Accurate and Robust Positive-Unlabeled Learning against Adversarial Perturbations

Ryo Shibazaki[1*], Kazuhiko Kawamoto[2] and Hiroshi Kera[2]

[1*]Graduate School of Science and Engineering, Chiba University, 1-33 Yayoichō, Inage-ku, Chiba-shi, 263-8522, Chiba, Japan.
[2]Graduate School of Informatics, Chiba University, 1-33 Yayoichō, Inage-ku, Chiba-shi, 263-8522, Chiba, Japan.

*Corresponding author(s). E-mail(s): ryo.shibazaki0517@chiba-u.jp;
Contributing authors: kawa@faculty.chiba-u.jp; kera@chiba-u.jp;

Abstract

Labeling costs are high in domains such as medical image analysis, where Positive and Unlabeled (PU) learning, which trains using only positive and unlabeled data, is effective. Medical images often contain small perturbations due to sensor noise and variations in acquisition conditions, which can cause a classifier to misclassify images that should be positive as negative. Therefore, in settings where even minor misclassifications may lead to critical misdiagnoses, high robustness is required. In this study, we focus on adversarial perturbations, which are known as worst-case noise among such perturbations, and aim to improve robustness within the PU learning framework. However, directly applying standard adversarial training methods to PU learning often severely degrades standard accuracy, making the trade-off between robustness and standard accuracy more pronounced. To address this issue, we propose PU-TRADES, a new learning method that extends the TRADES framework and integrates it with PU learning. Our method introduces label-independent adversarial perturbations and optimizes the balance between robustness and standard accuracy by combining a PU loss with a Kullback–Leibler loss. Furthermore, we theoretically derive an upper bound on the estimation error for the proposed loss and clarify conditions under which PU learning can outperform supervised learning when the number of unlabeled samples is sufficiently large. Finally, experiments on multiple benchmark datasets and a medical imaging dataset demonstrate that the proposed method provides an effective framework for robust learning in PU settings.

Keywords: positive-unlabeled learning, adversarial robustness, risk estimation, empirical risk minimization

1 Introduction

In recent years, machine learning has achieved remarkable success across a wide range of tasks, driven by advances in large-scale data and high-capacity models. However, in real-world applications, it is often difficult to obtain high-quality ...; moreover, in settings where the training set contains only positive and unlabeled data, one must learn from incomplete information in which the unlabeled set is a mixture of true positives and negatives. Therefore, unlike fully supervised learning, PU learning requires careful risk estimation and additional techniques to stabilize training.

Furthermore, in practical applications, robustness is an essential requirement in addition to label scarcity. In particular, medical images and sensor data are affected by variations in acquisition conditions, noise, and device ... We focus on adversarial perturbations[**?**] and aim to improve robustness within the PU learning framework.

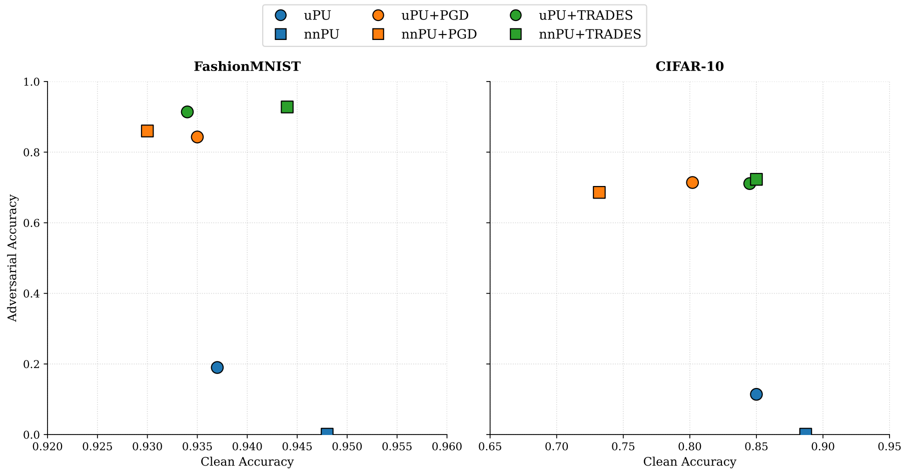
However, in preliminary experiments, directly applying adversarial training[**?**]—a standard approach to enhancing resilience against adversarial perturbations—to PU learning can substantially degrade classification performance on clean data. This is because, although unlabeled data contain a mixture of true positives and negatives, they are treated uniformly as negatives in the loss. As a result, the objectives of adversarial perturbation optimization and classification become misaligned, which destabilizes the updates associated with the unlabeled term. Figure **??** illustrates the relationship between clean accuracy and adversarial accuracy when PGD-based adversarial training is naively applied to PU learning. In particular, on CIFAR-10, adversarial accuracy improves while clean accuracy drops significantly. These observations indicate that, under the PU setting, introducing adversarial training may impair clean accuracy, which remains a key challenge.

In this study, we extend the TRADES framework[**?**], a representative adversarial training method, and propose a new learning method, PU-TRADES, by integrating it with PU learning. Our approach handles adversarial perturbations in a label-independent manner and combines a PU loss with a Kullback–Leibler loss, aiming to improve robustness while suppressing degradation in clean accuracy.

To validate the effectiveness of the proposed method, we conducted experiments on multiple benchmark datasets and medical imaging data. We evaluated both clean accuracy and ... and compared them with baseline methods. The results show that the proposed method substantially improves robustness to adversarial perturbations while maintaining accuracy on clean data.

In addition, we performed a theoretical analysis to better understand robustness in PU learning. Specifically, under binary classification with linear classifiers and adversarial perturbations, we derived upper bounds on the estimation error of the risks minimized by supervised learning and PU learning, thereby clarifying conditions under which PU learning can be advantageous over supervised learning. These conditions are consistent with practical scenarios, and they suggest that simply increasing the amount of unlabeled data can potentially achieve higher robustness than supervised learning.

Our contributions are summarized as follows.



	FashionMNIST			CIFAR-10		
Method	Clean	Adv	ΔClean	Clean	Adv	ΔClean
uPU	0.937	0.190	–	0.850	0.114	–
nnPU	**0.948**	0.001	–	**0.887**	0.001	–
uPU+PGD	0.935	0.843	-0.002	0.802	0.714	-0.048
nnPU+PGD	0.930	0.860	-0.018	0.732	0.686	**-0.155**

Fig. 1 Performance changes when PGD-based adversarial training is naively applied to PU learning (FashionMNIST / CIFAR-10). Left: a scatter plot showing the relationship between clean accuracy and adversarial accuracy for each method. Right: a numerical summary and the change in clean accuracy, ΔClean, before and after applying PGD (difference from the corresponding PU method without PGD). While PGD substantially improves adversarial accuracy, it degrades clean accuracy; this degradation is particularly pronounced for nnPU on CIFAR-10, where clean accuracy drops from 0.887 to 0.732 (ΔClean$= -0.155$).

- We propose a new learning framework (PU-TRADES) that integrates TRADES-style regularization into PU learning, improving robustness to adversarial perturbations while maintaining classification accuracy on clean samples.
- Assuming linear classifiers under adversarial perturbations, we derive upper bounds on the estimation error of the risks minimized by supervised learning and PU learning, and theoretically identify conditions under which PU learning becomes more favorable than supervised learning.
- Through experiments on benchmark datasets and medical imaging data, we empirically demonstrate that the proposed method acquires robustness to adversarial samples while preserving clean accuracy.

2 Related Work

In this chapter, we review prior studies related to this work, focusing on (i) Positive-Unlabeled (PU) learning and (ii) adversarial training. We then summarize existing research that combines PU learning with adversarial robustness.

