PILLAR: How to make semi private learning effective

Francesco Pinto*[1], Yaxi Hu*[1], Fanny Yang[1], Amartya Sanyal[1,2]

[1]ETH Zurich, [2]Max Planck Institute for Intelligent Systems

Differentially Private Learning

An algorithm $\mathcal{A}$ is said to be (ϵ, δ)-differentially private (DP) if

$$\mathbb{P}\left[\mathcal{A}(S_1) \in Q\right] \leq \exp(\epsilon)\mathbb{P}\left[\mathcal{A}(S_2) \in Q\right] + \delta$$

for all **neighbouring datasets** S_1, S_2 and **output sets** Q.

Existing Results: Sample complexity of DP algorithms are **dimension-dependent** in the worst case.

In **Semi-Private learning** [1], the learner accesses

- **Private Labelled** dataset,
- **Public Unlabelled** dataset from nearby distribution

This work: Design Semi-Private learner for linear half-spaces that

1. Is **Computationally Efficient**
2. Admits **Dimension Independent** sample complexity
3. Performs well in **Challenging Practical Applications**

Theoretical Results

We exploit two properties of data distribution μ (covariance Σ)

- **(A1) Large Margin**: μ admits a classifier w^* with margin γ
- **(A2) Low Rank**: Large Proj. of w^* on top-k components of Σ.

PILLAR 1 Unlabelled dataset ($\mathbf{X}_U \in$), Labelled dataset ($\mathbf{X_L}, Y_L$), k

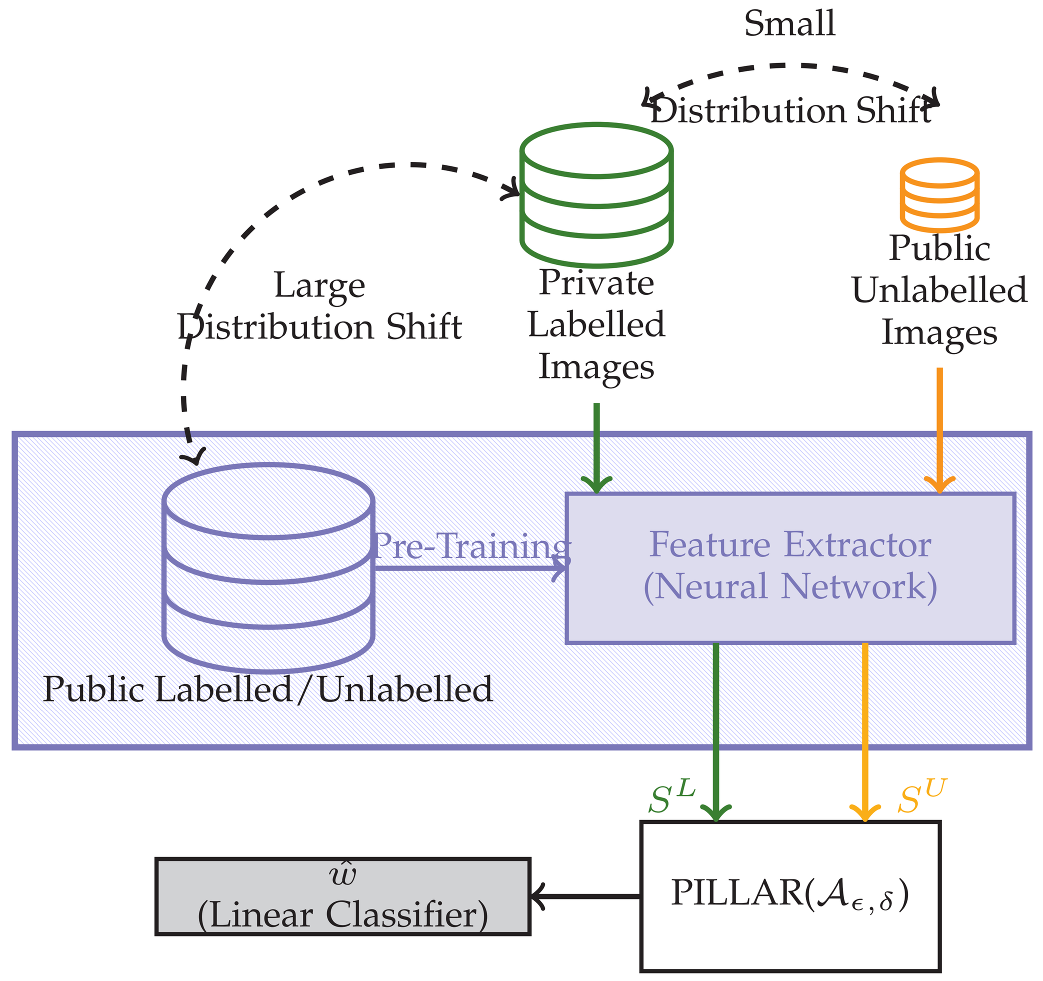
1: $\widehat{\Sigma} \leftarrow \sum_{\mathbf{x} \in S_U} \mathbf{x}\mathbf{x}^\top$, $\mathbf{A}_k \leftarrow$ top-k principal components of $\widehat{\Sigma}$.
2: $\mathbf{X}_L^{\text{Proj}} \leftarrow$ Project $\mathbf{X_L}$ on $\mathbf{A}_k$.
3: $\hat{\mathrm{w}}_{\epsilon,\delta} \leftarrow$ Run Noisy-SGD on $(\mathbf{X}_L^{\text{Proj}}, Y_L)$ with privacy parameters ϵ, δ

