A Bandit Framework for Strategic Regression

Yang Liu and Yiling Chen
School of Engineering and Applied Science, Harvard University
{yangl,yiling}@seas.harvard.edu

Abstract

We consider a learner's problem of acquiring data dynamically for training a regression model, where the training data are collected from strategic data sources. A fundamental challenge is to incentivize data holders to exert effort to improve the quality of their reported data, despite that the quality is not directly verifiable by the learner. In this work, we study a dynamic data acquisition process where data holders can contribute multiple times. Using a bandit framework, we leverage the long-term incentive of future job opportunities to incentivize high-quality contributions. We propose a Strategic Regression-Upper Confidence Bound (SR-UCB) framework, a UCB-style index combined with a simple payment rule, where the index of a worker approximates the quality of his past contributions and is used by the learner to determine whether the worker receives future work. For linear regression and a certain family of non-linear regression problems, we show that SR-UCB enables an $\mathrm{O}\big(\sqrt{\log T/T}\big)$-Bayesian Nash Equilibrium (BNE) where each worker exerts a target effort level that the learner has chosen, with T being the number of data acquisition stages. The SR-UCB framework also has some other desirable properties: (1) The indexes can be updated in an online fashion (hence computation is light). (2) A slight variant, namely Private SR-UCB (PSR-UCB), is able to preserve $\big(\mathrm{O}\big(\log^{-1} T\big), \mathrm{O}\big(\log^{-1} T\big)\big)$-differential privacy for workers' data, with only a small compromise on incentives (each worker exerting a target effort level is an $\mathrm{O}\big(\log^6 T/\sqrt{T}\big)$-BNE).

1 Introduction

More and more data for machine learning nowadays are acquired from distributed, unmonitored and strategic data sources and the quality of these collected data is often unverifiable. For example, in a crowdsourcing market, a data requester can pay crowd workers to label samples. While this approach has been widely adopted, crowdsourced labels have been shown to degrade the learning performance significantly, see e.g., [21], due to the low quality of the data. How to incentivize workers to contribute high-quality data is hence a fundamental question that is crucial to the long-term viability of this approach.

Recent works [2,4,11] have considered incentivizing data contributions for the purpose of estimating a regression model. For example Cai et al. [2] design payment rules so that workers are incentivized to exert effort to improve the quality of their contributed data, while Cummings et al. [4] design mechanisms to compensate privacy-sensitive workers for their privacy loss when contributing their data. These studies focus on a static data acquisition process, only considering one-time data acquisition from each worker. Hence, the incentives completely rely on the payment rule. However, in stable crowdsourcing markets, workers return to receive additional work. Future job opportunities are thus another dimension of incentives that can be leveraged to motive high-quality data contributions. In this paper, we study dynamic data acquisition from strategic agents for regression problems and explore the use of future job opportunities to incentivize effort exertion.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

In our setting, a learner has access to a pool of workers and in each round decides on which workers to ask for data. We propose a Multi-armed Bandit (MAB) framework, called Strategic Regression-Upper Confidence Bound (SR-UCB), that combines a UCB-style index rule with a simple per-round payment rule to align the incentives of data acquisition with the learning objective. Intuitively, each worker is an arm and has an index associated with him that measures the quality of his past contributions. The indexes are used by the learner to select workers in the next round. While MAB framework is natural for modeling selection problem with data contributors of potentially varying qualities, our setting has two challenges that are distinct from classical bandit settings. First, after a worker contributes his data, there is no ground-truth observation to evaluate how well the worker performs (or reward as commonly referred to in a MAB setting). Second, a worker's performance is a result of his strategic decision (e.g. how much effort he exerts), instead of being purely exogenously determined. Our SR-UCB framework overcomes the first challenge by evaluating the quality of an agent's contributed data against an estimator trained on data provided by all other agents to obtain an unbiased estimate of the quality, an idea inspired by the peer prediction literature [13, 18]. To address the second challenge, our SR-UCB framework enables a game-theoretic equilibrium with workers exerting target effort levels chosen by the learner. More specifically, in addition to proposing the SR-UCB framework, our contributions include:

- We show that SR-UCB helps simplify the design of payment, and successfully incentivizes effort exertion for acquiring data for linear regression. Every worker exerting a targeted effort level (for labeling and reporting the data) is an $\mathrm{O}\big(\sqrt{\log T/T}\big)$-Bayesian Nash Equilibrium (BNE). We can also extend the above results to a certain family of non-linear regression problems.
- SR-UCB indexes can be maintained in an online fashion, hence are computationally light.
- We extend SR-UCB to Private SR-UCB (PSR-UCB) to further provide privacy guarantees, with small compromise on incentives. PSR-UCB is $(\mathrm{O}\big(\log^{-1}T\big), \mathrm{O}\big(\log^{-1}T\big))$-differentially private and every worker exerting the targeted effort level is an $\mathrm{O}\big(\log^{6}T/\sqrt{T}\big)$-BNE.

2 Related work

Recent works have formulated various strategic learning settings under different objectives [2, 4, 11, 22]. Among these, payment based solutions are proposed for regression problems when data come from workers who are either effort sensitive [2] or privacy sensitive [4]. These solutions induce game-theoretic equilibria where high-quality data are contributed. The basic idea of designing the payment rules is inspired by the much mature literature of proper scoring rules [9] and peer prediction [18]. Both [2] and [4] consider a static data acquisition procedure, while our work focuses on a dynamic data acquisition process. Leveraging the long-term incentive of future job opportunities, our work has a much simpler payment rule than those of [2] and [4] and relaxes some of the restrictions on the learning objectives (e.g., well behaved in [2]), at the cost of a weaker equilibrium concept (approximate BNE in this work vs. dominate strategy in [2]).

Multi-armed Bandit (MAB) is a sequential decision making and learning framework which has been extensively studied. It is nearly impossible to survey the entire bandit literature. The seminal work by Lai et al [15] derived lower and upper bounds on asymptotic regret on bandit selection. More recently, finite-time algorithms have been developed for i.i.d. bandits [1] . Different from the classical settings, this work needs to deal with challenges such as no ground-truth observations for bandits and bandits' rewards being strategically determined. A few recent works [8, 17] also considered bandit settings with strategic arms. Our work differs from these in that we consider a regression learning setting without ground-truth observations, as well as we consider long-term workers whose decisions on reporting data can change over time.

Our work and motivations have some resemblance to online contract design problems for a principal-agent model [10]. But unlike the online contract design problems, our learner cannot verify the quality of finished work after each task assignment. In addition, instead of focusing on learning the optimal contract, we use bandits mainly to maintain a long-term incentive for inducing high-quality data.

3 Formulation

The learner observes a set of feature data X for training. To make our analysis tractable, we assume each $x \in X$ is sampled uniformly from a unit ball with dimension d: $x \in \mathbb{R}^d$ s.t. $||x||_2 \leq 1$. Each x associates with a ground-truth response (or label) $y(x)$, which cannot be observed directly by the learner. Suppose x and $y(x)$ are related through a function $f : \mathbb{R}^d \to \mathbb{R}$ that $y(x) = f(x) + z$, where z is a zero-mean noise with variance σ_z, and is independent of x. For example, for linear regression $f(x) = \theta^T x$ for some $\theta \in \mathbb{R}^d$. The learner would like to learn a good estimate $\tilde{f}$ of f. For the purpose of training, the learner needs to figure out $y(x)$ for different $x \in X$. To obtain an estimate $\tilde{y}(x)$ of $y(x)$, the learner assigns each x to a selected worker to obtain a label.

