Journal of Machine Learning Research 27 (2026) 1-58 Submitted 8/25; Revised 7/26; Published 8/26

Locally Private Estimation with Public Features

Yuheng Ma YHMA@SFS.ECNU.EDU.CN
KLATASDS-MOE, School of Statistics
East China Normal University
200062 Shanghai, China

Hanfang Yang* HYANG@RUC.EDU.CN
Center for Applied Statistics and School of Statistics
Renmin University of China
100872 Beijing, China

Ke Jia* JIAKE1999@RUC.EDU.CN
School of Statistics
Renmin University of China
100872 Beijing, China
* Corresponding authors*

Editor: Raef Bassily

Abstract

We initiate the study of locally differentially private (LDP) learning with public features. We define semi-feature LDP, where some features are publicly available while the remaining ones, along with the label, require protection under local differential privacy. Under semi-feature LDP, we consider three fundamental estimation problems: non-parametric density estimation, classification, and regression. Given the smoothness assumption, we show that the minimax convergence rate is significantly improved compared to classical LDP. Then, we propose `HistOfTree`, an estimator that fully leverages the information contained in both public and private features. Theoretically, `HistOfTree` reaches the minimax optimal convergence rate. Empirically, `HistOfTree` achieves superior performance on both synthetic and real data. We also explore scenarios where users have the flexibility to select features for protection manually. In such cases, we propose an estimator and a data-driven parameter tuning strategy, leading to analogous theoretical and empirical results.

Keywords: local differential privacy, public features, non-parametric estimation, regression, classification, density estimation

1. Introduction

Data privacy regulations such as Europe's General Data Protection Regulation (GDPR) (European Parliament and Council of the European Union, 2016) have led to a notable rise in the significance of privacy-preserving machine learning (Cummings and Desai, 2018). As a golden standard, local differential privacy (LDP) (Kairouz et al., 2014; Duchi et al., 2018), a variant of differential privacy (DP) (Dwork et al., 2006), has gained considerable attention in recent years, especially among industry experts (Erlingsson et al., 2014; Apple, 2017). LDP assumes that each sample is possessed by a data holder, who privatizes their

data before it is collected by the curator. Offering a stronger sense of privacy protection compared to central DP, LDP often encounters much more sophisticated challenges (Duchi et al., 2018), which obstruct both the theoretical analysis and practical implementation of LDP learning.

Fortunately, in some scenarios, a protocol with weakened protection can be adopted to privatize only the sensitive part of each sample. For instance, label differential privacy (Ghazi et al., 2021; Malek Esmaeili et al., 2021; Badanidiyuru Varadaraja et al., 2024) assumes only the labels contain sensitive information, while the features are freely accessible to the data users. Moreover, clear theoretical advantages have been established (Wang and Xu, 2019; Xu et al., 2023) for label LDP over classical LDP. In the context of central DP, recent works demonstrate the effectiveness of partially protecting a small portion of features (Curmei et al., 2023; Shen et al., 2023; Krichene et al., 2023; Chua et al., 2024), mostly from a methodological perspective. Yet, no theoretical framework has been established under this setting, which is the primary aim of this work.

As is addressed by GDPR, a properly functioning consent management process should provide granular consent options regarding the specific types of data being collected and the purposes for which it will be processed (Nouwens et al., 2020). In our case, to meet the requirements, users must be able to choose which parts of their data are collected under privacy protections and which are not. This necessitates the design of estimators with personalized privacy preferences, i.e., the private and public features for each user are manually selected rather than preset.

Given this context, we study the problem of statistical estimation under local differential privacy with user-personalized public features. We start with the most basic task of non-parametric distribution density estimation. Under the same heuristic, we solve non-parametric regression and classification, the two most common supervised learning problems. Our contributions are summarized as follows:

- We formalize the problem of LDP learning with public features. Specifically, we define semi-feature LDP, where a portion of the features is publicly available, while the remaining ones, along with the label (if there is any), require protection. We develop both a simplified aligned case (where the position of public features for all users is identical) and a general personalized case (where users select their own public features).

- Under semi-feature LDP, we provide theoretical minimax lower bounds for each considered problem, namely, density estimation, regression, and classification.

- We propose the `HistOfTree` estimator. Theoretically, we provide an excess risk upper bound for personalized privacy preferences. Together with the lower bound, we establish a minimax optimal convergence rate, up to logarithmic factors, for both aligned and personalized privacy preferences. We also provide a data-driven parameter selection rule to avoid tuning sensitive hyper-parameters.

- We demonstrate the superiority of `HistOfTree` through comparison of each task on both synthetic and real data sets. `HistOfTree` not only outperforms naive competitors but is also shown to be free of tuning hyper-parameters.

Section or result	AISTATS version	This version
Section 4	None	Adds density lower and upper bounds and a weighted estimator for aligned and personalized preferences.
Section 5 Theorems 5 and 6	An aligned lower bound and an unmatched personalized upper bound.	Extends the regression lower bound to personalized preferences and introduces inverse-variance weighting, yielding matching personalized lower and upper bounds up to logarithmic factors.
Section 5 Theorems 7 and 8	None	Adds classification results parallel to regression, with matching lower and upper bounds for aligned and personalized preferences under the margin condition.
Sections 6.2.2; Figures 3, 2(e), and 2(f)	None	Adds density simulations and regression experiments on the robustness of s and public–private feature correlation.

Table 1: Major changes between the conference version and the journal version.

In the following sections, we first introduce related work in Section 2. Next, we define the concept of semi-feature LDP in Section 3. For the unsupervised problem of density estimation, we present theoretical results and our proposed algorithm in Section 4. Similarly, we provide a parallel set of results for the supervised problems of classification and regression in Section 5. Experimental results are detailed in Section 6. Section 7 concludes the paper. All proofs and data set details are included in Appendices A and B, respectively.

A preliminary version of this paper was presented at the 28th International Conference on Artificial Intelligence and Statistics (Ma et al., 2025). Significant extensions have been made to the conference version, including the following key contributions:

- By introducing an inverse-variance weighted aggregation rule, we close the gap between the original personalized minimax lower bound and the high-probability upper bound left open in Ma et al. (2025).
- Under semi-feature LDP, we significantly broaden the problem scope, extending from non-parametric regression, a showcase example, to general non-parametric problems, including classification and density estimation. We provide parallel results comprising a minimax lower bound, an estimator, and the corresponding upper bound for the algorithm.
- We present new experimental results with a focus on density estimation.

2. Related Work

Public data and features There has been a long line of work on the benefits of public data (Papernot et al., 2017, 2018; Liu et al., 2021a,b; Yu et al., 2021, 2022; Nasr et al.,

2023; Gu et al., 2025; Fuentes et al., 2024). Public features, by contrast, have received less attention. Recently, several works have considered settings in which only a subset of features is sensitive (Curmei et al., 2023; Shen et al., 2023; Krichene et al., 2023; Chua et al., 2024). In recommendation and personalization, public item-side features are combined with private user information (Curmei et al., 2023; Krichene et al., 2023), whereas other works explicitly partition each record into sensitive and non-sensitive features (Shen et al., 2023; Chua et al., 2024; Ghazi et al., 2024; Mahloujifar et al., 2025; Jiang et al., 2026). Our setting is closely related to this line of work, but differs in that privacy is local and the set of private coordinates may vary across users.

An even more extreme setting is label differential privacy (Chaudhuri and Hsu, 2011; Busa-Fekete et al., 2021; Ghazi et al., 2021; Malek Esmaeili et al., 2021; Cunningham et al., 2022; Ghazi et al., 2022; Badanidiyuru Varadaraja et al., 2024; Zhao et al., 2025), where all features are treated as public. Under central DP, this relaxation has been shown to improve generalization error, with the privacy budget affecting only the constant factor (Badanidiyuru Varadaraja et al., 2024). Under LDP, however, the advantage of label LDP over standard LDP is more pronounced. Such a generalization gap has been established in both parametric (Wang and Xu, 2019) and non-parametric (Xu et al., 2023) settings.

Personalized privacy Our personalized privacy preferences exhibit heterogeneity w.r.t. both users and attributes, i.e., different attributes of the same user can be either private or public, and different users have different choices of which features to protect. Various works have studied the problem of learning under heterogeneous privacy preference w.r.t. users. Wang et al. (2015); Yang et al. (2021); Li et al. (2022) considered locally private learning, with each data holder possessing their own privacy budget. Sun et al. (2014); Song et al. (2019); Zhang et al. (2022a, 2024) assume that the privacy budget of each sample is related to its value. These studies neither provide theoretical guarantees nor cover our problem setting since there is no heterogeneity w.r.t. different features in their case.

Some related notions provide privacy guarantees at the attribute level. Ghazi et al. (2023) study partial DP in the central model, allowing individual attributes to receive stronger privacy guarantees than the complete record. Zhang et al. (2022b) propose attribute privacy, which protects sensitive global properties of a data set or of its underlying distribution. These notions differ from our local, individual-level constraint, which protects each user's selected private coordinates and label while treating the remaining coordinates as public features. Similar to our aligned setting, Amorino and Gloter (2025) proposed Componentwise LDP, where each attribute is released through independent private channels with different budgets. This is a more restrictive case than ours, as we require private features to be jointly protected. They also do not consider heterogeneous privacy w.r.t. users. Lu et al. (2026) also considered protecting different features separately. Moreover, their analysis does not allow for public features, i.e., features with infinitely large budgets. More recently, Aliakbarpour et al. (2025) proposed a Bayesian Coordinate LDP framework, which is more general than ours. This framework also presents a promising approach to addressing potential risks arising from inner correlations between features, a topic not covered in this paper.

3. Semi-Feature Local Differential Privacy

3.1 Notation

Throughout this paper, we use the notation $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ to denote that there exist positive constant n_1, c and c' such that $a_n \leq cb_n$ and $a_n \geq c'b_n$, for all $n \geq n_1$. In addition, we denote $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. Let $[n] = \{1, \ldots, n\}$. Let $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For any vector x, let x^i denote the i-th element of x. Recall that for $1 \leq T < \infty$, the L_p-norm of $x = (x^1, \ldots, x^d)$ is defined by $\|x\|_p := (|x^1|^p + \cdots + |x^d|^p)^{1/p}$. Let $x_{i:j} = (x_i, \cdots, x_j)$ be the slicing view of x in the $i, \cdots, j$ position. For any set $A \subset \mathbb{R}^d$, the diameter of A is defined by $\mathrm{diam}(A) := \sup_{x,x' \in A} \|x - x'\|_2$. Let σ denote the generation of a sigma algebra. Let $A \times B$ be the Cartesian product of sets where $A \in \mathcal{X}_1$ and $B \in \mathcal{X}_2$. For measure P on X_1 and Q on $\mathcal{X}_2$, define the product measure $\mathrm{P} \otimes \mathrm{Q}$ on $\mathcal{X}_1 \times \mathcal{X}_2$ as $\mathrm{P} \otimes \mathrm{Q}(A \times B) = \mathrm{P}(A) \cdot \mathrm{Q}(B)$. For an integer k, denote the k-fold product measure on $\mathcal{X}_1^k$ as P^k.

3.2 Definition

We formally set up private learning with public features. Definitions are made assuming there is a label Y and a feature X, while unsupervised learning (there is no Y) can be indicated by simply ignoring Y. For each (X_i, Y_i), the label Y_i and a subset of features of X_i require privacy protection, while the remaining features can be released freely. Let $W_i^\ell \in \{0, 1\}$ be the indicator of whether the ℓ-th feature of X_i should be protected. We assume that the privacy mask W is independent of the data (X, Y); the indicators W_i^ℓ need not be protected and are publicly known. The potential risk and solutions for such an assumption are discussed in Section 7. We consider the aligned (resp. personalized) case, where the private features are identical (resp. different) for each sample. The corresponding W are illustrated in Figure 1. Denote the number of private features of X_i by s_i^*.

With a slight abuse of notation, we write $(X_i, Y_i) = (X_i^{\mathrm{priv}}, X_i^{\mathrm{pub}}, Y_i)$, where X_i^{priv} and X_i^{pub} are the collection of private and public features for X_i, respectively. Let $\mathcal{X}_i^{\mathrm{priv}}$ and $\mathcal{X}_i^{\mathrm{pub}}$ be their corresponding domain space. Note that X_i^{priv} and X_i^{pub} are not necessarily ordered. For instance, we allow $X_i^{\mathrm{priv}} = (X_i^1, X_i^3)$ and $X_i^{\mathrm{pub}} = (X_i^2, X_i^4)$. Given the notations, we rigorously define our privacy constraint.

Definition 1 (Semi feature local differential privacy) *For each $i = 1, \ldots, n$, a privatized report Z_i is released based on (X_i, Y_i) and the previous reports $Z_1, \ldots, Z_{i-1}$. The report is a random variable on $\mathcal{S}$. Let $\sigma(\mathcal{S})$ be the σ-field on $\mathcal{S}$. Z_i is drawn conditional on (X_i, Y_i) and $Z_{1:i-1}$ via the distribution $\mathrm{R}\left(S \mid X_i^{priv}, X_i^{pub}, Y_i, Z_{1:i-1}\right)$ for $S \in \sigma(\mathcal{S})$. Then the mechanism $\mathrm{R} = \{\mathrm{R}_i\}_{i=1}^n$ is ε-semi feature local differential privacy (ε-semi feature LDP) if for all $1 \leq i \leq n, S \in \sigma(\mathcal{S})$, all $x^{priv}, x^{priv\prime} \in \mathcal{X}_i^{priv}$, $x^{pub} \in \mathcal{X}_i^{pub}$, $y, y' \in \mathcal{Y}$, and $z_{1:i-1} \in \mathcal{S}^{i-1}$, there holds*

$$\frac{\mathrm{R}_i\left(S \mid x^{priv}, x^{pub}, y, z_{1:i-1}\right)}{\mathrm{R}_i\left(S \mid x^{priv\prime}, x^{pub}, y', z_{1:i-1}\right)} \leq e^{\varepsilon} \tag{1}$$

Moreover, if for all $1 \le i \le n, S \in \sigma(\mathcal{S})$, *all* $x^{priv}, x^{priv\prime} \in \mathcal{X}_i^{priv}$, $x^{pub} \in \mathcal{X}_i^{pub}$, $y, y' \in \mathcal{Y}$, *there holds*

$$\frac{\mathrm{R}_i\left(S \mid x^{priv}, x^{pub}, y\right)}{\mathrm{R}_i\left(S \mid x^{priv\prime}, x^{pub}, y'\right)} \le e^{\varepsilon}, \tag{2}$$

then R *is non-interactive* ε*-semi feature LDP. Here, we let* $0/0 = 1$.

Similar definitions are proposed under central DP setting (Shen et al., 2023; Krichene et al., 2023; Chua et al., 2024). Semi-feature LDP requires individuals to guarantee their own privacy by considering the likelihood ratio of each $(X_i^{\mathrm{priv}}, Y_i)$ as in (1) and (2). The value of X_i^{pub} is regarded as known.

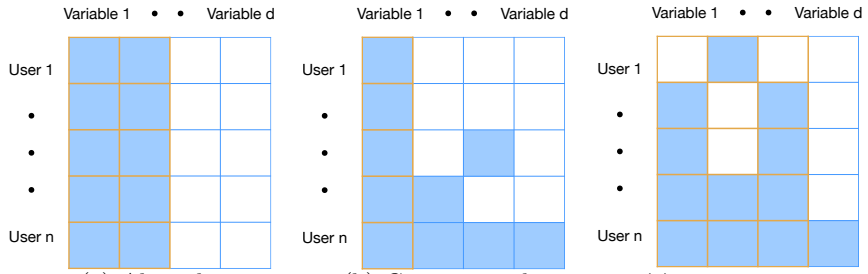
If $s_i^* = d, i \in [n]$, i.e. all features are private, our definition reduces to the classical LDP (Kairouz et al., 2014; Duchi et al., 2018), where non-parametric estimations have been well-explored (Berrett and Butucea, 2019; Berrett et al., 2021; Györfi and Kroll, 2025; Sart, 2023). If $s_i^* = 0, i \in [n]$, i.e. only the labels are private, the definition reduces to label local differential privacy (Busa-Fekete et al., 2021; Cunningham et al., 2022), which is shown to be an easier problem than pure LDP (Wang and Xu, 2019; Xu et al., 2023; Zhao et al., 2024). Our definition explores the intermediate phase, which is practically meaningful (Krichene et al., 2023; Shen et al., 2023; Chua et al., 2024).

We clarify the conditional nature of the semi-feature LDP guarantee under public–private feature correlations. Semi-feature LDP protects $U = (X_i^{\mathrm{priv}}, Y_i)$ conditional on the disclosed public coordinates $V = X_i^{\mathrm{pub}}$. Equivalently, for any prior distribution on (U, V), including arbitrarily correlated distributions, the privatized report changes posterior odds between two candidate private values by at most a multiplicative factor e^{ε} once V is known. In the non-interactive case, for any event $S \in \sigma(\mathcal{S})$ and any values for which the following ratios are well-defined, (2) implies

$$e^{-\varepsilon} \le \frac{\mathrm{P}(U = u \mid V = v, Z \in S)/\mathrm{P}(U = u' \mid V = v, Z \in S)}{\mathrm{P}(U = u \mid V = v)/\mathrm{P}(U = u' \mid V = v)} \le e^{\varepsilon}.$$

Thus, the guarantee controls only the additional Bayesian evidence contributed by Z after X_i^{pub} is already known. It does not control the information conveyed by X_i^{pub} itself, namely the update from $\mathrm{P}(U)$ to $\mathrm{P}(U \mid V = v)$. Therefore, public coordinates should be interpreted as coordinates already available to the analyst or explicitly authorized for exact disclosure, rather than as coordinates that are necessarily uninformative about private coordinates. If inference through correlations with public coordinates must also be protected, those correlated coordinates should be included in the protected set, or a prior-aware privacy definition such as Bayesian coordinate LDP (Aliakbarpour et al., 2025) should be adopted.

A further caveat is that the privacy mask W may be correlated with the realized data. Since our semi-feature LDP guarantee conditions on W, it does not cover information revealed by the mask choice itself. For example, choosing to hide a coordinate associated with a medical condition may already signal evidence about that condition. We discuss in Section 7 how such leakage can be mitigated methodologically in our setting, but a general privacy principle for value-dependent masks remains unexplored.



(a) Aligned (b) Concentrated (c) Heavy tailed

Figure 1: Illustration of different W, where blue means $W_i^\ell = 1$. In the aligned case (a), all users protect the first two features. In the personalized case, users specify different features, with the protected features being concentrated in (b) and spread in (c). The yellow boundaries represent the s selected private features.

3.3 Privacy Preference: Aligned and Personalized

The previous literature (Shen et al., 2023; Krichene et al., 2023; Chua et al., 2024) assumes that features needing protection are aligned, i.e. $W_i^\ell = W_{i'}^\ell$ for $i, i' \in [n]$. This pattern is illustrated in Figure 1(a). This assumption is natural because these studies consider central differential privacy, where all data are already collected. For the curator, the decision to protect or not protect a feature should be based on its potential benefits, regardless of the actual value of the data itself. Public features are typically pre-determined, such as those already known by adversaries or those with no clear privacy risks (e.g., properties of purchased items in recommendation systems). In this work, we consider the aligned case a useful special case and investigate it under LDP.

In practice, it is unlikely that all users share the same preferences regarding which features' privacy to prioritize. For instance, while most people may consider their favorite sports as insensitive information, some individuals feel uneasy about releasing such details due to concerns about targeted advertising, particularly from e-commerce platforms. We consider W_i to be different for each i. We assume the privacy budget ε is identical across users. This pattern is illustrated in Figure 1(b) and 1(c). To capture such privacy preferences, one can design simple surveys before collecting the actual data, for example, using hierarchical buttons similar to those in cookie consent prompts.

Although we assume that W is independent and publicly known, its distribution is worth investigating. Figures 1(b) and 1(c) illustrate two different patterns of W, respectively. In Figure 1(b), most users prioritize protecting the first feature, with the others being chosen less frequently. A W with this concentrated behavior will resemble the aligned case. In Figure 1(c), different users have varying preferences, although there is a small group of features that cover most of the masked positions. In this case, directly applying methods from the aligned case would likely result in poor performance.

4. A First Start: Distribution Density Estimation

In this section, we start with the task of distribution density estimation. In Section 4.1, we formulate the problem and establish minimax lower and upper bounds that match, up to logarithmic factors, under both aligned and personalized privacy preferences. In Section 4.2, we give our estimator for the aligned privacy preference. In Section 4.3, we give our estimator for the personalized privacy preference. In Section 4.4, we show how to automatically tune the hyper-parameters.

4.1 A Minimax Lower Bound

Let p_X be the density function of some probability measure P_X. The goal of density estimation is to estimate the true density function $p_X^*(x)$ at some point x based on a group of $D := \{X_i\}_{i=1}^n$ consisting of n i.i.d. observations drawn from the unknown probability measure P_X^* on $\mathcal{X} = [0,1]^d$. We evaluate L_1 loss function which is $\|p_X^* - \widehat{p}_X\|_{L_1} = \int_x |p_X^*(x) - \widehat{p}_X(x)| d\mathrm{P}_X^*(x)$ for estimates $\widehat{p}_X$. We first present a necessary assumption on the distribution P_X^*, which is a standard condition widely used in non-parametric statistics.

Assumption 1 *Let $\alpha \in (0,1]$. Assume $p_X^* : \mathcal{X} \to \mathbb{R}$ is α-Hölder continuous, i.e. there exists $c_L > 0$ s.t. for all $x_1, x_2 \in \mathcal{X}$, $|p_X^*(x_1) - p_X^*(x_2)| \le c_L \|x_1 - x_2\|^\alpha$.*

We present the following theorem that specifies the minimax lower bound with semi-feature LDP under the above assumptions. The proof is based on first constructing a finite class of hypotheses and then applying information inequalities for local privacy (Duchi et al., 2018) and Assouad's Lemma (Tsybakov, 2009).