Guarantees on $\hat{\mathrm{w}}_{\epsilon,\delta}$

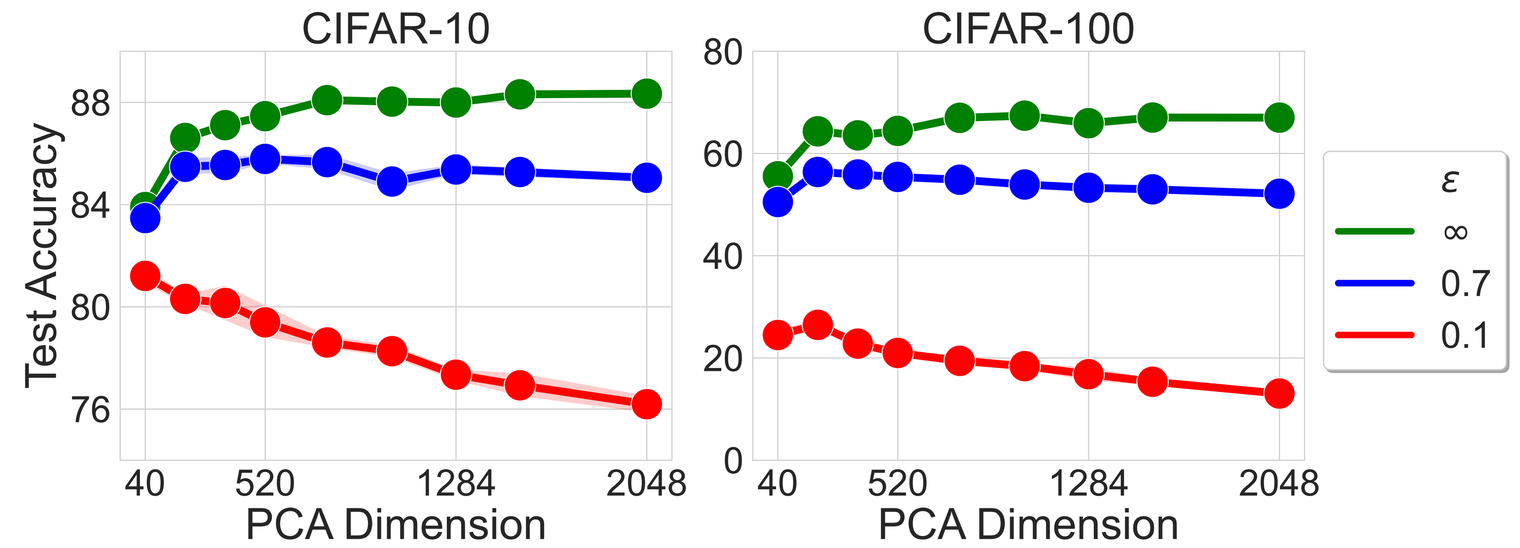
- **Privacy:** $\hat{\mathrm{w}}_{\epsilon,\delta}$ is (ϵ, δ)-DP.
- **Accuracy:** For $\alpha, \beta \geq 0$, $|\mathbf{X}_U| = O\left(\frac{1}{\gamma^2}\right)$ and $|\mathbf{X}_L| = \widetilde{O}\left(\frac{\sqrt{k}}{\alpha\epsilon\gamma}\right)$,

$$\mathbb{P}\left[\mathbf{Error}\left(\hat{\mathrm{w}}_{\epsilon,\delta}\right) \leq \alpha\right] \geq 1 - \beta$$