Agent model: Suppose we have a set of workers $\mathcal{U} = \{1, 2, ..., N\}$ with $N \geq 2$. After receiving the labeling task, each worker will decide on the effort level e he wants to exert to generate an outcome – higher effort leads to a better outcome, but is also associated with a higher cost. We assume e has bounded support $[0, \bar{e}]$ for all worker $i \in \mathcal{U}$. When deciding on an effort level, a worker wants to maximize his expected payment minus cost for effort exertion. The resulted label $\tilde{y}(x)$ will be given back to the learner. Denote by $\tilde{y}_i(x, e)$ the label returned by worker i for data instance x (if assigned) with chosen effort level e. We consider the following effort-sensitive agent model: $\tilde{y}_i(x, e) = f(x) + z + z_i(e)$, where $z_i(e)$ is a zero-mean noise with variance $\sigma_i(e)$. $\sigma_i(e)$ can be different for different workers, and $\sigma_i(e)$ decreases in $e, \forall i$. The z and z_i's have bounded support such that $|z|, |z_i| \leq Z$, $\forall i$. Without loss of generality, we assume that the cost for exerting effort e is simply e for every worker.

Learner's objective Suppose the learner wants to learn f with the set of samples X. Then the learner finds effort levels $\mathbf{e}^*$ for data points in X such that

$$\mathbf{e}^* \in \text{argmin}_{\{e(x)\}_{x \in X}} \texttt{ERROR}(\tilde{f}(\{x, \tilde{y}(x, e(x))\}_{x \in X})) + \lambda \cdot \texttt{PAYMENT}(\{e(x)\}_{x \in X}) \ ,$$

where $e(x)$ is the effort level for sample x, and $\{\tilde{y}(x, e(x))\}_{x \in X}$ is the set of labeled responses for training data X. $\tilde{f}(\cdot)$ is the regression model trained over this data. The learner assigns the data and pay appropriately to induce the corresponding effort level $\mathbf{e}^*$. This formulation resembles the one presented in [2]. The `ERROR` term captures the expected error of the trained model using collected data (e.g., measure in squared loss), while the `PAYMENT` term captures the total expected budget that the learner spends to receive the labels. This payment quantity depends on the mechanism that the learner chooses to use and is the expected payment of the mechanism to induce selected effort level for each data point $\{e(x)\}_{x \in X}$. $\lambda > 0$ is a weighting factor, which is a constant. It is clear that the objective function depends on σ_i's. We assume for now that the learner knows $\sigma_i(\cdot)$'s,[1] and the optimal $\mathbf{e}^*$ can be computed.

4 StrategicRegression-UCB (SR-UCB): A general template

We propose SR-UCB for solving the dynamic data acquisition problem. SR-UCB enjoys a bandit setting, where we borrow the idea from the classical UCB algorithm [1], which maintains an index for each arm (worker in our setting), balancing exploration and exploitation. While a bandit framework is not necessarily the best solution for our dynamic data acquisition problem, it is a promising option for the following reasons. First, as utility maximizers, workers would like to be assigned tasks as long as the marginal gain for taking a task is positive. A bandit algorithm can help execute the assignment process. Second, carefully designed indexes can potentially reflect the amount of effort exerted by the agents. Third, because the arm selection (of bandit algorithms) is based on the indexes of workers, it introduces competition among workers for improving their indexes.

SR-UCB contains the following two critical components:

Per-round payment For each worker i, once selected to label a sample x, we will assign a base payment $p_i = e_i + \gamma$, [2] after reporting the labeling outcome, where e_i is the desired effort level that we would like to induce from worker i (for simplicity we have assumed the cost for exerting effort e_i equals to the effort level), and $\gamma > 0$ is a small quantity. The design of this base payment is to ensure

[1] This assumption can be relaxed. See our supplementary materials for the case with homogeneous σ.

[2] We assume workers have knowledge of how the mechanism sets up this γ.

once selected, a worker's base cost will be covered. Note the above payment depends on neither the assigned data instance x nor the reported outcome $\tilde{y}$. Therefore such a payment procedure can be pre-defined after the learner sets a target effort level.

Assignment The learner assigns multiple task $\{x_i(t)\}_{i\in d(t)}$ at time t, with $d(t)$ denoting the set of workers selected at t. Denote by $e_i(t)$ the effort level worker i exerted for $x_i(t)$, if $i \in d(t)$. Note all $\{x_i(t)\}_{i\in d(t)}$ are different tasks, and each of them is assigned to exactly one worker. The selection of workers will depend on the notion of indexes. Details are given in Algorithm 1.

Algorithm 1 SR-UCB: Worker index & selection

Step 1. For each worker i, first train estimator $\tilde{f}_{-i,t}$ using data $\{x_j(n) : 1 \le n \le t-1,\ j \in d(n), j \neq i\}$, that is using the data collected from workers $j \neq i$ up to time $t-1$. When $t = 1$, we will initialize by sampling each worker at least once such that $\tilde{f}_{-i,t}$ can be computed.
Step 2. Then compute the following index for worker i at time t

$$I_i(t) = \frac{1}{n_i(t)} \sum_{n=1}^{t} 1(i \in d(n)) \left[a - b\left(\tilde{f}_{-i,t}(x_i(n)) - \tilde{y}_i(n, e_i(n)) \right)^2 \right] + c\sqrt{\frac{\log t}{n_i(t)}},$$

where $n_i(t)$ is the number of times worker i has been selected up to time t. a, b are two positive constants for "scoring", and c is a normalization constant. $\tilde{y}_i(n, e_i(n))$ is the corresponding label for task $x_i(n)$ with effort level $e_i(n)$, if $i \in d(n)$.
Step 3. Based on the above index, we select $d(t)$ at time t such that $d(t) := \{j : I_j(t) \ge \max_i I_i(t) - \tau(t)\}$, where $\tau(t)$ is a perturbation term decreasing in t.

Some remarks on SR-UCB: (1) Different from the classical bandit setting, when calculating the indexes, there is no ground-truth observation for evaluating the performance of each worker. Therefore we adopt the notion of scoring rule [9]. Particularly the one we used above is the well-known Brier scoring rule: $B(p,q) = a - b(p-q)^2$. (2) The scoring rule based index looks similar to the payment rules studied in [2, 4]. But as we will show later, under our framework the selection of a, b is much less sensitive to different problem settings, as with an index policy, only the relative values matter (ranking). This is another benefit of separating payment from selection. (3) Instead of only selecting the best worker with the highest index, we select workers whose index is within a certain range of the maximum one (a confidence region). This is because workers may have competing expertise level and hence selecting only one of them would de-incentivize workers' effort exertion.

