A. Interactive Model

A.1. Privacy Guarantee

We restate a version of the privacy theorem by (Gupta et al., 2011) in the context of this paper.

**Theorem 9** (Theorem 4.1 from Gupta et al. (2011)).

Let $T$ be the total number of queries and $B$ be the number of updates allowed in Algorithm 1, let $\epsilon_0 = \frac{\epsilon}{200\sqrt{BS\log(4/\delta)}}$ and $\sigma = \frac{\epsilon}{\epsilon_0}\log(2T/\beta)$, where $S$ is the maximum change in the output of a query (using $w^*$) when any one entry in the underlying data set is arbitrarily modified. Let $(\epsilon, \delta, \beta)$ be the privacy parameters and $\beta$ be the failure probability in Algorithm 1. Under this setting, Algorithm 1 is $(\epsilon, \delta, \beta)$-differentially private.

We now provide privacy proof of our PINP algorithm (Algorithm 1).

**Proof of Theorem 2.** The proof proceeds in two stages. In the first stage, we show that prediction function is relatively insensitive to change in the dataset. Specifically, we bound $|\langle w^*_G, \phi(z) \rangle - \langle w^*_G', \phi(z) \rangle|$, where $z \in \mathcal{X}$ and $G, G'$ are two datasets differing in exactly one data point. Here $w^*_G$ and $w^*_G'$ represent optimal solution to regularized ERM (2) when the underlying datasets are $G$ and $G'$ respectively. In the second stage, we invoke Theorem 9 with sensitive bound $|\langle w^*_G, \phi(z) \rangle - \langle w^*_G', \phi(z) \rangle|$ to complete the proof.

W.l.o.g. we can assume that the datasets $G$ and $G'$ differ in the $n$-th data point, i.e., $(x_n, y_n) \in G$ and $(x'_n, y'_n) \in G'$. Now, using optimality of $w^*_G$ and $w^*_G'$ for (2) (with dataset $G$ and $G'$ respectively) and strong convexity of the ERM (2):

$$1 \over n \sum_{i=1}^{n-1} \ell(\langle w^*_G, \phi(x_i) \rangle; y_i) + 1 \over n \ell(\langle w^*_G, \phi(x_n) \rangle; y_n) + \lambda \over 2 \|w^*_G\|^2_2 \geq \sum_{i=1}^{n-1} \ell(\langle w^*_G, \phi(x_i) \rangle; y_i) + 1 \over n \ell(\langle w^*_G, \phi(x_n) \rangle; y_n) + \lambda \over 2 \|w^*_G\|^2_2 + \lambda \over 2 \|w^*_G - w^*_G\|^2_2.$$  

Hence,

$$1 \over n \sum_{i=1}^{n-1} \ell(\langle w^*_G, \phi(x_i) \rangle; y_i) + 1 \over n \ell(\langle w^*_G, \phi(x'_n) \rangle; y'_n) + \lambda \over 2 \|w^*_G\|^2_2 \geq \sum_{i=1}^{n-1} \ell(\langle w^*_G, \phi(x_i) \rangle; y_i) + 1 \over n \ell(\langle w^*_G, \phi(x'_n) \rangle; y'_n) + \lambda \over 2 \|w^*_G\|^2_2 + \lambda \over 2 \|w^*_G - w^*_G\|^2_2.$$  

Adding the above two equations and using Lipschitz continuity of $\ell$:

$$\|w^*_G - w^*_G\|^2_2 \leq \frac{2LR_\theta}{\lambda n}. \tag{3}$$

Finally, using Cauchy-Schwarz inequality and the above inequality, we have,

$$|\langle w^*_G, \phi(z) \rangle - \langle w^*_G', \phi(z) \rangle| \leq \frac{2LR^2_\theta}{\lambda n}.$$  

With this bound in hand, we invoke Theorem 9 (Theorem 4.1 by (Gupta et al., 2011)) to complete the proof.

A.2. Utility Guarantee

In the following we restate a version of Theorem 5.2 from (Gupta et al., 2011) in the context of this paper.