Theorem 2 *Denote the function class satisfying Assumption 1 by $\mathcal{F}$. Define the quantity $\mathcal{E}^*$ by*

$$\mathcal{E}^* = \arg\max_{\mathcal{E}} \left(\sum_{i=1}^n \mathcal{E}^{\frac{2\alpha + d + \sum_{\ell=1}^d W_i^\ell}{\alpha}} \le \frac{1}{(e^\varepsilon - 1)^2} \right). \tag{3}$$

Then for any estimator $\tilde{p}_X$ that is sequentially-interactive ε-semi-feature LDP, there holds

$$\inf_{\tilde{p}_X} \sup_{p_X^* \in \mathcal{F}} \mathbb{E}_{\mathrm{P}_X} \left[\|p_X^* - \tilde{p}_X\|_{L_1} \right] \gtrsim \mathcal{E}^* \vee n^{-\frac{\alpha}{2\alpha+d}}. \tag{4}$$

This implies that when privacy preference is aligned, i.e. $W_i^\ell = 1$ if $\ell \le s^*$ and $W_i^\ell = 0$ otherwise, there holds

$$\inf_{\tilde{p}_X} \sup_{p_X^* \in \mathcal{F}} \mathbb{E}_{\mathrm{P}_X} \left[\|p_X^* - \tilde{p}_X\|_{L_1} \right] \gtrsim \left(n(e^\varepsilon - 1)^2 \right)^{-\frac{\alpha}{2\alpha + d + s^*}} \vee n^{-\frac{\alpha}{2\alpha+d}}. \tag{5}$$

Both lower bounds (4) and (5) each comprise two terms. The second term, $n^{-\frac{\alpha}{2\alpha+d}}$, corresponds to the classical minimax lower bound for non-private learners under Assumption 1. See e.g. Tsybakov (2009). Turning to the first term, (4) captures the personalized privacy preference and contains (5) as a special instance. We focus our discussion on (5) and postpone the comparison of (4) with the upper bound. When ε is large enough that

$(e^\varepsilon - 1) \gtrsim n^{\frac{s^*}{4\alpha+2d}}$, the first term in (5) is negligible compared to the second term. This scenario arises when the privacy level does not substantially impair the estimator. For constant level ε, the first term approaches the second term as s^* becomes smaller, as the learning essentially becomes non-private for $s^* = 0$. If $s^* = d$, our findings reduce to the LDP non-parametric density estimation setup that uses only private data. Previous work (Butucea et al., 2020) has constructed estimators achieving the minimax optimal convergence rate. Our results reveal the optimal behavior in the intermediate range. The advantage provided by public features grows at the rate $\frac{\alpha}{2\alpha+2d-(d-s^*)}$, which becomes increasingly substantial for larger $d - s^*$.

Technically, the above difference comes from the information inequality used in the Assouad argument. In the classical non-private lower bound, the distinguishability of two neighboring hypotheses is controlled directly by their raw KL divergence. For a bump construction with resolution k^{-1}, the bump height is of order $k^{-\alpha}$, while the mass of an active d-dimensional cell is of order k^{-d}. Thus, the KL divergence contributed by one raw sample is of order $k^{-2\alpha-d}$, leading to the usual balance $nk^{-2\alpha-d} \asymp 1$, and hence the non-private rate $n^{-\frac{\alpha}{2\alpha+d}}$. Under semi-feature LDP, the proof applies the information contraction inequality for locally private channels (Duchi et al., 2018). Conditional on the public features X_i^{pub}, the released variable Z_i is ε-LDP only with respect to the protected features X_i^{priv}. Consequently, for two neighboring hypotheses P^{σ_1} and P^{σ_2}, the KL divergence between the privatized conditional experiments is bounded by the squared conditional total variation distance,

$$KL\left(\left(\mathrm{R}_i\mathrm{P}^{\sigma_1}\right)_{Z_i|X_i^{\text{pub}}} \middle\| \left(\mathrm{R}_i\mathrm{P}^{\sigma_2}\right)_{Z_i|X_i^{\text{pub}}}\right) \lesssim (e^\varepsilon - 1)^2 V^2\left(\mathrm{P}^{\sigma_1 \text{priv}}, \mathrm{P}^{\sigma_2 \text{priv}}\right),$$

where V denotes total variation distance. This inequality is the source of the distinction between the ambient dimension and the private dimension. For user i, the conditional total variation over the s_i^* protected coordinates is of order $k^{-\alpha-s_i^*}$, whose square gives $k^{-2\alpha-2s_i^*}$. This conditional quantity is then averaged over the public coordinates, whose active cell mass is of order $k^{-(d-s_i^*)}$. Therefore, the KL contribution of user i after privatization is of order

$$(e^\varepsilon - 1)^2 k^{-2\alpha-2s_i^*} k^{-(d-s_i^*)} = (e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i^*}.$$

Hence the ambient dimension d still appears through the non-parametric cell mass, while only the private dimension s_i^* appears in the additional privacy contraction. In the aligned case, $s_i^* = s^*$ for all i, and the Assouad balance becomes $n(e^\varepsilon - 1)^2 k^{-2\alpha-d-s^*} \asymp 1$, which yields the first term in (5). In the personalized case, the same calculation is summed over users with heterogeneous s_i^*, leading to the quantity $\mathcal{E}^*$ in (3).

Theorem 3 *Let Assumption 1 hold. Suppose $0 < \varepsilon \le 1$. Let $s_i^* := \sum_{\ell=1}^d W_i^\ell$ and define*

$$\overline{\mathcal{E}}^* := \arg\max_{\mathcal{E}} \left\{ \sum_{i=1}^n \mathcal{E}^{\frac{2\alpha+d+s_i^*}{\alpha}} \le \frac{\log n}{\varepsilon^2} \right\}. \tag{6}$$

Then there exists a non-interactive ε-semi-feature LDP estimator $\tilde{p}_X$ such that, with probability at least $1 - 2/n^2$,

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \overline{\mathcal{E}}^* \vee \left(\frac{\log n}{n}\right)^{\frac{\alpha}{2\alpha+d}}. \tag{7}$$

If the privacy preference is aligned, i.e., $W_i^\ell = 1$ for $\ell \le s^*$ and $W_i^\ell = 0$ otherwise, then

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \left(\frac{\log n}{n\varepsilon^2}\right)^{\frac{\alpha}{2\alpha+d+s^*}} \vee \left(\frac{\log n}{n}\right)^{\frac{\alpha}{2\alpha+d}}.$$

The upper bound in Theorem 3 has the same two-term structure as the lower bound in Theorem 2. The non-private terms match up to a logarithmic factor. For the privacy-dependent terms, (6) and (3) involve the same positive moment of the heterogeneous privacy counts s_i^*. Their thresholds differ only by the factor $\log n$ and by replacing $(e^\varepsilon - 1)^2$ with ε^2. Since $e^\varepsilon - 1 \asymp \varepsilon$ for $0 < \varepsilon \le 1$, the personalized upper and lower bounds therefore match up to logarithmic factors. The same conclusion holds in the aligned case, where $s_i^* = s^*$ for every user. Moreover, the upper bound is attained by a non-interactive estimator, whereas the lower bound allows the larger class of sequentially interactive mechanisms. Finally, in the aligned setting the aggregation weights introduced below are uniform, so the weighted estimator reduces to the original aligned construction.

4.2 Methodology for Aligned Privacy Preference

In this section, we introduce the `HistOfTree` density estimator under the aligned private features.

4.2.1 Privacy Mechanism

We adopt a partition-based estimator, which creates disjoint partitions and estimates density based on empirical density in each grid. Specifically, let $\pi^{\text{priv}} = \{A_j\}$ be a partition of $\mathcal{X}^{\text{priv}}$, i.e. $\cup A_j = \mathcal{X}^{\text{priv}}$ and $A_{j_1} \cap A_{j_2} = \emptyset$, $j_1 \ne j_2$. Similarly, $\pi^{\text{pub}} = \{B_k\}$ on $\mathcal{X}^{\text{pub}}$. The partition-based estimator (which is non-private) of x in $A_j \times B_k$ is

$$\widehat{p}(x) = \frac{\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{A_j}(X_i^{\text{priv}})\mathbf{1}_{B_k}(X_i^{\text{pub}})}{\mu(A_j \times B_k)},$$

where $\mu(\cdot)$ is the Lebesgue measure. The numerator represents the empirical probability mass in $A_j \times B_k$, while the denominator is the volume of $A_j \times B_k$. Thus, their division estimates the density. To perform such an estimation, two pieces of information are necessary from each data holder: $\mathbf{1}_{A_j}(X_i^{\text{priv}})$ and $\mathbf{1}_{B_k}(X_i^{\text{pub}})$. The first one requires protection, while the second is regarded as publicly available. We use the randomized response mechanism (Warner, 1965) and let

$$\tilde{U}_i^j = \begin{cases} C_\varepsilon\left(\mathbf{1}_{A_j}(X_i^{\text{priv}}) - \frac{1}{1+e^{\varepsilon/2}}\right) & \text{w.p. } \frac{e^{\varepsilon/2}}{1+e^{\varepsilon/2}}, \\ C_\varepsilon\left(\mathbf{1}_{A_j^c}(X_i^{\text{priv}}) - \frac{1}{1+e^{\varepsilon/2}}\right) & \text{w.p. } \frac{1}{1+e^{\varepsilon/2}}, \end{cases} \tag{8}$$

where $C_\varepsilon = \frac{e^{\varepsilon/2}+1}{e^{\varepsilon/2}-1}$. Here, A_j^c denotes the complement of A_j. Note that $\mathbb{E}_{\text{R}}\left[\tilde{U}_i^j\right] = \mathbf{1}_{A_j}(X_i^{\text{priv}})$ and thus we have an unbiased estimator of $\mathbf{1}_{A_j}(X_i^{\text{priv}})$.

4.2.2 Partition

To formalize the private estimator, we also need to formalize the partitions π^{priv} and π^{pub}. The choice of π^{priv} is restrictive due to privacy constraints. We adopt the histogram

partition, a classical technique in LDP learning (Berrett et al., 2021; Györfi and Kroll, 2025). Specifically, consider $0 = a_0 < a_1 < \cdots < a_t = 1$, an equal length partition of $[0,1]$. Then a histogram partition of $\mathcal{X}^{\text{priv}} = [0,1]^{s^*}$ is $\left\{\otimes_{k=1}^{s^*}[a_{\sigma_k}, a_{\sigma_k+1}) | 0 \le \sigma_k \le t\right\}$. For π^{pub}, we can fully utilize the information contained in the data set. Given the public feature vectors X_i^{pub}, various candidate methods are available depending on the sample size, computational power, and prior information. We choose a decision tree partition for its merit of interpretability, efficiency, stability, extensiveness to multiple feature types, and resistance to the curse of dimensionality. A decision tree partition is obtained by recursively splitting grids into subgrids using criteria such as variance reduction. Since it can be challenging to use the original CART rule (Breiman, 1984) for theoretical analysis, we adopt the *max-edge* rule following Ma et al. (2023); Ma and Yang (2024); Ma et al. (2026). This rule is amenable to theoretical analysis and can also achieve satisfactory practical performance. For each grid, the partition rule selects the midpoint of the longest edges that achieves the largest entropy reduction (Györfi et al., 2022). This is defined by $g^{ent}(B, D) = -\frac{1}{n}\sum_{X_i \in D} \mathbf{1}(X_i \in B) \log\left(\frac{1}{n}\sum_{X_i \in D} \mathbf{1}(X_i \in B)\right)$ for cell B and data set D. This procedure continues until the depth of the tree reaches its predetermined limit. The detailed algorithm is in Algorithm 1. Together, the `HistOfTree` estimator is

$$\tilde{p}(x) = \sum_{A_j, B_k} \mathbf{1}_{A_j \times B_k}(x) \frac{\frac{1}{n}\sum_{i=1}^{n} \tilde{U}_i^j \mathbf{1}_{B_k}(X_i^{\text{pub}})}{\mu(A_j \times B_k)}. \tag{9}$$

Estimator (9) differs from classical non-parametric LDP histogram and tree-based estimators in two essential aspects (Györfi and Kroll, 2025; Ma et al., 2023). Classical LDP histogram estimators treat all coordinates as private, and therefore the full cell membership information must be reported through a local privacy channel. In contrast, (9) only privatizes the histogram membership of the protected features, namely $\mathbf{1}_{A_j}(X_i^{\text{priv}})$, while the public-cell membership $\mathbf{1}_{B_k}(X_i^{\text{pub}})$ is observed exactly. Thus, the public features route each sample to the corresponding public cell without injecting privacy noise. This separation reduces the variance of the privatized cell-count estimator and is the main source of the rate improvement of semi-feature LDP over classical LDP, as formalized in Theorem 2.

The second distinction is that public features make it possible to construct an informative, data-dependent partition without consuming privacy budget. In classical LDP tree-based estimators, split decisions involving private coordinates must either be predetermined or learned from privatized and hence noisy summaries. By contrast, (9) uses a fixed histogram partition on $\mathcal{X}^{\text{priv}}$, where privacy constraints are restrictive, and builds a tree partition on $\mathcal{X}^{\text{pub}}$, where the raw feature information is available. This hybrid "histogram-of-tree" structure is fundamentally enabled by the public features: the private part controls the amount of injected noise, while the public tree refines the partition according to the observed geometry of the data. Although the public-tree refinement does not further improve the minimax rate under our current analysis, it produces a more informative partition and is important for empirical performance (Ma et al., 2023).

Our estimator also implies a novel non-parametric LDP density estimator. When $s^* = d$, `HistOfTree` becomes an LDP density estimator. Previously, the problem has been studied by Butucea et al. (2020) using wavelets, while it requires the existence of such wavelet basis. Other methods either allow only a single predetermined test point (Kroll, 2021; Sart, 2023)

Algorithm 1 Max-edge partition rule

Input: Public data $\{X_i^{\text{pub}}\}_{i=1}^n$, depth T, (optional) privatized labels $\{\tilde{Y}_i\}_{i=1}^n$.
Initialization: $B_{0,0} = [0,1]^d$, $D_{0,0} = \{X_i^{\text{pub}}, \tilde{Y}_i\}_{i=1}^n$, $\pi_k = \emptyset$ for $1 \le k \le T$.
for $k \in 1, \ldots, T$ **do**
  for $j \in 0, \ldots, 2^{k-1} - 1$ **do**
    Suppose $B_{k-1,j} = \times_{\ell=1}^d [a_\ell, b_\ell]$, let $\mathcal{M}_{k-1}^j = \{i \mid |b_k - a_k| = \max_{\ell=1,\cdots,d} |b_\ell - a_\ell|\}$. `# Longest edges.`
    Let $B_{k-1,j}^0(\ell) = \left\{x \mid x \in B_{k-1,j}, x^\ell < \frac{a_\ell + b_\ell}{2}\right\}$ and $B_{k-1,j}^1(\ell) = B_{k-1,j} / B_{k-1,j}^0(\ell)$.
    Let $D_{k-1,j}(\ell) = \left\{(X_i^{\text{pub}}, \tilde{Y}_i) \in D_{k-1,j} \mid W_i^\ell = 0\right\}$. `# Identify available samples along each axis.`
    Select ℓ as

$$\arg\min_{\ell \in \mathcal{M}_{i-1}^j} \; g(B_{k-1,j}^0(\ell), D_{k-1,j}(\ell)) + g(B_{k-1,j}^1(\ell), D_{k-1,j}(\ell)). \tag{10}$$

    $B_{k,2j-1} = B_{k-1,j}^0(\ell)$, $B_{k,2j} = B_{k-1,j}^1(\ell)$, $\pi_k = \pi_k \cup \{B_{k,2j-1}, B_{k,2j}\}$.
    $D_{k,2j-1} = \{(X_i^{\text{pub}}, \tilde{Y}_i) \in D_{k-1,j}(\ell) \mid X_i^{\text{pub}\ell} < \frac{a_\ell + b_\ell}{2}\}$, $D_{k,2j} = D_{k-1,j}(\ell) / D_{k,2j-1}$. `# Allocate samples.`
  end for
end for
Output: Partition π_T

or leverage kernel approximation (Zhou et al., 2024), whereas `HistOfTree` accommodates arbitrarily many test points and avoids additional approximation error.

4.3 Methodology for Personalized Privacy Preference

Following the last section, we adapt the partition as well as the privacy mechanism to accommodate the personalized privacy preference case.

4.3.1 Partition

The first step is to manually select the features to be treated as private. Suppose $\sum_{i=1}^n W_i^\ell$ is ranked in decreasing order, i.e., the first feature is the most protected. We select the first s features with the largest $\sum_{i=1}^n W_i^\ell$, i.e. $1, \cdots, s$. The amount of public entries in these dimensions is not sufficient to guide the partition and thus is ignored. See the yellow boxes in Figure 1(b) and 1(c) for an illustration. The selection of parameter s is discussed in Section 4.4. We create a histogram partition π^{priv} on these dimensions. On the rest of the dimensions, we use all public items to create a decision tree partition π^{pub}. Specifically, when splitting each grid, the entropy reduction along the ℓ-th dimension (10) is calculated by samples with $W_i^\ell = 0$. We give the detailed algorithm for such a partition rule in Algorithm 1.

4.3.2 Privacy Mechanism

Injecting noise into the grid indices is trickier. We first define the potential grids for a sample. For X_i, let

$$\mathcal{V}_i = \left\{ A \times B \in \pi^{\text{priv}} \otimes \pi^{\text{pub}} \mid \exists \overline{X} \in A \times B \text{ s.t. } X_i^\ell = \overline{X}^\ell \text{ if } W_i^\ell = 0 \right\}. \tag{11}$$

Intuitively, $\mathcal{V}_i$ is the collection of grids where X_i could potentially be located, given no information about its private features. For simplicity, let j be the index of $A \times B$ in the combined partition $\pi^{\text{priv}} \otimes \pi^{\text{pub}}$. We then define the privacy mechanism:

$$\tilde{V}_i^j = \begin{cases} 0 & \text{if } A \times B \notin \mathcal{V}_i, \\ C_\varepsilon \left(\mathbf{1}_{A\times B}(X_i) - \frac{1}{1+e^{\varepsilon/2}} \right) & \text{else w.p. } \frac{e^{\varepsilon/2}}{1+e^{\varepsilon/2}}, \\ C_\varepsilon \left(\mathbf{1}_{(A\times B)^c}(X_i) - \frac{1}{1+e^{\varepsilon/2}} \right) & \text{else w.p. } \frac{1}{1+e^{\varepsilon/2}}. \end{cases} \tag{12}$$

In this case, the indices of potential grids are estimated analogously to $\tilde{U}_i^j$, while the indices for the remaining grids, which are already known without private features, are zeroed out. Furthermore, we have $\mathbb{E}_{\text{P,R}}\left[\tilde{V}_i^j\right] = \mathbb{E}_{\text{P}}\left[\mathbf{1}_{A\times B}(X_i)\right]$. For a candidate pair (s, T) with $s < d$, let

$$q_i(s) := \sum_{\ell=s+1}^{d} W_i^\ell, \qquad h_{s,T} := 2^{-T/(d-s)}, \qquad S_{s,T} := \sum_{k=1}^{n} h_{s,T}^{q_k(s)}.$$

We assign user i the deterministic weight

$$\omega_i(s,T) := \frac{h_{s,T}^{q_i(s)}}{S_{s,T}}, \qquad \sum_{i=1}^{n} \omega_i(s,T) = 1, \tag{13}$$

and define the personalized estimator

$$\tilde{p}(x) = \sum_{A\times B} \mathbf{1}_{A\times B}(x) \frac{\sum_{i=1}^n \omega_i(s,T)\tilde{V}_i^j}{\mu(A \times B)}. \tag{14}$$

Here j is the index of $A \times B$.

The weights have a direct inverse-variance interpretation. By Lemma 24, user i is compatible with at most order $t^s h_{s,T}^{-q_i(s)}$ cells. If arbitrary coefficients $a_i \geq 0$ with $\sum_i a_i = 1$ are used, the user-dependent part of the privatization variance is proportional to $\sum_{i=1}^n a_i^2 h_{s,T}^{-q_i(s)}$. Minimizing this expression over the simplex gives $a_i \propto h_{s,T}^{q_i(s)}$, which yields exactly (13). In particular,

$$\sum_{i=1}^{n} \omega_i(s,T)^2 h_{s,T}^{-q_i(s)} = \frac{1}{S_{s,T}}. \tag{15}$$

Thus $S_{s,T}$ acts as an effective information size at resolution $h_{s,T}$. Each additional private tree coordinate enlarges the set of cells compatible with a user by a factor $h_{s,T}^{-1}$ and, correspondingly, reduces that user's aggregation weight by a factor $h_{s,T}$.

The weights depend only on the publicly known privacy masks and on (s, T), so their use is post-processing and incurs no additional privacy loss. Moreover, because W is independent of the observations and the weights sum to one,

$$\mathbb{E}\left[\sum_{i=1}^{n} \omega_i(s,T)\mathbf{1}_{A\times B}(X_i)\right] = \mathrm{P}_X(A \times B).$$

The weighted estimator therefore preserves the same population target while downweighting reports with larger privacy-induced variance. The following proposition establishes the privacy guarantee of (12).

Proposition 4 *Let $\pi = \{A \times B \mid A \in \pi^{priv}, B \in \pi^{pub}\}$ be any partition of $\mathcal{X}$. Then the privacy mechanism* (12) *satisfies non-interactive ε-semi-feature LDP.*

In the aligned case where the public features are shared across users, i.e. $\sum_{i=1}^{n} W_i^{\ell} = n$ for $1 \le \ell \le s$ and $\sum_{i=1}^{n} W_i^{\ell} = 0$ otherwise, mechanism (12) reduces to that in (8). The privacy guarantees of these mechanisms are thus implied by Proposition 4. In the aligned case, $q_i(s)$ is constant over i, so $\omega_i(s, T) = 1/n$. Hence the estimator (14) also recovers (9).

4.4 Selection of s

In this section, we explore the selection of the parameter s within the context of personalized privacy preferences. Unlike scenarios with aligned preferences, where an intrinsic value $s = s^*$ is implied, the personalized case requires the explicit specification of s as a hyper-parameter. From the proof of Theorem 3, the quantity in brackets below is a high-probability upper bound on the estimation error. We therefore choose (s, T) by

$$s^{\dagger}, T^{\dagger} = \arg\min_{s,T} \left\{ \sqrt{\frac{2^{T(d+s)/(d-s)} \log n}{\varepsilon^2 \sum_{i=1}^{n} 2^{-q_i(s)T/(d-s)}}} + 2^{-T/(d-s)} \right\}, \tag{16}$$

where the unknown α is replaced by 1, i.e. the Lipschitz continuous case. Since there are at most $\lceil d \log n \rceil$ possible candidates for $(s^{\dagger}, T^{\dagger})$, this can be done via a brute force search. The choice $s = 0$ gives the matched minimax rate. Allowing $s > 0$ can nevertheless be useful in finite samples because it changes which coordinates are assigned to the histogram part and which public observations are available for constructing the tree partition.

5. Supervised Learning: Regression and Classification

In this section, we extend the weighted personalized construction to regression and classification. In Section 5.1, we formulate the two problems and establish minimax lower and upper bounds that match, up to logarithmic factors, under both aligned and personalized privacy preferences. In Section 5.2, we present the corresponding estimators and comment on hyper-parameter tuning.

5.1 Theoretical Results

The goal of regression is to predict the value of an unobserved label of a given input X, based on a data set $D := \{(X_i, Y_i)\}_{i=1}^{n}$ consisting of n i.i.d. observations drawn from an

unknown probability measure P on $\mathcal{X} \times \mathcal{Y} = [0,1]^d \times [-M, M]$. It is legitimate to consider the least square loss $L : \mathcal{X} \times \mathcal{Y} \times \mathcal{Y} \to [0, \infty)$ defined by $L(x, y, f(x)) := (y - f(x))^2$ for our target of regression. Then, for a measurable decision function $f : \mathcal{X} \to \mathcal{Y}$, the risk is defined by $\mathcal{R}_{\mathrm{P}}(f) := \int_{\mathcal{X}\times\mathcal{Y}} L(x, y, f(x))\, d\mathrm{P}(x, y)$. The Bayes risk is the smallest possible risk with respect to P and L. The function that achieves the Bayes risk is called Bayes function, namely, $f^*(x) := \mathbb{E}\big(Y|X = x\big)$. For classification, the label Y belongs to $\{0, 1\}$. The conditional relationship becomes $f^*(x) = \mathbb{P}(Y = 1|X = x)$ and the loss should be the misclassification error $L(x, y, f(x)) = \mathbf{1}\left(y \neq \mathbf{1}\left(f(x) \geq 1/2\right)\right)$.