4.1 Solution concept

Denote by $\mathbf{e}(n) := \{e_1(n), ..., e_N(n)\}$, and $e_{-i}(n) = \{e_j(n)\}_{j\neq i}$. We define approximate Bayesian Nash Equilibrium as our solution concept:

Definition 1. Suppose SR-UCB runs for T stages. $\{e_i(t)\}_{i=1,t=1}^{N,T}$ is a π-BNE if $\forall i, \{\tilde{e}_i(t)\}_{t=1}^{T}$:

$$\frac{1}{T}\mathbb{E}[\sum_{t=1}^{T}(p_i - e_i(t))1(i \in d(t))\big|\{\mathbf{e}(n)\}_{n\le t}] \ge \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T}(p_i - \tilde{e}_i(t))1(i \in d(t))\big|\{\tilde{e}_i(n), e_{-i}(n)\}_{n\le t}] - \pi.$$

This is to say by deviating, each worker will gain no more than π net-payment per around. We will establish our main results in terms of π-BNE. The reason we adopt such a notion is that in a sequential setting it is generally hard to achieve strict BNE or other stronger notion as any one-step deviation may not affect a long-term evaluation by much.[3] Approximate BNE is likely the best solution concept we can hope for.

5 Linear regression

5.1 Settings and a warm-up scenario

In this section we present our results for a simple linear regression task where the feature x and observation y are linearly related via an unknown θ: $y(x) = \theta^T x + z,\ \forall x \in X$. Let's start with assuming

[3] Certainly, we can run mechanisms that induce BNE or dominant-strategy equilibrium for one-shot setting, e.g. [2], for every time step. But such solution does not incorporate long-term incentives.

all workers are statistically identical such that $\sigma_1 = \sigma_2 = ... = \sigma_N$. This is an easier case that serves as a warm-up. It is known that given training data, we can find an estimation $\tilde{\theta}$ that minimizes a non-regularized empirical risk function: $\tilde{\theta} = \text{argmin}_{\hat{\theta}\in\mathbb{R}^d} \sum_{x\in X}(y(x) - \hat{\theta}^T x)^2$ (linear least square). To put this model into SR-UCB, denote $\tilde{\theta}_{-i}(t)$ as the linear least square estimator trained using data from workers $j \neq i$ up to time $t-1$. And $I_i(t) := S_i(t) + c\sqrt{\log t / n_i(t)}$, with

$$S_i(t) := \frac{1}{n_i(t)} \sum_{n=1}^{t-1} 1(i \in d(n)) \left[a - b\left(\tilde{\theta}_{-i}^T(t) x_i(n) - \tilde{y}_i(n, e_i(n)) \right)^2 \right]. \tag{5.1}$$

Suppose $||\theta||_2 \leq M$. Given $||x||_2 \leq 1$ and $|z|$, $|z_i| \leq Z$, we then prove that $\forall t, n, i$, $(\tilde{\theta}_{-i}^T(t) x_i(n) - \tilde{y}_i(n, e_i(n)))^2 \leq 8M^2 + 2Z^2$. Choose a, b such that $a - (8M^2 + 2Z^2)b \geq 0$, then we have $0 \leq S_i(t) \leq a$, $\forall i, t$. For the perturbation term, we set $\tau(t) := \mathrm{O}\big(\sqrt{\log t / t}\big)$. The intuition is that with t samples, the uncertainties in the indexes, coming from both the score calculation and the bias term, can be upper bounded at the order of $\mathrm{O}\big(\sqrt{\log t / t}\big)$. Thus, to not miss a competitive worker, we set the tolerance to be at the same order.

We now develop the formal equilibrium result of SR-UCB for linear least square. Our analysis requires the following assumption on the smoothness of σ.

Assumption 1. *We assume $\sigma(e)$ is convex on $e \in [0, \bar{e}]$, with gradient $\sigma'(e)$ being both upper bounded, and lower bounded away from 0, i.e., $\overline{L} \geq |\sigma'(e)| \geq \underline{L} > 0$, $\forall e$.*

The learner wants to learn f with a total of NT ($= |X|$ or $\lceil NT \rceil = |X|$) samples. Since workers are statistically equivalent, ideally the learner would like to run SR-UCB for T steps and collect a label for a unique sample from each worker at each step. Hence, the learner would like to elicit a single target effort level e^* from all workers and for all samples:

$$e^* \in \text{argmin}_e \mathbb{E}_{x,y,\tilde{y}} \left[\theta^T (\{x_i(n), \tilde{y}_i(n,e)\}_{i=1,n=1}^{N,T}) \cdot x - y \right]^2 + \lambda \cdot (e + \gamma) NT. \tag{5.2}$$

Due to the uncertainty in worker selection, it is highly likely that after step T, there will be tasks left unlabelled. We can let the mechanism go for extra steps to complete labelling of these tasks. But due to the bounded number of missed selections as we will show later, stopping at step T won't affect the accuracy in the model trained.

Theorem 1. *Under SR-UCB for linear least square, set fixed payment $p_i = e^* + \gamma$ for all i, where $\gamma = \Omega(\sqrt{\log T / T})$, choose c to be a large enough constant, $c \geq$ Const.(M, Z, N, b), and let $\tau(t) := \mathrm{O}\big(\sqrt{\log t / t}\big)$. Workers have full knowledge of the mechanism and the values of the parameters. Then at an $\mathrm{O}\big(\sqrt{\log T / T}\big)$-BNE, workers, whenever selected, exert effort $e_i(t) \equiv e^*$ for all i and t.*

The net payment (payment minus the cost of effort) per task can be made arbitrarily small by setting γ exactly on the order of $\mathrm{O}\big(\sqrt{\log T / T}\big)$, and $p_i - e^* = \gamma = \mathrm{O}\big(\sqrt{\log T / T}\big) \to 0$, as $T \to \infty$.

Our solution heavily relies on forming a race among workers. By establishing the convergence of bandit indexes to a function of effort (via $\sigma(\cdot)$), we show that when other workers $j \neq i$ follow the equilibrium strategy, worker i will be selected w.h.p. at each round, if he also puts in the same amount of effort. On the other hand, if worker i shirks from doing so by as much as $(\mathrm{O}\big(\sqrt{\log T / T}\big))$, his number of selection will go down in order. This establishes the π-BNE. As long as there exists one competitive worker, all others will be incentivized to exert effort. Though as will be shown in the next section, all workers shirking from exerting effort is also an $\mathrm{O}\big(\sqrt{\log T / T}\big)$-BNE. This equilibrium can be removed by adding some uncertainty on top of the bandit selection procedure. When there are ≥ 2 workers being selected in SR-UCB, each of them will be assigned a task with certain probability $0 < p_s < 1$. While when there is a single selected worker, the worker is assigned a task w.p. 1. Set $p_s := 1 - \mathrm{O}\big(\sqrt{\log T / T / \gamma}\big)$. So with probability $1 - p_s = \mathrm{O}\big(\sqrt{\log T / T / \gamma}\big)$, even the "winning"workers will miss the selection. With this change, exerting e^* still forms an $\mathrm{O}\big(\sqrt{\log T / T}\big)$-BNE, while every worker exerting any effort level that is $\Delta e > \mathrm{O}\big(\gamma\big)$ lower than the target effort level is not a π-BNE with $\pi \leq \mathrm{O}\big(\sqrt{\log T / T}\big)$.

