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 $\epsilon_0 = \frac{\epsilon}{200\sqrt{B}S\log(4/\delta)}$ and $\sigma = \frac{4}{\epsilon_0}\log(2T/\beta)$, where S is the maximum change in the output of a query (using $\mathbf{w}^*$) when any one entry in the underlying data set is arbitrarily modified. Let (ϵ, δ) be the privacy parameters and β be the failure probability in Algorithm 1. Under this setting, Algorithm 1 is (ϵ, δ)-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 $|\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{z})\rangle - \langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{z})\rangle|$, where $\mathbf{z} \in \mathcal{X}$ and $\mathcal{G}, \mathcal{G}'$ are two datasets differing in exactly one data point. Here $\mathbf{w}^*_{\mathcal{G}}$ and $\mathbf{w}^*_{\mathcal{G}'}$ represent optimal solution to regularized ERM (2) when the underlying datasets are $\mathcal{G}$ and $\mathcal{G}'$, respectively. In the second stage, we invoke Theorem 9 with sensitive bound $|\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{z})\rangle - \langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{z})\rangle|$ to complete the proof.

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

$$\frac{1}{n}\sum_{i=1}^{n-1}\ell(\langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{x}_i)\rangle; y_i) + \frac{1}{n}\ell(\langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{x}_n)\rangle; y_n) + \frac{\lambda}{2}\|\mathbf{w}^*_{\mathcal{G}'}\|_2^2$$
$$\geq \frac{1}{n}\sum_{i=1}^{n-1}\ell(\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{x}_i)\rangle; y_i) + \frac{1}{n}\ell(\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{x}_n)\rangle; y_n) + \frac{\lambda}{2}\|\mathbf{w}^*_{\mathcal{G}}\|_2^2 + \frac{\lambda}{2}\|\mathbf{w}^*_{\mathcal{G}} - \mathbf{w}^*_{\mathcal{G}'}\|_2^2.$$

Hence,

$$\frac{1}{n}\sum_{i=1}^{n-1}\ell(\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{x}_i)\rangle; y_i) + \frac{1}{n}\ell(\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{x}'_n)\rangle; y'_n) + \frac{\lambda}{2}\|\mathbf{w}^*_{\mathcal{G}}\|_2^2$$
$$\geq \frac{1}{n}\sum_{i=1}^{n-1}\ell(\langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{x}_i)\rangle; y_i) + \frac{1}{n}\ell(\langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{x}'_n)\rangle; y'_n) + \frac{\lambda}{2}\|\mathbf{w}^*_{\mathcal{G}'}\|_2^2 + \frac{\lambda}{2}\|\mathbf{w}^*_{\mathcal{G}} - \mathbf{w}^*_{\mathcal{G}'}\|_2^2.$$

Adding the above two equations and using Lipschitz continuity of ℓ:

$$\|\mathbf{w}^*_{\mathcal{G}} - \mathbf{w}^*_{\mathcal{G}'}\|_2 \leq \frac{2LR_\phi}{\lambda n}. \quad (3)$$

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

$$|\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{z})\rangle - \langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{z})\rangle| \leq \frac{2LR_\phi^2}{\lambda n}.$$

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. Setting the parameters as in Theorem 3 gives us the desired utility guarantee.

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

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

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 $\mathcal{G}$ and $\mathcal{G}'$ differing in exactly one entry, the following is true:

$$\|\mathbf{w}^*_{\mathcal{G}} - \mathbf{w}^*_{\mathcal{G}'}\|_2 \leq \frac{2LR_\phi}{\lambda n}.$$

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

$$|\langle \mathbf{w}^*_{\mathcal{G}}, \phi(\mathbf{z})\rangle - \langle \mathbf{w}^*_{\mathcal{G}'}, \phi(\mathbf{z})\rangle| \leq \frac{2LR_\phi^2}{\lambda n}.$$

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 ϵ-differentially private mechanisms satisfy (ϵ', δ')-differential privacy under k-fold adaptive composition for:*

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

B.2. Utility Guarantee of Test Data Dependent Learner

Proof of Theorem 5. Let,

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

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

$$\sum_{t=1}^{T}(\langle \hat{\mathbf{w}}, \phi(\mathbf{z}_t)\rangle - \langle \mathbf{w}^*, \phi(\mathbf{z}_t)\rangle)^2 \leq 2\sum_{t=1}^{T}\langle \hat{\mathbf{w}} - \mathbf{w}^*, \phi(\mathbf{z}_t)\rangle\, b_t.$$