5.1.1 Regression

We first investigate the theory of non-parametric regression, which requires the following assumption.

Assumption 2 *Assume that the marginal density function is upper and lower bounded, i.e.* $\underline{c} \leq p_X^*(x) \leq \overline{c}$ *for some* $\overline{c} \geq \underline{c} > 0$. *Let* $\alpha \in (0, 1]$. *Assume* $f^* : \mathcal{X} \to \mathbb{R}$ *is* α*-Hölder continuous, i.e. there exists* $c_L > 0$ *s.t. for all* $x_1, x_2 \in \mathcal{X}$, $|f^*(x_1) - f^*(x_2)| \leq c_L \|x_1 - x_2\|^\alpha$.

The assumption is the most basic and well adopted in the literature of non-parametric regression. See Cai et al. (2023); Berrett et al. (2021); Cai and Pu (2024). We present the following theorem that specifies the minimax lower bound with semi-feature LDP under the above assumptions.

Theorem 5 *Denote the function class of* P *satisfying Assumption 2 by* $\mathcal{F}$. *Define the quantity* $\underline{\mathcal{E}}^*$ *by*

$$\underline{\mathcal{E}}^* = \arg\max_{\mathcal{E}} \left(\sum_{i=1}^{n} \mathcal{E}^{\frac{2\alpha + d + \sum_{\ell=1}^{d} W_i^\ell}{2\alpha}} \leq \frac{1}{(e^\varepsilon - 1)^2} \right). \tag{17}$$

Then for any estimator $\tilde{f}$ *that is sequentially-interactive* ε*-semi feature LDP, there holds*

$$\inf_{\tilde{f}} \sup_{\mathcal{F}} \mathbb{E}_{\mathrm{P}} \left[\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \right] \gtrsim \underline{\mathcal{E}}^* \vee n^{-\frac{2\alpha}{2\alpha + d}}. \tag{18}$$

This implies that when privacy preference is aligned, i.e. $W_i^\ell = 1$ *if* $\ell \leq s^*$ *and* $W_i^\ell = 0$ *otherwise, there holds*

$$\inf_{\tilde{f}} \sup_{\mathcal{F}} \mathbb{E}_{\mathrm{P}} \left[\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \right] \gtrsim \left(n(e^\varepsilon - 1)^2 \right)^{-\frac{2\alpha}{2\alpha + d + s^*}} \vee n^{-\frac{2\alpha}{2\alpha + d}}.$$

The findings brought by this theorem are similar to those by Theorem 2 since the rates are merely squared. If $s^* = d$, our results reduce to the special case of LDP non-parametric regression with only private data. Previous studies (Berrett et al., 2021; Györfi and Kroll, 2025) have provided estimators that achieve the minimax optimal convergence rate. The following result establishes corresponding upper bounds.

Theorem 6 *Let Assumption 2 hold. Suppose* $0 < \varepsilon \le 1$. *Let* $s_i^* := \sum_{\ell=1}^{d} W_i^\ell$ *and define*

$$\overline{\mathcal{E}}^* := \arg\max_{\mathcal{E}} \left\{ \sum_{i=1}^{n} \mathcal{E}^{\frac{2\alpha+d+s_i^*}{2\alpha}} \le \frac{\log n}{\varepsilon^2} \right\}. \tag{19}$$

Then there exists a non-interactive ε*-semi-feature LDP weighted* `HistOfTree` *estimator* $\tilde{f}$ *such that, with probability at least* $1 - 8/n^2$,

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \overline{\mathcal{E}}^* \vee \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2\alpha+d}}.$$

If the privacy preference is aligned, i.e., $W_i^\ell = 1$ *for* $\ell \le s^*$ *and* $W_i^\ell = 0$ *otherwise, then*

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \left(\frac{\log n}{n\varepsilon^2}\right)^{\frac{2\alpha}{2\alpha+d+s^*}} \vee \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2\alpha+d}}. \tag{20}$$

Comparing (19) with (17), the personalized upper and lower bounds are governed by the same positive moment of s_i^*. Therefore, for $0 < \varepsilon \le 1$, they match up to the logarithmic factor introduced by the high-probability analysis. In the aligned case, the positive moment reduces to the common count s^*, yielding the rate in (20). These rates are the squared counterparts of the density-estimation rates. In the next part, we present the theoretical results for classification.

5.1.2 Classification

Besides smoothness, classification needs an additional assumption on the hardness of the decision boundary, which is commonly used in the statistical learning literature (Audibert and Tsybakov, 2007; Samworth, 2012; Chaudhuri and Dasgupta, 2014).

Assumption 3 *Let* $\beta \ge 0, C_\beta > 0$. *Assume*

$$\mathrm{P}_X\left(0 < \left|f^*(X) - \frac{1}{2}\right| \le \gamma\right) \le C_\beta \gamma^\beta$$

for $\gamma > 0$.

Under such conditions, we establish the following lower bound.

Theorem 7 *Denote the function class of* P *satisfying Assumption 2 and 3 by* $\mathcal{F}$. *Define the quantity* $\mathcal{E}^*$ *by*

$$\mathcal{E}^* = \arg\max_{\mathcal{E}} \left(\sum_{i=1}^{n} \mathcal{E}^{\frac{2\alpha+d+\sum_{\ell=1}^{d} W_i^\ell}{\alpha}} \le \frac{1}{(e^\varepsilon - 1)^2} \right). \tag{21}$$

Then for any estimator $\tilde{f}$ *that is sequentially-interactive* ε*-semi feature LDP, there holds*

$$\inf_{\tilde{f}} \sup_{\mathcal{F}} \mathbb{E}_{\mathrm{P}} \left[\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \right] \gtrsim (\mathcal{E}^*)^{1+\beta} \vee n^{-\frac{\alpha(1+\beta)}{2\alpha+d}}. \tag{22}$$

This implies that when privacy preference is aligned, i.e. $W_i^\ell = 1$ *if* $\ell \le s^*$ *and* $W_i^\ell = 0$ *otherwise, there holds*

$$\inf_{\tilde{f}} \sup_{\mathcal{F}} \mathbb{E}_{\mathrm{P}} \left[\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \right] \gtrsim \left(n(e^\varepsilon - 1)^2 \right)^{-\frac{\alpha(1+\beta)}{2\alpha+d+s^*}} \vee n^{-\frac{\alpha(1+\beta)}{2\alpha+d}}.$$

Theorem 8 *Let Assumptions 2 and 3 hold. Suppose* $0 < \varepsilon \le 1$. *Let* $s_i^* := \sum_{\ell=1}^d W_i^\ell$ *and define*

$$\overline{\mathcal{E}}^* := \arg\max_{\mathcal{E}} \left\{ \sum_{i=1}^n \mathcal{E}^{\frac{2\alpha+d+s_i^*}{\alpha}} \le \frac{\log n}{\varepsilon^2} \right\}. \tag{23}$$

Then there exists a non-interactive ε*-semi-feature LDP weighted* `HistOfTree` *estimator* $\tilde{f}$ *such that, with probability at least* $1 - 8/n^2$,

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \left(\overline{\mathcal{E}}^* \right)^{1+\beta} \vee \left(\frac{\log n}{n} \right)^{\frac{\alpha(1+\beta)}{2\alpha+d}}.$$

If the privacy preference is aligned, i.e., $W_i^\ell = 1$ *for* $\ell \le s^*$ *and* $W_i^\ell = 0$ *otherwise, then*

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \left(\frac{\log n}{n\varepsilon^2} \right)^{\frac{\alpha(1+\beta)}{2\alpha+d+s^*}} \vee \left(\frac{\log n}{n} \right)^{\frac{\alpha(1+\beta)}{2\alpha+d}}.$$

Together, Theorems 5, 6, 7, and 8 give matching minimax rates, up to logarithmic factors, for both supervised tasks and for both privacy-preference regimes. Under personalized preferences, the regression risk is governed by the positive moment in (19), while the pointwise classification scale is governed by (23) and is converted into excess classification risk through the margin exponent $1 + \beta$. Under aligned preferences, both expressions reduce to the common privacy count s^*. We remark that the corresponding rates for classical LDP were established for regression (Berrett et al., 2021; Györfi and Kroll, 2025) and classification (Berrett and Butucea, 2019; Berrett et al., 2021). Our estimator recovers these rates if $s^* = d$. Similarly, we include the rate in Xu et al. (2023) for label LDP regression if $s^* = 0$.

5.2 Methodology

In this section, we follow the strategy in Section 4 to describe the estimator used for regression and classification. Note that given the developed density estimation tool, we have two options. One is to conduct two density estimations for both P_X and $\mathrm{P}_{X,Y}$, and then use their ratio as the estimation of the conditional relationship. Another one is to mask Y separately. Both approaches are able to achieve the established theoretical rate. However, the first approach does not allow the utilization of information in Y to create the partition, and thus is practically unfavorable. We take the second approach. Recall that in Section 4, the personalized estimator contains the aligned one as a special case. Thus, in this section, we develop only the personalized estimator for regression and classification.

The weighted partition-based estimator (which is non-private) of x is

$$\widehat{f}_\omega(x) = \sum_{A \times B} \mathbf{1}_{A \times B}(x) \frac{\sum_{i=1}^n \omega_i(s,T) Y_i \cdot \mathbf{1}_{A \times B}(X_i)}{\sum_{i=1}^n \omega_i(s,T) \mathbf{1}_{A \times B}(X_i)}. \tag{24}$$

In each grid $A \times B$ that x belongs to, the denominator represents the weighted sample marginal distribution in $A \times B$, while the numerator corresponds to the weighted first moment. Thus, their division estimates their conditional relationship.

To perform such an estimation, two pieces of information are necessary from each data holder: Y_i and $\mathbf{1}_{A\times B}(X_i)$. We employ the standard Laplace mechanism (Dwork et al., 2006) to protect Y_i, specifically:

$$\tilde{Y}_i = Y_i + \frac{4M}{\varepsilon}\xi_i, \tag{25}$$

where the variables ξ_i are i.i.d. standard Laplace random variables. Note that we can take $M = 1/2$ for classification. For indicator functions, we recall the definition in (11) and have

$$\tilde{V}_i^j = \begin{cases} 0 & \text{if } A \times B \notin \mathcal{V}_i, \\ C_\varepsilon\left(\mathbf{1}_{A\times B}(X_i) - \frac{1}{1+e^{\varepsilon/4}}\right) & \text{else w.p. } \frac{e^{\varepsilon/4}}{1+e^{\varepsilon/4}}, \\ C_\varepsilon\left(\mathbf{1}_{(A\times B)^c}(X_i) - \frac{1}{1+e^{\varepsilon/4}}\right) & \text{else w.p. } \frac{1}{1+e^{\varepsilon/4}}. \end{cases} \tag{26}$$

Note that (25) and (26) take half of the privacy budget, respectively. Consequently, using the weights in (13), we define the personalized estimator:

$$\tilde{f}(x) = \sum_{A\times B} \mathbf{1}_{A\times B}(x) \frac{\sum_{i=1}^n \omega_i(s,T)\tilde{Y}_i\tilde{V}_i^j}{\sum_{i=1}^n \omega_i(s,T)\tilde{V}_i^j}. \tag{27}$$

Here j is the index of $A \times B$. For regression, the predicted value at x is $\tilde{f}(x)$. For classification, the prediction is $\mathbf{1}(\tilde{f}(x) > 1/2)$.

The following proposition establishes the privacy guarantee of (27).

Proposition 9 *Let $\pi = \{A \times B \mid A \in \pi^{priv}, B \in \pi^{pub}\}$ be any partition of $\mathcal{X}$. Then the privacy mechanisms* (25) *and* (26) *jointly satisfy non-interactive ε-semi-feature LDP.*

As for the partition process, we adopt exactly the same algorithm as in Algorithm 1, while we leverage both released private labels and public features. Moreover, we use different information criteria for classification and regression. For both tasks, the criterion g is taken to be the sample variance, i.e. $g^{var}(B, D) = \sum_{\tilde{Y}_i \in D}(\tilde{Y}_i - \sum_{\tilde{Y}_i \in D} \tilde{Y}_i)^2$. For classification, we need to design a private substitute for classification criteria such as gini impurity, since $\tilde{Y}$ is not necessarily an integer. This can be resolved by truncation $\overline{p} = (0 \vee \frac{1}{|D|}\sum_{\tilde{Y}_i \in D} \tilde{Y}_i) \wedge 1$ and $g^{impurity}(B, D) = -\overline{p}\log(\overline{p}) - (1-\overline{p})\log(1-\overline{p})$.

For supervised tasks, we use the same weighted optimization objective (16) to search over s and T. The density, regression, and classification bounds have the same pointwise bias–variance balance; regression squares this pointwise rate, while classification converts it through the margin exponent $1 + \beta$.

5.3 Complexity Analysis

We analyze the complexity of `HistOfTree` from perspectives of computation, memory, and communication. The overall communication rounds for the personalized estimator (27) are

two. Each user first sends its privatized label $\tilde{Y}_i$ and public features X_i^{pub}, which costs communication capacity of at most $d+1$ real numbers. Based on the created partition, each user then sends the binary encodings $\tilde{V}_i^j$, totaling $t^s 2^T$ Boolean values.

The computation complexity of `HistOfTree` consists of two parts, creating the partition and computing the node values. The partition procedure takes $\mathcal{O}(pnd)$ time and the computation of (27) takes $\mathcal{O}(n)$ time. Thus the training stage complexity is at most $\mathcal{O}(n + nd \log n\varepsilon^2)$, which is linear in n. Since each prediction of the decision tree takes $\mathcal{O}(T)$ time, the test time for each test instance is $\mathcal{O}(\log n\varepsilon^2)$. As for storage complexity, since `HistOfTree` only requires storage of the tree structure and the node values, the space complexity is also $\mathcal{O}(t^s 2^T)$. In short, `HistOfTree` is an efficient method in terms of computation, memory, and communication.

6. Experiments

To demonstrate the superiority of proposed methods and to validate our theoretical findings, we conduct experiments on both synthetic and real data sets in Section 6.2 and 6.3, respectively.

For supervised problems, the tested methods include: **(i)** `HistOfTree` is the proposed estimator. In addition to the partition rule proposed in Algorithm 1, we boost the performance by incorporating the criterion reduction scheme from the original CART (Breiman, 1984), similar to Ma et al. (2023); Ma and Yang (2024); **(ii)** `AdHistOfTree` is identical to `HistOfTree` except that the values of s, T, and t are selected according to Sections 4.4 and 5.2. **(iii)** `Hist` is the locally private histogram estimation (Berrett et al., 2021; Györfi and Kroll, 2025) which treats all features and the label as private; **(iv)** `KRR` is the obfuscation approach that directly perturbs X_i^ℓ using k-generalized randomized response if $W_i^\ell = 1$. It then fits a decision tree. **(v)** `ParDT` estimates by adding Laplace noise to labels and then fitting a decision tree based on the noisy label and the public part of the samples. **(vi)** `LabelDT` estimates by adding Laplace noise to labels and then using a decision tree. It serves as the label LDP benchmark. **(vii)** `DT` is the non-private decision tree and serves as the non-private benchmark. For unsupervised problems, we compare **(i)** the basic `HistOfTree` and `AdHistOfTree` density estimators and **(ii)** `Hist`, a locally private histogram estimator based on the mechanisms in Györfi and Kroll (2025). Implementation details are in Section 6.1.

For each model, we report the best result over its parameter grids, with the best result determined based on the average of at least 50 replications. The size of the parameter grids is selected based on running time to ensure that each method incurs an equal amount of computation. All experiments are conducted on a machine with 72-core Intel Xeon 2.60GHz and 128GB of main memory.

6.1 Implementation Details

For `HistOfTree`, we adopt the generalized randomized response mechanism introduced in Section 6.1.1. We add one more parameter adjusting the allocation of the privacy budget on the numerator and denominator. Specifically, let the mechanism for $\tilde{Y}_i$ and $\tilde{U}_i$ be respectively $\rho\varepsilon$-LDP and $(1-\rho)\varepsilon$-semi-feature LDP for $\rho \in [0,1]$. By composition, the hybrid

mechanism is still ε-LDP. We select $\rho \in \{0.5, 0.7, 0.9\}$. For other parameters, we select $p \in \{1, 2, 4, 6\}$ and $t \in \{1, 2, 3\}$.

For `AdHistOfTree`, we also adopt the generalized randomized response mechanism. For better empirical performance, the second term of the objective function is multiplied with a constant selected in $\{0.01, 0.1, 1\}$. Also, we allow the selection $t \in \{t^* - 1, t^*, t^* + 1\}$, where t^* is the default value derived in Section 5.2.

For `Hist`, we implement the private histogram proposed by Berrett et al. (2021); Györfi and Kroll (2025) in Python. `Hist` applies the Laplacian mechanism to privatize the estimation of marginal and joint probabilities for a cubic histogram partition with the number of bins t. In cells with a private marginal probability estimation less than ζ, estimation is truncated to 0. We let $t \in \{1, 2, 3, 4\}$. We set the truncation parameter $\zeta \in \{0.01, 0.05\}$.

For `ParDT` and `LabelDT`, we add Laplace noise with scale ξ/ε where ξ is the range of the label. For `ParDT`, `LabelDT`, `KRR` and `DT`, we use the implementation by Scikit-Learn (Pedregosa et al., 2011). We select *max_depth* in $\{1, 2, 4, 6, 8\}$ and *min_sample_leaf* in $\{1, 10, 100\}$. For `KRR`, we select k in $\{2, 3, 4, 5\}$.

6.1.1 Generalized Randomized Response Mechanism

In this section, we introduce the generalized randomized response mechanism (Warner, 1965; Kairouz et al., 2016; Ghazi et al., 2022) (or direct encoding (Wang et al., 2017)). Specifically, for the aligned privacy preference case, the mechanism calculates (26) by

$$\Pr[U_i = j] = \begin{cases} \frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+|\pi^{\text{priv}}|-1}, & \text{if } \mathbf{1}_{A_j}(X_i^{\text{priv}}) = 1, \\ \frac{1}{e^{\varepsilon/2}+|\pi^{\text{priv}}|-1}, & \text{otherwise.} \end{cases}$$

It then uses $\tilde{U}_i^j = \frac{e^{\varepsilon/2}+|\pi^{\text{priv}}|-1}{e^{\varepsilon/2}-1}\left(\mathbf{1}_{U_i=j} - \frac{1}{e^{\varepsilon/2}+|\pi^{\text{priv}}|-1}\right)$ as an estimator of $\mathbf{1}_{A_j}(X_i^{\text{priv}})$, since $\mathbb{E}_{\text{P,R}}\left[\tilde{U}_i^j\right] = \mathbb{E}_{\text{P}}\left[\mathbf{1}_{A_j}(X_i^{\text{priv}})\right]$. The estimator (27) remains identical. For the personalized case, the mechanism computes (26) by

$$\tilde{Y}_i = Y_i + \frac{4M}{\varepsilon}\xi_i, \quad \Pr[V_i = j] = \begin{cases} 0 & \text{if } A \times B \notin \mathcal{V}_i, \\ \frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+|\mathcal{V}_i|-1} & \text{if } \mathbf{1}_{A\times B}(X_i) = 1, \\ \frac{1}{e^{\varepsilon/2}+|\mathcal{V}_i|-1}, & \text{otherwise.} \end{cases} \tag{28}$$

To estimate indicator function $\mathbf{1}_{A\times B}(X_i)$, we use

$$\tilde{V}_i^j = \frac{e^{\varepsilon/2}+|\mathcal{V}_i|-1}{e^{\varepsilon/2}-1}\left(\mathbf{1}_{V_i=j} - \frac{1}{e^{\varepsilon/2}+|\mathcal{V}_i|-1}\right) \cdot \mathbf{1}_{A\times B\in\mathcal{V}_i}.$$

In this case, the indices of potential grids are estimated analogously to $\tilde{U}_i^j$, while the remaining grids are free of privacy concern and zeroed out. The following proposition establishes the privacy guarantee of (28).

Proposition 10 *Let $\pi = \{A \times B \mid A \in \pi^{priv}, B \in \pi^{pub}\}$ be any partition of $\mathcal{X}$. Then the privacy mechanism* (28) *satisfies ε-semi-feature LDP.*

The reason we introduce this mechanism is due to the fact that when $|\mathcal{V}_i|$ is small, it has been shown to be quite efficient (Wang et al., 2017; Kairouz et al., 2016; Xiong et al., 2020). In experiments where sample sizes $n \sim 1e3$ and $|\mathcal{V}_i| \sim 1e1$, we find that the generalized randomized response mechanism performs slightly better than the randomized responses introduced in the main text. However, the variance of such a mechanism grows linearly with $|\mathcal{V}_i|$, which can be troublesome in theoretical analysis where $|\mathcal{V}_i|$ is n raised to some power. Consequently, in such cases, the convergence rate fails to attain optimality.

6.2 Simulation

We conduct two parallel sets of simulation for regression (Section 6.2.1) and density estimation (Section 6.2.2).

6.2.1 Regression

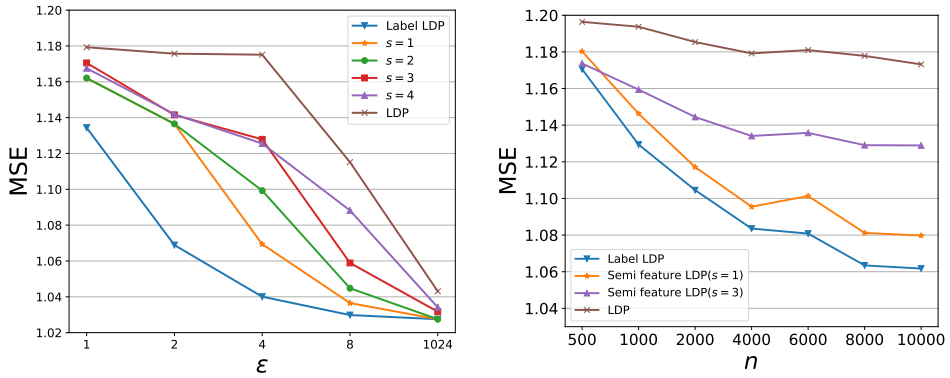
In the simulation, we consider the following distribution P. The marginal distribution P_X is uniform on $\mathcal{X} = [0,1]^5$. The label is generated via $Y = \sin(X^1)+\sin(X^2)+\sin(X^5)+\sigma$, where σ is a standard Gaussian random variable. It can be easily verified that the distribution above satisfies Assumption 2. For mask matrix W, we always let $\sum_{i=1}^n W_i^\ell \geq \sum_{i=1}^n W_i^{\ell'}$ if $\ell \leq \ell'$. In this case, useful information is contained in both public and private features, which necessitates our method. The evaluation metric in all experiments is the mean squared error (MSE).