5.2 Linear regression with different σ

Now we consider the more realistic case that different workers have different noise-effort function σ's. W.l.o.g., we assume $\sigma_1(e) < \sigma_2(e) < ... < \sigma_N(e), \forall e$.[4] In such a setting, ideally we would always like to collect data from worker 1 since he has the best expertise level (lowest variance in labeling noise). Suppose we are targeting an effort level e_1^* from data source 1 (the best data source). We first argue that we also need to incentivize worker 2 to exert competitive effort level e_2^* such that $\sigma_1(e_1^*) = \sigma_2(e_2^*)$, and we assume such an e_2^* exists.[5] This also naturally implies that $e_2^* > e_1^*$ as worker 1 contributes data with less variance in noise at the same effort level. The reason is similar to the homogeneous setting—over time workers form a competition on $\sigma_i(e_i)$. Having a competitive peer will motivate workers to exert as much effort as he can (up to the payment). Therefore the goal for such a learner (with $2T$ samples to assign) is to find an effort level e^* such that [6]

$$e^* \in \text{argmin}_{e_2:\sigma_1(e_1)=\sigma_2(e_2)} \mathbb{E}_{x,y,\tilde{y}} \left[\theta^T (\{x_i(n), \tilde{y}_i(n, e_i)\}_{i=1,n=1}^{2,T}) x - y \right]^2 + \lambda \cdot (e_2 + \gamma) 2T.$$

Set the one-step payment to be $p_i = e^* + \gamma, \forall i$. Let e_1^* be the solution to $\sigma_1(e_1^*) = \sigma_2(e^*)$ and let $e_i^* = e^*$ for $i \geq 2$. Note for $i > 2$ we have $\sigma_i(e_i^*) - \sigma_1(e_1^*) > 0$. While we have argued about the necessity for choosing the top two most competitive workers, we have not mentioned the optimality of doing so. In fact selecting the top two is the best we can do. Suppose on the contrary, the optimal solution is by selecting top $k > 2$ workers, at effort level e_k. According to our solution, we targeted the effort level that leads to variance of noise $\sigma_k(e_k)$ (so the least competitive worker will be incentivized). Then we can simply target the same effort level e_k, but migrating the task loads to only the top two workers – this keeps the payment the same, but the variance of noise now becomes $\sigma_2(e_k) < \sigma_k(e_k)$, which leads to better performance. Denote $\Delta_1 := \sigma_3(e^*) - \sigma_1(e_1^*) > 0$ and assume Assumption 1 applies to all σ_i's. We prove:

Theorem 2. *Under SR-UCB for linear least square, set* $c \geq Const.(M, Z, b, \Delta_1)$, $\Omega(\sqrt{\log T/T}) = \gamma \leq \frac{\Delta_1}{2L}$, $\tau(t) := O\big(\sqrt{\log t/t}\big)$. *Then, each worker* i *exerting effort* e_i^* *once selected forms an* $O\big(\sqrt{\log T/T}\big)$*-BNE.*

Performance with acquired data If workers follow the π-BNE, the contributed data from the top two workers (who have been selected the most number of times) will have the same variance $\sigma_1(e_1^*)$. Then following results in [4], w.h.p. the performance of the trained classifier is bounded by $O\big(\sigma_1(e_1^*)/(\sum_{i=1,2} n_i(T))^2\big)$. Ideally we want to have $\sum_{i=1,2} n_i(T) = 2T$, such that an upper bound of $O\big(\sigma_1(e_1^*)/(2T)^2\big)$ can be achieved. Compared to the bound $O\big(\sigma_1(e_1^*)/(2T)^2\big)$, SR-UCB's expected performance loss (due to missed sampling & wrong selection, which is bounded at the order of $O\big(\log T\big)$) is bounded by $\mathbb{E}[\sigma_1(e_1^*)/(\sum_{i=1,2} n_i(T))^2 - \sigma_1(e_1^*)/(2T)^2] \leq O\big(\sigma_1(e_1^*) \log T/T^3\big)$ w.h.p. .

Regularized linear regression Ridge estimator has been widely adopted for solving linear regression. The objective is to find a linear model $\tilde{\theta}$ that minimizes the following regularized empirical risk: $\tilde{\theta} = \text{argmin}_{\hat{\theta} \in \mathbb{R}^d} \sum_{x \in X} (y(x) - \hat{\theta}^T x)^2 + \rho ||\hat{\theta}||_2^2$, with $\rho > 0$ being the regularization parameter. We claim that simply changing the $\tilde{f}_{-i,t}(\cdot)$ in SR-UCB to the output from the above ridge regression, the $O\big(\sqrt{\log T/T}\big)$-BNE for inducing an effort level e^* will hold. Different from the non-regularized case, the introduction of the regularization term will add bias in $\tilde{\theta}_{-i}^T(t)$, which gives a biased evaluation of indexes. However, we prove the convergence of $\tilde{\theta}_{-i}^T(t)$ (so again the indexes will converge properly) in the following lemma, which enables an easy adaption of our previous results for non-regularized case to ridge regression:

Lemma 1. *With n i.i.d. samples, w.p.* $\geq 1 - e^{-Kn}$ *(*$K > 0$ *is a constant),* $||\tilde{\theta}_{-i}(t) - \theta||_2^2 \leq O\big(\frac{1}{n^2}\big)$.

Non-linear regression The basic idea for extending the results to non-linear regression is inspired by the consistency results on *M*-estimator [16], when the error of training data satisfies zero mean. Similar to the reasoning for Lemma 1, if $(\tilde{f}_{-i,t}(x) - f(x))^2 \to 0$, we can hope for an easy adaptation

[4] Combing with the results for homogeneous workers, we can again easily extend our results to the case where there are a mixture of homogeneous and heterogenous workers.

[5] It exists when the supports for $\sigma_1(\cdot), \sigma_2(\cdot)$ overlap for a large support range.

[6] Since we only target the top two workers, we can limit the number of acquisitions on each stage to be no more than two, so the number of query does not go beyond $2T$.

of our previous results. Suppose the non-linear regression model can be characterized by a parameter family Θ, where f is characterized by parameter θ, and $\tilde{f}_{-i,t}$ by $\tilde{\theta}_i(t)$. Due to the consistency of M-estimator we will have $||\tilde{\theta}_i(t) - \theta||_2 \to 0$. More specifically, according to the results from [20], for the non-linear regression model we can establish an $\mathrm{O}\big(1/\sqrt{n}\big)$ convergence rate with n training samples. When f is Lipschitz in parameter space, i.e. there exists a constant $L_N > 0$ such that $|\tilde{f}_{-i,t}(x) - f(x)| \leq L_N||\tilde{\theta}_i(t) - \theta||_2$, by dominated convergence theorem we also have $(\tilde{f}_{-i,t}(x) - f(x))^2 \to 0$, and $(\tilde{f}_{-i,t}(x) - f(x))^2 \leq \mathrm{O}\big(1/t\big)$. The rest of the proof can then follow.