**Theorem 10** (Theorem 5.2 from Gupta et al. (2011)).

Let $T$ be the total number of queries and $B$ be the number of updates allowed in Algorithm 1, let $\epsilon_0 = \frac{\epsilon}{200\sqrt{BS\log(4/\delta)}}$ and $\sigma = \frac{\epsilon}{\epsilon_0}\log(2T/\beta)$, where $S$ is the maximum change in the output of a query (using $w^*$) when any one entry in the underlying data set is arbitrarily modified. Let $(\epsilon, \delta, \beta)$ be the privacy parameters and $\beta$ be the failure probability in Algorithm 1. As long as the variable counter in Algorithm 1 is less than $B$, for each query $z_t$, with probability at least $1 - \beta$, the following is true:

$$|\hat{v}_t - \langle \phi(z_t), w^* \rangle| = O \left( \frac{S\sqrt{B\log(1/\delta)\log(T/\beta)}}{\epsilon} \right)$$

B. Test Data Dependent Learner

(Semi-interactive model)

B.1. Privacy Guarantee of Test Data Dependent Learner

**Proof of Theorem 4.** From (3), we know that for any two training data sets $G$ and $G'$ differing in exactly one entry, the following is true:

$$\|w^*_G - w^*_G\|^2_2 \leq \frac{2LR_\theta}{\lambda n}.$$  

Therefore by Cauchy-Schwarz inequality, for any $z \in \mathcal{X}$ we have

$$|\langle w^*_G, \phi(z) \rangle - \langle w^*_G', \phi(z) \rangle| \leq \frac{2LR^2_\theta}{\lambda n}.$$  

Theorem now follows by using the above given bound with the following composition theorem.
Theorem 11 (Composition Theorem from (Dwork et al., 2010)). Let $\epsilon', \delta' > 0$. The class of $\epsilon'$-differentially private mechanisms satisfy $(\epsilon', \delta')$-differential privacy under $k$-fold adaptive composition for:

$$\epsilon' = \sqrt{2k\log(1/\delta')} + \varepsilon(\epsilon' - 1).$$

B.2. Utility Guarantee of Test Data Dependent Learner

Proof of Theorem 5. Let,

$$J(w) = \frac{1}{T} \sum_{t=1}^{T} \langle (w, \phi(z_t)) - \langle w^*, \phi(z_t) \rangle - b_t \rangle^2.$$ 

Since $\hat{w} = \arg \min_{w \in \mathcal{C}} J(w)$ and by assumption $w^* \in \mathcal{C}$, the following holds:

$$\sum_{t=1}^{T} \langle \hat{w}, \phi(z_t) \rangle - \langle w^*, \phi(z_t) \rangle \leq \sum_{t=1}^{T} \langle \hat{w} - w^*, \phi(z_t) \rangle b_t.$$ 

Let $b = \langle b_1, \ldots, b_T \rangle$. Using Cauchy-Schwarz inequality and the fact that $\|v\|_2 \leq \sqrt{T}\|v\|_2$, we get:

$$\sum_{t=1}^{T} |\langle \hat{w}, \phi(z_t) \rangle - \langle w^*, \phi(z_t) \rangle | \leq 2\sqrt{T}|b|_2.$$ 

Therefore, using (4) and (5), we get (w.p. $\geq 1 - \beta$):

$$\mathbb{E}_{z \sim \mathcal{P}} [g(\hat{w}; z)] \leq C_1 \sqrt{T} \frac{LR^2}{\lambda n} \sqrt{\frac{\log(\frac{1}{\beta})}{\delta}} + C_2 \frac{\|C\|_2 R^2 \log(\frac{1}{\beta})}{\sqrt{T}},$$

where $C_1, C_2 > 0$ are global constants.

Theorem now follows by setting $T$ as mentioned in the theorem along with using Lipschitz property of $\ell$.