Privacy utility trade-off We analyze how privacy budget ε influences the prediction quality under different numbers of public features. For $1 \leq s^* \leq 4$, we consider $n = 10000$ and vary ε. The results are displayed in Figure 2(a). When ε increases, all MSEs decrease, while the performance steadily improves as the number of public features increases. This observation is compatible with both Theorem 6 and the notion that semi-feature LDP serves as the intermediate stage of LDP and label LDP.

Consistency We analyze the consistency of (27) as n grows. We take $\varepsilon = 4$. The results are displayed in Figure 2(b), showing that the MSE is diminishing as n grows, while the rate is faster for small s.

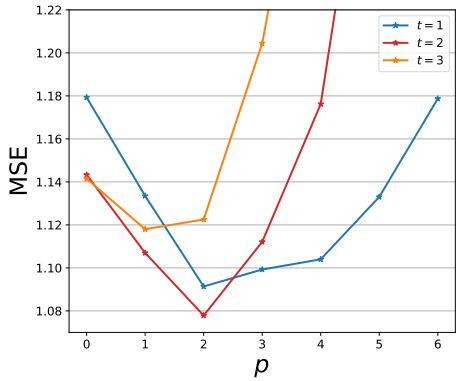
Parameter analysis We conduct experiments to investigate the influence of key parameters T and t. We consider $n = 10000$, $\varepsilon = 4$, and $s^* = 2$. As shown in Figure 2(c), for each t, as T increases, MSE exhibits a U-shaped curve. Also, the best performance is achieved with $t = 2$, indicating such a phenomenon is also true for t. This further emphasizes the importance of choosing correct hyper-parameters as outlined in Theorem 6. Moreover, the depth T at which the test error is minimized increases as t increases. This is compatible with theory since $t \asymp 2^{T/(d-s^*)}$.

Selection of s under different behaviors of W We investigate how the behavior of W affects the selection of s. Specifically, we use W_γ parameterized by γ, where $W_{\gamma i}^\ell = 1$ if $i \leq n/\ell^\gamma$. A smaller γ implies a thicker tail of W. We compare the performance of `HistOfTree` with different choices of s, as well as `AdHistOfTree` whose s is selected automatically. In Figure 2(d), the observations are two-fold. First, MSE decreases w.r.t. γ, which aligns with Theorem 6 that a thicker tail of W leads to a larger upper bound. On the other hand, one

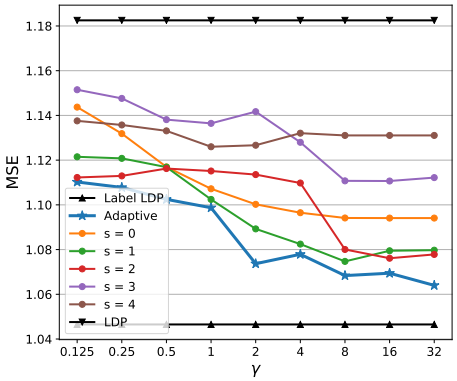


(a) Privacy utility trade off

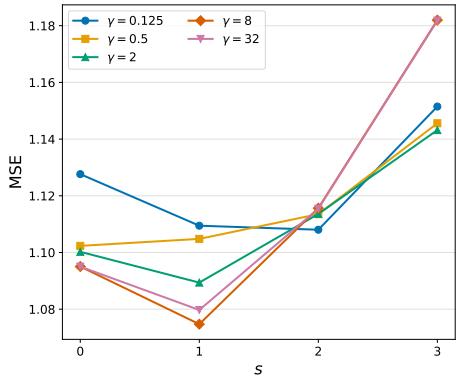
(b) Consistency



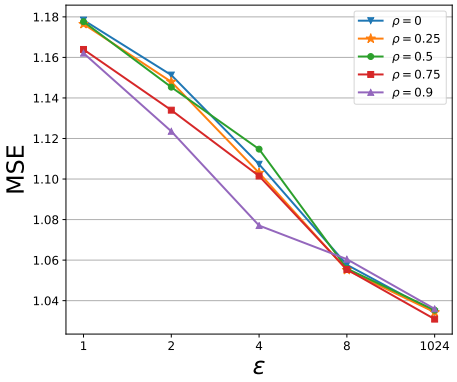
(c) Parameter analysis



(d) Selection of s



(e) Robustness to s



(f) Public–private correlation

Figure 2: Regression results on synthetic data. `LabelDT` and `Hist` are captioned as label LDP and LDP, respectively. `HistOfTree` is captioned with specific choice of parameters s and t. In 2(d), `AdHistOfTree` is captioned as adaptive.

s does not consistently outperform another under different thicknesses of the tail, which necessitates the discussions about W in Section 4.4. The best s lies between 1 and 2, while

`AdHistOfTree` always achieves approximately the best performance. This illustrates the effectiveness of the selection rule in (16).

Robustness of s Figure 2(e) gives a complementary view of Figure 2(d) by examining the robustness of the choice of s for `HistOfTree` under selected values of γ. For heavy-tailed W, although the best fixed s changes with γ, the loss surface remains relatively flat around its minimizer. For example, when $\gamma = 0.125$, the choices $s = 1$ and $s = 2$ are nearly indistinguishable, with $s = 2$ performing slightly better. For moderate or thin tails, choosing $s = 1$ is consistently near-optimal, and its advantage becomes more pronounced as the tail becomes thinner, since unnecessary privatization increases the variance. This supports the adaptive rule in (16): the procedure does not rely on a single universally optimal s, but tracks the local bias–variance trade-off induced by W.

Public–private correlation We explore how the correlation between private and public features affects the performance of `HistOfTree`. We keep the same experimental setting while varying the latent Gaussian-copula correlation ρ between (X^1, X^3) and (X^2, X^4). In Figure 2(f), we observe that this correlation yields some performance improvement when ε is small. This can be explained by the shared information between private and public features. When the privacy constraint is strong, `HistOfTree` can implicitly exploit this information through public features, thereby improving performance. When ε is very large, the curves become close because the privacy noise on private coordinates is already negligible.

6.2.2 Density Estimation

We consider the distribution P_X to be a 5-dimensional marginally independent distribution with each feature being Beta(2,2). It can be easily verified that the constructed distributions above satisfy Assumption 1. For mask matrix W, we always let $\sum_{i=1}^n W_i^\ell \geq \sum_{i=1}^n W_i^{\ell'}$ if $\ell \leq \ell'$. Similar settings (sample sizes, hyper-parameters) are chosen as in regression, except that the evaluation metric in all experiments is the L_1 distance. Observing Figure 3, we arrive at similar conclusions as in regression.

6.3 Real Data

In the real data experiments, we focus on regression. We conduct experiments on 10 real-world data sets that are available online. These data sets cover a wide range of sample sizes and feature dimensions commonly encountered in real-world scenarios. Some of these data sets contain both sensitive and insensitive features, making them suitable for our case. Privacy preferences follow two paradigms. For both paradigms, we rank the features subjectively according to their level of sensitivity, from sensitive to non-sensitive. In the aligned case, we select the first $s^* = \lceil \log \sqrt{d} \rceil$ features. In the personalized case, we set $W_i^\ell = 1$ if $i \leq n/10^{\lfloor \ell/s^* \rfloor}$ and 0 otherwise. A summary of key information for these data sets and pre-processing details is in Appendix B.

We first compute the mean squared error over at least 50 random train-test splits for $\varepsilon = 1, 2, 4$. To standardize the scale across data sets, we report the ratio of each MSE to that of a non-private decision tree fitted on the full training set. The results are displayed in Table 2. For all privacy budgets, the proposed methods significantly outperform competitors in terms of both average performance (rank sum) and the number of best results achieved.

$\varepsilon = 1$

	Aligned					Personalized						
	HistOfTree		ParDT	Hist	KRR	HistOfTree		AdHistOfTree		ParDT	Hist	KRR
	ME	CART				ME	CART	ME	CART			
ABA	2.04	**1.80***	2.04	2.02	2.74	2.05	2.05	**2.01***	**2.01***	2.26	2.02	2.71
AQU	**1.71***	**1.71***	2.86	1.80	3.35	1.72	1.74	**1.70***	**1.70***	3.03	1.80	3.35
BUI	**1.59***	**1.59***	3.18	5e5	6.70	1.64	1.64	**1.61***	**1.61***	3.28	5e5	6.57
DAK	6.69	**6.65***	8.03	8.43	22.6	7.95	7.95	**7.47**	**7.47**	19.0	8.43	22.6
FIS	**2.22***	2.32	3.22	2.45	4.60	2.32	2.32	**2.26***	**2.26***	3.49	2.45	4.59
MUS	1.17	**1.17**	1.89	9e5	2.72	1.17	1.17	**1.16***	**1.16***	1.99	9e5	2.74
POR	**4.15***	**4.15***	4.67	7.22	7.89	4.04	4.04	**3.69***	**3.69***	4.72	7.22	7.70
PYR	**2.31***	**2.31***	2.98	1e5	5.61	2.33	2.33	**1.76***	**1.76***	3.18	1e5	5.64
RED	**1.48***	**1.48***	2.02	1.81	2.53	1.49	1.53	**1.48***	**1.48***	2.08	1.81	2.53
WHI	**1.46***	**1.46***	1.61	1.61	1.85	1.47	1.47	**1.46***	**1.46***	1.70	1.61	1.84
Rank Sum	21	**18**	35	38	47	33	33	**13**	**13**	56	55	67

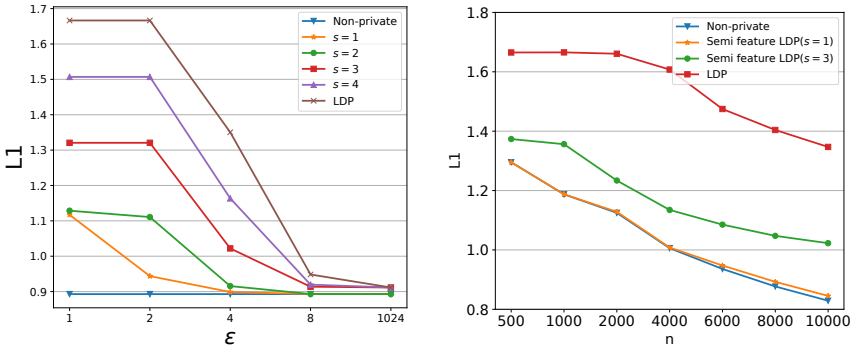
$\varepsilon = 2$

	Aligned					Personalized						
	HistOfTree		ParDT	Hist	KRR	HistOfTree		AdHistOfTree		ParDT	Hist	KRR
	ME	CART				ME	CART	ME	CART			
ABA	1.85	1.65	**1.63**	1.98	1.81	1.70	1.77	1.71	1.77	**1.65**	1.98	1.77
AQU	1.55	**1.47***	1.71	1.63	2.05	1.60	1.60	**1.60***	**1.60***	1.75	1.63	2.05
BUI	1.48	**1.27***	1.51	1e4	2.43	**1.41***	1.47	1.47	1.47	1.59	1e4	2.41
DAK	6.69	**6.65***	8.03	8.34	9.57	6.73	6.73	**6.58**	**6.58**	8.85	8.34	9.57
FIS	**1.88***	2.01	2.14	2.16	2.35	**2.05***	2.16	2.08	2.14	2.23	2.16	2.42
MUS	**1.13***	**1.13***	1.26	2e5	1.44	**1.12***	1.13	1.13	1.13	1.26	2e5	1.45
POR	3.15	**3.04***	3.32	4.23	4.02	3.08	3.08	**2.98***	**2.98***	3.35	4.23	3.92
PYR	**1.48***	**1.48***	1.65	3e3	2.29	1.49	1.49	**1.34***	**1.34***	1.77	3e3	2.30
RED	1.44	**1.37***	1.45	1.55	1.67	1.44	**1.43**	1.44	1.44	1.46	1.55	1.67
WHI	1.42	**1.41**	**1.41**	1.53	1.50	**1.39**	1.39	1.44	1.44	1.42	1.53	1.49
Rank Sum	22	**15**	28	45	43	22	33	**21**	26	47	63	62

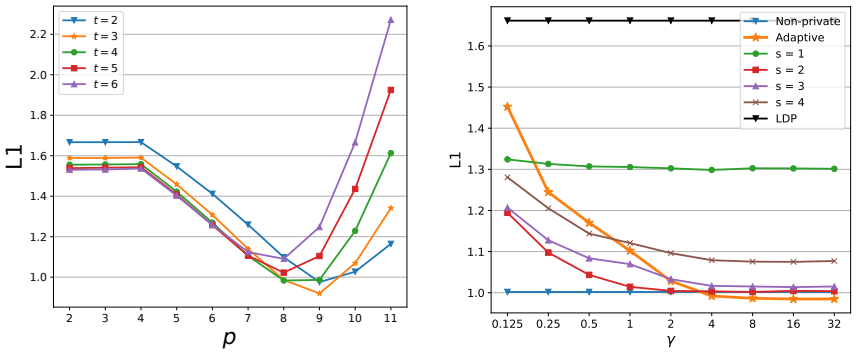
$\varepsilon = 4$

	Aligned					Personalized						
	HistOfTree		ParDT	Hist	KRR	HistOfTree		AdHistOfTree		ParDT	Hist	KRR
	ME	CART				ME	CART	ME	CART			
ABA	1.72	1.59	**1.58**	1.97	1.64	1.63	**1.56**	1.71	1.66	1.61	1.97	1.97
AQU	1.45	**1.35***	1.56	1.58	1.64	1.50	**1.44**	1.54	1.49	1.68	1.58	1.65
BUI	1.23	**1.16***	1.32	3e4	1.30	12.8	12.4	14.3	14.3	1.36	3e4	**1.29**
DAK	4.67	**4.48***	8.03	6.32	6.57	4.63	**4.51***	6.14	6.34	6.01	6.32	6.58
FIS	1.66	**1.64**	1.65	2.12	1.82	**1.63**	1.63	1.86	1.98	1.68	2.12	1.77
MUS	**1.12**	**1.12**	1.15	6e3	1.36	**1.10**	1.12	1.12	1.12	1.22	6e3	1.14
POR	2.89	**2.40***	2.98	3.21	3.09	**2.73**	2.77	2.77	2.78	2.97	3.31	3.02
PYR	1.26	**1.16***	1.33	7e2	1.49	1.29	1.29	**1.26***	**1.26***	1.43	7e2	1.50
RED	**1.27**	1.47	1.29	1.49	1.44	**1.27**	1.28	1.43	1.32	1.28	1.49	1.44
WHI	**1.24**	**1.24**	1.26	1.51	1.42	1.25	**1.23***	1.29	1.23	1.30	1.51	1.42
Rank Sum	22	**16**	29	47	38	22	**20**	38	37	41	65	53

Table 2: Real data performance for $\varepsilon = 1, 2$, and 4. We employ the Wilcoxon signed-rank test (Wilcoxon, 1992) at significance level 0.05. Within the aligned and personalized blocks separately, the best results are **bolded**; a $*$ marks a best result that is significant against the compared methods in the same block, except that in the personalized block a best `HistOfTree`/`AdHistOfTree` result is marked if it is significant against the external competitors. ME and CART stand for max-edge and the classical CART rule, respectively.



(a) Privacy utility trade off (b) Consistency



(c) Parameter analysis (d) Selection of s

Figure 3: Density estimation results on synthetic data. We use `Hist` as the LDP benchmark, captioned as LDP. We use `HistOfTree` with $\varepsilon = 1024$ as the non-private benchmark, captioned as non-private. Other `HistOfTree` results are captioned with the specific choices of parameters s and t. In 3(d), `AdHistOfTree` is captioned as adaptive.

The data-driven selected parameters achieve comparable performance to the best parameter, indicating the effectiveness of the selection rule. In some cases, `AdHistOfTree` performs slightly better, largely due to the incompleteness of parameter grids of `HistOfTree`. Moreover, the relative performance of the ME and CART rules depends on the privacy-preference regime and privacy budget; CART is strong in the aligned setting, while the personalized setting does not exhibit a uniform dominance pattern between the two rules.

7. Conclusion and Discussion

In this work, we for the first time investigate the problem of LDP learning with public features. We provide a comprehensive understanding of this problem within the nonparametric framework, specifically addressing density estimation, regression, and classification tasks. For each of these three problems, we establish minimax lower bounds and propose `HistOfTree`, a unified algorithmic framework that achieves near-tight upper bounds. Our

analysis reveals that this problem represents an intermediate stage between standard local differential privacy and label local differential privacy, requiring specialized techniques for effective estimation.

A remaining practical correlation issue concerns the privacy mask: the current results rely on the assumption that W and (X, y) are independent. In some cases, the two are dependent, which brings additional privacy leakage. For instance, AIDS carriers may choose to conceal their status if they are diagnosed with the condition. If the carriers are aware of the mechanism and believe it could leak information, they may provide incorrect results (choose not to protect this feature and claim no AIDS). We propose two solutions. (i) If the curator wishes to prevent such wrong data collection, it should classify highly correlated features as private beforehand, including these features among the first s^* private features. This approach requires certain prior knowledge and does not completely resolve the privacy accounting issue. (ii) Observing the estimator in (27), there is no need to release W to the server. The server only needs to receive a binary vector of V. To maintain the utility of the proposed methodology, the only remaining issue is the tree partition, which could be mitigated by using random selection instead of information criterion assisted selection (10). Although this reduces the effectiveness of the partition, it entirely eliminates the risk associated with W.

Acknowledgments

The authors would like to thank the reviewers for their constructive comments, which led to a significant improvement in this work. The research is supported by the Special Funds of the National Natural Science Foundation of China (Grant No. 72342010). This research is also supported by Public Computing Cloud, Renmin University of China. The authors declare that they have no competing interests.

Appendix A. Proofs

A.1 Proof of Proposition 4 and 9

Proof of Propositions 4 and 9. We first prove Proposition 9. Since our mechanism (25) and (26) do not rely on previous private information $Z_1, \cdots, Z_{i-1}$ when drawing Z_i, it is a non-interactive mechanism. Each user reveals only two pieces of private information, namely $\tilde{Y}_i$ and $\tilde{V}_i$. Note that they are conditionally independent, thus for each conditional distribution $\mathrm{R}_i\left(\tilde{Y}_i, \tilde{V}_i, |X_i^{\mathrm{priv}} = x_i^{\mathrm{priv}}, X_i^{\mathrm{pub}} = x_i^{\mathrm{pub}}, Y_i = y\right)$, we can compute the density ratio as

$$\begin{aligned}
&\sup_{x_i^{\mathrm{priv}} x_i^{\mathrm{priv}\prime}, x_i^{\mathrm{pub}}, y, y'} \frac{\mathrm{R}_i\left(\tilde{Y}_i, \tilde{V}_i | X_i^{\mathrm{priv}} = x_i^{\mathrm{priv}}, X_i^{\mathrm{pub}} = x_i^{\mathrm{pub}}, Y_i = y\right)}{\mathrm{R}_i\left(\tilde{Y}_i, \tilde{V}_i | X_i^{\mathrm{priv}} = x_i^{\mathrm{priv}\prime}, X_i^{\mathrm{pub}} = x_i^{\mathrm{pub}}, Y_i = y'\right)} \\
\leq &\sup_{x_i^{\mathrm{priv}} x_i^{\mathrm{priv}\prime}, x_i^{\mathrm{pub}}} \frac{\mathrm{R}_i(\tilde{V}_i | X_i^{\mathrm{priv}} = x_i^{\mathrm{priv}}, X_i^{\mathrm{pub}} = x_i^{\mathrm{pub}})}{\mathrm{R}_i(\tilde{V}_i | X_i^{\mathrm{priv}} = x_i^{\mathrm{priv}\prime}, X_i^{\mathrm{pub}} = x_i^{\mathrm{pub}})} \cdot \sup_{y, y'} \frac{\mathrm{R}_i(\tilde{Y}_i | Y_i = y)}{\mathrm{R}_i(\tilde{Y}_i | Y_i = y')}.
\end{aligned} \tag{29}$$

(i) The first part corresponds to the randomized response mechanism. Since the conditional density is identical if x^{priv} and $x^{\text{priv}\prime}$ belongs to a same A_j, we have

$$\frac{\mathrm{R}_i(\tilde{V}_i|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})}{\mathrm{R}_i(\tilde{V}_i|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})}=\frac{\prod_j \mathrm{R}_i(\tilde{V}_i^j|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})}{\prod_j \mathrm{R}_i(\tilde{V}_i^j|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})}.$$

By definition, for two distinct values of x, only two indicators $\mathbf{1}_{A\times B}(x)$ differ over all $A\times B$. Without loss of generality, assume they differ on the first two elements. Then we have

$$\begin{aligned}&\sup_{x_i^{\text{priv}}x_i^{\text{priv}\prime},x_i^{\text{pub}}}\frac{\mathrm{R}_i(\tilde{V}_i|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})}{\mathrm{R}_i(\tilde{V}_i|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})}\\ =&\sup_{x_i^{\text{priv}}x_i^{\text{priv}\prime},x_i^{\text{pub}}}\frac{\prod_{j=1,2}\mathrm{R}_i(\tilde{V}_i^j|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})}{\prod_{j=1,2}\mathrm{R}_i(\tilde{V}_i^j|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})}\le e^{\varepsilon/4+\varepsilon/4}=e^{\varepsilon/2}.\end{aligned}\tag{30}$$

(ii) For the second part, there holds

$$\begin{aligned}\sup_{y,y'}\frac{\mathrm{R}_i(\tilde{Y}_i|Y_i=y)}{\mathrm{R}_i(\tilde{Y}_i|Y_i=y')}\le\sup_{y,y'}\frac{d\mathrm{R}_i(\tilde{Y}_i|Y_i=y)}{d\mathrm{R}_i(\tilde{Y}_i|Y_i=y')}=&\sup_{y,y'}\frac{\exp\left(-\frac{\varepsilon}{4M}|\tilde{Y}_i-y|\right)}{\exp\left(-\frac{\varepsilon}{4M}|\tilde{Y}_i-y'|\right)}\\ \le&\sup_{y,y'}\exp\left(\frac{\varepsilon}{4M}|y-y'|\right)\le e^{\varepsilon/2}.\end{aligned}\tag{31}$$

Bringing (30) and (31) into (29) yields the desired conclusion. Note that (30) yields (12) is $\varepsilon/2$ semi-feature LDP. Substituting $\varepsilon/2$ to ε yields Proposition 4. ■

Proof of Proposition 10. We only need to modify (i) of the previous proof, which is now characterized by the generalized randomized response mechanism. For each grid $A\times B$ with index j, we consider the three cases in (26). The first case is $A\times B\notin\mathcal{V}_i$. The definition of $\mathcal{V}_i$ yields that, for any $X_i'=(x_i^{\text{priv}\prime},x_i^{\text{pub}})$, its set of potential grids $\mathcal{V}_i'$ is exactly the same as $\mathcal{V}_i$. Thus, we have $\mathrm{R}_i(V_i=k|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})=\mathrm{R}_i(V_i=k|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})=0$. And thus the likelihood ratio is bounded by $e^{\varepsilon/2}$ as we define $0/0=1$. The second case is $A\times B\in\mathcal{V}_i$ and $X_i\in A\times B$. In this case, any $X_i'=(x_i^{\text{priv}\prime},x_i^{\text{pub}})$ will have $A\times B\in\mathcal{V}_i'$. As a result,

$$\frac{\mathrm{R}_i(V_i=k|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})}{\mathrm{R}_i(V_i=k|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})}\le\frac{\frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+|\mathcal{V}_i|-1}}{\frac{1}{e^{\varepsilon/2}+|\mathcal{V}_i|-1}}=e^{\varepsilon/2}.$$

The third case is the symmetric situation of the second case. Three cases together, we have

$$\frac{\mathrm{R}_i(V_i|X_i^{\text{priv}}=x_i^{\text{priv}},X_i^{\text{pub}}=x_i^{\text{pub}})}{\mathrm{R}_i(V_i|X_i^{\text{priv}}=x_i^{\text{priv}\prime},X_i^{\text{pub}}=x_i^{\text{pub}})}\le e^{\varepsilon/2}.$$

■

A.2 Proof of Theorem 2

Proof of Theorem 2. It suffices to prove (4). The lower bound contains two terms. The second term $n^{-\frac{\alpha}{2\alpha+d}}$ is the classical minimax lower bound for non-private learners under Assumption 1. See Tsybakov (2009) for a comprehensive discussion. In the following, we prove the conclusion for the first term. The overall proof strategy mimics that of Butucea et al. (2020), while the utilization of information inequality is different. We first construct a finite set of hypotheses. We then combine the information inequality under privacy constraint (Duchi et al., 2018) and Assouad's Lemma (Tsybakov, 2009).