Example 1. Logistic function $f(x) = \frac{1}{1+e^{-\theta^T x}}$ satisfies Lipschitz condition with $L_N = 1/4$.

6 Computational issues

In order to update the indexes and select workers adaptively, we face a few computational challenges. First, in order to update the index for each worker at any time t, a new estimator $\tilde{\theta}_{-i}(t)$ (using data from all other workers $j \neq i$ up to time $t-1$) needs to be re-computed. Second, we need to re-apply $\tilde{\theta}_{-i}(t)$ to every collected sample from worker $i, \{(x_i(n), \tilde{y}_i(n, e_i(n)) : i \in d(n), n = 1, 2, ...t-1\}$ from previous rounds. We propose online variants of SR-UCB to address these challenges.

Online update of $\tilde{\theta}_{-i}(\cdot)$ Inspired by the online learning literature, instead of re-computing $\tilde{\theta}_{-i}(t)$ at each step, which involves re-calculating the inverse of a covariance matrix (e.g., $(\rho I + X^T X)^{-1}$ for ridge regression) whenever there is a new sample point arriving, we can update $\tilde{\theta}_{-i}(t)$ in an online fashion, which is computationally much more efficient. We demonstrate our results with ridge linear regression. Start with an initial model $\tilde{\theta}^{\text{online}}_{-i}(1)$. Denote by $(x_{-i}(t), \tilde{y}_{-i}(t))$ any newly arrived sample at time t from worker $j \neq i$. Update $\tilde{\theta}^{\text{online}}_{-i}(t+1)$ (for computing $I_i(t+1)$) as [19]:

$$\tilde{\theta}^{\text{online}}_{-i}(t+1) := \tilde{\theta}^{\text{online}}_{-i}(t) - \eta_t \cdot \nabla_{\tilde{\theta}^{\text{online}}_{-i}(t)}[(\theta^T x_{-i}(t) - \tilde{y}_{-i}(t))^2 + \rho||\theta||_2^2] \,,$$

Notice there could be multiple such data points arriving at each time – in which case we will update sequentially in an arbitrarily order. It is also possible that there is no sample point arriving from workers other than i at a time t, in which case we simply do not perform an update. Name this online updating SR-UCB as OSR1-UCB. With online updating, the accuracy of trained model $\tilde{\theta}^{\text{online}}_{-i}(t+1)$ converges slower, so is the accuracy in the index for characterizing worker's performance. Nevertheless we prove exerting targeted effort exertion e^* is $\mathrm{O}\big(\sqrt{\log T/T}\big)$-BNE under OSR1-UCB for ridge regression, using convergence results for $\tilde{\theta}^{\text{online}}_{-i}(t)$ proved in [19].

Online score update Online updating can also help compute $S_i(t)$ (in $I_i(t)$) efficiently. Instead of repeatedly re-calculating the score for each data point (in $S_i(t)$), we only update the newly assigned samples which has not been evaluated yet, by replacing $\tilde{\theta}^{\text{online}}_{-i}(t)$ with $\tilde{\theta}^{\text{online}}_{-i}(n)$ in $S_i(t)$:

$$S^{\text{online}}_i(t) := \frac{1}{n_i(t)} \sum_{n=1}^{t} 1(i \in d(n))[a - b((\tilde{\theta}^{\text{online}}_{-i}(n))^T x_i(n) - \tilde{y}_i(n, e_i(n)))^2]. \tag{6.1}$$

With less aggressive update, again the index term's accuracy converges slower than before, which is due to the fact the older data is scored using an older (and less accurate) version of $\tilde{\theta}^{\text{online}}_{-i}$ without being further updated. We propose OSR2-UCB where we change the index SR-UCB to: $S^{\text{online}}_i(t) + c\sqrt{(\log t)^2/n_i(t)}$, to accommondate the slower convergence. We establish an $\mathrm{O}\big(\log T/\sqrt{T}\big)$-BNE for workers exerting target effort—the change is due to the change of the bias term.

7 Privacy preserving SR-UCB

With a repeated data acquisition setting, workers' privacy in data may leak repeatedly. In this section we study an extension of SR-UCB to preserve privacy of each individual worker's contributed data. Denote the training data collected as $\mathcal{D} := \{\tilde{y}_i(t, e_i(t))\}_{i \in d(t), t}$. We quantify privacy using differential privacy [6], and we adopt (ε, δ)-differential privacy (DP) [7], which for our setting is defined below:

Definition 2. A mechanism $\mathcal{M} : (X \times \mathbb{R})^{|\mathcal{D}|} \to O$ is (ε, δ)-differentially private if for any $i \in d(t), t$, any two distinct $\tilde{y}_i(t, e_i(t)), \tilde{y}'_i(t, e'_i(t))$, and for every subset of possible outputs $\mathcal{S} \subseteq O$, $\Pr[\mathcal{M}(\mathcal{D}) \in \mathcal{S}] \leq \exp(\varepsilon) \Pr[\mathcal{M}(\mathcal{D} \backslash \{\tilde{y}_i(t, e_i(t))\}, \tilde{y}'_i(t, e'_i(t))) \in \mathcal{S}] + \delta$.

An outcome $o \in O$ of a mechanism contains two parts, both of which can contribute to privacy leakage: (1) The learned regression model $\tilde{\theta}(T)$, which is trained using all data collected after T rounds. Suppose after learning the regression model $\tilde{\theta}(T)$, this information will be released for public usage or monitoring. This information contains each individual worker's private information. Note this is a one-shot leak of privacy (published at the end of the training (step T)). (2) The indexes can reveal private information. Each worker i's data will be utilized towards calculating other workers' indexes $I_j(t), j \neq i$, as well as his own $I_i(t)$, which will be published.[7] Note this type of leakage occurs at each step. The lemma below allows us to focus on the privacy losses in $S_j(t)$, instead of $I_j(t)$, as both $I_j(t)$ and $n_i(t)$ are functions of $\{S_j(n)\}_{n \leq t}$.

Lemma 2. *At any time t, $\forall i$, $n_i(t)$ can be written as a function of $\{S_j(n), n < t\}_j$.*

Preserving privacy in $\tilde{\theta}(T)$ To protect privacy in $\tilde{\theta}(T)$, following standard method [7], we add a Laplacian noise vector $\mathbf{v}_\theta$ to it: $\tilde{\theta}^p(T) = \tilde{\theta}(T) + \mathbf{v}_\theta$, where $\Pr(\mathbf{v}_\theta) \propto \exp(-\varepsilon_\theta ||\mathbf{v}_\theta||_2)$. $\varepsilon_\theta > 0$ is a parameter controlling the noise level.