Theorem 12 (Theorem 1 from (Shalev-Shwartz et al., 2009)). Let $\mathcal{C} = \{w : \|w\|_2 \leq B\}$ be a convex set, let $\phi : \mathcal{X} \to \mathbb{R}^{d_\phi}$ be a feature map with the image of $\phi$ has $L_2$-norm of at most $R_\phi$, and let $f : \mathcal{X} \to \mathbb{R}$ be a $L_1$-Lipschitz continuous convex cost function in its first parameter. Then for any $\mathcal{P}$ over the domain $\mathcal{X}$, and for $Z = \{z_1, \ldots, z_T\}$ drawn i.i.d. from $\mathcal{P}$, the following is true with probability at least $1 - \beta$.

$$\sup_{w \in \mathcal{C}} \left| \mathbb{E}_{z \sim \mathcal{P}} [f((w, \phi(z)) ; z)] - \frac{1}{T} \sum_{t=1}^{T} f((w, \phi(z_t)) ; z_t) \right| \leq O \left( \frac{\sqrt{B^2 (R_\phi L)^2 (1/\beta) \log(1/\beta)}}{T} \right).$$

B.3. Generalization Bound for Test Data Dependent Learner

Theorem 13 (Error Bound over Test Distribution). Let $\mathcal{P}$ be a fixed test distribution and let $Z = \{z_1, \ldots, z_T\}$ be sampled uniformly from $\mathcal{P}$. If $T = O \left( \frac{\|C\|_2 \sqrt{n\lambda}}{LR_\phi \log(1/\delta)} \right)$ and $w^* \in \mathcal{C}$ in Algorithm 2, then w.p. $1 - \beta$,

$$\mathbb{E}_{z \sim \mathcal{P}} [\ell((\hat{w}, \phi(q)); y_{q_t})] = \mathbb{E}_{z \sim \mathcal{P}} [\ell((w^*, \phi(q)); y_{q_t})] + O \left( \frac{(LR_\phi)^{3/2} \sqrt{\|C\|_2 \log(1/\delta) \log(T/\delta)}}{\sqrt{n\lambda}} \right).$$

C. Test Data Independent Learner (Non-interactive model)

Proof sketch of Theorem 7. For a given dataset $\mathcal{G}$, let $f(\mathcal{G}) = \left( \frac{\epsilon_m}{2LR_\phi^2} \left\langle \phi(z), w_r - w^*(\mathcal{G}) \right\rangle \right)$. Using the fact that $|w^*(\mathcal{G}) - w^*(\mathcal{G}')|_2 \leq \frac{2LR_\phi}{\alpha}$ for any two datasets $\mathcal{G}$ and $\mathcal{G}'$ differing in exactly one entry (see Theorem 2 from Section 5), it directly follows that $|f(\mathcal{G}) - f(\mathcal{G}')| \leq \frac{\epsilon}{4}$. Hence, it follows that each iteration of Line 3 in Algorithm 3 is $\epsilon_0/2$-differentially private. Now from the analysis of Theorem 2 (from Section 5), it follows (5) that Algorithm 3 is $(\epsilon, \delta)$-differentially private.
Proof of Theorem 8. Intuition: The proof of this theorem goes via the following key insight: if we can make almost every round of Algorithm 3 an update round, then the iterates $w_t$ will become representative of $w^*$ as time $t$ progresses. This can be formalized via a simple potential argument. (See (Gupta et al., 2011) for the exact formalization.) The way we ensure that each iteration is an update round is by finding a $z$ (via exponential mechanism) such that it can distinguish between $w_t$ and $w^*$ with high probability, (i.e., the value of $\langle \phi(z), w_t - w^* \rangle$ is greater than $\frac{\gamma}{2}$).

Main Proof: We apply exponential mechanism to a finite set $S = \{ x : x$ is the center of the $\nu$-net$\}$, where $\nu$ is as given in the Theorem. That is, we divide the entire space into (overlapping) $L_2$ balls of radius $\nu$ and $S$ is the collection of centers of all such balls. Also, it is known that $|S| = \frac{1}{\nu^d}$.