2.1 Positive-Unlabeled (PU) Learning

PU learning is a classification framework in which only positive-labeled and unlabeled data are available. A representative line of work is risk-estimation-based PU learning, which constructs an (unbiased) estimator of the supervised classification risk using the class prior and the mixture structure of the unlabeled set. In particular, uPU (unbiased PU learning) estimates the true risk without bias, while nnPU (non-negative PU learning) introduces a non-negativity constraint to prevent the empirical risk from becoming negative, thereby mitigating overfitting. Many extensions have also been proposed, including methods that exploit high-confidence samples from the unlabeled set and approaches that incorporate various correction mechanisms.

2.2 Adversarial Training

Adversarial examples are inputs that are intentionally perturbed to cause a model to misclassify, and they have attracted extensive attention as a major threat to machine learning systems. A standard defense is adversarial training, which improves robustness by training the model on adversarially perturbed samples. Representative methods such as PGD-based adversarial training can be formulated as a min–max optimization problem consisting of an outer minimization (parameter optimization) and an inner maximization (perturbation generation). In addition, TRADES [**?**] is a prominent method that introduces a regularization term based on the Kullback–Leibler divergence between the model outputs on clean and perturbed inputs, aiming to balance clean accuracy and adversarial robustness.

3 Preliminaries

In this chapter, we introduce PU learning, adversarial examples and representative attacks, and adversarial training. Hereafter, we refer to learning a binary classifier from fully labeled positive and negative data as Positive–Negative (PN) learning.

3.1 Positive-Unlabeled (PU) Learning

We denote the input space by $\mathcal{X} \subseteq \mathbb{R}^d$ and the label space by $\mathcal{Y} = \{-1, +1\}$. Let $p(\boldsymbol{x}, y)$ be the joint distribution over $(\mathcal{X}, \mathcal{Y})$. Let the total number of samples be $n \in \mathbb{N}$, and let n_{P} and n_{N} denote the numbers of positive (P) and negative (N) samples, respectively. Each set is represented as follows:

$$\begin{aligned} \mathscr{X}_{\mathrm{P}} &= \{\boldsymbol{x}_i^{\mathrm{P}}\}_{i=1}^{n_{\mathrm{P}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{P}}(\boldsymbol{x}), \\ \mathscr{X}_{\mathrm{N}} &= \{\boldsymbol{x}_i^{\mathrm{N}}\}_{i=1}^{n_{\mathrm{N}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{N}}(\boldsymbol{x}). \end{aligned} \tag{1}$$

Here, $p_{\mathrm{P}}(\boldsymbol{x})$ and $p_{\mathrm{N}}(\boldsymbol{x})$ denote the class-conditional densities for the positive and negative classes, respectively. The full dataset $\mathscr{X} = \mathscr{X}_{\mathrm{P}} \cup \mathscr{X}_{\mathrm{N}}$ is written as

$$\begin{aligned} \mathscr{X} &= \{\boldsymbol{x}_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{x}), \\ p(\boldsymbol{x}) &= \pi_{\mathrm{P}}\, p_{\mathrm{P}}(\boldsymbol{x}) + \pi_{\mathrm{N}}\, p_{\mathrm{N}}(\boldsymbol{x}), \end{aligned} \tag{2}$$

where $\pi_{\mathrm{P}} = p(y=+1)$ and $\pi_{\mathrm{N}} = ...p(y=-1)$ are the class priors satisfying $\pi_{\mathrm{P}} + \pi_{\mathrm{N}} = 1$.

In PU learning, the training set consists of positive (P) samples and unlabeled (U) samples. Since the marginal distribution of unlabeled data is $p_{\mathrm{U}}(\boldsymbol{x}) = \pi_{\mathrm{P}} p_{\mathrm{P}}(\boldsymbol{x}) + \pi_{\mathrm{N}} p_{\mathrm{N}}(\boldsymbol{x})$, letting n_{U} be the number of unlabeled samples, the unlabeled set is given by

$$\mathscr{X}_{\mathrm{U}} = \{\boldsymbol{x}_i^{\mathrm{U}}\}_{i=1}^{n_{\mathrm{U}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{U}}(\boldsymbol{x}) = p(\boldsymbol{x}). \tag{3}$$

That is, the unlabeled samples are drawn i.i.d. from the marginal distribution of inputs, which is a mixture of positive and negative class-conditional distributions.
Unbiased PU Learning (uPU). uPU assumes that the positive class prior π_{P} is known and estimates the negative risk indirectly from the unlabeled data. Specifically, it minimizes the following empirical risk:

$$\widehat{R}_{\mathrm{uPU}}(g) = \frac{\pi_{\mathrm{P}}}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \tilde{\ell}\big(g\big(\boldsymbol{x}_i^{\mathrm{P}}\big), +1\big) \; + \; \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \ell\big(g\big(\boldsymbol{x}_i^{\mathrm{U}}\big), -1\big). \tag{4}$$

Here, the composite loss $\tilde{\ell}(g(\boldsymbol{x}), y)$ is defined by $\tilde{\ell}(g(\boldsymbol{x}), y) = \ell(g(\boldsymbol{x}), y) - \ell(g(\boldsymbol{x}), -y)$.
Non-Negative PU Learning (nnPU). nnPU was introduced to address the overfitting issue in uPU, where the empirical risk can take negative values[**?**]. Specifically, when the estimated term involving the negative risk in PU learning becomes negative, nnPU clips its negative contribution to zero, yielding a non-negative risk estimator that keeps the overall estimate bounded below by 0. This modification prevents the empirical risk from diverging to negative values during empirical minimization and enables stable training. The empirical risk of nnPU is defined as follows:

$$\begin{aligned}\widehat{R}_{\mathrm{nnPU}}(g) = {} & \frac{\pi_{\mathrm{P}}}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\left(g\left(\boldsymbol{x}_i^{\mathrm{P}}\right), +1\right) \\ & + \max\left\{0, -\frac{\pi_{\mathrm{P}}}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\left(g\left(\boldsymbol{x}_i^{\mathrm{P}}\right), -1\right) + \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \ell\left(g\left(\boldsymbol{x}_i^{\mathrm{U}}\right), -1\right)\right\}.\end{aligned} \tag{5}$$

3.2 Adversarial Examples

Adversarial examples are inputs constructed by adding small, carefully designed perturbations to an image that a classifier originally classified correctly, thereby intentionally causing misclassification...w2014ExplainingAH. For example, Fig. **??** adds a tiny perturbation to a cat image and generates an adversarial example...

3.3 Adversarial attack: FGSM and PGD

We refer to methods for generating adversarial examples as *adversarial attacks*, and many variants have been studied. In this work, we focus on representative first-order attacks: the Fast Gradient Sign Method (FGSM) [**?**] and Projected Gradient Descent

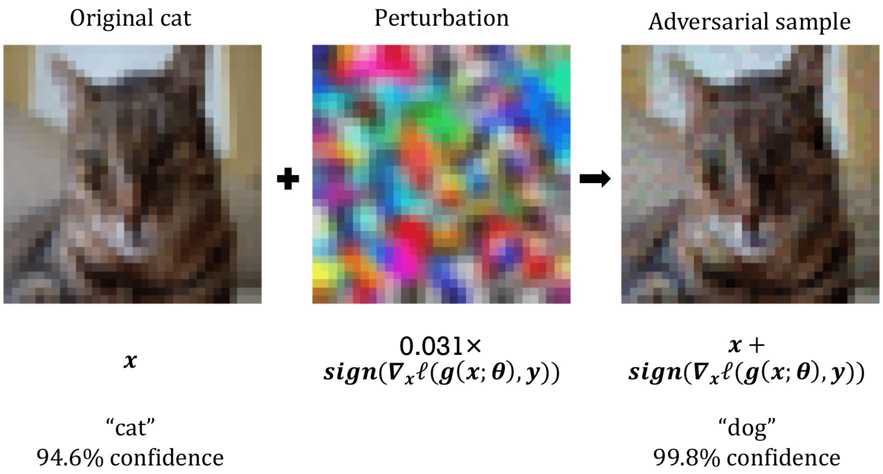


Fig. 2 An example of generating adversarial examples. Starting from the clean image on the left, we add a small perturbation using PGD to obtain an adversarial input... The classifier correctly predicts the clean image as a cat (confidence 94.6%), while it misclassifies the adversarial example as a dog (confidence 99.8%). This illustrates that predictions can change drastically even when the input appears almost identical to humans[**?**].