Lemma 3. *Set $\varepsilon_\theta = 2\sqrt{T}$, the output $\tilde{\theta}^p(T)$ of SR-UCB for linear regression preserves $(\mathrm{O}(T^{-1/2}), \exp(-\mathrm{O}(T)))$-DP. Further w.p. $\geq 1 - 1/T^2$, $||\tilde{\theta}^p(T) - \tilde{\theta}(T)||_2 = ||\mathbf{v}_\theta||_2 \leq \log T/\sqrt{T}$.*

Preserving privacy in $\{I_i(t)\}_{i,t}$: a continual privacy preserving model For indexes $\{I_i(t)\}_i$, it is also tempting to add $v_i(t)$ to each index, i.e. $I_i(t) := I_i(t) + v_i(t)$, where again $v_i(t)$ is a zero-mean Laplacian noise. However releasing $\{I_i(t)\}_i$ at each step will release a noisy version of each $\tilde{y}_i(n, e_i(n)), i \in d(n), \forall n < t$. The composition theory in differential privacy [14] implies that the preserved privacy level will grow in time t, unless we add significant noise on each stage, which will completely destroy the informativeness of our index policy. We borrow the partial sum idea for continual observations [3]. The idea is when releasing continual data, instead of inserting noise at every step, the current to-be-released data will be decoupled into sum of partial sums, and we only add noise to each partial sum and this noisy version of the partial sums can be re-used repeatedly.

We consider adding noise to a modified version of the online indexes $\{S_i^{\text{online}}(t)\}_{i,t}$ as defined in Eqn. (6.1), with $\tilde{\theta}_{-i}^{\text{online}}(t)$ replaced by $\sum_{n=1}^t \tilde{\theta}_{-i}(n)/t$, where $\tilde{\theta}_{-i}(n)$ is the regression model we estimated using all data from worker $j \neq i$ up to time n. For each worker i, his contributed data appear in both $\{S_i^{\text{online}}(t)\}_t$ and $\{S_j^{\text{online}}(t)\}_t, j \neq i$. For $S_j^{\text{online}}(t), j \neq i$, we want to preserve privacy in $\sum_{n=1}^t \tilde{\theta}_{-j}(n)/t$, which contains information of $\tilde{y}_i(n, e_i(n))$.

We first apply the partial sums idea to $\sum_{n=1}^t \tilde{\theta}_{-j}(n)/t$. Write down t as a binary string and find the rightmost digit that is a 1, then flip that digit to 0: convert is back to decimal gives $q(t)$. Take the sum from $q(t)+1$ to t: $\sum_{n=q(t)+1}^t \tilde{\theta}_{-j}(n)$ as one partial sum. Repeat above for $q(t)$, to get $q(q(t))$, and the second partial sum $\sum_{n=q(q(t))+1}^{q(t)} \tilde{\theta}_{-j}(n)$, until we reach $q(\cdot) = 0$. So

$$\sum_{n=1}^t \tilde{\theta}_{-j}(n)/t = \frac{1}{t}\Big(\sum_{n=q(t)+1}^t \tilde{\theta}_{-j}(n) + \sum_{n=q(q(t))+1}^{q(t)} \tilde{\theta}_{-j}(n) + \ldots + \sum_{n=0}^{0} \tilde{\theta}_{-j}(n) \Big) . \tag{7.1}$$

Add noise $\mathbf{v}_{\tilde{\theta}}$ with $\Pr(\mathbf{v}_{\tilde{\theta}}) \propto e^{-\varepsilon||\mathbf{v}_{\tilde{\theta}}||_2}$ to each partial sum. The number of noise terms is bounded by $\leq \lceil \log t \rceil$ at time t. So is the number of appearance of each private data in the partial sums [3]. Denote the noisy version of $\sum_{n=1}^t \tilde{\theta}_{-j}(n)/t$ as $\tilde{\tilde{\theta}}_{-i}^{\text{online}}(n)$. Each $S_i^{\text{online}}(t)$ is computed using $\tilde{\tilde{\theta}}_{-i}^{\text{online}}(n)$.

For $S_i^{\text{online}}(t)$, we also want to preserve privacy in $\tilde{y}_i(n, e_i(n))$. Clearly $S_i^{\text{online}}(t)$ can be written as sum of partial sums of terms involving $\tilde{y}_i(n, e_i(n))$: write $S_i^{\text{online}}(t)$ as a summation: $\sum_{n=1}^{n_i(t)} dS(n)/n_i(t)$ (short-handing $dS(n) := a - b((\tilde{\tilde{\theta}}_{-i}^{\text{online}}(t(n)))^T x_i(t(n)) - \tilde{y}_i(t(n), e_i(t(n))))^2$, where $t(n)$ denotes the time of worker i being sampled the n-th time.). Decouple $S_i^{\text{online}}(t)$ into partial sums using the same technique. For each partial sum, add a noise v_S with distribution $\Pr(v_S) \propto e^{-\varepsilon|v_S|}$.

We then show that with the above two noise exertion procedures, our index policy SR-UCB will not lose its value in incentivizing effort. In order to prove similar convergence results, we need to modify SR-UCB by changing the index to the following format:

$$I_i(t) = \hat{S}_i^{\text{online}}(t) + c(\log^3 t \log^3 T)/\sqrt{n_i(t)},\ \tau(t) = \mathrm{O}\big((\log^3 t \log^3 T)/\sqrt{t}\big) ,$$

[7] It is debatable whether the indexes should be published or not. But revealing decisions on worker selection will also reveal information on the indexes. We consider the more direct scenario – indexes are published.

where $\hat{S}_i^{\text{online}}(t)$ denotes the noisy version of $S_i^{\text{online}}(t)$ with added noises ($v_S, \mathbf{v}_{\tilde{\theta}}$ etc). The change of bias is mainly to incorporate the increased uncertainty level (due to added privacy preserving noise). Denote this mechanism as PSR-UCB, we have:

Theorem 3. *Set* $\varepsilon := 1/\log^3 T$ *for added noises (both* v_S *and* $\mathbf{v}_{\tilde{\theta}}$*), PSR-UCB preserves* $(\mathrm{O}(\log^{-1} T), \mathrm{O}(\log^{-1} T))$*-DP for linear regression.*

With homogeneous workers, we similarly can prove exerting effort $\{e_i^*\}_i$ (optimal effort level) is $\mathrm{O}(\log^6 T/\sqrt{T})$-BNE. We can see that, in order to protect privacy in the bandit setting, the approximation term of BNE is worse than before.

Acknowledgement: We acknowledge the support of NSF grant CCF-1301976.

References

[1] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[2] Yang Cai, Constantinos Daskalakis, and Christos H Papadimitriou. Optimum statistical estimation with strategic data sources. *arXiv preprint arXiv:1408.2539*, 2014.

[3] T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):26, 2011.

[4] Rachel Cummings, Stratis Ioannidis, and Katrina Ligett. Truthful linear regression. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015*, pages 448–483, 2015.

[5] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[6] Cynthia Dwork. Differential privacy. In *Automata, languages and programming*. 2006.

[7] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy.

[8] Arpita Ghosh and Patrick Hummel. Learning and incentives in user-generated content: Multi-armed bandits with endogenous arms. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 233–246. ACM, 2013.

[9] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[10] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In *Proceedings of the fifteenth ACM EC*, pages 359–376. ACM, 2014.

[11] Stratis Ioannidis and Patrick Loiseau. Linear regression as a non-cooperative game. In *Web and Internet Economics*, pages 277–290. Springer, 2013.