Experimental Setting



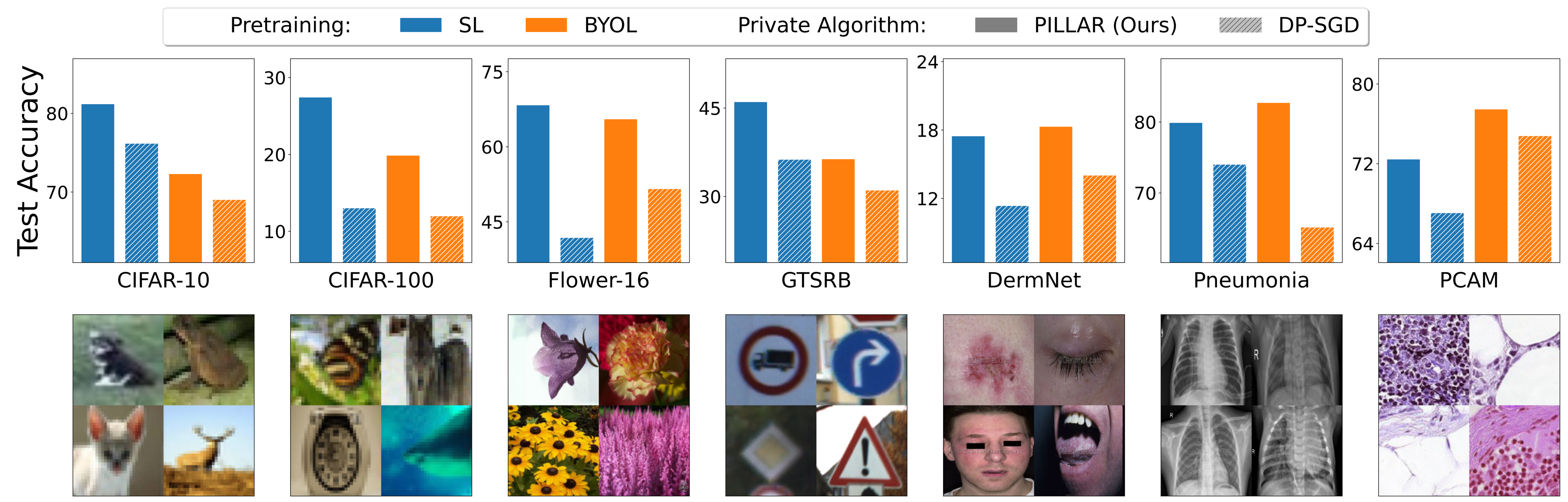
Experiments I: Reducing dimensions



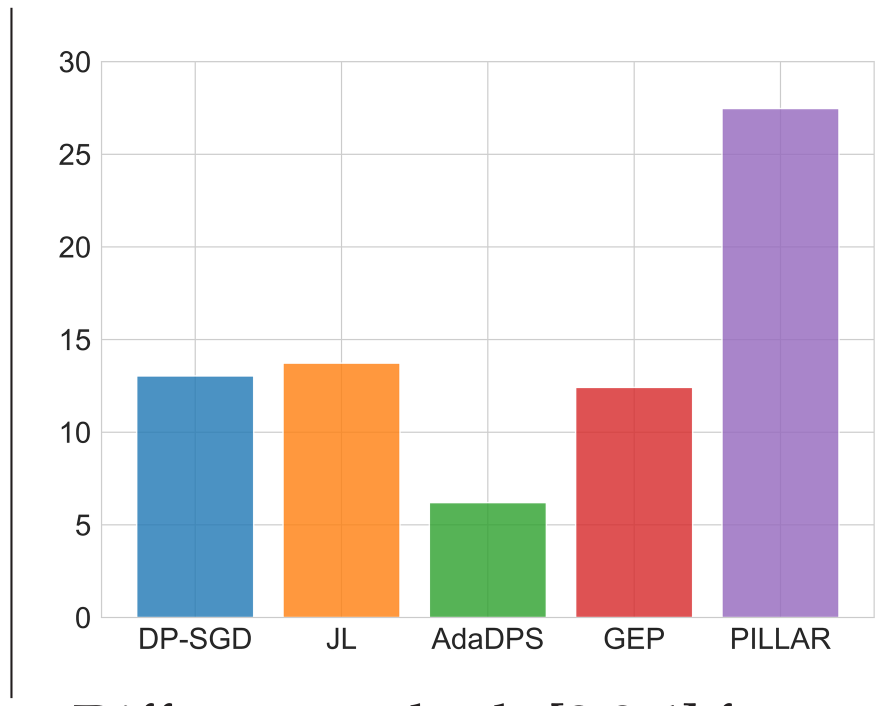
Takeaway:

- **Strict privacy** ($\epsilon = 0.1$): Dimension $\Downarrow \implies$ Accuracy $\Uparrow$.
- **Without privacy** ($\epsilon = \infty$): Dimension $\Downarrow \implies$ Accuracy $\Downarrow$.

Experiments II: PILLAR outperforms other algorithms across datasets



Comparison across datasets and pre-training for $\epsilon = 0.1$.



Different methods [2,3,4] for $\epsilon = 0.1$.

Experiments III: Distribution Shift

- Public and private data may come from different distributions.
- PILLAR's performance is robust to using CIFAR-10v1 for public data and CIFAR10/100 for private data.

Pre-training / PCA Data	CIFAR10 SL	CIFAR10 BYOL	CIFAR100 SL	CIFAR100 BYOL
In-distribution	81.21	72.33	27.47	19.89
CIFAR-10v1	81.18	73.24	27.18	19.21

QR Code for paper

[1] Alon, et al. "Limits of private learning with access to public data." NeurIPS (2019).

[2] Lê Nguyên, et al. "Efficient private algorithms for learning large-margin halfspaces." ALT (2020).

[3] Li, Tian, et al. "Private adaptive optimization with side information." ICML (2022).

[4] Yu, Da, et al. "Do not let privacy overbill utility: Gradient embedding perturbation for private learning." ICLR (2021).