(PGD) [**?**]. Here, sign : $\mathbb{R} \to [-1, 1]$ is applied element-wise to the argument vector.

$$\boldsymbol{x}' = \boldsymbol{x} + \epsilon \cdot \mathrm{sign}\left(\nabla_{\boldsymbol{x}} \ell(g(\boldsymbol{x}; \boldsymbol{\theta}), y)\right) \tag{6}$$

This produces an input that increases the loss when fed into the model. A stronger iterative variant of this method is PGD [**?**], which updates the input in the direction that increases the loss with step size α, similarly to FGSM, and then applies the projection $\Pi_{\dots}$. In particular,

$$\mathcal{B}_{\infty}(\boldsymbol{x}, \epsilon) := \{\boldsymbol{z} \in \mathbb{R}^d \mid \|\boldsymbol{z} - \boldsymbol{x}\|_{\infty} \leq \epsilon\}$$

Then, $\Pi_{\mathcal{B}_{\infty}(\boldsymbol{x},\epsilon)}$ denotes the projection onto the ℓ_{∞}-ball, which guarantees $\|\boldsymbol{x}' - \boldsymbol{x}\|_{\infty} \leq \epsilon$. Under this setting, the PGD update is given by:

$$\boldsymbol{x}' \leftarrow \Pi_{\mathcal{B}_{\infty}(\boldsymbol{x},\epsilon)} \left[\boldsymbol{x} + \alpha \ \mathrm{sign}\left(\nabla_{\boldsymbol{x}} \ell(g(\boldsymbol{x}; \boldsymbol{\theta}), y)\right)\right]. \tag{7}$$

By repeating Eq. (**??**) multiple times, we can generate samples that more strongly increase the loss within the ϵ-ball.

3.4 Adversarial Training

Adversarial training improves model robustness by training on adversarial examples[**?**]. It is typically formulated as the following min–max optimization problem: one first generates adversarial examples for each input, and then learns parameters that minimize the average loss over these adversarial inputs.

$$\min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \max_{\|\boldsymbol{x}'_i - \boldsymbol{x}_i\|_\infty \le \epsilon} \ell(g(\boldsymbol{x}'_i; \boldsymbol{\theta}), y_i) \,. \tag{8}$$

In addition, as a method to further enhance robustness against adversarial examples, TRADES (TRadeoff-inspired Adversarial DEfense via Surrogate-loss minimization), proposed by Zhang *et al.*[**?**], is a prominent approach. TRADES explicitly models the trade-off between accuracy on clean samples and robustness to adversarial samples, and aims to minimize the following loss function:

$$\begin{aligned} \mathcal{L}_{\mathrm{TR}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \Big[& \ell\big(g(\boldsymbol{x}_i; \boldsymbol{\theta}), y_i\big) \\ & + \beta \cdot \max_{\|\boldsymbol{x}'_i - \boldsymbol{x}_i\|_\infty \le \epsilon} \ell_{\mathrm{KL}}\big(g(\boldsymbol{x}_i; \boldsymbol{\theta}), g(\boldsymbol{x}'_i; \boldsymbol{\theta})\big) \Big]. \end{aligned} \tag{9}$$

Here, $\ell_{\mathrm{KL}}(\cdot, \cdot)$ denotes the Kullback–Leibler (KL) divergence between predictive distributions, i.e.,

$$\ell_{\mathrm{KL}}\big(g(\boldsymbol{x}_i; \boldsymbol{\theta}),\, g(\boldsymbol{x}'_i; \boldsymbol{\theta})\big) := \mathrm{KL}\big(p_{\boldsymbol{\theta}}(\cdot \mid \boldsymbol{x}_i) \,\big\|\, p_{\boldsymbol{\theta}}(\cdot \mid \boldsymbol{x}'_i)\big) \,.$$

The first term of Eq. (**??**), $\ell\big(g(\boldsymbol{x}_i; \boldsymbol{\theta}), y_i\big)$, is the standard classification loss on the clean input $\boldsymbol{x}_i$.

On the other hand, the second term $\ell_{\mathrm{KL}}\big(g(\boldsymbol{x}_i; \boldsymbol{\theta}), g(\boldsymbol{x}'_i; \boldsymbol{\theta})\big)$ constrains the model so that the output distributions for $\boldsymbol{x}_i$ and its perturbed version $\boldsymbol{x}'_i$ are close, and this term plays a key role in improving robustness. Thus, TRADES is designed to enhance robustness while maintaining classification accuracy. In this study, we apply this framework to PU learning to achieve both high performance on clean samples and robustness to adversarial examples.

4 Accurate and Robust PU Learning

In this chapter, we first clarify the issues that arise when uPU learning is naively combined with PGD-based adversarial training. We then propose a new learning method, PU+TRADES, which adapts the TRADES framework to PU learning.

4.1 uPU+PGD

In uPU learning, the empirical risk $\widehat{R}_{\mathrm{uPU}}(g)$ is estimated by minimizing

$$\widehat{R}_{\mathrm{uPU}}(g) = \frac{\pi_{\mathrm{P}}}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \tilde{\ell}\left(g\big(\boldsymbol{x}_i^{\mathrm{P}}\big), +1\right) + \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \ell\left(g\big(\boldsymbol{x}_i^{\mathrm{U}}\big), -1\right). \tag{10}$$

Here, the loss is taken differently for P and U samples, which can be summarized as

$$\mathscr{L}(\boldsymbol{x}) := \begin{cases} \tilde{\ell}(g(\boldsymbol{x}), +1), & \boldsymbol{x} \in \mathscr{X}_{\mathrm{P}}, \\ \ell(g(\boldsymbol{x}), -1), & \boldsymbol{x} \in \mathscr{X}_{\mathrm{U}}. \end{cases} \tag{11}$$

Using this loss $\mathscr{L}$, we generate an adversarial example for each sample via PGD. A single PGD update step is given by

$$\boldsymbol{x}' \leftarrow \mathrm{Clip}_{(\boldsymbol{x}-\epsilon,\ \boldsymbol{x}+\epsilon)}\Big[\boldsymbol{x}' + \alpha\ \mathrm{sign}\big(\nabla_{\boldsymbol{x}'}\mathscr{L}(\boldsymbol{x}')\big)\Big]. \tag{12}$$

Issues with uPU+PGD

In uPU, the loss for unlabeled data is computed as if the label were always $y = -1$. However, in reality, the unlabeled set contains a mixture of positives and negatives. This property is incompatible with PGD-based adversarial training.

- **Negative U samples.** Since the loss $\ell(g(\boldsymbol{x}^{\mathrm{U}}), -1)$ is consistent with the true label, it pushes the input in a direction that increases the loss for the negative class. Consequently, PGD generates appropriate adversarial perturbations, contributing to improved robustness.
- **Positive U samples.** If perturbations are generated using $\ell(g(\boldsymbol{x}^{\mathrm{U}}), -1)$ even though the sample is truly positive, PGD updates the input so as to maximize the *negative-class* loss. As a result, the input may be pushed not toward the decision boundary, but rather toward a region where it is classified as positive with higher confidence. Therefore, PGD fails to produce perturbations in the "most misclassifiable direction," and the training can break down.

Hence, to generate adversarial perturbations appropriately in PU learning, it is essential to use a *label-independent* perturbation generation mechanism. This motivates PU+TRADES, introduced in the next section.