[12] Svante Janson. Tail bounds for sums of geometric and exponential variables. 2014.

[13] Radu Jurca and Boi Faltings. Collusion-resistant, incentive-compatible feedback payments. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 200–209. ACM, 2007.

[14] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. *arXiv preprint arXiv:1311.0776*, 2013.

[15] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[16] Guy Lebanon. m-estimators and z-estimators.

[17] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. In *Proceedings of the Sixteenth ACM EC*, pages 565–582. ACM, 2015.

[18] Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

[19] Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. *arXiv preprint arXiv:1109.5647*, 2011.

[20] BLS Prakasa Rao. The rate of convergence of the least squares estimator in a non-linear regression model with dependent errors. *Journal of Multivariate Analysis*, 1984.

[21] Victor S Sheng, Foster Provost, and Panagiotis G Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008.

[22] Panos Toulis, David C. Parkes, Elery Pfeffer, and James Zou. Incentive-Compatible Experimental Design. *Proceedings 16th ACM EC'15*, pages 285–302, 2015.

[23] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

APPENDIX

8 Proofs for Section 5

8.1 Boundedness for indexes

Proof. We prove the indexes have bounded support:

$$\begin{aligned}
&(\tilde{\theta}_{-i}^T(t)x_i(n) - \tilde{y}_i(n,e))^2 \\
&\leq (\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n)) - z_i(e))^2 \\
&\leq 2(\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n)))^2 + 2z^2 \\
&\leq 2(\tilde{\theta}_{-i}^T(t)x_i(n) - \theta^T x_i(n))^2 + 2Z^2 \\
&\leq 2||\tilde{\theta}_{-i}^T(t) - \theta||_2^2 ||x_i(n)||_2^2 + 2Z^2 \\
&\leq 8M^2 + 2Z^2
\end{aligned}$$

□

8.2 Intuitions and some results that are needed for proving Theorem 1

In order to analyze our bandit setting, we need to track the evolution of the indexes, which are mainly affected by the change of the "scoring" term $S_i(t)$. In analogy to classical bandit setting, we are hoping to establish a convergence result for $S_i(t)$. Specifically we prove the following results:

Lemma 4. *Suppose we have n i.i.d. samples to construct $\tilde{\theta}_{-i}(t)$ in $S_i(t)$. Then*

$$|\mathbb{E}[S_i(t)] - (a - b(\sigma_z + \sigma(e_i)))| \leq \mathrm{O}\big(\frac{1}{n^2}\big). \tag{8.1}$$

And w.p. being at least $1 - e^{-Kn}$ for some $K > 0$,

$$|S_i^1(t)| \leq \mathrm{O}\big(\frac{1}{n^2}\big)\ .$$

Proof. To give some intuition, we first decouple the quadratic term $(\tilde{\theta}_{-i}^T(t)x_i(n) - \tilde{y}_i(n,e))^2$ in each $S_i(t)$, for any time t, and any data sample $x_i(n)$ that is collected before t ($n \leq t$):

$$\begin{aligned}
&\left(\tilde{\theta}_{-i}^T(t)x_i(n) - \tilde{y}_i(n,e)\right)^2 = \left(\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n)) + y(x_i(n)) - \tilde{y}_i(n,e)\right)^2 \\
&= (\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n)))^2 + (\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n)))(y(x_i(n)) - \tilde{y}_i(n,e)) + (y(x_i(n)) - \tilde{y}_i(n,e))^2 \\
&= \underbrace{((\tilde{\theta}_{-i}(t) - \theta)^T x_i(n))^2}_{l_{i,t}^1(n)} + \underbrace{(\tilde{\theta}_{-i}(t) - \theta)^T x_i(n) \cdot z(n)}_{l_{i,t}^2(n)} + \underbrace{(\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n)))(y(x_i(n)) - \tilde{y}_i(n,e_i))}_{l_{i,t}^3(n)} \\
&\quad + \underbrace{z^2(n)}_{l_{i,t}^4(n)} + \underbrace{(y(x_i(n)) - \tilde{y}_i(n,e))^2}_{l_{i,t}^5(n)}\ .
\end{aligned} \tag{8.2}$$

With above decoupling we can re-write $S_i(t)$ as

$$S_i(t) := a + \sum_{k=1}^{5} S_i^k(t), \text{ where } S_i^k(t) = -b\frac{\sum_{n=1}^{t-1} 1(i \in d(n)) l_{i,t}^k(n)}{n_i(t)} .$$

We analyze each of the five terms $S_i^k(t), k = 1,2,...,5$. For the first term $l_{i,t}^1(n)$ first notice $\forall n$

$$((\tilde{\theta}_{-i}(t) - \theta)^T x_i(n))^2 \leq ||\tilde{\theta}_{-i}(t) - \theta||_2^2 ||x_i(n)||_2^2 \leq ||\tilde{\theta}_{-i}(t) - \theta||_2^2 .$$

We have the following lemma:

Lemma 5. *Suppose we have n i.i.d. samples to construct $\tilde{\theta}_{-i}(t)$, then w.p. being at least $1 - e^{-Kn}$ where $K > 0$ is a constant,*

$$||\tilde{\theta}_{-i}(t) - \theta||_2^2 \leq Z^4(d+2)^6 \frac{(1+\xi)^2}{(1-\xi)^4} \frac{1}{n^2}, \text{ with } \xi \in (0,1) \text{ being a constant.}$$

Using above lemma we know w.p. being at least $1 - e^{-Kn}$,

$$|S_i^1(t)| \leq b||\tilde{\theta}_{-i}(t) - \theta||_2^2 \leq bZ^4(d+2)^6 \frac{(1+\xi)^2}{(1-\xi)^4} \frac{1}{n^2} .$$

For the second term $l_{i,t}^2(n)$, consider its expectation. Due to independence between any data, and the independence between data and noise z, we have

$$\mathbb{E}[l_{i,t}^2(n))] = \mathbb{E}(\tilde{\theta}_{-i}(t) - \theta)^T x_i(n))\mathbb{E}(z) = 0 .$$

Similarly for $l_{i,t}^3(n)$, since the noise term in $\tilde{\theta}_{-i}^T(t)$ is independent from the one in $y(x_i(n)) - \tilde{y}_i(n,e)$, again we have

$$\mathbb{E}[l_{i,t}^3(n)] = \mathbb{E}[\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n))] \cdot \mathbb{E}[y(x_i(n)) - \tilde{y}_i(n,e_i)] = 0 .$$

The second equality follows as $\mathbb{E}[y(x_i(n)) - \tilde{y}_i(n,e_i)] = 0$. Also we would like to note that due to the boundedness of $(\tilde{\theta}_{-i}(t) - \theta)^T x_i(n)$ and $\tilde{\theta}_{-i}^T(t)x_i(n) - y(x_i(n))$, the convergence of $S_i^2(t), S_i^3(t)$can be established using Hoeffding bound [5].