Let $\mathbf{b} = \langle b_1, \cdots, b_T\rangle$. Using Cauchy-Schwarz inequality and the fact that $\|\mathbf{v}\|_1 \leq \sqrt{T}\|\mathbf{v}\|_2$, we get:

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

Since ν is the scaling parameter for the Laplace distribution from which each b_t are drawn, therefore by the tail property of Laplace distribution it follows that w.p. $\geq 1-\beta$,

$$\sum_{t=1}^{T}|\, \langle \hat{\mathbf{w}}, \phi(\mathbf{z}_t)\rangle - \langle \mathbf{w}^*, \phi(\mathbf{z}_t)\rangle\, | \leq 2\sqrt{2}T\nu\log(T/\beta)$$

Plugging in the value of $\nu = O\left(\frac{L{R_\phi}^2\sqrt{T\log(1/\delta)}}{\lambda n\epsilon}\right)$, we have

$$\sum_{t=1}^{T}|\, \langle \hat{\mathbf{w}}, \phi(\mathbf{z}_t)\rangle - \langle \mathbf{w}^*, \phi(\mathbf{z}_t)\rangle\, | = O\left(\frac{T^{3/2}LR_\phi^2\log(T/\beta)\sqrt{\log(1/\delta)}}{n\epsilon\lambda}\right). \quad (4)$$

Now, define $g(\mathbf{w};\mathbf{z}_t) = |\,\langle \mathbf{w}-\mathbf{w}^*, \phi(\mathbf{z}_t)\rangle\,|$; note that $g(\mathbf{w};\mathbf{z}_t)$ is a convex cost functions in $\mathbf{w}$. Now, using Theorem 1 from (Shalev-Shwartz et al., 2009) (stated below) we obtain the following:.

$$\mathbb{E}_{\mathbf{z}\sim\mathcal{P}}\left[g(\hat{\mathbf{w}};\mathbf{z})\right] \leq \frac{1}{T}\sum_{t=1}^{T}|g(\hat{\mathbf{w}};\mathbf{z}_t)| + O\left(\frac{\|\mathcal{C}\|_2 R_\phi\sqrt{\log(1/\beta)}}{\sqrt{T}}\right). \quad (5)$$

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

$$\mathbb{E}_{\mathbf{z}\sim\mathcal{P}}\left[g(\hat{\mathbf{w}};\mathbf{z})\right] \leq \frac{C_1\sqrt{T}LR_\phi^2\log(T/\beta)\sqrt{\log(1/\delta)}}{n\epsilon\lambda} + \frac{C_2\|\mathcal{C}\|_2 R_\phi\sqrt{\log(1/\beta)}}{\sqrt{T}},$$

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 ℓ. ☐

Theorem 12 (Theorem 1 from (Shalev-Shwartz et al., 2009)). *Let $\mathcal{C} = \{\mathbf{w} : \|\mathbf{w}\|_2 \leq B\}$ be a convex set, let $\phi : \mathcal{X} \to \mathbb{R}^{d_\phi}$ be a feature map with the image of ϕ has L_2-norm of at most R_ϕ, and let $f : \mathbb{R}\times\mathcal{X}\to\mathbb{R}$ be a L_f-Lipschitz continuous convex cost function in its first parameter. Then for any $\mathcal{P}$ over the domain $\mathcal{X}$, and for $Z = \{\mathbf{z}_1, \cdots, \mathbf{z}_T\}$ drawn i.i.d. from $\mathcal{P}$, the following is true with probability at least $1-\beta$.*

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

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 $Z = \{\mathbf{z}_1, \cdots, \mathbf{z}_T\}$ be sampled uniformly from $\mathcal{P}$. If $T = O\left(\frac{\|\mathcal{C}\|_2 n\epsilon\lambda}{LR_\phi\sqrt{\log(1/\delta)}}\right)$ and $\mathbf{w}^* \in \mathcal{C}$ in Algorithm 2, then w.p. $1-\beta$,*

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

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

Proof sketch of Theorem 7. For a given dataset $\mathcal{G}$, let $f(\mathcal{G}) = \left(\frac{\epsilon_0 n\lambda}{8LR_\phi^2}|\,\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^*(\mathcal{G})\rangle\,|\right)$. Using the fact that $\|\mathbf{w}^*(\mathcal{G}) - \mathbf{w}^*(\mathcal{G}')\|_2 \leq \frac{2LR_\phi}{n\lambda}$ for any two datasets $\mathcal{G}$ and $\mathcal{G}'$ differing in exactly one entry (see Theorem 2 from Section 5), it directly follows that $|f(\mathcal{G}) - f(\mathcal{G}')| \leq \frac{\epsilon_0}{4}$. 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 that Algorithm 3 is (ϵ, δ)-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 $\mathbf{w}_t$ will become representative of $\mathbf{w}^*$ 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 $\mathbf{z}$ (via exponential mechanism) such that it can distinguish between $\hat{\mathbf{w}}_t$ and $\mathbf{w}^*$ with high probability, (*i.e.*, the value of $\langle \phi(\mathbf{z}), \hat{\mathbf{w}} - \mathbf{w}^* \rangle$ is greater than $\frac{\sigma}{4}$).