Now, using the exponential distribution specified in Step 3 of Algorithm 3, we get:

$$\Pr [ z \text{ s.t. } |\langle \phi(z), w_t - w^* \rangle| \leq OPT_\nu - \gamma \leq |S| e^{-\lambda \gamma},$$

where $OPT_\nu = \max_{z \in S} |\langle \phi(z), w_t - w^* \rangle|$ and $\lambda = \frac{\rho}{8L_2^2}$. Hence, w.p. at least $1 - \beta$, a $z$ is sampled s.t.,

$$|\langle \phi(z), w_t - w^* \rangle| \geq OPT_\nu - \frac{\ln(|S|/\beta)}{\lambda}.$$

Now, let $OPT^*$ be the maximum value of $|\langle \phi(z), w_t - w^* \rangle|$ over the input space $X$, i.e.,

$$OPT^* = \max_{x \in X} |\langle \phi(z), w_t - w^* \rangle|.$$

Also, $|z^* - z_\nu|_2 \leq 2\nu$ where $z_\nu = \arg \max_{z \in S} |\langle \phi(z), w_t - w^* \rangle|$ is the optimal over $S$. Hence, using Lipschitz continuity of the mapping $\phi$, we obtain a sample $z$ w.p. at least $1 - \beta$ s.t.:

$$|\langle \phi(z), w_t - w^* \rangle| \geq OPT^* - \frac{\ln(|S|/\beta)}{\lambda} - 2\nu L_2 R_2 L.$$

Hence, selecting $\nu = \frac{d R_2}{\rho n L_2}$, we get

$$|\langle \phi(z), w_t - w^* \rangle| \geq OPT^* - \Omega \left( \frac{d R_2^2 \ln(L_2) \ln(1/\beta)}{\rho n L_2} \right).$$

Now, using Theorem 7.3 of (Gupta et al., 2011) (see Theorem 14), we get,

$$|\langle \phi(z), w_t - w^* \rangle| \leq \sigma = \max \left( \frac{\|w^*\| R_2 d L^2 R_2^2 \ln(L_2) \ln \frac{1}{\beta} \|w^*\|^2}{2\sigma^2 \lambda^2 n^2 \epsilon}, \right),$$

for all $z \in \mathbb{R}^d$ and $\|z\|_2 \leq 1$. Hence, minimizing over $\sigma$, we get

$$|\langle \phi(z), w_t - w^* \rangle| = O \left( \frac{\|w^*\| R_2 d^{1/3} L^{2/3} \ln(L_2) \log^2 \frac{1}{\beta} \ln(1/\beta)}{\lambda n^{2/3} \sqrt{\epsilon}} \right),$$

for all $z \in \mathbb{R}^d$ and $\|z\|_2 \leq 1$. The theorem now follows using Lipschitz continuity of the loss function $\ell$ and using the bound $\|w^*\|_2 \leq 2LR_2 / \lambda$. 

Theorem 14 (Modified Theorem 7.3 from (Gupta et al., 2011)). If the distinguisher in Line 3 of Algorithm 3 outputs a $z$ (with $\|z\|_2 \leq 1$) at each step $t \in \{1, \ldots, B\}$ such that with probability at least $1 - \beta$ (over all the $B$-steps),

$$|\langle \phi(z), w_t - w^* \rangle| = \max_{z \in \mathcal{X}, \|z\|_2 \leq 1} |\langle \phi(z), w_t - \phi(z_1) \rangle| - \Omega \left( \frac{d R_2^2 \ln(L_2) \ln(1/\beta)}{\lambda n} \right),$$

then for all $z \in \mathbb{R}^d$ with $\|z\|_2 \leq 1$, with probability at least $1 - \beta$ (over all the $B$-steps),

$$|\langle \phi(z), w_t - w^* \rangle| \leq \mu,$$

where $\mu = \max \left( \frac{\|w^*\| R_2 \ln(L_2) \ln \frac{1}{\beta} \|w^*\|^2}{2\sigma^2 \lambda L^2 n^2 \epsilon}, \right).$