For each user, write

$$S_i = \{\ell \in [d] : W_i^\ell = 1\}, \qquad s_i = |S_i| = \sum_{\ell=1}^{d} W_i^\ell.$$

The assumption in Theorem 2 ensures that S_i is nonempty for every i.

Let $\psi_0 : \mathbb{R} \to \mathbb{R}$ be an α-Hölder continuous function supported on $(0,1)$ such that

$$\int_0^1 \psi_0(u)du = 0, \qquad \int_0^1 |\psi_0(u)|du > 0, \qquad \|\psi_0\|_\infty < \infty.$$

For $u = (u^1, \ldots, u^d) \in \mathbb{R}^d$, define

$$\psi(u) = \prod_{\ell=1}^{d} \psi_0(u^\ell).$$

Then ψ is supported on $(0,1)^d$ and integrates to zero over every nonempty subset of coordinates. Let

$$\mathcal{J}_k = \{0, \ldots, k-1\}^d, \qquad C_j = \prod_{\ell=1}^{d} \left[j^\ell/k, (j^\ell+1)/k \right], \qquad j \in \mathcal{J}_k.$$

For $x \in C_j$, define

$$\psi_j(x) = k^{-\alpha}\psi(kx - j),$$

and set $\psi_j(x) = 0$ for $x \notin C_j$. For $\theta = (\theta_j)_{j\in\mathcal{J}_k} \in \{-1, +1\}^{\mathcal{J}_k}$, define

$$p_X^\theta(x) = 1 + \rho \sum_{j\in\mathcal{J}_k} \theta_j \psi_j(x),$$

where $\rho > 0$ is a sufficiently small constant. Since the supports of the ψ_j's are disjoint and each ψ_j has zero integral, p_X^θ integrates to one. For ρ sufficiently small and k sufficiently large, p_X^θ is nonnegative and satisfies Assumption 1. Moreover, p_X^θ is bounded from below and above by positive numerical constants.

Let us now assume that the raw data have been anonymized by means of an arbitrary privacy mechanism R that is sequentially-interactive ε-semi feature LDP. R maps each X_i to Z_i based on $Z_1, \cdots, Z_{i-1}$. Let $T_i = (X_i^{\text{pub}}, Z_i)$ be the information available from user

i, and let $\mathrm{RP}^{\theta n}$ denote the joint distribution of $T_{1:n}$. Let $\widehat{p}_X$ be any estimator measurable with respect to $T_{1:n}$. For each fixed $\widehat{p}_X$, define

$$\widehat{\theta} \in \arg\min_{\vartheta \in \{-1,+1\}^{\mathcal{J}_k}} \|p_X^{\vartheta} - \widehat{p}_X\|_{L_1(dx)}.$$

Since p_X^{θ} is uniformly bounded from below, the density-estimation loss satisfies

$$\|p_X^{\theta} - \widehat{p}_X\|_{L_1(\mathrm{P}_X^{\theta})} \gtrsim \|p_X^{\theta} - \widehat{p}_X\|_{L_1(dx)}.$$

By the triangular decomposition and the definition of $\widehat{\theta}$,

$$\|p_X^{\theta} - \widehat{p}_X\|_{L_1(dx)} \geq \frac{1}{2}\|p_X^{\theta} - p_X^{\widehat{\theta}}\|_{L_1(dx)}.$$

Because the bumps have disjoint supports,

$$\begin{aligned}\|p_X^{\theta} - p_X^{\vartheta}\|_{L_1(dx)} &= \rho \sum_{j \in \mathcal{J}_k} |\theta_j - \vartheta_j| \int_{C_j} |\psi_j(x)| dx \\ &= 2\rho k^{-\alpha-d} \|\psi\|_1 Ham(\theta, \vartheta),\end{aligned}$$

where Ham denotes the hamming distance between vectors. Therefore,

$$\sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[\|p_X^{\theta} - \widehat{p}_X\|_{L_1(\mathrm{P}_X^{\theta})} \right] \gtrsim k^{-\alpha-d} \inf_{\widehat{\theta}} \sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[Ham(\widehat{\theta}, \theta) \right]. \tag{32}$$

It remains to control the KL divergence between neighboring hypotheses. For $j \in \mathcal{J}_k$, let $\theta^{(j)}$ be the vector obtained by flipping the j-th coordinate of θ. Let $H_{i-1} = T_{1:i-1}$. For every history h in the support of H_{i-1}, let $M_{i,\theta}^h$ denote the conditional law of T_i given $H_{i-1} = h$ under P^{θ}. By the chain rule for KL divergence,

$$KL(M_{i,\theta}^h \| M_{i,\theta^{(j)}}^h) = KL\left(\mathrm{P}_i^{\theta,\mathrm{pub}} \| \mathrm{P}_i^{\theta^{(j)},\mathrm{pub}}\right) \tag{33}$$

$$+ \mathbb{E}_{X_i^{\mathrm{pub}} \sim \mathrm{P}_i^{\theta,\mathrm{pub}}} KL\left((M_{i,\theta}^h)_{Z_i|X_i^{\mathrm{pub}}} \middle\| (M_{i,\theta^{(j)}}^h)_{Z_i|X_i^{\mathrm{pub}}} \right). \tag{34}$$

Since $s_i \geq 1$, the first term is zero. Indeed, by the product construction of ψ, integrating ψ_j over the nonempty collection of private coordinates of user i gives zero, and hence the public marginal distributions under θ and $\theta^{(j)}$ coincide. Conditional on $H_{i-1} = h$ and $X_i^{\mathrm{pub}} = x^{\mathrm{pub}}$, Definition 1 implies that Z_i is ε-LDP with respect to X_i^{priv}. Thus, the symmetrized information contraction inequality for locally private channels (Duchi et al., 2018) implies the following directional bound:

$$KL\left((M_{i,\theta}^h)_{Z_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \middle\| (M_{i,\theta^{(j)}}^h)_{Z_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \right) \tag{35}$$

$$\leq C(e^{\varepsilon}-1)^2 V^2 \left(\mathrm{P}^{\theta}_{X_i^{\mathrm{priv}}|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}, \mathrm{P}^{\theta^{(j)}}_{X_i^{\mathrm{priv}}|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \right). \tag{36}$$

For x^{pub} outside the public projection of C_j, this conditional total variation distance is zero. On the public projection of C_j, the conditional densities differ only through the j-th bump. Since p_X^{θ} is uniformly bounded above and below,

$$V\left(\mathrm{P}^{\theta}_{X_i^{\mathrm{priv}}|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}, \mathrm{P}^{\theta^{(j)}}_{X_i^{\mathrm{priv}}|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \right) \leq Ck^{-\alpha-s_i}.$$

The public projection of C_j has Lebesgue measure of order $k^{-(d-s_i)}$. Combining this display with (33) and (35) yields

$$KL(M_{i,\theta}^h \| M_{i,\theta^{(j)}}^h) \leq C(e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i}.$$

The bound is uniform in h. Applying the KL chain rule to the interactive transcript gives

$$KL(\mathrm{RP}^{\theta n} \| \mathrm{RP}^{\theta^{(j)} n}) = \sum_{i=1}^n \mathbb{E}_{\mathrm{RP}^{\theta n}} KL\left(M_{i,\theta}^{H_{i-1}} \middle\| M_{i,\theta^{(j)}}^{H_{i-1}}\right) \tag{37}$$

$$\leq C \sum_{i=1}^n (e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i}. \tag{38}$$

Choose k such that the right-hand side of (37) is bounded by a sufficiently small constant. Assouad's Lemma then yields

$$\inf_{\widehat{\theta}} \sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[Ham(\widehat{\theta}, \theta) \right] \gtrsim k^d.$$

Bringing this result into (32) leads to

$$\sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[\| p_X^\theta - \widehat{p}_X \|_{L_1(\mathrm{P}_X^\theta)} \right] \gtrsim k^{-\alpha}.$$

Letting $\mathcal{E} = k^{-\alpha}$, up to the constant change caused by rounding k to an integer, the KL condition becomes

$$\sum_{i=1}^n (e^\varepsilon - 1)^2 \mathcal{E}^{\frac{2\alpha+d+s_i}{\alpha}} \leq c.$$

The exponents in this display range over a fixed compact interval depending only on α and d. Consequently, multiplying the supremum in (3) by a sufficiently small constant makes this condition hold. This proves the first term in (4) up to the constants hidden by $\gtrsim$. The aligned result (5) follows by setting $s_i = s^*$ for all i. ■

A.3 Proof of Theorem 5

Proof of Theorem 5. The lower bound contains two terms. The second term $n^{-\frac{2\alpha}{2\alpha+d}}$ is the classical minimax lower bound for non-private learners under Assumption 2. See Tsybakov (2009) for a comprehensive discussion. In the following, we prove the conclusion for the first term. The overall proof strategy mimics that of Györfi and Kroll (2025), while the utilization of information inequality is different. We first construct a finite set of hypotheses. We then combine the information inequality under privacy constraint (Duchi et al., 2018) and Assouad's Lemma (Tsybakov, 2009).

For each user, write

$$S_i = \{\ell \in [d] : W_i^\ell = 1\}, \qquad s_i = |S_i| = \sum_{\ell=1}^d W_i^\ell.$$

Let $K : [0,1]^d \to [0,\infty)$ be an α-Hölder continuous function supported on $(0,1)^d$ such that

$$\|K\|_\infty \le 1, \qquad \int_{[0,1]^d} K^2(u)du > 0.$$

Let

$$\mathcal{J}_k = \{0,\dots,k-1\}^d, \qquad C_j = \prod_{\ell=1}^d \left[j^\ell/k, (j^\ell+1)/k\right], \qquad j \in \mathcal{J}_k.$$

For $x \in C_j$, define

$$K^j(x) = \rho k^{-\alpha} K(kx - j),$$

and set $K^j(x) = 0$ for $x \notin C_j$, where $\rho > 0$ is sufficiently small. For $\theta = (\theta_j)_{j\in\mathcal{J}_k} \in \{-1,+1\}^{\mathcal{J}_k}$, define

$$f^\theta(x) = \sum_{j\in\mathcal{J}_k} \theta_j K^j(x).$$

We consider the following class of distributions P^θ. The marginal distribution P^θ_X is uniform on $[0,1]^d$, and $Y \mid X = x$ is the uniform distribution on $[f^\theta(x) - M/2, f^\theta(x) + M/2]$. Then $f^\theta(x) = \mathbb{E}_\theta(Y \mid X = x)$. For ρ sufficiently small, $|f^\theta(x)| \le M/2$, so the conditional support of Y is contained in $[-M, M]$. The same choice ensures that f^θ satisfies Assumption 2; the marginal density is also bounded above and below. Thus, all P^θs belong to $\mathcal{F}$.

Let us now assume that the raw data have been anonymized by means of an arbitrary privacy mechanism R that is sequentially-interactive ε-semi feature LDP. R maps each (X_i, Y_i) to Z_i based on $Z_1, \cdots, Z_{i-1}$. Let $T_i = (X_i^{\mathrm{pub}}, Z_i)$, and let $\mathrm{RP}^{\theta n}$ denote the joint distribution of $T_{1:n}$. Let $\widehat{f}$ be any estimator measurable with respect to $T_{1:n}$. Since the marginal distribution of X is uniform, the excess risk under the square loss satisfies

$$\mathcal{R}_{\mathrm{P}^\theta}(\widehat{f}) - \mathcal{R}^*_{\mathrm{P}^\theta} = \int_{[0,1]^d} \left(\widehat{f}(x) - f^\theta(x)\right)^2 dx.$$

For each fixed $\widehat{f}$, define

$$\widehat{\theta} \in \arg\min_{\vartheta\in\{-1,+1\}^{\mathcal{J}_k}} \int_{[0,1]^d} \left(\widehat{f}(x) - f^\vartheta(x)\right)^2 dx.$$

By the triangular decomposition in $L_2([0,1]^d)$,

$$\int_{[0,1]^d} \left(\widehat{f}(x) - f^\theta(x)\right)^2 dx \ge \frac{1}{4}\int_{[0,1]^d} \left(f^{\widehat{\theta}}(x) - f^\theta(x)\right)^2 dx.$$

Because the bumps have disjoint supports,

$$\begin{aligned}\int_{[0,1]^d} \left(f^\vartheta(x) - f^\theta(x)\right)^2 dx &= \sum_{j\in\mathcal{J}_k} |\vartheta_j - \theta_j|^2 \int_{C_j} \left(K^j(x)\right)^2 dx \\ &= 4\rho^2 k^{-2\alpha-d}\|K\|_2^2 Ham(\vartheta,\theta).\end{aligned}$$

Therefore,

$$\sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[\mathcal{R}_{\mathrm{P}^\theta}(\widehat{f}) - \mathcal{R}^*_{\mathrm{P}^\theta} \right] \gtrsim k^{-2\alpha-d} \inf_{\widehat{\theta}} \sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[Ham(\widehat{\theta}, \theta) \right]. \tag{39}$$

It remains to control the KL divergence between neighboring hypotheses. For $j \in \mathcal{J}_k$, let $\theta^{(j)}$ be the vector obtained by flipping the j-th coordinate of θ. Let $H_{i-1} = T_{1:i-1}$. For every history h in the support of H_{i-1}, let $M^h_{i,\theta}$ denote the conditional law of T_i given $H_{i-1} = h$ under P^θ. By the chain rule for KL divergence,

$$KL(M^h_{i,\theta} \| M^h_{i,\theta^{(j)}}) = KL\left(\mathrm{P}^{\theta,\mathrm{pub}}_i \| \mathrm{P}^{\theta^{(j)},\mathrm{pub}}_i \right) \tag{40}$$

$$+ \mathbb{E}_{X_i^{\mathrm{pub}} \sim \mathrm{P}_i^{\theta,\mathrm{pub}}} KL\left((M^h_{i,\theta})_{Z_i|X_i^{\mathrm{pub}}} \middle\| (M^h_{i,\theta^{(j)}})_{Z_i|X_i^{\mathrm{pub}}} \right). \tag{41}$$

The first term is zero because the marginal distribution of X is uniform for every θ. Conditional on $H_{i-1} = h$ and $X_i^{\mathrm{pub}} = x^{\mathrm{pub}}$, Definition 1 implies that Z_i is ε-LDP with respect to $(X_i^{\mathrm{priv}}, Y_i)$. Thus, the symmetrized information contraction inequality for locally private channels (Duchi et al., 2018) implies the following directional bound:

$$KL\left((M^h_{i,\theta})_{Z_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \middle\| (M^h_{i,\theta^{(j)}})_{Z_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \right) \tag{42}$$

$$\leq C(e^\varepsilon - 1)^2 V^2 \left(\mathrm{P}^\theta_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}, \mathrm{P}^{\theta^{(j)}}_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \right). \tag{43}$$

For x^{pub} outside the public projection of C_j, the total variation distance is zero. On the public projection of C_j, the two conditional laws differ only through the conditional distribution of Y on the j-th private slice. The total variation distance between two length-M uniform distributions whose centers differ by Δ is at most $|\Delta|/M$. Hence

$$V\left(\mathrm{P}^\theta_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}, \mathrm{P}^{\theta^{(j)}}_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \right) \leq Ck^{-\alpha-s_i}.$$

The public projection of C_j has Lebesgue measure of order $k^{-(d-s_i)}$. Combining this display with (40) and (42) gives

$$KL(M^h_{i,\theta} \| M^h_{i,\theta^{(j)}}) \leq C(e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i}.$$

The bound is uniform in h. Applying the KL chain rule to the interactive transcript gives

$$KL(\mathrm{RP}^{\theta n} \| \mathrm{RP}^{\theta^{(j)} n}) = \sum_{i=1}^n \mathbb{E}_{\mathrm{RP}^{\theta n}} KL\left(M^{H_{i-1}}_{i,\theta} \middle\| M^{H_{i-1}}_{i,\theta^{(j)}} \right) \tag{44}$$

$$\leq C \sum_{i=1}^n (e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i}. \tag{45}$$

Choose k such that the right-hand side of (44) is bounded by a sufficiently small constant. Assouad's Lemma then yields

$$\inf_{\widehat{\theta}} \sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[Ham(\widehat{\theta}, \theta) \right] \gtrsim k^d.$$

Combined with (39), this gives

$$\sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}} \left[\mathcal{R}_{\mathrm{P}^\theta}(\widehat{f}) - \mathcal{R}^*_{\mathrm{P}^\theta} \right] \gtrsim k^{-2\alpha}.$$

Letting $\mathcal{E} = k^{-2\alpha}$, up to the constant change caused by rounding k to an integer, the KL condition becomes

$$\sum_{i=1}^{n} (e^\varepsilon - 1)^2 \mathcal{E}^{\frac{2\alpha+d+s_i}{2\alpha}} \leq c.$$

The exponents in this display range over a fixed compact interval depending only on α and d. Consequently, multiplying the supremum in (17) by a sufficiently small constant makes this condition hold. This proves the first term in (18) up to the constants hidden by $\gtrsim$. The aligned result follows by setting $s_i = s^*$ for all i. ■

A.4 Proof of Theorem 7

Proof of Theorem 7. The lower bound contains two terms. The second term $n^{-\frac{\alpha(1+\beta)}{2\alpha+d}}$ is the classical minimax lower bound for non-private learners under Assumption 2 and 3. See Audibert and Tsybakov (2007); Samworth (2012); Chaudhuri and Dasgupta (2014) for a comprehensive discussion. In the following, we prove the conclusion for the first term. The overall proof strategy mimics that of Berrett and Butucea (2019), while the utilization of information inequality is different. We first construct a finite set of hypotheses. We then combine the information inequality under privacy constraint (Duchi et al., 2018) and Assouad's Lemma (Tsybakov, 2009).

For each user, write

$$S_i = \{\ell \in [d] : W_i^\ell = 1\}, \qquad s_i = |S_i| = \sum_{\ell=1}^{d} W_i^\ell.$$

We use an Assouad construction adapted to the strong density assumption. The assumption $\alpha\beta \leq 1$ ensures that the Hölder boundary layers of the bumps satisfy the margin condition. Let $m = \max\{1, \lfloor c_m k^{d-\alpha\beta} \rfloor\}$ for a sufficiently small constant $c_m > 0$. Choose m disjoint cells indexed by $\sigma_{(1)}, \ldots, \sigma_{(m)} \in \{0, \ldots, k-1\}^d$. Let $\theta \in \{-1, +1\}^m$. For $u \in \mathbb{R}^d$, define

$$K_0(u) = c_K \operatorname{dist}\left(u, \left((0,1)^d\right)^c\right)^\alpha,$$

where $c_K > 0$ is sufficiently small. The function K_0 is nonnegative, vanishes outside $(0,1)^d$, and is α-Hölder. Moreover, for $0 < t \leq \|K_0\|_\infty$,

$$\lambda\left(0 < K_0(u) \leq t\right) \leq C t^{1/\alpha} \leq C t^\beta,$$

where the last inequality uses $\alpha\beta \leq 1$. For x in the cell indexed by $\sigma_{(a)}$, define

$$K^{\sigma_{(a)}}(x) = \rho k^{-\alpha} K_0(kx - \sigma_{(a)}),$$

and set $K^{\sigma_{(a)}}(x) = 0$ outside that cell, where $\rho > 0$ is a sufficiently small constant. The bumps have disjoint supports and satisfy

$$0 \le K^{\sigma_{(a)}}(x) \le Ck^{-\alpha}, \qquad \int K^{\sigma_{(a)}}(x)dx \asymp k^{-\alpha-d},$$

as well as

$$\lambda\left(0 < K^{\sigma_{(a)}}(x) \le \gamma\right) \le Ck^{-d}\left(\frac{\gamma}{k^{-\alpha}}\right)^{\beta}, \qquad 0 < \gamma \le Ck^{-\alpha}.$$

For $\theta \in \{-1,+1\}^m$, define the regression function

$$f^{\theta}(x) = \frac{1}{2} + \frac{1}{2}\sum_{a=1}^{m} \theta_a K^{\sigma_{(a)}}(x).$$

We consider the class of distributions P^{θ} where P_X^{θ} is the uniform distribution on $[0,1]^d$ and $Y \mid X = x$ is binary with

$$\mathrm{P}^{\theta}(Y = 1 \mid X = x) = f^{\theta}(x).$$

For sufficiently small constants in the construction, f^{θ} is α-Hölder and takes values in $[0,1]$. For $0 < \gamma \le Ck^{-\alpha}$, the level-set bound and $mk^{-d} \asymp k^{-\alpha\beta}$ give

$$\mathrm{P}_X^{\theta}\left(0 < \left|f^{\theta}(X) - \frac{1}{2}\right| \le \gamma\right) \le C_{\beta}\gamma^{\beta}.$$

For larger γ, the same inequality follows because the probability of the union of the active cells is of order $k^{-\alpha\beta} \lesssim \gamma^{\beta}$. Thus, all P^{θ}s belong to $\mathcal{F}$.

Let us now assume that the raw data have been anonymized by means of an arbitrary privacy mechanism R that is sequentially-interactive ε-semi feature LDP. R maps each (X_i, Y_i) to Z_i based on $Z_1, \cdots, Z_{i-1}$. Let $T_i = (X_i^{\mathrm{pub}}, Z_i)$, and let $\mathrm{RP}^{\theta n}$ denote the joint distribution of $T_{1:n}$. Let $\tilde{f}$ be any estimator measurable with respect to $T_{1:n}$, and let $\tilde{g}(x) = \mathbf{1}\{\tilde{f}(x) \ge 1/2\}$ be the induced classifier. For each $a \in [m]$, define the weighted decoder

$$\widehat{\theta}_a \in \arg\min_{v\in\{-1,+1\}} \int K^{\sigma_{(a)}}(x)\,|\tilde{g}(x) - \mathbf{1}\{v = +1\}|\,dx.$$

If $\widehat{\theta}_a \ne \theta_a$, the weighted disagreement of $\tilde{g}$ with the Bayes classifier on the a-th cell is at least one half of $\int K^{\sigma_{(a)}}(x)dx$. Since the excess classification risk equals

$$\int_{[0,1]^d} \left|2f^{\theta}(x) - 1\right| \mathbf{1}\left\{\tilde{g}(x) \ne \mathbf{1}\left\{f^{\theta}(x) \ge \frac{1}{2}\right\}\right\} dx,$$

the reduction to hamming loss gives

$$\sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}}\left[\mathcal{R}_{\mathrm{P}^{\theta}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}^{\theta}}\right] \gtrsim k^{-\alpha-d} \inf_{\widehat{\theta}} \sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}}\left[Ham(\widehat{\theta}, \theta)\right]. \tag{46}$$

Here we used $\int K^{\sigma_{(a)}}(x)dx \asymp k^{-\alpha-d}$ for every active cell.