4.2 PU+TRADES

In this work, we propose **uPU+TRADES** and **nnPU+TRADES**, which integrate TRADES into uPU and nnPU, respectively. By introducing the TRADES framework into PU learning, we endow the model with robustness.

The objective function of uPU+TRADES is given by

$$\min_g \left[\widehat{R}_{\mathrm{uPU}}(g) + \beta \cdot \frac{1}{n} \sum_{i=1}^{n} \max_{\|\boldsymbol{x}'-\boldsymbol{x}_i\|_\infty \le \epsilon} \ell_{\mathrm{KL}}\left(g\left(\boldsymbol{x}_i\right)\ \|\ g\left(\boldsymbol{x}'\right)\right)\right], \tag{13}$$

and the objective function of nnPU+TRADES is given by

$$\min_{g}\left[\widehat{R}_{\mathrm{nnPU}}(g)+\beta\cdot\frac{1}{n}\sum_{i=1}^{n}\max_{\|\boldsymbol{x}'-\boldsymbol{x}_i\|_\infty\leq\epsilon}\ell_{\mathrm{KL}}\left(g\left(\boldsymbol{x}_i\right)\ \|\ g\left(\boldsymbol{x}'\right)\right)\right]. \tag{14}$$

In binary classification, the network outputs a one-dimensional logit $g(\boldsymbol{x};\boldsymbol{\theta})$. We convert it into a Bernoulli probability vector and compute the KL divergence:

$$p(\boldsymbol{x})=\big[\sigma(g(\boldsymbol{x};\boldsymbol{\theta})),\,1-\sigma(g(\boldsymbol{x};\boldsymbol{\theta}))\big], \tag{15}$$

where $\sigma(\cdot)$ denotes the sigmoid function. The KL loss is then defined as

$$\begin{aligned}\ell_{\mathrm{KL}}\left(g\left(\boldsymbol{x}_i;\boldsymbol{\theta}\right),g\left(\boldsymbol{x}_i';\boldsymbol{\theta}\right)\right)&=\mathrm{KL}\big(p(\boldsymbol{x}_i)\,\|\,p(\boldsymbol{x}_i')\big)\\&=\sum_{c\in\{0,1\}}p_c(\boldsymbol{x}_i)\,\log\frac{p_c(\boldsymbol{x}_i)}{p_c(\boldsymbol{x}_i')}.\end{aligned} \tag{16}$$

This term encourages the model outputs to remain stable under small perturbations of $\boldsymbol{x}_i$, thereby providing robustness against adversarial perturbations.

5 Theoretical Analysis

This section presents generalization (estimation-error) bounds for TRADES in supervised and PU settings, and derives a sufficient condition on the number of unlabeled samples under which PU+TRADES achieves a tighter bound than supervised TRADES.

Problem Setting 5.1. (Adversarial Binary Classification Setting)

Let the input space be $\mathcal{X}\subseteq\mathbb{R}^d$, the input data $\boldsymbol{x}\in\mathcal{X}$ are bounded,

$$\|\boldsymbol{x}\|_\infty\leq C_x$$

Assume this condition holds.

$$p_{\mathrm{P}}(\boldsymbol{x})=p(\boldsymbol{x}\mid y=+1),\quad p_{\mathrm{N}}(\boldsymbol{x})=p(\boldsymbol{x}\mid y=-1)$$

We define it as follows. $\pi_{\mathrm{P}}=p(y=+1)$ $\pi_{\mathrm{N}}=p(y=-1)$

In supervised learning, we use labeled positive (P) and negative (N) samples.

$$\begin{aligned}\mathscr{X}_{\mathrm{P}}&=\{\boldsymbol{x}_i^{\mathrm{P}}\}_{i=1}^{n_{\mathrm{P}}}\overset{\text{i.i.d.}}{\sim}p_{\mathrm{P}}(\boldsymbol{x}),\\\mathscr{X}_{\mathrm{N}}&=\{\boldsymbol{x}_i^{\mathrm{N}}\}_{i=1}^{n_{\mathrm{N}}}\overset{\text{i.i.d.}}{\sim}p_{\mathrm{N}}(\boldsymbol{x})\end{aligned} \tag{17}$$

In contrast, in PU learning we use positive and unlabeled (U) data.

$$p_{\mathrm{U}}(\boldsymbol{x})=\pi_{\mathrm{P}}p_{\mathrm{P}}(\boldsymbol{x})+\pi_{\mathrm{N}}p_{\mathrm{N}}(\boldsymbol{x})$$

$$\mathscr{X}_{\mathrm{U}} = \{\boldsymbol{x}_i^{\mathrm{U}}\}_{i=1}^{n_{\mathrm{U}}} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{U}}(\boldsymbol{x}) \tag{18}$$

.

$$g(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x}$$
$$\mathscr{G} = \{g(\boldsymbol{x}) : \|\boldsymbol{w}\|_p \le W\}$$

We define it as follows.

$$\ell_{\mathrm{KL}}\big(g(\boldsymbol{x}), g(\boldsymbol{x}')\big)$$

5.1 Preliminaries

We summarize the basic tools used in the analysis, including Rademacher complexity, contraction inequalities, and McDiarmid's inequality.

5.1.1 Rademacher Complexity

$$\mathfrak{R}(S) = \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left[\sup_{(s_1,\ldots,s_n)\in S} \frac{1}{n}\sum_{i=1}^n \sigma_i s_i\right]$$
$$\mathscr{G} \circ S_n = \{(g(\boldsymbol{x}_1), \ldots, g(\boldsymbol{x}_n)) \mid g \in \mathscr{G}\}$$
$$\mathfrak{R}_{S_n}(\mathscr{G}) = \mathfrak{R}(\mathscr{G} \circ S_n) = \mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left[\sup_{g\in\mathscr{G}} \frac{1}{n}\sum_{i=1}^n \sigma_i g(\boldsymbol{x}_i)\right]$$

Definition 5.1 (Rademacher complexity)**.**

$$\mathfrak{R}_{n,\nu}(\mathscr{G}) = \mathbb{E}_{S_n\sim\nu^n}\left[\mathfrak{R}_{S_n}(\mathscr{G})\right] = \mathbb{E}_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\left[\sup_{g\in\mathscr{G}} \frac{1}{n}\sum_{i=1}^n \sigma_i g(\boldsymbol{x}_i)\right] \tag{19}$$

Lemma 5.2 (Talagrand's contraction lemma)**.** *Let* $S_n = \{\boldsymbol{x}_i\}_{i=1}^n \overset{i.i.d.}{\sim} \nu(\boldsymbol{x})$.

$$\mathfrak{R}_{n,\nu}(f \circ \mathscr{G}) \le L_f\, \mathfrak{R}_{n,\nu}(\mathscr{G}) \tag{20}$$

holds.

5.1.2 Upper Bound on the Rademacher Complexity

Theorem 5.3 (Upper bound on the Rademacher complexity)**.** $\mathscr{G} = \{\boldsymbol{x} \mapsto \boldsymbol{w}^\top\boldsymbol{x} \mid \|\boldsymbol{w}\|_\infty \le W\}$

$$\mathfrak{R}_{n,\nu}(\mathscr{G}) \le \frac{C_x W}{\sqrt{n}} \tag{21}$$

holds.

