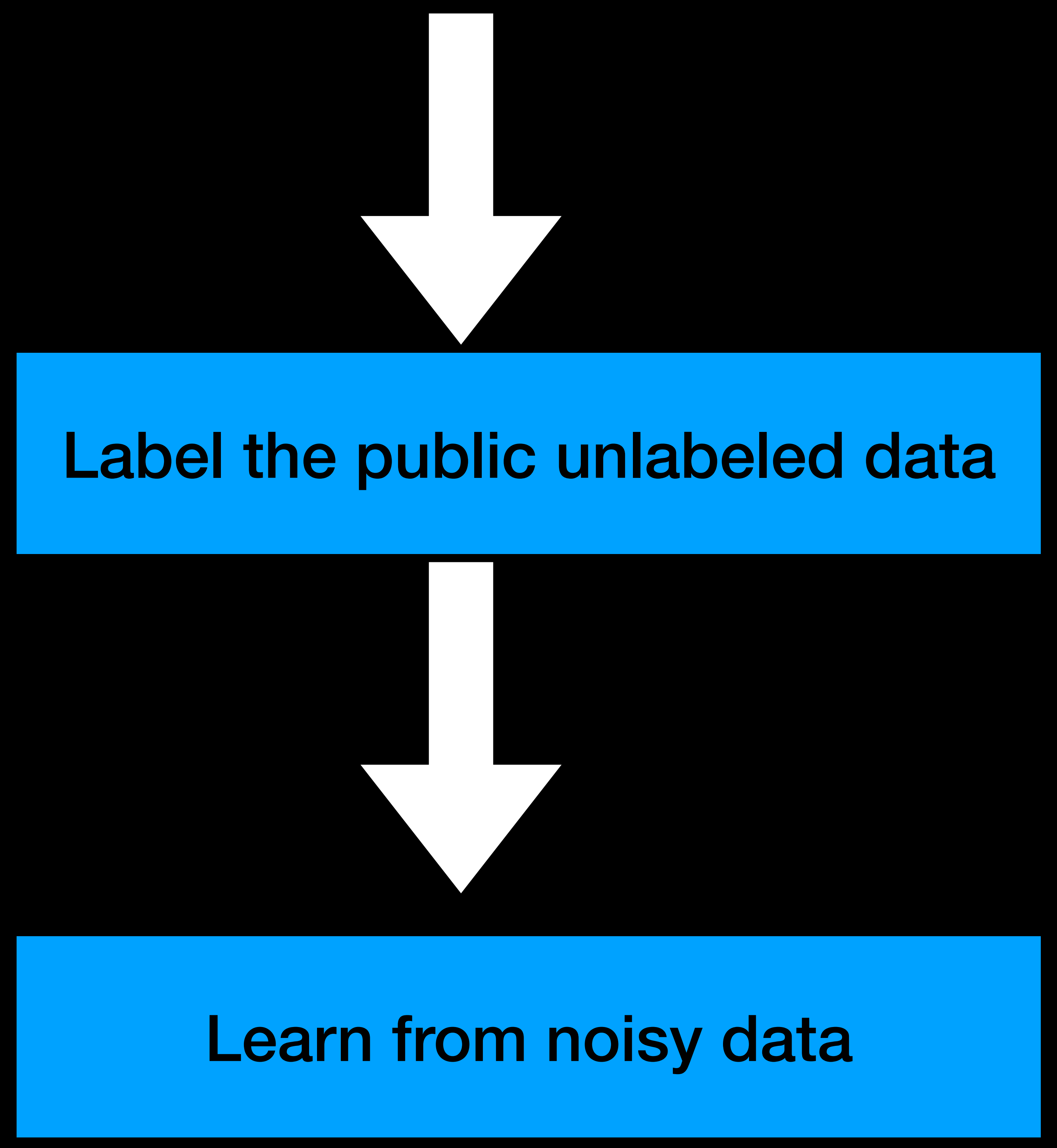
Background

- **Learning half spaces**: $y = \text{sign}(\langle w^*, x \rangle + \theta^*)$
- (α, β)**-PAC learner:**
  $\underset{(x,y)\sim\mathscr{P}}{Pr}[y \neq \text{sign}(\langle \hat{w}, x \rangle)] \leq \alpha$ w.p. at least $1 - \beta$
- **Local Differential Privacy**: $Pr[\mathscr{A}(x) \in E] \leq e^{\epsilon} Pr[\mathscr{A}(x^{'}) \in E] + \delta$
- **Non-interactive:** interact only once $(T = 1)$
- **Goal:**
  - Design private (α, β)-PAC learner in NLDP model with public unlabeled data
  - Sample complexity as low as possible

Additional public data $q \sim \mathscr{P}_x$

Motivation

- The public unlabeled data is cheap
  Learn a weak/noisy labeler



Via Massart Noise model

Massart Noise

Label are flipped w.p. $\leq \lambda$

1. Private part:
   Construct **Massart Noise example oracle** $\hat{f}$

   - Devide Private data into k **disjoint** groups
   - Using NLDP algorithm in [Wang et al.(2020)] for each groups of data
   - Boost accuracy by majority voting $\Rightarrow \hat{f}$ **with** $\lambda = \frac{3}{16}$

Private data: $\tilde{O}(d\text{Ploy}(\frac{1}{\epsilon}, \frac{1}{\delta}))$

Public data: $O(\frac{d}{\alpha^4})$

2. Non-private part:

   - Label public data with $\hat{f}$
   - Invoke Non-private algorithm for **learning half spaces with Massart Noise**

Via Self-supervised learning

1. Private part

    - Use **Logistic Loss** NLDP$\Rightarrow w^{priv}$

Pseudo labeler

Sufficiently small but constant error C_{err}

2. Non-private part

    Convert weak learner to strong learner$\Rightarrow$**Self-training**

    - Label unlabeled data with pseudo labeler

    - **Gradient descent** on pseudo labeled data$\Rightarrow$ **Update pseudo labeler**

Private data: $\tilde{O}(d\text{Ploy}(\frac{1}{\epsilon}, \frac{1}{\delta}))$

Public data: $O(\frac{d}{\alpha^2})$

Comparison to previous result:

Methods	Private data	Public data	With Public data?	Assume Large margin?
Prior method [Daniely and Feldman (2019)]	$\tilde{O}\left(\frac{d^{10}}{\epsilon^2 \cdot \gamma^{12}\alpha^6}\right)$	$\tilde{O}\left(\frac{d^{10}}{\epsilon^2 \cdot \gamma^{12}\alpha^6}\right)$	Yes	Yes
This paper (via Massart noise model)	$\tilde{O}(d\text{Ploy}(\frac{1}{\epsilon},\frac{1}{\delta}))$	$O(\frac{d}{\alpha^4})$	Yes	No
This paper (via self-training)	$\tilde{O}(d\text{Ploy}(\frac{1}{\epsilon},\frac{1}{\delta}))$	$O(\frac{d}{\alpha^2})$	Yes	No

Limitations and Future direction

- Distribution dependent (bounded distribution/ mixture distribution)
- Find efficient PAC learning algorithm with Massart Noise