It remains to control the KL divergence between neighboring hypotheses. For $a \in [m]$, let $\theta^{(a)}$ be the vector obtained by flipping the a-th coordinate of θ. Let $H_{i-1} = T_{1:i-1}$. For every history h in the support of H_{i-1}, let $M^h_{i,\theta}$ denote the conditional law of T_i given $H_{i-1} = h$ under P^θ. By the chain rule for KL divergence,

$$KL(M^h_{i,\theta} \| M^h_{i,\theta^{(a)}}) = KL\left(\mathrm{P}_i^{\theta,\mathrm{pub}} \| \mathrm{P}_i^{\theta^{(a)},\mathrm{pub}}\right) \tag{47}$$

$$+ \mathbb{E}_{X_i^{\mathrm{pub}} \sim \mathrm{P}_i^{\theta,\mathrm{pub}}} KL\left((M^h_{i,\theta})_{Z_i|X_i^{\mathrm{pub}}} \middle\| (M^h_{i,\theta^{(a)}})_{Z_i|X_i^{\mathrm{pub}}}\right). \tag{48}$$

The first term is zero because the marginal distribution of X is uniform for every θ. Conditional on $H_{i-1} = h$ and $X_i^{\mathrm{pub}} = x^{\mathrm{pub}}$, Definition 1 implies that Z_i is ε-LDP with respect to $(X_i^{\mathrm{priv}}, Y_i)$. Thus, the symmetrized information contraction inequality for locally private channels (Duchi et al., 2018) implies the following directional bound:

$$KL\left((M^h_{i,\theta})_{Z_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}} \middle\| (M^h_{i,\theta^{(a)}})_{Z_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}\right) \tag{49}$$

$$\leq C(e^\varepsilon - 1)^2 V^2 \left(\mathrm{P}^\theta_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}, \mathrm{P}^{\theta^{(a)}}_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}\right). \tag{50}$$

For x^{pub} outside the public projection of the a-th active cell, the total variation distance is zero. On the public projection of the active cell, the two conditional laws differ only through the Bernoulli label probability on a private slice of volume of order k^{-s_i}. Since the two Bernoulli probabilities differ by at most $Ck^{-\alpha}$, we have

$$V\left(\mathrm{P}^\theta_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}, \mathrm{P}^{\theta^{(a)}}_{X_i^{\mathrm{priv}},Y_i|X_i^{\mathrm{pub}}=x^{\mathrm{pub}}}\right) \leq Ck^{-\alpha-s_i}.$$

The public projection of the active cell has Lebesgue measure of order $k^{-(d-s_i)}$. Combining this display with (47) and (49) gives

$$KL(M^h_{i,\theta} \| M^h_{i,\theta^{(a)}}) \leq C(e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i}.$$

The bound is uniform in h. Applying the KL chain rule to the interactive transcript gives

$$KL(\mathrm{RP}^{\theta n} \| \mathrm{RP}^{\theta^{(a)} n}) = \sum_{i=1}^n \mathbb{E}_{\mathrm{RP}^{\theta n}} KL\left(M^{H_{i-1}}_{i,\theta} \middle\| M^{H_{i-1}}_{i,\theta^{(a)}}\right) \tag{51}$$

$$\leq C \sum_{i=1}^n (e^\varepsilon - 1)^2 k^{-2\alpha-d-s_i}. \tag{52}$$

Choose k such that the right-hand side of (51) is bounded by a sufficiently small constant. Assouad's Lemma gives

$$\inf_{\widehat{\theta}} \sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}}\left[Ham(\widehat{\theta}, \theta)\right] \gtrsim m.$$

Combined with (46) and $m \asymp k^{d-\alpha\beta}$, this leads to

$$\sup_{\theta} \mathbb{E}_{\mathrm{RP}^{\theta n}}\left[\mathcal{R}_{\mathrm{P}^\theta}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}^\theta}\right] \gtrsim k^{-\alpha-d} m \asymp k^{-\alpha(1+\beta)}.$$

Letting $\mathcal{E} = k^{-\alpha}$, up to the constant change caused by rounding k to an integer, the KL condition becomes

$$\sum_{i=1}^{n} (e^{\varepsilon} - 1)^2 \mathcal{E}^{\frac{2\alpha+d+s_i}{\alpha}} \leq c.$$

The exponents in this display range over a fixed compact interval depending only on α and d. Consequently, multiplying the supremum in (21) by a sufficiently small constant makes this condition hold. This proves the first term in (22) up to the constants hidden by $\gtrsim$. The aligned result follows by setting $s_i = s^*$ for all i. ■

A.5 Proof of Theorem 3

We first define the intermediate quantity

$$\overline{p}_X(x) = \sum_{A_j, B_k} \mathbf{1}_{A_j \times B_k}(x) \frac{\int_{A_j \times B_k} d\mathrm{P}_X(x')}{\mu(A_j \times B_k)}. \tag{53}$$

For the same partition and the weights in (13), define the weighted non-private estimator

$$\widehat{p}_{X,\omega}(x) = \sum_{A_j, B_k} \mathbf{1}_{A_j \times B_k}(x) \frac{\sum_{i=1}^{n} \omega_i(s,T) V_i^{j,k}}{\mu(A_j \times B_k)}, \tag{54}$$

where $V_i^{j,k} := \mathbf{1}_{A_j \times B_k}(X_i)$. We rely on the following decomposition.

$$\|p_X^* - \tilde{p}_X\|_{L_1} \leq \underbrace{\|\tilde{p}_X - \widehat{p}_{X,\omega}\|_{L_1}}_{\textbf{Privatized Error}} + \underbrace{\|\widehat{p}_{X,\omega} - \overline{p}_X\|_{L_1}}_{\textbf{Sample Error}} + \underbrace{\|\overline{p}_X - p_X^*\|_{L_1}}_{\textbf{Approximation Error}}. \tag{55}$$

where the `HistOfTree` density estimator $\tilde{p}_X$, the weighted non-private partition-based density estimator $\widehat{p}_{X,\omega}$, and the population partition-based density estimator $\overline{p}_X$ are defined in (14), (54), and (53), respectively. Loosely speaking, the first error term quantifies the degradation brought by adding privacy noises to the estimator, which we call the privatized error. The second term corresponds to the expected estimation error brought by the randomness of the data, which we call the sample error. The last term is called approximation error, which arises due to the limited approximation capacity of piecewise constant functions. The following three lemmas provide bounds for each of the three errors.

Lemma 11 (Bounding privatised error) *Let $\tilde{p}_X$ be the `HistOfTree` estimator defined in (14). Let $\widehat{p}_{X,\omega}$ be the weighted non-private partition-based estimator defined in (54). For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Let Assumption 1 hold. Write $h = 2^{-T/(d-s)}$, $q_i(s) = \sum_{\ell=s+1}^{d} W_i^{\ell}$, and $S_{s,T} = \sum_i h^{q_i(s)}$. Suppose $t^s 2^T \log n \lesssim S_{s,T}$. Then, there holds*

$$\|\tilde{p}_X - \widehat{p}_{X,\omega}\|_{L_1} \lesssim \sqrt{\frac{t^{2s} 2^T \log n}{\varepsilon^2 S_{s,T}}}$$

with probability $\mathrm{P}^n \otimes \mathrm{R}^n$ at least $1 - 1/n^2$.

Lemma 12 (Bounding sample error) *Let the weighted non-private partition-based estimator $\widehat{p}_{X,\omega}$ and the population partition-based estimator $\overline{p}_X$ be defined in (54) and (53), respectively. For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Let Assumption 1 hold. Suppose $t^s 2^T \log n \lesssim S_{s,T}$. Then, there holds*

$$\|\widehat{p}_{X,\omega} - \overline{p}_X\|_{L_1} \lesssim \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}}.$$

with probability $\mathrm{P}^n \otimes \mathrm{R}^n$ at least $1 - 1/n^2$.

Lemma 13 (Bounding approximation error) *Let the population partition-based estimator $\overline{p}_X$ be defined in (53). For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Let Assumption 1 hold. Then there holds*

$$\|\overline{p}_X - p_X^*\|_{L_1} \lesssim\ t^{-\alpha} + 2^{-\alpha T/(d-s)}.$$

Proof of Theorem 3.

The privacy of $\tilde{p}_X$ follows from Proposition 4, since the deterministic weights are post-processing. We then focus on the excess risk upper bound. Bringing Proposition 11, 12, and 13 into the decomposition (55), we have

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \sqrt{\frac{t^{2s} 2^T \log n}{\varepsilon^2 S_{s,T}}} + \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}} + t^{-\alpha} + 2^{-\alpha T/(d-s)} \tag{56}$$

holds with probability $\mathrm{P}^n \otimes \mathrm{R}^n$ at least $1 - 2/n^2$. We first consider the case where the first term dominates the second term. This is always the relevant case for $0 < \varepsilon \leq 1$, since $t^s \geq 1$. We take $t \asymp 2^{T/(d-s)}$. In this case, we have

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \sqrt{\frac{2^{T(d+s)/(d-s)} \log n}{\varepsilon^2 \sum_{i=1}^n 2^{-q_i(s)T/(d-s)}}} + 2^{-\alpha T/(d-s)},$$

which recovers equation (7). To show the matched personalized rate, we further take $s = 0$ and $t \asymp 2^{T/d}$. Then, (56) becomes

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \sqrt{\frac{\log n}{\varepsilon^2 h^d \sum_{i=1}^n h^{s_i^*}}} + h^\alpha, \qquad h := 2^{-T/d}.$$

The minimum is attained, up to constants, when the two terms are of the same order. Thus, writing $\overline{\mathcal{E}}^* \asymp h^\alpha$, some algebra yields

$$\sum_{i=1}^n \overline{\mathcal{E}}^{*\frac{2\alpha+d+s_i^*}{\alpha}} \asymp \frac{\log n}{\varepsilon^2},$$

which is (6) and has the same positive-moment structure as the lower bound. In the aligned case, we take $s = s^*$ and the weights reduce to $1/n$. The bound becomes

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \sqrt{\frac{2^{T(d+s)/(d-s)} \cdot \log n}{n\varepsilon^2}} + 2^{-\alpha T/(d-s)},$$

where taking $2^T = (n\varepsilon^2)^{\frac{d-s}{d+s+2\alpha}}$ yields

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \left(\frac{\log n}{n\varepsilon^2}\right)^{\frac{\alpha}{2\alpha+d+s^*}}.$$

If the sample term dominates, we take $t \asymp 2^{T/(d-s)}$ and $2^T \asymp (n/\log n)^{\frac{d-s}{2\alpha+d}}$, as in the original proof. This yields

$$\|\tilde{p}_X - p_X^*\|_{L_1} \lesssim \left(\frac{\log n}{n}\right)^{\frac{\alpha}{2\alpha+d}}.$$

This leads to the non-private bound in both aligned and personalized cases. ■

A.6 Proof of Theorem 6

We first define the intermediate quantity

$$\overline{f}(x) = \sum_{A_j, B_k} \mathbf{1}_{A_j \times B_k}(x) \frac{\int_{A_j \times B_k} f^*(x') d\mathrm{P}_X(x')}{\int_{A_j \times B_k} d\mathrm{P}_X(x')}. \tag{57}$$

We rely on the following decomposition.

$$\left\|\tilde{f} - f^*\right\|_\infty \leq \underbrace{\left\|\tilde{f} - \widehat{f}_\omega\right\|_\infty}_{\textbf{Privatized Error}} + \underbrace{\left\|\widehat{f}_\omega - \overline{f}\right\|_\infty}_{\textbf{Sample Error}} + \underbrace{\left\|\overline{f} - f^*\right\|_\infty}_{\textbf{Approximation Error}}. \tag{58}$$

where the `HistOfTree` estimator $\tilde{f}$, the weighted non-private partition-based estimator $\widehat{f}_\omega$, and the population partition-based estimator $\overline{f}$ are defined in (27), (24), and (57), respectively. Loosely speaking, the first error term quantifies the degradation brought by adding privacy noises to the estimator, which we call the privatized error. The second term corresponds to the expected estimation error brought by the randomness of the data, which we call the sample error. The last term is called approximation error, which arises due to the limited approximation capacity of piecewise constant functions. The following three lemmas provide bounds for each of the three errors.

Lemma 14 (Bounding privatised error) *Let $\tilde{f}$ be the `HistOfTree` estimator defined in (27). Let $\widehat{f}_\omega$ be the weighted non-private partition-based estimator defined in (24). For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by*

Algorithm 1 with depth T. Let Assumption 2 hold. Suppose $t^{2s}2^T \log n \lesssim \varepsilon^2 S_{s,T}$. Then, there holds

$$\|\tilde{f} - \widehat{f}_\omega\|_\infty \lesssim \sqrt{\frac{t^{2s}2^T \log n}{\varepsilon^2 S_{s,T}}}.$$

with probability $\mathrm{P}^n \otimes \mathrm{R}^n$ at least $1 - 4/n^2$.

Lemma 15 (Bounding sample error) *Let the weighted non-private partition-based estimator $\widehat{f}_\omega$ and the population partition-based estimator $\overline{f}$ be defined in (24) and (57), respectively. For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Let Assumption 2 hold. Suppose $t^s 2^T \log n \lesssim S_{s,T}$. Then, there holds*

$$\|\widehat{f}_\omega - \overline{f}\|_\infty \lesssim \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}}.$$

with probability $\mathrm{P}^n \otimes \mathrm{R}^n$ at least $1 - 4/n^2$.

Lemma 16 (Bounding approximation error) *Let the population partition-based estimator $\overline{f}$ be defined in (57). For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Let Assumption 2 hold. Then there holds*

$$\|\overline{f} - f^*\|_\infty \lesssim\ t^{-\alpha} + 2^{-\alpha T/(d-s)}.$$

Proof of Theorem 6. The privacy of $\tilde{f}$ follows from Proposition 9, since the deterministic weights are post-processing. We then focus on the excess risk upper bound. Bringing Proposition 14, 15, and 16 into the decomposition (58), we have

$$\begin{aligned}
\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} =& \int_{\mathcal{X}} |\tilde{f}(x) - f^*(x)|^2 d\mathrm{P}_X(x) \\
\leq& \|\tilde{f} - f^*\|^2_\infty \\
\leq& 3\left(\|\tilde{f} - \widehat{f}_\omega\|^2_\infty + \|\widehat{f}_\omega - \overline{f}\|^2_\infty + \|\overline{f} - f^*\|^2_\infty\right) \\
\lesssim& \frac{t^{2s}2^T \log n}{\varepsilon^2 S_{s,T}} + \frac{t^s 2^T \log n}{S_{s,T}} + t^{-2\alpha} + 2^{-2\alpha T/(d-s)}
\end{aligned} \tag{59}$$

holds with probability $\mathrm{P}^n \otimes \mathrm{R}^n$ at least $1 - 8/n^2$. We first consider when the first term dominates the second term. For $0 < \varepsilon \leq 1$, this follows from $t^s \geq 1$. We take $t \asymp 2^{T/(d-s)}$. Thus we have $t^{-2\alpha} \asymp 2^{-2\alpha T/(d-s)}$ and $t^{2s} \cdot 2^T \asymp 2^{T(d+s)/(d-s)}$. Therefore, (59) becomes

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} \lesssim \frac{2^{T(d+s)/(d-s)} \log n}{\varepsilon^2 \sum_{i=1}^n 2^{-q_i(s)T/(d-s)}} + 2^{-2\alpha T/(d-s)},$$

This is the oracle upper bound for a general candidate pair (s, T). For the matched personalized rate, take $s = 0$, write $h = 2^{-T/d}$, and balance the two terms. If $\overline{\mathcal{E}}^* \asymp h^{2\alpha}$, then

$$\sum_{i=1}^{n} \left(\overline{\mathcal{E}}^*\right)^{\frac{2\alpha+d+s_i^*}{2\alpha}} \asymp \frac{\log n}{\varepsilon^2},$$

which gives (19). In the aligned case, we take $s = s^*$, and the weights are uniform. The bound becomes

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \frac{2^{T(d+s)/(d-s)} \cdot \log n}{n\varepsilon^2} + 2^{-2\alpha T/(d-s)},$$

where taking $2^T = (n\varepsilon^2)^{\frac{d-s}{d+s+2\alpha}}$ yields

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \left(\frac{\log n}{n\varepsilon^2}\right)^{\frac{2\alpha}{2\alpha+d+s^*}}.$$

If the sample term dominates, the same choice as in the original proof gives

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}_{\mathrm{P}}^* \lesssim \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2\alpha+d}}.$$

This leads to the non-private bound in both aligned and personalized cases. ■

A.7 Proof of Theorem 8

Proof of Theorem 8. The privacy of $\tilde{f}$ follows from Proposition 9. We then focus on the excess risk upper bound. Denote the region

$$\Omega_\delta = \{x \in \mathcal{X} \mid |f^*(x) - 1/2| \geq \delta\}.$$

and

$$\overline{\Omega}_f = \{x \in \mathcal{X} \mid \mathbf{1}(f^*(x) > 1/2) = \mathbf{1}(f(x) > 1/2)\}.$$

Bringing Proposition 14, 15, and 16 into the decomposition (58), we have

$$\begin{aligned}\|\tilde{f} - f^*\|_\infty \leq & \|\tilde{f} - \widehat{f}_\omega\|_\infty + \|\widehat{f}_\omega - \overline{f}\|_\infty + \|\overline{f} - f^*\|_\infty \\ \leq & c\left(\sqrt{\frac{t^{2s}2^T \log n}{\varepsilon^2 S_{s,T}}} + \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}} + t^{-\alpha} + 2^{-\alpha T/(d-s)}\right) \end{aligned} \tag{60}$$

for some constant c. Thus, if $f^*(x) > 1/2$, for any $x \in \Omega_{(60)}$,

$$\tilde{f}(x) - 1/2 = \tilde{f}(x) - f^*(x) + f^*(x) - 1/2 \geq -(60) + (60) = 0.$$

The opposite conclusion for $f^*(x) < 1/2$ holds analogously. Thus, we know $\Omega_{(60)} \subset \overline{\Omega}_{\tilde{f}}$. Turning to the risk, we have

$$\begin{aligned}\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} &= \int_{\mathcal{X}} \mathbf{1}\left(y \neq \mathbf{1}\left(f(x) \geq 1/2\right)\right) d\mathrm{P}(x,y) - \int_{\mathcal{X}} \mathbf{1}\left(y \neq \mathbf{1}\left(f^*(x) \geq 1/2\right)\right) d\mathrm{P}(x,y) \\ &= \int_{\overline{\Omega}^c_{\tilde{f}}} 2|f^*(x) - 1/2| d\mathrm{P}_X(x) \\ &\leq \int_{\Omega^c_{(60)}} 2|f^*(x) - 1/2| d\mathrm{P}_X(x) \\ &\leq 2(60) \int_{\Omega^c_{(60)}} d\mathrm{P}_X(x) \leq 2(60)^{1+\beta}.\end{aligned}$$

The first and second equalities are by definition. The third inequality is by $\Omega_{(60)} \subset \overline{\Omega}_{\tilde{f}}$. The fourth inequality is by the definition of $\Omega_{(60)}$. The last inequality is by Assumption 3. Observing the form of $(60)^{1+\beta}$, we first consider when the first term dominates the second term. We take $t \asymp 2^{T/(d-s)}$. Thus we have $t^{-2\alpha} \asymp 2^{-2\alpha T/(d-s)}$ and $t^{2s} \cdot 2^T \asymp 2^{T(d+s)/(d-s)}$. Therefore, we have

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} \lesssim \left(\frac{2^{T(d+s)/(d-s)} \log n}{\varepsilon^2 \sum_{i=1}^n 2^{-q_i(s)T/(d-s)}} + 2^{-2\alpha T/(d-s)} \right)^{(1+\beta)/2},$$

This is the oracle upper bound for a general candidate pair (s, T). For $s = 0$, write $h = 2^{-T/d}$ and balance the pointwise private and approximation terms. Writing $\overline{\mathcal{E}}^* \asymp h^\alpha$ gives

$$\sum_{i=1}^n \left(\overline{\mathcal{E}}^*\right)^{\frac{2\alpha+d+s_i^*}{\alpha}} \asymp \frac{\log n}{\varepsilon^2},$$

The margin argument above then yields the matched personalized rate $\left(\overline{\mathcal{E}}^*\right)^{1+\beta}$. In the aligned case, we take $s = s^*$, and the weights are uniform. The bound becomes

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} \lesssim \left(\frac{2^{T(d+s)/(d-s)} \cdot \log n}{n\varepsilon^2} + 2^{-2\alpha T/(d-s)} \right)^{(1+\beta)/2},$$

where taking $2^T = (n\varepsilon^2)^{\frac{d-s}{d+s+2\alpha}}$ yields

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} \lesssim \left(\frac{\log n}{n\varepsilon^2} \right)^{\frac{\alpha(1+\beta)}{2\alpha+d+s^*}}.$$

If the sample term dominates, the same calculation as in the original proof yields

$$\mathcal{R}_{\mathrm{P}}(\tilde{f}) - \mathcal{R}^*_{\mathrm{P}} \lesssim \left(\frac{\log n}{n} \right)^{\frac{\alpha(1+\beta)}{2\alpha+d}}.$$

This leads to the non-private bound in both aligned and personalized cases. ■

A.8 Proofs of Results in Sections A.5 and A.6

To prove lemmas in Section A.5 and A.6, we first present several technical results.

Lemma 17 *Suppose $\zeta_i, i = 1, \cdots, n$ are independent random variables such that $a_i \le \zeta_i \le b_i$. Then there holds*

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i - \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right| \ge t\right] \le 2e^{-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}}$$

for any $t > 0$.

Proof of Lemma 17. The conclusion follows from Example 2.4 and Proposition 2.5 in Wainwright (2019). ■

Lemma 18 *Suppose $\xi_i, i = 1, \cdots, n$ are independent sub-exponential random variables with parameters (ν_i, β_i) (Wainwright, 2019, Definition 2.9). Then there holds*

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\xi_i\right| \ge t\right] \le 2e^{-\frac{nt^2}{2\nu_*^2}}.$$

for any $0 < t < \nu_*^2/(n\beta_*)$, where $\nu_* = \sqrt{\sum_{i=1}^{n}\nu_i^2}$ and $\beta_* = \max_{i=1,\cdots,n}\beta_i$. Moreover, a standard Laplace random variable is sub-exponential with parameters $(\sqrt{2}, 1)$, and $a\xi_i$ has $(a\nu_i, a\beta_i)$ for any positive a.