Lemma 5.4 (Rademacher complexity under adversarial perturbations)**.** $\mathscr{G} = \{\boldsymbol{x} \mapsto \boldsymbol{w}^\top \boldsymbol{x} : \ \|\boldsymbol{w}\|_p \leq W\}$

$$\mathfrak{R}_{n,\nu}\Big(\{\, \boldsymbol{x} \mapsto g(\boldsymbol{x}+\boldsymbol{\eta}) : \ \|\boldsymbol{\eta}\|_\infty \leq \varepsilon, \ g \in \mathscr{G}\,\}\Big) \ \leq \ \mathfrak{R}_{n,\nu}(\mathscr{G}) \ + \ \frac{\varepsilon W \, d^{1/q}}{\sqrt{n}}.$$

Lemma 5.5 (Vector contraction)**.**

$$\begin{aligned}\mathbb{E}_\sigma\Big[\sup_h \frac{1}{n}\sum_{i=1}^n \sigma_i\, \ell\big(f_h(\boldsymbol{x}_i),\, g_h(\boldsymbol{x}_i)\big)\Big] \leq 2L_\ell \Bigg\{ & \mathbb{E}_\sigma\Big[\sup_h \frac{1}{n}\sum_{i=1}^n \sigma_i f_h(\boldsymbol{x}_i)\Big] \\ & + \mathbb{E}_\sigma\Big[\sup_h \frac{1}{n}\sum_{i=1}^n \sigma_i g_h(\boldsymbol{x}_i)\Big]\Bigg\}.\end{aligned} \tag{22}$$

5.1.3 McDiarmid's Inequality

Theorem 5.6 (McDiarmid's inequality)**.**

$$|f\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n\right) - f\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i', \ldots, \boldsymbol{x}_n\right)| \leq c_i \tag{23}$$

$$\Pr(f\left(X_1, \ldots, X_n\right) - \mathbb{E}\left[f\left(X_1, \ldots, X_n\right)\right] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right) \tag{24}$$

holds.

5.2 Estimation Error Bound for Supervised TRADES

We first derive a uniform deviation bound and an estimation-error bound for the supervised TRADES objective.

Supervised TRADES risk (population and empirical risks)

$$\begin{aligned}R_{\text{PN-TR}}(g) := \pi_{\text{P}}\mathbb{E}_{\text{P}}\Big[&\ell\big(g(\boldsymbol{x}), +1\big) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \leq \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}), g(\boldsymbol{x}+\boldsymbol{\eta})\big)\Big] \\ &+ \pi_{\text{N}}\mathbb{E}_{\text{N}}\Big[\ell\big(g(\boldsymbol{x}), -1\big) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \leq \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}), g(\boldsymbol{x}+\boldsymbol{\eta})\big)\Big]\end{aligned} \tag{25}$$

$$\begin{aligned}\widehat{R}_{\text{PN-TR}}(g) := \frac{\pi_{\text{P}}}{n_{\text{P}}}\sum_{i=1}^{n_{\text{P}}}&\Big[\ell\big(g(\boldsymbol{x}_i^{\text{P}}), +1\big) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \leq \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}_i^{\text{P}}), g(\boldsymbol{x}_i^{\text{P}}+\boldsymbol{\eta})\big)\Big] \\ &+ \frac{\pi_{\text{N}}}{n_{\text{N}}}\sum_{i=1}^{n_{\text{N}}}\Big[\ell\big(g(\boldsymbol{x}_i^{\text{N}}), -1\big) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \leq \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}_i^{\text{N}}), g(\boldsymbol{x}_i^{\text{N}}+\boldsymbol{\eta})\big)\Big]\end{aligned} \tag{26}$$

$$\widehat{g}_{\text{PN-TR}} := \arg\min_{g \in \mathscr{G}} \widehat{R}_{\text{PN-TR}}(g)$$

$$g^* \in \arg\min_{g\in\mathscr{G}} R_{\text{PN-TR}}(g)$$

Theorem 5.7 (Estimation error bound for supervised TRADES)**.** •
(Bounded losses)
for any $\widehat{y} \in \mathbb{R}$ and $y \in \mathcal{Y}$, $\|g\|_\infty = \sup_{\boldsymbol{x}\in\mathcal{X}} |g(\boldsymbol{x})| \le C_g$ $\ell(\widehat{y}, y) \le C_\ell$ holds.

- ***(Lipschitz continuity of losses)***
- ***(Regularity of the KL term)*** *$\ell_{\text{KL}}(u, v) \le C_{\text{KL}}$ satisfies this condition.*

$$\begin{aligned} R_{\text{PN-TR}}(\widehat{g}_{\text{PN-TR}}) - R_{\text{PN-TR}}(g^*) \ &\le 4\big(L_\ell + 4\beta L_{\text{KL}}\big)\Big(\pi_{\text{P}}\mathfrak{R}_{n_{\text{P}},p_{\text{P}}}(\mathscr{G}) + \pi_{\text{N}}\mathfrak{R}_{n_{\text{N}},p_{\text{N}}}(\mathscr{G})\Big) \\ &\quad + 8\beta L_{\text{KL}}\, \varepsilon W d^{1/q}\left(\frac{\pi_{\text{P}}}{\sqrt{n_{\text{P}}}} + \frac{\pi_{\text{N}}}{\sqrt{n_{\text{N}}}}\right) \\ &\quad + \sqrt{2\ln\frac{2}{\delta}}\,(C_\ell + \beta C_{\text{KL}})\left(\frac{\pi_{\text{P}}}{\sqrt{n_{\text{P}}}} + \frac{\pi_{\text{N}}}{\sqrt{n_{\text{N}}}}\right). \end{aligned} \tag{27}$$

Auxiliary lemmas

Lemma 5.8 (Uniform deviation bound for supervised TRADES)**.**

$$\begin{aligned} \sup_{g\in\mathscr{G}} \left|\widehat{R}_{\text{PN-TR}}(g) - R_{\text{PN-TR}}(g)\right| &\le 2\big(L_\ell + 4\beta L_{\text{KL}}\big)\Big(\pi_{\text{P}}\mathfrak{R}_{n_{\text{P}},p_{\text{P}}}(\mathscr{G}) + \pi_{\text{N}}\mathfrak{R}_{n_{\text{N}},p_{\text{N}}}(\mathscr{G})\Big) \\ &\quad + 4\beta L_{\text{KL}}\, \varepsilon W d^{1/q}\left(\frac{\pi_{\text{P}}}{\sqrt{n_{\text{P}}}} + \frac{\pi_{\text{N}}}{\sqrt{n_{\text{N}}}}\right) \\ &\quad + \sqrt{\frac{1}{2}\ln\frac{2}{\delta}}\,(C_\ell + \beta C_{\text{KL}})\left(\frac{\pi_{\text{P}}}{\sqrt{n_{\text{P}}}} + \frac{\pi_{\text{N}}}{\sqrt{n_{\text{N}}}}\right). \end{aligned} \tag{28}$$

holds.

5.3 Estimation Error Bound for uPU+TRADES

We extend the analysis to the uPU+TRADES objective, which combines unbiased PU risk estimation with the TRADES regularization.

uPU+TRADES risk (population and empirical risks)

$$\begin{aligned} R_{\text{uPU-TR}}(g) := \pi_{\text{P}}\mathbb{E}_{\text{P}}\Big[\tilde{\ell}(g(\boldsymbol{x}), +1) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \le \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}+\boldsymbol{\eta}), g(\boldsymbol{x})\big)\Big] \\ + \mathbb{E}_{\text{U}}\Big[\ell\big(g(\boldsymbol{x}), -1\big) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \le \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}+\boldsymbol{\eta}), g(\boldsymbol{x})\big)\Big] \end{aligned} \tag{29}$$

$$\widehat{R}_{\text{uPU-TR}}(g) := \frac{\pi_{\text{P}}}{n_{\text{P}}}\sum_{i=1}^{n_{\text{P}}}\left[\tilde{\ell}(g(\boldsymbol{x}_i^{\text{P}}), +1) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \le \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}_i^{\text{P}} + \boldsymbol{\eta}), g(\boldsymbol{x}_i^{\text{P}})\big)\right]$$
$$+ \frac{1}{n_{\text{U}}}\sum_{i=1}^{n_{\text{U}}}\left[\ell\big(g(\boldsymbol{x}_i^{\text{U}}), -1\big) + \beta \max_{\|\boldsymbol{\eta}\|_\infty \le \epsilon} \ell_{\text{KL}}\big(g(\boldsymbol{x}_i^{\text{U}} + \boldsymbol{\eta}), g(\boldsymbol{x}_i^{\text{U}})\big)\right] \tag{30}$$

$$\widehat{g}_{\text{uPU-TR}} := \arg\min_{g\in\mathscr{G}} \widehat{R}_{\text{uPU-TR}}(g)$$
$$g^* \in \arg\min_{g\in\mathscr{G}} R_{\text{uPU-TR}}(g)$$