Main Proof: We apply exponential mechanism to a finite set $S = \{\mathbf{z} : \mathbf{z} \text{ is the center of the } \nu\text{-net}\}$, where ν is as given in the Theorem. That is, we divide the entire space into (overlapping) L_2 balls of radius ν and S is the collection of centers of all such balls. Also, it is known that $|S| = \left(\frac{4}{\nu}\right)^d$.

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

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

where $OPT_\nu = \max_{\mathbf{z}\in S} |\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle|$ and $\Lambda = \frac{\epsilon_0 n \lambda}{8LR_\phi^2}$. Hence, w.p. at least $1-\beta$, a $\mathbf{z}$ is sampled s.t.,

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

Now, let OPT^* be the maximum value of $|\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle|$ over the input space $\mathcal{X}$, i.e., $OPT^* = \max_{\mathbf{z}\in\mathcal{X}} |\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle|$. Also, $\|\mathbf{z}^* - \mathbf{z}_\nu\|_2 \leq 2\nu$ where $z_\nu = \arg\max_{\mathbf{z}\in S} |\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle|$ is the optimal over S. Hence, using Lipschitz continuity of the mapping ϕ, we obtain a sample $\mathbf{z}$ w.p. at least $1-\beta$ s.t.:

$$|\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle| \geq OPT^* - \frac{\ln(|S|/\beta)}{\Lambda} - \frac{2\nu L_\phi R_\phi L}{\lambda}.$$

Hence, selecting $\nu = \frac{dR_\phi}{\epsilon_0 n L_\phi}$, we get $|\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle| \geq OPT^* - \Omega\left(\frac{dLR_\phi^2 \ln(L_\phi)\ln(1/\beta)}{\lambda\epsilon_0 n}\right)$. Now, using Theorem 7.3 of (Gupta et al., 2011) (see Theorem 14), we get,

$$\begin{aligned} &|\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle| \leq \sigma \\ &\quad = \max\left(\frac{\|\mathbf{w}^*\| R_\phi}{2\sigma\epsilon}, \frac{dL^2 R_\phi^6 \ln(L_\phi) \ln\frac{1}{\beta} \|\mathbf{w}^*\|^2}{\sigma^2\lambda^2 n^2 \epsilon}\right), \end{aligned}$$

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

$$\begin{aligned} &|\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle| \\ &= O\left(\frac{\|\mathbf{w}^*\| R_\phi^2 d^{1/3} L^{2/3} \ln L_\phi \log^2 1/\delta \ln(1/\beta)}{(\lambda n)^{2/3}\sqrt{\epsilon}}\right), \end{aligned}$$

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

Theorem 14 (Modified Theorem 7.3 from (Gupta et al., 2011))**.** *If the distinguisher in Line 3 of Algorithm 3 outputs a* $\mathbf{z}$ *(with* $\|\mathbf{z}\|_2 \leq 1$*) at each step* $t \in \{1, \cdots, B\}$ *such that with probability at least* $1-\beta$ *(over all the* B*-steps),* $|\langle \mathbf{w}^* - \mathbf{w}_t, \phi(\mathbf{z}) \rangle| = \max_{\mathbf{z}_1\in\mathcal{X}, \|\mathbf{z}_1\|_2\leq 1} |\langle \mathbf{w}^* - \mathbf{w}_t, \phi(\mathbf{z}_1) \rangle| - \Omega\left(\frac{dLR_\phi^2 \ln(L_\phi)\ln(1/\beta)}{\lambda\epsilon_0 n}\right)$*, then for all* $\mathbf{z} \in \mathbb{R}^d$ *with* $\|\mathbf{z}\|_2 \leq 1$*, with probability at least* $1-\beta$ *(over all the* B*-steps),* $|\langle \phi(\mathbf{z}), \mathbf{w}_t - \mathbf{w}^* \rangle| \leq \mu$*, where* $\mu = \max\left(\frac{\|\mathbf{w}^*\| R_\phi}{2\sigma\epsilon}, \frac{dL^2 R_\phi^6 \ln(L_\phi) \ln\frac{1}{\beta}\|\mathbf{w}^*\|^2}{\sigma^2\lambda^2 n^2\epsilon}\right)$.

□