Proof of Lemma 18. The conclusion follows from (2.18) in Wainwright (2019). ■

Lemma 19 *Let $\tilde{\mathcal{A}}$ be the collection of all cells $\times_{i=1}^{d}[a_i, b_i]$ in $\mathbb{R}^d$. The VC index of $\tilde{\mathcal{A}}$ equals $2d + 1$. Moreover, for all $0 < \varepsilon < 1$, there exists a universal constant C such that*

$$\mathcal{N}(\mathbf{1}_{\tilde{\mathcal{A}}}, \|\cdot\|_{L_1(Q)}, \varepsilon) \le C(2d+1)(4e)^{2d+1}(1/\varepsilon)^{2d}.$$

Proof of Lemma 19. The first result of the VC index follows from Example 2.6.1 in van der Vaart and Wellner (1996). The second result of covering number follows directly from Theorem 9.2 in Kosorok (2008). ■

Lemma 20 *For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Suppose Assumption 1 holds. Let $\omega_1, \ldots, \omega_n$ be nonnegative weights that are independent of the observations and satisfy $\sum_{i=1}^{n}\omega_i = 1$. Then, for any $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ drawn independently from P and any $A \in \pi$, there holds*

$$\begin{aligned}&\left|\sum_{i=1}^{n}\omega_i \mathbf{1}\{X_i \in A\} - \int_A d\mathrm{P}_X(x')\right| \\ &\quad\le \sqrt{2\bar{c}\, t^{-s}2^{-T}(4d+5)\|\omega\|_2^2 \log n} + \frac{2(4d+5)}{3}\|\omega\|_\infty \log n + \frac{4}{n},\end{aligned} \tag{61}$$

with probability P^n *at least* $1 - 1/n^2$*, where* $\|\omega\|_2^2 := \sum_i \omega_i^2$ *and* $\|\omega\|_\infty := \max_i \omega_i$*. Here we use* A *instead of* $A \times B$ *for notational simplicity.*

Proof of Lemma 20. Condition on the privacy masks $W_1, \ldots, W_n$, so that the weights are deterministic. In the subsequent proof, we define the weighted empirical measure

$$\mathrm{D}_\omega := \sum_{i=1}^n \omega_i \delta_{(X_i, Y_i)}$$

given the samples $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where δ is the Dirac measure. Let $\widetilde{\mathcal{A}}$ be the collection of all cells $\times_{i=1}^d [a_i, b_i]$ in $\mathbb{R}^d$. Applying Lemma 19 with $Q := (\mathrm{D}_{\omega,X} + \mathrm{P}_X)/2$, there exists an η-net $\{\widetilde{A}_k\}_{k=1}^K \subset \widetilde{\mathcal{A}}$ with

$$K \le C(2d+1)(4e)^{2d+1}(1/\eta)^{2d} \tag{62}$$

such that, for any $A \in \pi$, there exists some $k \in \{1, \ldots, K\}$ satisfying

$$\|\mathbf{1}(x \in A) - \mathbf{1}(x \in \widetilde{A}_k)\|_{L_1((\mathrm{D}_{\omega,X} + \mathrm{P}_X)/2)} \le \eta.$$

Since

$$\begin{aligned}
&\|\mathbf{1}(x \in A) - \mathbf{1}(x \in \widetilde{A}_k)\|_{L_1((\mathrm{D}_{\omega,X} + \mathrm{P}_X)/2)} \\
=&\frac{1}{2}\|\mathbf{1}(x \in A) - \mathbf{1}(x \in \widetilde{A}_k)\|_{L_1(\mathrm{D}_{\omega,X})} + \frac{1}{2}\|\mathbf{1}(x \in A) - \mathbf{1}(x \in \widetilde{A}_k)\|_{L_1(\mathrm{P}_X)},
\end{aligned}$$

we get

$$\|\mathbf{1}(x \in A) - \mathbf{1}(x \in \widetilde{A}_k)\|_{L_1(\mathrm{D}_{\omega,X})} \le 2\eta, \qquad \|\mathbf{1}(x \in A) - \mathbf{1}(x \in \widetilde{A}_k)\|_{L_1(\mathrm{P}_X)} \le 2\eta. \tag{63}$$

Consequently, by the definition of the covering number and the triangle inequality, for any $A \in \pi$,

$$\begin{aligned}
&\left|\sum_{i=1}^n \omega_i \mathbf{1}(X_i \in A) - \mathrm{P}_X(A)\right| \\
\le&\left|\sum_{i=1}^n \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) - \mathrm{P}_X(\widetilde{A}_k)\right| + \|\mathbf{1}_A - \mathbf{1}_{\widetilde{A}_k}\|_{L_1(\mathrm{D}_{\omega,X})} + \|\mathbf{1}_A - \mathbf{1}_{\widetilde{A}_k}\|_{L_1(\mathrm{P}_X)} \\
\le&\left|\sum_{i=1}^n \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) - \mathrm{P}_X(\widetilde{A}_k)\right| + 4\eta.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{A \in \pi}\left|\sum_{i=1}^n \omega_i \mathbf{1}(X_i \in A) - \mathrm{P}_X(A)\right| \\
&\quad\le \sup_{1 \le k \le K}\left|\sum_{i=1}^n \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) - \mathrm{P}_X(\widetilde{A}_k)\right| + 4\eta.
\end{aligned} \tag{64}$$

For any fixed $1 \le k \le K$, let $\xi_i := \mathbf{1}(X_i \in \widetilde{A}_k) - \mathrm{P}_X(\widetilde{A}_k)$. Then $\mathbb{E}\xi_i = 0$, $|\xi_i| \le 1$, and $\mathbb{E}\xi_i^2 \le \mathrm{P}_X(\widetilde{A}_k)$. According to Assumption 1, $\mathbb{E}\xi_i^2 \le \bar{c}\, t^{-s} 2^{-T}$. Applying Bernstein's inequality to the independent variables $\omega_i \xi_i$, we obtain

$$\left| \sum_{i=1}^{n} \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) - \mathrm{P}_X(\widetilde{A}_k) \right| \le \sqrt{2\bar{c}\, t^{-s} 2^{-T} \|\omega\|_2^2 \tau} + \frac{2}{3}\|\omega\|_\infty \tau$$

with probability at least $1 - 2e^{-\tau}$. The union bound together with the covering-number estimate (62) implies that

$$\begin{aligned} &\sup_{1 \le k \le K} \left| \sum_{i=1}^{n} \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) - \mathrm{P}_X(\widetilde{A}_k) \right| \\ &\quad \le \sqrt{2\bar{c}\, t^{-s} 2^{-T} \|\omega\|_2^2 (\tau + \log(2K))} + \frac{2}{3}\|\omega\|_\infty (\tau + \log(2K)) \end{aligned}$$

with probability at least $1 - e^{-\tau}$. Let $\tau = 2\log n$ and $\eta = 1/n$. As in the original proof, $\tau + \log(2K) \le (4d+5)\log n$ for all sufficiently large n. Combining this fact with (64) proves (61). ∎

Lemma 21 *Under the conditions of Lemma 20, for any $A \in \pi$,*

$$\begin{aligned} &\left| \sum_{i=1}^{n} \omega_i Y_i \mathbf{1}\{X_i \in A\} - \int_A f^*(x') d\mathrm{P}_X(x') \right| \\ &\quad \le M \sqrt{2\bar{c}\, t^{-s} 2^{-T} (4d+5) \|\omega\|_2^2 \log n} + \frac{2M(4d+5)}{3}\|\omega\|_\infty \log n + \frac{4M}{n}, \end{aligned} \tag{65}$$

with probability P^n *at least* $1 - 1/n^2$.

Proof of Lemma 21. Let $\widetilde{\mathcal{A}}$ be the collection of all cells $\times_{i=1}^{d}[a_i, b_i]$ in $\mathbb{R}^d$. As in the proof of Lemma 20, there exists an η-net $\{\widetilde{A}_k\}_{k=1}^{K} \subset \widetilde{\mathcal{A}}$ with K bounded by (62) such that (63) holds for every $A \in \pi$ and some $k \in \{1, \ldots, K\}$. Consequently, by the definition of the covering number and the triangle inequality,

$$\begin{aligned} &\left| \sum_{i=1}^{n} \omega_i \mathbf{1}(X_i \in A) Y_i - \int_A f^*(x') d\mathrm{P}_X(x') \right| \\ &\le \left| \sum_{i=1}^{n} \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) Y_i - \int_{\widetilde{A}_k} f^*(x') d\mathrm{P}_X(x') \right| \\ &\quad + \max_{1 \le i \le n} |Y_i| \|\mathbf{1}_A - \mathbf{1}_{\widetilde{A}_k}\|_{L_1(\mathrm{D}_{\omega,X})} + M \|\mathbf{1}_A - \mathbf{1}_{\widetilde{A}_k}\|_{L_1(\mathrm{P}_X)} \\ &\le \left| \sum_{i=1}^{n} \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) Y_i - \int_{\widetilde{A}_k} f^*(x') d\mathrm{P}_X(x') \right| + 4M\eta. \end{aligned} \tag{66}$$

For any fixed $1 \le k \le K$, let

$$\widetilde{\xi}_i := Y_i \mathbf{1}(X_i \in \widetilde{A}_k) - \int_{\widetilde{A}_k} f^*(x') d\mathrm{P}_X(x').$$

Then $\mathbb{E}\widetilde{\xi}_i = 0$, $|\widetilde{\xi}_i| \le 2M$, and $\mathbb{E}\widetilde{\xi}_i^2 \le M^2 \mathrm{P}_X(\widetilde{A}_k) \le M^2 \bar{c}\, t^{-s} 2^{-T}$. Applying Bernstein's inequality to $\omega_i \widetilde{\xi}_i$ gives

$$\begin{aligned}\left| \sum_{i=1}^n \omega_i \mathbf{1}(X_i \in \widetilde{A}_k) Y_i - \int_{\widetilde{A}_k} f^*(x') d\mathrm{P}_X(x') \right| \\ \le M \sqrt{2\bar{c}\, t^{-s} 2^{-T} \|\omega\|_2^2 \tau} + \frac{2M}{3} \|\omega\|_\infty \tau\end{aligned}$$

with probability at least $1 - 2e^{-\tau}$. The same covering-number and union-bound calculation as in Lemma 20, with $\tau = 2 \log n$ and $\eta = 1/n$, together with (66), proves (65). ■

Lemma 22 *Let $h = 2^{-T/(d-s)}$, $q_i(s) = \sum_{\ell=s+1}^d W_i^\ell$, $S_{s,T} = \sum_i h^{q_i(s)}$, and $\omega_i = h^{q_i(s)}/S_{s,T}$. For a cell $C_j \in \pi$, define*

$$v_{j,\omega} := \sum_{i=1}^n \omega_i^2 \mathbf{1}\{C_j \in \mathcal{V}_i\}.$$

If $S_{s,T} h^{d-s} \gtrsim \log n$, then, uniformly over the cells of the possibly data-dependent partition,

$$v_{j,\omega} \lesssim \frac{h^{d-s}}{S_{s,T}} \tag{67}$$

with probability at least $1 - 1/n^2$.

Proof of Lemma 22. Condition on the privacy masks. For a fixed mask, the event $C_j \in \mathcal{V}_i$ is the event that the public coordinates of X_i lie in the corresponding coordinate projection of C_j. These projected cells form a VC class of axis-aligned cylinder rectangles, so the weighted empirical- process argument of Lemma 20 applies uniformly, even though the tree partition is data-dependent. A cell constrains $d - s - q_i(s)$ public tree coordinates and therefore

$$\mathrm{P}_X(C_j \in \mathcal{V}_i) \lesssim h^{d-s-q_i(s)}.$$

Consequently,

$$\mathbb{E} v_{j,\omega} \lesssim \frac{1}{S_{s,T}^2} \sum_{i=1}^n h^{2q_i(s)} h^{d-s-q_i(s)} = \frac{h^{d-s}}{S_{s,T}}.$$

For completeness, apply Bernstein's inequality to $\omega_i^2\{\mathbf{1}(C_j \in \mathcal{V}_i) - \mathrm{P}_X(C_j \in \mathcal{V}_i)\}$. Its variance proxy is at most $S_{s,T}^{-2} \mathbb{E} v_{j,\omega}$ and its envelope is at most $S_{s,T}^{-2}$. The same VC covering and union-bound calculation as in Lemma 20 therefore gives, uniformly in j,

$$v_{j,\omega} \lesssim \frac{h^{d-s}}{S_{s,T}} + \sqrt{\frac{h^{d-s} \log n}{S_{s,T}^3}} + \frac{\log n}{S_{s,T}^2}.$$

Under $S_{s,T}h^{d-s} \gtrsim \log n$, the last two terms are absorbed by the first, which proves (67). ■

Lemma 23 *For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. Then for any $A \times B := \times_{i=1}^{d}[a_i, b_i] \in \pi$, there holds*

$$2^{-2}\left(\sqrt{d-s}\cdot 2^{-T/(d-s)} + \sqrt{s}\,t^{-1}\right) \le \operatorname{diam}(A\times B) \le 2\left(\sqrt{d-s}\cdot 2^{-T/(d-s)} + \sqrt{s}\,t^{-1}\right).$$

Proof of Lemma 23. For each $A \times B$, s edges of A have length t^{-1}. The rest $d-s$ edges of B are decided by the max-edge partition rule. When the depth of the tree T is a multiple of dimension $d-s$, each cell of the tree partition is a high-dimensional cube with a side length $2^{-T/(d-s)}$. On the other hand, when the depth of the tree T is not a multiple of dimension $d-s$, we consider the max-edge tree partition with depth $\lfloor T/(d-s)\rfloor$ and $\lceil T/(d-s)\rceil$, whose corresponding side length of the higher dimensional cube is $2^{-\lfloor T/(d-s)\rfloor}$ and $2^{-\lceil T/(d-s)\rceil}$. Note that in the splitting procedure of max-edge partition, the side length of each sub-rectangle decreases monotonically with the increase of T, so the side length of a random tree partition cell is between $2^{-\lceil T/(d-s)\rceil}$ and $2^{-\lfloor T/(d-s)\rfloor}$. This implies that

$$\sqrt{d-s}\cdot 2^{-\lceil T/(d-s)\rceil} \le \operatorname{diam}(B) \le \sqrt{d-s}\cdot 2^{-\lfloor T/(d-s)\rfloor}$$

Since $T/(d-s)-1 \le \lfloor T/(d-s)\rfloor \le \lceil T/(d-s)\rceil \le T/(d-s)+1$, we immediately get $2^{-1}\sqrt{d-s}\cdot 2^{-T/(d-s)} \le \operatorname{diam}(B) \le 2\sqrt{d-s}\cdot 2^{-p/(d-s)}$. Together, we have $\operatorname{diam}(A\times B) \le \operatorname{diam}(A)+\operatorname{diam}(B) \le \sqrt{s}t^{-1} + 2\sqrt{d-s}\cdot 2^{-p/(d-s)}$ and $\operatorname{diam}(A\times B) \ge (\operatorname{diam}(A)+\operatorname{diam}(B))/2 \ge (\sqrt{d-s}\cdot 2^{-T/(d-s)} + \sqrt{s}t^{-1})/4$. ■

Lemma 24 *For $\pi = \pi^{priv} \otimes \pi^{pub}$, let π^{priv} be a histogram partition with t bins and π^{pub} be generated by Algorithm 1 with depth T. For each i, the number of non-zero elements among $V_i^1, \cdots, V_i^{t^s\cdot 2^T}$ is at most*

$$t^s \cdot 2^{T-\lfloor T/(d-s)\rfloor\cdot\left(d-s-\sum_{\ell=s+1}^{d} W_i^{\ell}\right)}.$$

Proof of Lemma 24. The number of non-zero elements is the cardinality of potential grids $|\mathcal{V}_i|$. For $A\times B$, since the first s dimensions are all private, all t^s possibilities for A appear in $\mathcal{V}_i$. For B, consider the process of Algorithm 1 that gradually grows 2^T grids. At any step $k = 1, \cdots, T$, the algorithm splits along the ℓ_k-th dimension. If $W_i^{\ell_k}$ is 1, i.e. the feature is private, the number of potential grids will double. Otherwise, it remains the same. Each feature is split at least $\lfloor T/(d-s)\rfloor$ times, i.e. there are at most $p - \lfloor T/(d-s)\rfloor \cdot \left(d-s-\sum_{\ell=s+1}^{d} W_i^{\ell}\right)$ splits that cause the number of potential grids to double. This quantity is upper bounded by $\sum_{\ell=s+1}^{d} p\cdot W_i^{\ell}/(d-s) + d - s$. Multiplying potential possibilities of A and B leads to $t^s \cdot 2^{\sum_{\ell=s+1}^{d} p\cdot W_i^{\ell}/(d-s)+d-s}$. ■

Proof of Lemma 11. We intend to bound

$$\|\tilde{p}_X - \widehat{p}_{X,\omega}\|_{L_1} = \int_{\mathcal{X}} |\tilde{p}(x) - \widehat{p}_{X,\omega}(x)|\, d\mathrm{P}_X(x) \lesssim \sum_j \left|\sum_{i=1}^{n} \omega_i(\tilde{V}_i^j - V_i^j)\right|, \tag{68}$$

where j is the index of $A \times B$ and $V_i^j = \mathbf{1}_{A\times B}(X_i)$. The last inequality follows from Assumption 1. For each cell, let

$$v_{j,\omega} := \sum_{i=1}^{n} \omega_i^2 \mathbf{1}\{A \times B \in \mathcal{V}_i\}.$$

Applying Lemma 17 with the deterministic coefficients ω_i and taking a union bound over the cells yields

$$\left|\sum_{i=1}^{n} \omega_i(\tilde{V}_i^j - V_i^j)\right| \lesssim \sqrt{\frac{v_{j,\omega}\log n}{\varepsilon^2}} \tag{69}$$

with probability at least $1 - 1/n^2$. Thus, by the Cauchy–Schwarz inequality and Lemma 24,

$$\begin{aligned}
\sum_j \sqrt{v_{j,\omega}} &\le \sqrt{t^s 2^T}\sqrt{\sum_j \sum_{i=1}^{n} \omega_i^2 \mathbf{1}\{A \times B \in \mathcal{V}_i\}} \\
&= \sqrt{t^s 2^T}\sqrt{\sum_{i=1}^{n} \omega_i^2 |\mathcal{V}_i|} \\
&\lesssim \sqrt{t^s 2^T}\sqrt{t^s \sum_{i=1}^{n} \omega_i^2 h^{-q_i(s)}} = \sqrt{\frac{t^{2s}2^T}{S_{s,T}}},
\end{aligned}$$

where the last equality follows from (15). Bringing this bound into (68) and (69) gives the desired conclusion. ∎

Proof of Lemma 12. We intend to bound

$$\|\widehat{p}_{X,\omega} - \overline{p}_X\|_{L_1} \le \sum_j \left|\sum_{i=1}^{n} \omega_i V_i^j - \int_{A\times B} d\mathrm{P}_X(x)\right|.$$

By Lemma 20, uniformly over the cells,

$$\left|\int_{A\times B} d\mathrm{P}_X(x') - \sum_{i=1}^{n} \omega_i V_i^j\right| \lesssim \sqrt{t^{-s}2^{-T}\|\omega\|_2^2 \log n} + \|\omega\|_\infty \log n + \frac{1}{n}.$$

Therefore,

$$\begin{aligned}
\|\widehat{p}_{X,\omega} - \overline{p}_X\|_{L_1} &\lesssim \sqrt{t^s 2^T \|\omega\|_2^2 \log n} + t^s 2^T \|\omega\|_\infty \log n + \frac{t^s 2^T}{n} \\
&\lesssim \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}} + \frac{t^s 2^T \log n}{S_{s,T}},
\end{aligned}$$

where we used $\|\omega\|_2^2 \le S_{s,T}^{-1}$, $\|\omega\|_\infty \le S_{s,T}^{-1}$, and $S_{s,T} \le n$. Under $t^s 2^T \log n \lesssim S_{s,T}$, the second term is bounded by the first, which proves the claim. ∎

Proof of Lemma 13. Let $A \times B(x)$ be the grid of which x belongs to. We intend to bound

$$\|\overline{p}_X - p_X^*\|_{L_1} = \int_{x\in\mathcal{X}} \left| \frac{\int_{A\times B(x)} p_X^*(x')dx'}{\int_{A\times B(x)} dx'} - p_X^*(x) \right| d\mathrm{P}_X(x).$$

For each x, if Assumption 1 holds, the point-wise error can be bounded by

$$\begin{aligned}\frac{\int_{A\times B(x)} p_X^*(x')dx'}{\int_{A\times B(x)} dx'} - p_X^*(x) \le& \frac{\int_{A\times B(x)} |p_X^*(x') - p_X^*(x)|dx'}{\int_{A\times B(x)} dx'} \\ \le& \frac{\int_{A\times B(x)} c_L\|x'-x\|^\alpha dx'}{\int_{A\times B(x)} dx'} \le c_L \mathrm{diam}(A\times B(x))^\alpha\end{aligned}$$

Then we can bound the excess risk using Lemma 23,

$$\|\overline{p}_X - p_X^*\|_{L_1} \lesssim \mathrm{diam}(A\times B(x))^\alpha \lesssim 2^{-T\alpha/(d-s)} + t^{-\alpha}.$$

∎

Proof of Lemma 14. We intend to bound

$$\|\tilde{f} - \widehat{f}_\omega\|_\infty = \max_j \left| \frac{\sum_i \omega_i \tilde{Y}_i \tilde{V}_i^j}{\sum_i \omega_i \tilde{V}_i^j} - \frac{\sum_i \omega_i Y_i V_i^j}{\sum_i \omega_i V_i^j} \right|.$$

Here j is the index of $A\times B$, $V_i^j := \mathbf{1}_{A\times B}(X_i)$, and, throughout this proof, $\omega_i := \omega_i(s,T)$. For each term, we have the same triangular decomposition as in the unweighted proof,

$$\begin{aligned}\left| \frac{\sum_i \omega_i \tilde{Y}_i \tilde{V}_i^j}{\sum_i \omega_i \tilde{V}_i^j} - \frac{\sum_i \omega_i Y_i V_i^j}{\sum_i \omega_i V_i^j} \right| \le& \underbrace{\left| \frac{\sum_{i=1}^n \omega_i (\tilde{Y}_i - Y_i)\tilde{V}_i^j}{\sum_{i=1}^n \omega_i \tilde{V}_i^j} \right|}_{(I)} \\ + \underbrace{\left| \frac{\sum_{i=1}^n \omega_i Y_i \tilde{V}_i^j - \sum_{i=1}^n \omega_i Y_i V_i^j}{\sum_{i=1}^n \omega_i \tilde{V}_i^j} \right|}_{(II)} &+ \underbrace{\left| \frac{\left(\sum_{i=1}^n \omega_i Y_i V_i^j\right)\left(\sum_{i=1}^n \omega_i (V_i^j - \tilde{V}_i^j)\right)}{\left(\sum_{i=1}^n \omega_i \tilde{V}_i^j\right)\left(\sum_{i=1}^n \omega_i V_i^j\right)} \right|}_{(III)}.\end{aligned}$$

We bound the three parts separately. Define the weighted potential-cell variance proxy

$$v_{j,\omega} := \sum_{i=1}^n \omega_i^2 \mathbf{1}\{A\times B \in \mathcal{V}_i\}.$$

(i) For the numerator of (I), the same sub-exponential calculation as in the original proof, with the deterministic coefficient $1/n$ replaced by ω_i, and Lemma 18 yield

$$\left| \sum_{i=1}^n \omega_i (\tilde{Y}_i - Y_i)\tilde{V}_i^j \right| \lesssim \sqrt{\frac{M^2 v_{j,\omega} \log n}{\varepsilon^2}} \tag{70}$$

with probability at least $1 - 1/n^2$. For the denominator, applying the same Hoeffding argument with coefficients ω_i gives

$$\left|\sum_{i=1}^n \omega_i \tilde{V}_i^j - \sum_{i=1}^n \omega_i V_i^j\right| \lesssim \sqrt{\frac{v_{j,\omega} \log n}{\varepsilon^2}} \tag{71}$$

with probability at least $1 - 2/n^2$.