Theorem 5.9 (Estimation error bound for uPU+TRADES)**.** • ***(Bounded losses)*** *There exist constants $C_\ell, C_{\text{KL}} > 0$ such that, for any $y \in \mathcal{Y}$ and any input, the following holds:*

$$\ell_{\text{KL}}(u, v) \le C_{\text{KL}}$$

- ***(Lipschitz continuity of losses)*** *There exist constants $L_\ell, L_{\text{KL}} > 0$ such that,*

$$\begin{aligned} R_{\text{uPU-TR}}(\widehat{g}_{\text{uPU-TR}}) - R_{\text{uPU-TR}}(g^*) \ &\le 8\pi_{\text{P}}\big(L_\ell + 2\beta L_{\text{KL}}\big)\mathfrak{R}_{n_{\text{P}},p_{\text{P}}}(\mathscr{G}) \\ &+ 4\big(L_\ell + 4\beta L_{\text{KL}}\big)\mathfrak{R}_{n_{\text{U}},p_{\text{U}}}(\mathscr{G}) \\ &+ 8\beta L_{\text{KL}}\,\varepsilon W d^{1/q}\left(\frac{\pi_{\text{P}}}{\sqrt{n_{\text{P}}}} + \frac{1}{\sqrt{n_{\text{U}}}}\right) \\ &+ \sqrt{2\ln\frac{2}{\delta}}\left(\frac{\pi_{\text{P}}(2C_\ell + \beta C_{\text{KL}})}{\sqrt{n_{\text{P}}}} + \frac{C_\ell + \beta C_{\text{KL}}}{\sqrt{n_{\text{U}}}}\right). \end{aligned} \tag{31}$$

Auxiliary lemmas

Lemma 5.10 (Uniform deviation bound for uPU+TRADES)**.** *For any $\delta > 0$, with probability at least $1 - \delta$*

$$\begin{aligned} \sup_{g\in\mathscr{G}}\left|\widehat{R}_{\text{uPU-TR}}(g) - R_{\text{uPU-TR}}(g)\right| \ &\le 4\pi_{\text{P}}\big(L_\ell + 2\beta L_{\text{KL}}\big)\mathfrak{R}_{n_{\text{P}},p_{\text{P}}}(\mathscr{G}) \\ &+ 2\big(L_\ell + 4\beta L_{\text{KL}}\big)\mathfrak{R}_{n_{\text{U}},p_{\text{U}}}(\mathscr{G}) \\ &+ 4\beta L_{\text{KL}}\,\varepsilon W d^{1/q}\left(\frac{\pi_{\text{P}}}{\sqrt{n_{\text{P}}}} + \frac{1}{\sqrt{n_{\text{U}}}}\right) \\ &+ \sqrt{\frac{1}{2}\ln\frac{2}{\delta}}\left(\frac{\pi_{\text{P}}(2C_\ell + \beta C_{\text{KL}})}{\sqrt{n_{\text{P}}}} + \frac{C_\ell + \beta C_{\text{KL}}}{\sqrt{n_{\text{U}}}}\right) \end{aligned} \tag{32}$$

(**??**) holds.

5.4 Estimation Error Bound for nnPU+TRADES

We further analyze nnPU+TRADES, which enforces non-negativity to mitigate overfitting caused by negative empirical risk estimates.

nnPU+TRADES risk (population and empirical risks)

$$\psi(g, \boldsymbol{x}) := \max_{\|\boldsymbol{\eta}\|_\infty \le \varepsilon} \ell_{\mathrm{KL}}\big(g(\boldsymbol{x} + \boldsymbol{\eta}),\, g(\boldsymbol{x})\big)$$

$$\begin{aligned} R_{\text{nnPU-TR}}(g) := \pi_{\mathrm{P}}\, \mathbb{E}_{\mathrm{P}}\big[\ell\big(g(\boldsymbol{x}), +1\big) + \beta\, \psi(g, \boldsymbol{x})\big] \\ + \max\Big\{0,\, -\pi_{\mathrm{P}}\, \mathbb{E}_{\mathrm{P}}\big[\ell\big(g(\boldsymbol{x}), -1\big)\big] + \mathbb{E}_{\mathrm{U}}\big[\ell\big(g(\boldsymbol{x}), -1\big)\big]\Big\} \\ + \beta\, \mathbb{E}_{\mathrm{U}}[\psi(g, \boldsymbol{x})] \end{aligned} \tag{33}$$

$$\begin{aligned} \widehat{R}_{\text{nnPU-TR}}(g) := \frac{\pi_{\mathrm{P}}}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \Big[\ell\big(g(\boldsymbol{x}_i^{\mathrm{P}}), +1\big) + \beta\, \psi(g, \boldsymbol{x}_i^{\mathrm{P}})\Big] \\ + \max\left\{0,\, -\frac{\pi_{\mathrm{P}}}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} \ell\big(g(\boldsymbol{x}_i^{\mathrm{P}}), -1\big) + \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \ell\big(g(\boldsymbol{x}_i^{\mathrm{U}}), -1\big)\right\} \\ + \frac{\beta}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \psi(g, \boldsymbol{x}_i^{\mathrm{U}}) \end{aligned} \tag{34}$$

$$\widehat{g}_{\text{nnPU-TR}} := \arg\min_{g \in \mathscr{G}} \widehat{R}_{\text{nnPU-TR}}(g)$$

$$g^* \in \arg\min_{g \in \mathscr{G}} R_{\text{nnPU-TR}}(g)$$

Theorem 5.11 (Estimation error bound for nnPU+TRADES)**.** • ***(Bounded losses)*** *There exist constants $C_\ell, C_{\mathrm{KL}} > 0$ such that, for any $y \in \mathcal{Y}$ and any input, the following holds:*

$\ell_{\mathrm{KL}}(u, v) \le C_{\mathrm{KL}}$

- ***(Lipschitz continuity of losses)*** *There exist constants $L_\ell, L_{\mathrm{KL}} > 0$ such that,*

$$\begin{aligned} R_{\text{nnPU-TR}}(\widehat{g}_{\text{nnPU-TR}}) - R_{\text{nnPU-TR}}(g^*) \le\ & 8\pi_{\mathrm{P}}\big(L_\ell + 2\beta L_{\mathrm{KL}}\big)\mathfrak{R}_{n_{\mathrm{P}}, p_{\mathrm{P}}}(\mathscr{G}) \\ & + 4\big(L_\ell + 4\beta L_{\mathrm{KL}}\big)\mathfrak{R}_{n_{\mathrm{U}}, p_{\mathrm{U}}}(\mathscr{G}) \\ & + 8\beta L_{\mathrm{KL}}\, \varepsilon W d^{1/q} \left(\frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}} + \frac{1}{\sqrt{n_{\mathrm{U}}}}\right) \\ & + \sqrt{2 \ln \frac{2}{\delta}} \left(\frac{\pi_{\mathrm{P}}(2C_\ell + \beta C_{\mathrm{KL}})}{\sqrt{n_{\mathrm{P}}}} + \frac{C_\ell + \beta C_{\mathrm{KL}}}{\sqrt{n_{\mathrm{U}}}}\right). \end{aligned} \tag{35}$$

Auxiliary lemmas

Lemma 5.12 (Uniform deviation bound for nnPU+TRADES)**.** *For any* $\delta > 0$, *with probability at least* $1 - \delta$

$$\begin{aligned}\sup_{g\in\mathscr{G}}\left|\widehat{R}_{\text{nnPU-TR}}(g) - R_{\text{nnPU-TR}}(g)\right| \le\ & 4\pi_{\mathrm{P}}\big(L_\ell + 2\beta L_{\mathrm{KL}}\big)\mathfrak{R}_{n_{\mathrm{P}},p_{\mathrm{P}}}(\mathscr{G}) \\ & + 2\big(L_\ell + 4\beta L_{\mathrm{KL}}\big)\mathfrak{R}_{n_{\mathrm{U}},p_{\mathrm{U}}}(\mathscr{G}) \\ & + 4\beta L_{\mathrm{KL}}\,\varepsilon W d^{1/q}\left(\frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}} + \frac{1}{\sqrt{n_{\mathrm{U}}}}\right) \\ & + \sqrt{\frac{1}{2}\ln\frac{2}{\delta}}\left(\frac{\pi_{\mathrm{P}}(2C_\ell + \beta C_{\mathrm{KL}})}{\sqrt{n_{\mathrm{P}}}} + \frac{C_\ell + \beta C_{\mathrm{KL}}}{\sqrt{n_{\mathrm{U}}}}\right). \end{aligned} \tag{36}$$

5.5 Sufficient Condition on the Number of Unlabeled Samples for PU+TRADES to Outperform Supervised TRADES

Finally, we compare the derived bounds and obtain a sufficient unlabeled-sample size condition ensuring that PU+TRADES yields a tighter estimation-error bound than supervised TRADES.

Rademacher complexity of a linear hypothesis class

$$\mathfrak{R}_{n,\nu}(\mathscr{G}) = \mathbb{E}_{\boldsymbol{x}_{1:n}\sim\nu}\,\mathbb{E}_{\boldsymbol{\sigma}}\Big[\sup_{\|\boldsymbol{w}\|_p\le W}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\boldsymbol{w}^\top\boldsymbol{x}_i\Big] \;\le\; \frac{W\,\sup_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{x}\|_q}{\sqrt{n}} \;\le\; \frac{WC_x\,d^{1/q}}{\sqrt{n}} \tag{37}$$

$$\kappa_\delta := \sqrt{2\ln\frac{2}{\delta}} \tag{38}$$

$$\Gamma_\delta := 4\big(L_\ell + 4\beta L_{\mathrm{KL}}\big)\,WC_x d^{1/q} + 8\beta L_{\mathrm{KL}}\,\varepsilon W d^{1/q} + \kappa_\delta\,(C_\ell + \beta C_{\mathrm{KL}}) \tag{39}$$

(1) Comparing PN+TRADES and uPU+TRADES

$$R_{\text{PN-TR}}(\widehat{g}_{\text{PN-TR}}) - R_{\text{PN-TR}}(g^*) \;\le\; \Gamma_\delta\Big(\frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}} + \frac{\pi_{\mathrm{N}}}{\sqrt{n_{\mathrm{N}}}}\Big) \tag{40}$$

$$R_{\text{uPU-TR}}(\widehat{g}_{\text{uPU-TR}}) - R_{\text{uPU-TR}}(g^*) \;\le\; \frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}}\Big(\Gamma_\delta + 4L_\ell\,WC_x d^{1/q} + \kappa_\delta C_\ell\Big) + \frac{\Gamma_\delta}{\sqrt{n_{\mathrm{U}}}} \tag{41}$$

Main result: a sufficient condition on the number of unlabeled samples

Theorem 5.13 (Sufficient unlabeled-sample condition for PU+TRADES to outperform supervised TRADES)**.** *We compare the estimation-error bounds in Theorems* **??**, **??**, *and* **??**.

$$\Gamma_\delta \frac{\pi_{\mathrm{N}}}{\sqrt{n_{\mathrm{N}}}} \;>\; \frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}} \Big(4L_\ell\, W C_x d^{1/q} + \kappa_\delta C_\ell\Big) \tag{42}$$

(i) PN+TRADES vs uPU+TRADES

$$\boxed{n_{\mathrm{U}} \;>\; \left(\frac{\Gamma_\delta}{\Gamma_\delta \frac{\pi_{\mathrm{N}}}{\sqrt{n_{\mathrm{N}}}} - \frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}}\big(4L_\ell\, W C_x d^{1/q} + \kappa_\delta C_\ell\big)}\right)^2} \tag{43}$$

*The right-hand side of the PN+TRADES bound (Eq. (***??***))*

(ii) PN+TRADES vs nnPU+TRADES

$$\boxed{n_{\mathrm{U}} \;>\; \left(\frac{\Gamma_\delta}{\Gamma_\delta \frac{\pi_{\mathrm{N}}}{\sqrt{n_{\mathrm{N}}}} - \frac{\pi_{\mathrm{P}}}{\sqrt{n_{\mathrm{P}}}}\big(4L_\ell\, W C_x d^{1/q} + \kappa_\delta C_\ell\big)}\right)^2} \tag{44}$$

Table 1 Caption text

Column 1	Column 2	Column 3	Column 4
row 1	data 1	data 2	data 3
row 2	data 4	data 5[1]	data 6
row 3	data 7	data 8	data 9[2]

Source: This is an example of table footnote. This is an example of table footnote.

[1]Example for a first table footnote. This is an example of table footnote.

[2]Example for a second table footnote. This is an example of table footnote.

The input format for the above table is as follows:

```
\begin{table}[<placement-specifier>]
\caption{<table-caption>}\label{<table-label>}%
\begin{tabular}{@{}llll@{}}
\toprule
Column 1 & Column 2 & Column 3 & Column 4\\
```

```
\midrule
row 1 & data 1 & data 2  & data 3 \\
row 2 & data 4 & data 5\footnotemark[1] & data 6 \\
row 3 & data 7 & data 8  & data 9\footnotemark[2]\\
\botrule
\end{tabular}
\footnotetext{Source: This is an example of table footnote.
This is an example of table footnote.}
\footnotetext[1]{Example for a first table footnote.
This is an example of table footnote.}
\footnotetext[2]{Example for a second table footnote.
This is an example of table footnote.}
\end{table}
```

Table 2 Example of a lengthy table which is set to full textwidth

	Element 1[1]			Element 2[2]		
Project	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 4	500 A	961	922 ± 10	900 A	1268	1092 ± 40

Note: This is an example of table footnote. This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote.

[1]Example for a first table footnote.

[2]Example for a second table footnote.

In case of double column layout, tables which do not fit in single column width should be set to full text width. For this, you need to use `\begin{table*} ... \end{table*}` instead of `\begin{table} ... \end{table}` environment. Lengthy tables which do not fit in textwidth should be set as rotated table. For this, you need to use `\begin{sidewaystable} ... \end{sidewaystable}` instead of `\begin{table*} ... \end{table*}` environment. This environment puts tables rotated to single column width. For tables rotated to double column width, use `\begin{sidewaystable*} ... \end{sidewaystable*}`.

6 Figures

As per the LaTeX standards you need to use eps images for LaTeX compilation and `pdf/jpg/png` images for `PDFLaTeX` compilation. This is one of the major difference between LaTeX and `PDFLaTeX`. Each image should be from a single input .eps/vector image file. Avoid using subfigures. The command for inserting images for LaTeX and `PDFLaTeX` can be generalized. The package used to insert images in `LaTeX/PDFLaTeX`

Table 3 Tables which are too long to fit, should be written using the "sidewaystable" environment as shown here

	Element 1[1]			Element[2]		
Projectile	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 4	500 A	961	922 ± 10	900 A	1268	1092 ± 40
Element 5	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 6	500 A	961	922 ± 10	900 A	1268	1092 ± 40

Note: This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote.