Moreover, by Lemma 20,

$$\begin{aligned}\sum_{i=1}^n \omega_i V_i^j &\geq \int_{A\times B} d\mathrm{P}_X(x) - \left|\int_{A\times B} d\mathrm{P}_X(x) - \sum_{i=1}^n \omega_i V_i^j\right| \\ &\gtrsim \frac{1}{t^s 2^T} - \sqrt{\frac{\log n}{t^s 2^T S_{s,T}}} - \frac{\log n}{S_{s,T}} \gtrsim \frac{1}{t^s 2^T},\end{aligned}$$

where the last inequality follows from $t^s 2^T \log n \lesssim S_{s,T}$. In addition, Lemma 22 gives

$$v_{j,\omega} \lesssim \frac{h^{d-s}}{S_{s,T}}. \tag{72}$$

Combining (71) and (72), the condition $t^{2s} 2^T \log n \lesssim \varepsilon^2 S_{s,T}$ guarantees that

$$\left|\sum_{i=1}^n \omega_i \tilde{V}_i^j\right| \gtrsim \frac{1}{t^s 2^T} - \sqrt{\frac{h^{d-s} \log n}{\varepsilon^2 S_{s,T}}} \gtrsim \frac{1}{t^s 2^T}. \tag{73}$$

This together with (70) yields

$$(I) \lesssim t^s 2^T \sqrt{\frac{M^2 v_{j,\omega} \log n}{\varepsilon^2}}.$$

(*ii*) For (II), we apply the weighted Hoeffding inequality and obtain

$$\left|\sum_{i=1}^n \omega_i Y_i (\tilde{V}_i^j - V_i^j)\right| \lesssim M \sqrt{\frac{v_{j,\omega} \log n}{\varepsilon^2}},$$

with probability at least $1 - 2/n^2$, since $|Y_i| \leq M$. Combining this with (73), we get

$$(II) \lesssim t^s 2^T \sqrt{\frac{M^2 v_{j,\omega} \log n}{\varepsilon^2}}.$$

(*iii*) For (III), since $V_i^j \in \{0, 1\}$ and $|Y_i| \leq M$, we have $|\sum_i \omega_i Y_i V_i^j| \leq M \sum_i \omega_i V_i^j$. Thus,

$$\begin{aligned}(III) &\leq M \left|\frac{\sum_{i=1}^n \omega_i V_i^j - \sum_{i=1}^n \omega_i \tilde{V}_i^j}{\sum_{i=1}^n \omega_i \tilde{V}_i^j}\right| \\ &\lesssim t^s 2^T \sqrt{\frac{M^2 v_{j,\omega} \log n}{\varepsilon^2}},\end{aligned}$$

where the last inequality follows from (71) and (73).

Combining the bounds for (I)–(III), using (72), and recalling that $2^T = h^{-(d-s)}$, we obtain, uniformly over the cells,

$$\begin{aligned}\|\tilde{f} - \widehat{f}_\omega\|_\infty &\lesssim t^s 2^T \sqrt{\frac{h^{d-s}\log n}{\varepsilon^2 S_{s,T}}} \\ &= \sqrt{\frac{t^{2s}2^T \log n}{\varepsilon^2 S_{s,T}}}.\end{aligned}$$

This is the desired conclusion. ∎

Proof of Lemma 15. We intend to bound

$$\begin{aligned}\|\widehat{f}_\omega - \overline{f}\|_\infty \leq & \max_j \left| \frac{\sum_{i=1}^n \omega_i Y_i V_i^j}{\sum_{i=1}^n \omega_i V_i^j} - \frac{\int_{A\times B} f^*(x')d\mathrm{P}_X(x')}{\int_{A\times B} d\mathrm{P}_X(x)} \right| \\ \leq & \underbrace{\left| \frac{\left(\sum_{i=1}^n \omega_i Y_i V_i^j - \int_{A\times B} f^*(x')d\mathrm{P}_X(x')\right)\int_{A\times B} d\mathrm{P}_X(x')}{\left(\sum_{i=1}^n \omega_i V_i^j\right)\int_{A\times B} d\mathrm{P}_X(x)} \right|}_{(I)} \\ & + \underbrace{\left| \frac{\int_{A\times B} f^*(x')d\mathrm{P}_X(x')\left(\int_{A\times B} d\mathrm{P}_X(x') - \sum_{i=1}^n \omega_i V_i^j\right)}{\left(\sum_{i=1}^n \omega_i V_i^j\right)\int_{A\times B} d\mathrm{P}_X(x)} \right|}_{(II)}.\end{aligned}$$

We bound the two terms separately. For (I), Lemma 20 and the condition $t^s 2^T \log n \lesssim S_{s,T}$ yield

$$\left(\sum_{i=1}^n \omega_i V_i^j\right) \int_{A\times B} d\mathrm{P}_X(x) \gtrsim t^{-2s} 2^{-2T} \tag{74}$$

with probability at least $1 - 1/n^2$. For the numerator, Lemma 21 gives

$$\begin{aligned}&\left| \sum_{i=1}^n \omega_i Y_i V_i^j - \int_{A\times B} f^*(x')d\mathrm{P}_X(x') \right| \left| \int_{A\times B} d\mathrm{P}_X(x') \right| \\ &\quad \lesssim \left(\sqrt{\frac{t^{-s}2^{-T}\log n}{S_{s,T}}} + \frac{\log n}{S_{s,T}} \right) t^{-s}2^{-T}.\end{aligned}$$

Together with (74), this yields

$$(I) \lesssim \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}} + \frac{t^s 2^T \log n}{S_{s,T}} \lesssim \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}}.$$

Analogously, by Lemma 20,

$$\left|\int_{A\times B} d\mathrm{P}_X(x') - \sum_{i=1}^{n} \omega_i V_i^j\right| \left|\int_{A\times B} f^*(x') d\mathrm{P}_X(x')\right| \lesssim M\left(\sqrt{\frac{t^{-s}2^{-T}\log n}{S_{s,T}}} + \frac{\log n}{S_{s,T}}\right) t^{-s}2^{-T}.$$

This together with (74) yields

$$(II) \lesssim \sqrt{\frac{t^s 2^T \log n}{S_{s,T}}}.$$

The bounds for (I) and (II) together give the desired conclusion. ■

Proof of Lemma 16. Let $A \times B(x)$ be the grid of which x belongs to. We intend to bound

$$\|\overline{f} - f^*\|_\infty = \max_x \left|\frac{\int_{A\times B(x)} f(x') d\mathrm{P}_X(x')}{\int_{A\times B(x)} d\mathrm{P}_X(x')} - f(x)\right|.$$

For each x, if Assumption 2 holds, the point-wise error can be bounded by

$$\begin{aligned}\frac{\int_{A\times B(x)} f(x') d\mathrm{P}_X(x')}{\int_{A\times B(x)} d\mathrm{P}_X(x')} - f(x) \leq & \frac{\int_{A\times B(x)} |f(x') - f(x)| d\mathrm{P}_X(x')}{\int_{A\times B(x)} d\mathrm{P}_X(x')} \\ \leq & \frac{\int_{A\times B(x)} c_L \|x' - x\|^\alpha d\mathrm{P}_X(x')}{\int_{A\times B(x)} d\mathrm{P}_X(x')} \leq c_L \mathrm{diam}(A \times B(x))^\alpha\end{aligned}$$

Then we can bound the excess risk using Lemma 23,

$$\|\overline{f} - f^*\|_\infty \lesssim \mathrm{diam}(A \times B(x))^\alpha \lesssim 2^{-T\alpha/(d-s)} + t^{-\alpha}.$$

■

Appendix B. Data Set Details

For all data sets, each feature is min-max scaled to the range $[0, 1]$ individually.

ABA: The *Abalone* data set originally comes from biological research (Nash et al., 1994) and now it is accessible on UCI Machine Learning Repository (Dua and Graff, 2017). ABA contains 4177 observations of one target variable and 8 attributes related to the physical measurements of abalone.

AQU: The *QSAR aquatic toxicity* data set contains 546 chemicals, each represented by 8 molecular descriptors, and a response measuring acute aquatic toxicity toward *Daphnia magna*.

BUI: The *Residential Building* data set on UCI Machine Learning Repository includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran. It contains 372 instances with 103 input attributes and 2 output attributes.

DAK: The *Istanbul Stock Exchange* data set includes returns of the Istanbul stock exchange with seven other international indexes. It has 536 instances with 10 features.

FIS: The *QSAR fish toxicity* data set on UCI Machine Learning Repository was used to develop quantitative regression QSAR models to predict acute aquatic toxicity toward the fish *Pimephales promelas* (fathead minnow). It contains 908 instances with 6 input features and 1 output variable.

MUS: The *The Geographical Original of Music* data set was built from a personal collection of 1059 tracks covering 33 countries/areas. The program MARSYAS (Zhou et al., 2014) was used to extract audio features from the wave files. We used the default MARSYAS settings in single vector format (68 features) to estimate the performance with basic timbal information covering the entire length of each track.

POR: The *Stock Portfolio Performance* data set of performances of weighted scoring stock portfolios are obtained with mixture design from the US stock market historical database (Liu and Yeh, 2017). It has 315 samples with 12 features.

PYR: The *Pyrimidines* data set is a subset of the Qualitative Structure Activity Relationships data set on UCI data sets. This sub-data set has 74 instances of dimension 27.

RED: This data set contains the information on red wine of the *Wine Quality* data set (Cortez et al., 2009) on UCI Machine Learning Repository. There are 11 input variables to predict the output variable wine quality. The data set contains 1599 instances.

WHI: This data set also originates from the *Wine Quality* data set (Cortez et al., 2009) on the UCI Machine Learning Repository. It contains 4898 white-wine instances with 11 input features and wine quality as the output variable.

References

Maryam Aliakbarpour, Syomantak Chaudhuri, Thomas Courtade, Alireza Fallah, and Michael Jordan. Enhancing feature-specific data protection via bayesian coordinate differential privacy. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *PMLR*, pages 4069–4077, 2025.

Chiara Amorino and Arnaud Gloter. Minimax rate for multivariate data under componentwise local differential privacy constraints. *The Annals of Statistics*, 53(3), 2025. doi: 10.1214/25-AOS2497.

Apple. Differential privacy technical overview. Technical report, Apple Inc., 2017. URL `https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf`.

Jean-Yves Audibert and Alexandre B. Tsybakov. Fast learning rates for plug-in classifiers. *Annals of statistics*, 35(2):608–633, 2007.

Ashwinkumar Badanidiyuru Varadaraja, Badih Ghazi, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Avinash V. Varadarajan, and Chiyuan Zhang. Optimal

unbiased randomizers for regression with label differential privacy. *Advances in Neural Information Processing Systems*, 36, 2024.

Thomas Berrett and Cristina Butucea. Classification under local differential privacy. *arXiv preprint arXiv:1912.04629*, 2019.

Thomas B. Berrett, László Györfi, and Harro Walk. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15:2430–2453, 2021.

Leo Breiman. Classification and regression trees. *The Wadsworth & Brooks/Cole*, 1984.

Robert Istvan Busa-Fekete, Umar Syed, Sergei Vassilvitskii, et al. Population level privacy leakage in binary classification wtih label noise. In *NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021.

Cristina Butucea, Amandine Dubois, Martin Kroll, and Adrien Saumard. Local differential privacy: Elbow effect in optimal density estimation and adaptation over besov ellipsoids. *Bernoulli*, 26(3):1727–1764, 2020.

T. Tony Cai and Hongming Pu. Transfer learning for nonparametric regression: Non-asymptotic minimax analysis and adaptive procedure. *arXiv preprint arXiv:2401.12272*, 2024.

Yuchao Cai, Yuheng Ma, Yiwei Dong, and Hanfang Yang. Extrapolated random tree for regression. In *Proceedings of the 40th International Conference on Machine Learning*, pages 3442–3468. PMLR, 2023.

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for nearest neighbor classification. *Advances in Neural Information Processing Systems*, 27, 2014.

Kamalika Chaudhuri and Daniel Hsu. Sample complexity bounds for differentially private learning. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 155–186. JMLR Workshop and Conference Proceedings, 2011.

Lynn Chua, Qiliang Cui, Badih Ghazi, Charlie Harrison, Pritish Kamath, Walid Krichene, Ravi Kumar, Pasin Manurangsi, Krishna Giri Narra, Amer Sinha, et al. Training differentially private ad prediction models with semi-sensitive features. *arXiv preprint arXiv:2401.15246*, 2024.

Paulo Cortez, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision support systems*, 47(4):547–553, 2009.

Rachel Cummings and Deven Desai. The role of differential privacy in gdpr compliance. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, page 20, 2018.

Teddy Cunningham, Konstantin Klemmer, Hongkai Wen, and Hakan Ferhatosmanoglu. Geopointgan: Synthetic spatial data with local label differential privacy. *arXiv preprint arXiv:2205.08886*, 2022.

Mihaela Curmei, Walid Krichene, Li Zhang, and Mukund Sundararajan. Private matrix factorization with public item features. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 805–812, 2023.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

John Duchi, Michael Jordan, and Martin Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521): 182–201, 2018.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council, 2016. URL `https://data.europa.eu/eli/reg/2016/679/oj`.

Miguel Fuentes, Brett C. Mullins, Ryan McKenna, Gerome Miklau, and Daniel Sheldon. Joint selection: Adaptively incorporating public information for private synthetic data. In *International Conference on Artificial Intelligence and Statistics*, pages 2404–2412. PMLR, 2024.

Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, and Chiyuan Zhang. Deep learning with label differential privacy. *Advances in neural information processing systems*, 34:27131–27145, 2021.

Badih Ghazi, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Avinash Varadarajan, and Chiyuan Zhang. Regression with label differential privacy. In *The Eleventh International Conference on Learning Representations*, 2022.

Badih Ghazi, Ravi Kumar, Pasin Manurangsi, and Thomas Steinke. Algorithms with more granular differential privacy guarantees. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *LIPIcs*, pages 54:1–54:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi: 10.4230/LIPIcs.ITCS.2023.54.

Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Raghu Meka, and Chiyuan Zhang. On convex optimization with semi-sensitive features. In *Proceedings of the Thirty-Seventh Conference on Learning Theory*, volume 247 of *PMLR*, pages 1916–1938, 2024.

Xin Gu, Gautam Kamath, and Zhiwei Steven Wu. Choosing public datasets for private machine learning via gradient subspace distance. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 879–900, 2025. doi: 10.1109/SaTML64287.2025.00054.

László Györfi and Martin Kroll. On rate optimal private regression under local differential privacy. *Statistica Sinica*, 35:613–627, 2025. doi: 10.5705/ss.202022.0186.

László Györfi, Aryeh Kontorovich, and Roi Weiss. Tree density estimation. *IEEE Transactions on Information Theory*, 69(2):1168–1176, 2022.

Shuli Jiang, Walid Krichene, and Nicolas Mayoraz. Private learning with public feature conditioning. *arXiv preprint arXiv:2606.18773*, 2026.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *Advances in neural information processing systems*, 27, 2014.

Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*, pages 2436–2444. PMLR, 2016.

Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York, 2008.

Walid Krichene, Nicolas Mayoraz, Steffen Rendle, Shuang Song, Abhradeep Thakurta, and Li Zhang. Private learning with public features. *arXiv preprint arXiv:2310.15454*, 2023.

Martin Kroll. On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics*, 15:1783–1813, 2021.

Xiaoguang Li, Haonan Yan, Zelei Cheng, Wenhai Sun, and Hui Li. Protecting regression models with personalized local differential privacy. *IEEE Transactions on Dependable and Secure Computing*, 20(2):960–974, 2022.

Terrance Liu, Giuseppe Vietri, Thomas Steinke, Jonathan Ullman, and Steven Wu. Leveraging public data for practical private query release. In *International Conference on Machine Learning*, pages 6968–6977. PMLR, 2021a.

Terrance Liu, Giuseppe Vietri, and Steven Z. Wu. Iterative methods for private synthetic data: Unifying framework and new methods. *Advances in Neural Information Processing Systems*, 34:690–702, 2021b.

Yi-Cheng Liu and I-Cheng Yeh. Using mixture design and neural networks to build stock selection decision support systems. *Neural Computing and Applications*, 28:521–535, 2017.

Qilong Lu, Songxi Chen, and Yumou Qiu. Versatile differentially private learning for general loss functions. *The Annals of Statistics*, 54(2), 2026. doi: 10.1214/25-AOS2583.

Yuheng Ma and Hanfang Yang. Optimal locally private nonparametric classification with public data. *Journal of Machine Learning Research*, 25(167):1–62, 2024.

Yuheng Ma, Han Zhang, Yuchao Cai, and Hanfang Yang. Decision tree for locally private estimation with public data. In *Advances in Neural Information Processing Systems*, volume 36, pages 43676–43705, 2023. doi: 10.52202/075280-1894.

Yuheng Ma, Ke Jia, and Hanfang Yang. Locally private estimation with public features. In *International Conference on Artificial Intelligence and Statistics*, pages 55–63. PMLR, 2025.

Yuheng Ma, Feiyu Jiang, Zifeng Zhao, Hanfang Yang, and Yi Yu. Locally private nonparametric contextual multi-armed bandits with transfer learning. *Journal of the American Statistical Association*, pages 1–104, 2026. doi: 10.1080/01621459.2026.2721731.

Saeed Mahloujifar, Chuan Guo, G. Edward Suh, and Kamalika Chaudhuri. Machine learning with privacy for protected attributes. In *2025 IEEE Symposium on Security and Privacy*, pages 2640–2657, 2025.

Mani Malek Esmaeili, Ilya Mironov, Karthik Prasad, Igor Shilov, and Florian Tramer. Antipodes of label differential privacy: Pate and alibi. *Advances in Neural Information Processing Systems*, 34:6934–6945, 2021.

Warwick J. Nash, Tracy L. Sellers, Simon R. Talbot, Andrew J. Cawthorn, and Wes B. Ford. The population biology of abalone (haliotis species) in tasmania. i. blacklip abalone (h. rubra) from the north coast and islands of bass strait. *Sea Fisheries Division, Technical Report*, 48:p411, 1994.

Milad Nasr, Saeed Mahloujifar, Xinyu Tang, Prateek Mittal, and Amir Houmansadr. Effectively using public data in privacy preserving machine learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202. PMLR, 2023.

Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark patterns after the gdpr: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In *International Conference on Learning Representations*, 2017.

Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Ulfar Erlingsson. Scalable private learning with pate. In *International Conference on Learning Representations*, 2018.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Richard J. Samworth. Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, pages 2733–2763, 2012.

Mathieu Sart. Density estimation under local differential privacy and hellinger loss. *Bernoulli*, 29(3):2318–2341, 2023.

Zeyu Shen, Anilesh Krishnaswamy, Janardhan Kulkarni, and Kamesh Munagala. Classification with partially private features. *arXiv preprint arXiv:2312.07583*, 2023.

Haina Song, Tao Luo, Xun Wang, and Jianfeng Li. Multiple sensitive values-oriented personalized privacy preservation based on randomized response. *IEEE transactions on information forensics and security*, 15:2209–2224, 2019.

Chongjing Sun, Yan Fu, Junlin Zhou, Hui Gao, et al. Personalized privacy-preserving frequent itemset mining using randomized response. *The Scientific World Journal*, 2014, 2014.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. doi: 10.1007/b13794. URL `http://hfhhc391f4815d8064db7sqbf60x0xf60c6kpo.faxb.libproxy.ruc.edu.cn/10.1007/b13794`. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

Martin J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Di Wang and Jinhui Xu. On sparse linear regression in the local differential privacy model. In *International Conference on Machine Learning*, pages 6628–6637. PMLR, 2019.

Shaowei Wang, Liusheng Huang, Miaomiao Tian, Wei Yang, Hongli Xu, and Hansong Guo. Personalized privacy-preserving data aggregation for histogram estimation. In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2015.

Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 729–745, 2017.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer, 1992.

Xingxing Xiong, Shubo Liu, Dan Li, Zhaohui Cai, and Xiaoguang Niu. A comprehensive survey on local differential privacy. *Security and Communication Networks*, 2020:1–29, 2020.

Shirong Xu, Chendi Wang, Will Wei Sun, and Guang Cheng. Binary classification under local label differential privacy using randomized response mechanisms. *Transactions on Machine Learning Research*, 2023.

Ge Yang, Shaowei Wang, and Haijie Wang. Federated learning with personalized local differential privacy. In *2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS)*, pages 484–489. IEEE, 2021.

Da Yu, Huishuai Zhang, Wei Chen, Jian Yin, and Tie-Yan Liu. Large scale private learning via low-rank reparametrization. In *International Conference on Machine Learning*, pages 12208–12218. PMLR, 2021.

Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A. Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, Sergey Yekhanin, and Huishuai Zhang. Differentially private fine-tuning of language models. In *The Tenth International Conference on Learning Representations*, 2022.

Dongyan Zhang, Lili Zhang, Zhiyong Zhang, and Zhongya Zhang. Adaptive personalized randomized response method based on local differential privacy. *International Journal of Information Security and Privacy (IJISP)*, 18(1):1–19, 2024.

Guopeng Zhang, Xuebin Chen, Yuanli Jia, and Ran Zhai. Support personalized weighted local differential privacy skyline query. *Security and Communication Networks*, 2022, 2022a.

Wanrong Zhang, Olga Ohrimenko, and Rachel Cummings. Attribute privacy: Framework and mechanisms. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 757–766. ACM, 2022b. doi: 10.1145/3531146.3533139.

Puning Zhao, Huiwen Wu, Jiafei Wu, and Zhe Liu. On theoretical limits of learning with label differential privacy, 2024.

Puning Zhao, Jiafei Wu, Zhe Liu, Li Shen, Zhikun Zhang, Rongfei Fan, Le Sun, and Qingming Li. Enhancing learning with label differential privacy by vector approximation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Fang Zhou, Claire Elizabeth Q, and Ross D. King. Predicting the geographical origin of music. In *2014 IEEE International Conference on Data Mining*, pages 1115–1120. IEEE, 2014.

Yi Zhou, Yanhao Wang, Long Teng, Qiang Huang, and Cen Chen. Approximate kernel density estimation under metric-based local differential privacy. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.