[1]This is an example of table footnote.

is the graphicx package. Figures can be inserted via the normal figure environment as shown in the below example:

```
\begin{figure}[<placement-specifier>]
\centering
\includegraphics{<eps-file>}
\caption{<figure-caption>}\label{<figure-label>}
\end{figure}
```

Fig. 3 This is a widefig. This is an example of long caption this is an example of long caption this is an example of long caption this is an example of long caption

In case of double column layout, the above format puts figure captions/images to single column width. To get spanned images, we need to provide `\begin{figure*}` ... `\end{figure*}`.

For sample purpose, we have included the width of images in the optional argument of `\includegraphics` tag. Please ignore this.

7 Algorithms, Program codes and Listings

Packages `algorithm`, `algorithmicx` and `algpseudocode` are used for setting algorithms in LaTeX using the format:

```
\begin{algorithm}
\caption{<alg-caption>}\label{<alg-label>}
\begin{algorithmic}[1]
. . .
\end{algorithmic}
\end{algorithm}
```

You may refer above listed package documentations for more details before setting `algorithm` environment. For program codes, the "verbatim" package is required and the command to be used is `\begin{verbatim} ... \end{verbatim}`.

Similarly, for `listings`, use the `listings` package. `\begin{lstlisting} ... \end{lstlisting}` is used to set environments similar to `verbatim` environment. Refer to the `lstlisting` package documentation for more details.

A fast exponentiation procedure:

```
begin
  for i := 1 to 10 step 1 do
      expt(2, i);
      newline() od                    Comments will be set flush to the right margin
where
proc expt(x, n) ≡
  z := 1;
  do if n = 0 then exit fi;
     do if odd(n) then exit fi;
        comment: This is a comment statement;
        n := n/2; x := x * x od;
     { n > 0 };
     n := n − 1; z := z * x od;
  print(z).
end
```

Algorithm 1 Calculate $y = x^n$

Require: $n \geq 0 \vee x \neq 0$
Ensure: $y = x^n$

$y \Leftarrow 1$
if $n < 0$ **then**
    $X \Leftarrow 1/x$
    $N \Leftarrow -n$
else
    $X \Leftarrow x$
    $N \Leftarrow n$
end if
while $N \neq 0$ **do**
    if N is even **then**
        $X \Leftarrow X \times X$
        $N \Leftarrow N/2$
    else[N is odd]
        $y \Leftarrow y \times X$
        $N \Leftarrow N - 1$
    end if
end while

```
for i:=maxint to 0 do
begin
{ do nothing }
end;
Write('Case-insensitive-');
Write('Pascal-keywords.');
```

8 Cross referencing

Environments such as figure, table, equation and align can have a label declared via the `\label{#label}` command. For figures and table environments use the `\label{}` command inside or just below the `\caption{}` command. You can then use the `\ref{#label}` command to cross-reference them. As an example, consider the label declared for Figure **??** which is `\label{fig1}`. To cross-reference it, use the command `Figure \ref{fig1}`, for which it comes up as "Figure **??**".

To reference line numbers in an algorithm, consider the label declared for the line number 2 of Algorithm **??** is `\label{algln2}`. To cross-reference it, use the command `\ref{algln2}` for which it comes up as line **??** of Algorithm **??**.

8.1 Details on reference citations

Standard LaTeX permits only numerical citations. To support both numerical and author-year citations this template uses `natbib` LaTeX package. For style guidance please refer to the template user manual.

Here is an example for `\cite{...}`: [**?**]. Another example for `\citep{...}`: [**?**]. For author-year citation mode, `\cite{...}` prints Jones et al. (1990) and `\citep{...}` prints (Jones et al., 1990).

All cited bib entries are printed at the end of this article: [**?**], [**?**], [**?**], [**?**], [**?**], [**?**], [**?**], [**?**], [**?**], [**?**] and [**?**].

9 Examples for theorem like environments

For theorem like environments, we require `amsthm` package. There are three types of predefined theorem styles exists—`thmstyleone`, `thmstyletwo` and `thmstylethree`

`thmstyleone`	Numbered, theorem head in bold font and theorem text in italic style
`thmstyletwo`	Numbered, theorem head in roman font and theorem text in italic style
`thmstylethree`	Numbered, theorem head in bold font and theorem text in roman style

For mathematics journals, theorem styles can be included as shown in the following examples:

Theorem 1 (Theorem subhead) *Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text.*

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Proposition 2 *Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text.*

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Example 1 Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem.

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Remark 1 Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem.

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Definition 1 (Definition sub head) Example definition text. Example definition text. Example definition text. Example definition text. Example definition text. Example definition text. Example definition text. Example definition text.

Additionally a predefined “proof” environment is available: `\begin{proof} ... \end{proof}`. This prints a “Proof” head in italic font style and the “body text” in roman font style with an open square at the end of each proof environment.

Proof Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. $\square$

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Proof of Theorem **??** Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. □

For a quote environment, use `\begin{quote}...\end{quote}`

> Quoted text example. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text (refer Figure **??**). Sample body text. Sample body text. Sample body text (refer Table **??**).

10 Methods

Topical subheadings are allowed. Authors must ensure that their Methods section includes adequate experimental and characterization data necessary for others in the field to reproduce their work. Authors are encouraged to include RIIDs where appropriate.

Ethical approval declarations (only required where applicable) Any article reporting experiment/s carried out on (i) live vertebrate (or higher invertebrates), (ii) humans or (iii) human samples must include an unambiguous statement within the methods section that meets the following requirements:

1. Approval: a statement which confirms that all experimental protocols were approved by a named institutional and/or licensing committee. Please identify the approving body in the methods section
2. Accordance: a statement explicitly saying that the methods were carried out in accordance with the relevant guidelines and regulations
3. Informed consent (for experiments involving humans or human tissue samples): include a statement confirming that informed consent was obtained from all participants and/or their legal guardian/s

If your manuscript includes potentially identifying patient/participant information, or if it describes human transplantation research, or if it reports results of a clinical trial then additional information will be required. Please visit (https://www.nature.com/nature-research/editorial-policies) for Nature Portfolio journals, (https://www.springer.com/gp/authors-editors/journal-author/journal-author-helpdesk/publishing-ethics/14214) for Springer Nature journals, or (https://www.biomedcentral.com/getpublished/editorial-policies#ethics+and+consent) for BMC.

11 Discussion

Discussions should be brief and focused. In some disciplines use of Discussion or 'Conclusion' is interchangeable. It is not mandatory to use both. Some journals prefer a section 'Results and Discussion' followed by a section 'Conclusion'. Please refer to Journal-level guidance for any specific requirements.

12 Conclusion

Conclusions may be used to restate your hypothesis or research question, restate your major findings, explain the relevance and the added value of your work, highlight any limitations of your study, describe future directions for research and recommendations.

In some disciplines use of Discussion or 'Conclusion' is interchangeable. It is not mandatory to use both. Please refer to Journal-level guidance for any specific requirements.

Supplementary information. If your article has accompanying supplementary file/s please state so here.

Authors reporting data from electrophoretic gels and blots should supply the full unprocessed scans for key as part of their Supplementary information. This may be requested by the editorial team/s if it is missing.

Please refer to Journal-level guidance for any specific requirements.

Acknowledgements. Acknowledgements are not compulsory. Where included they should be brief. Grant or contribution numbers may be acknowledged.

Please refer to Journal-level guidance for any specific requirements.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval and consent to participate
- Consent for publication
- Data availability
- Materials availability
- Code availability
- Author contribution

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Editorial Policies for:

Springer journals and proceedings: https://www.springer.com/gp/editorial-policies

Nature Portfolio journals: https://www.nature.com/nature-research/editorial-policies

Scientific Reports: https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals: https://www.biomedcentral.com/getpublished/editorial-policies

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.