For $l_{i,t}^4(n), l_{i,t}^5(n)$ we have (suppose worker i exerts consistent effort e_i)

$$\mathbb{E}[l_{i,t}^4(n)] = \mathbb{E}[z^2(n)] = \sigma_z, \ \mathbb{E}[l_{i,t}^5(n)]^2 = \mathbb{E}[z_i(e_i)]^2 = \sigma(e_i) .$$

The convergence rate is depending on how many samples worker i has been assigned. To summarize we know

$$\mathbb{E}[\sum_{k=2}^{5} S_j^k(t)] = -b(\sigma_z + \sigma(e_i)). \tag{8.3}$$

And the expected scoring term S_i for each user will roughly converge to

$$a - b(\sigma_z + \sigma(e_i)) + \mathrm{O}\big(\frac{1}{n^2}\big), , \tag{8.4}$$

with $\mathrm{O}\big(\frac{1}{n^2}\big)$ being an additional bias term, where n is the number of samples contributed by other workers. This also implies that

$$|\mathbb{E}[S_i(t)] - (a - b(\sigma_z + \sigma(e_i)))| \leq \mathrm{O}\big(\frac{1}{n^2}\big).$$

□

With above preparation we see if every worker is exerting the same level of efforts, $e_i \equiv e$, the expected scoring function for workers will become equivalent. Then in order to be selected, workers will *race* with each other on $\sigma(e_i)$[8] and be incentivized to exert efforts.

[8] Or average over $\sigma(e_i(n))$ when different effort levels are chosen at different steps.

8.3 Proof for Lemma 5

Proof. Denote the stacked data in a matrix form as $\mathbf{X} \in \mathbb{R}^{n\times d}$, and the corresponding labeling outcome $\tilde{y} \in \mathbb{R}^n$. Then it is well known that the optimal estimator from minimizing a non-regularized empirical loss function is given by $\tilde{\theta}_{-i}(t) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\tilde{y}$. Denote y as the true labels. Consider the following facts.

$$\begin{aligned}
&||\tilde{\theta}_{-i}(t) - \theta||_2^2 \\
=& ||(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\tilde{y} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y||_2^2 \\
=& \text{trace}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) \\
=& ||(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}||_2^2 \\
\leq& ||(\mathbf{X}^T\mathbf{X})^{-1}||_2^2 \cdot ||\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}||_2^2 \cdot ||(\mathbf{X}^T\mathbf{X})^{-1}||_2^2 \,. \qquad (8.5)
\end{aligned}$$

Since xs are sampled uniformly from a unit ball, by Theorem 7 in [4] (adapted from Corollary 5.52 in [23]), $||(\mathbf{X}^T\mathbf{X})^{-1}||_2^2$ can be bounded at the order of $\mathrm{O}\big(\frac{1}{n^2}\big)$ w.h.p. ($> 1 - \mathrm{O}\big(e^{-Kn}\big)$):

Theorem 4. *Let $\xi \in (0,1)$, and $t \geq 1$. Let $||\cdot||$ denote the spectral norm. If $\{x_i\}_{i=1}^n$ are i.i.d. and sample uniformly from the unit ball (with dimension d), then w.p. being at least $1 - d^{-t^2}$, when $n \geq C(\frac{t}{\xi})^2(d+2)\log d$, for some constant C, then*

$$||\mathbf{X}^T\mathbf{X}|| \leq \frac{1+\xi}{2+d}n,\ ||(\mathbf{X}^T\mathbf{X})^{-1}|| \leq \frac{1}{(1-\xi)\frac{1}{2+d}n}\,. \qquad (8.6)$$

We will be repeatedly using this lemma. Then the first and third term in Eqn. (8.5) can be well bounded: $||(\mathbf{X}^T\mathbf{X})^{-1}||_2^2 \leq \frac{1}{(1-\xi)\frac{1}{2+d}n}$. Consider the second term. For $||\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}||_2^2$, w.h.p.,

$$\begin{aligned}
&||\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}||_2^2 = \text{trace}(\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}) \\
\leq& \max_i z_i^2 \cdot \sum_i x_i^T x_i = \max_i z_i^2 \cdot \text{trace}(\mathbf{X}^T\mathbf{X}) \\
\leq& Z^2 \cdot \text{trace}(\mathbf{X}^T\mathbf{X}) = Z^2|||\mathbf{X}^T\mathbf{X}||_2^2 \\
\leq& Z^2\frac{1+\xi}{2+d}n\,,
\end{aligned}$$

Combining above argument, we establish that w.h.p.,

$$\begin{aligned}
||\tilde{\theta}_{-i}(t) - \theta||_2^2 &\leq (\frac{1}{(1-\xi)\frac{1}{2+d}n})^2 \cdot (Z^2\frac{1+\xi}{2+d}n)^2 \cdot (\frac{1}{(1-\xi)\frac{1}{2+d}n})^2 \\
&= Z^4(d+2)^6\frac{(1+\xi)^2}{(1-\xi)^4}\frac{1}{n^2} \to 0, \text{ as } n \to \infty.
\end{aligned}$$

□

8.4 Proof for Theorem 1

Proof. To prove the theorem, we proceed in the following ways. We first prove the following lemma:

Lemma 6. *If every worker exerts effort level $e_i(t) = e^*, \forall t$, there exists a constant $\delta_U > 0$ such that for any i, j that $i \neq j$ we have probability at least $1 - \mathrm{O}\big(\frac{1}{t^2}\big)$, $n_i(t) \leq (1+\delta_U)n_j(t)$.*

What this lemma is implying is that w.h.p., one worker cannot be selected more than another by a constant fraction. This result is crucial for us to establish the index analysis for bandits – different from classical bandit, due to the lack of ground-truth, the evaluation of each worker's index does

not only depend on the number of samples from worker himself, but also on the ones from other workers. With above results at hand, and using union bound we know w.p. being at least $1-\mathrm{O}\big(\frac{N}{t^2}\big)$,

$$(\frac{N-1}{1+\delta_U}+1)n_i(t) \le \sum_j n_j(t) \le [(1+\delta_U)(N-1)+1]n_i(t).$$

Since $\sum_j n_j(t) \ge t$ (at least one selection at each time) we must have $n_i(t) \ge \frac{t}{(1+\delta_U)(N-1)+1}$. Based on this we can now establish the following lemma.

Lemma 7. *If every worker exerts effort level $e_i(t) = e^*, \forall t$, we have $\mathbb{E}[n_i(t)] \ge t - const.$.*

With this lemma we are most ready to prove the first part of the $\pi-$BNE. First of all, for any worker, there is no reason to deviate to $e > e^*$. This is due to the fact with exerting e^* each worker has already guaranteed nearly T number of selection. Further exerting effort, while will decrease the net payment at each step, will at most bring in $\mathrm{O}\big(\frac{1}{T}\big)$ gain per round (a constant number more selections).

Now we show deviating to $e < e^* - \mathrm{O}\big(\sqrt{\frac{\log T}{T^z}}\big), \forall\, 0 \le z < 1$ will also be non-profitable. For any such z, we can always find a $z' = z+\varsigma < 1,\ \varsigma > 0$ such that $\sqrt{\frac{\log t}{t^{z'}}} < \sqrt{\frac{\log t}{t^z}}$. Denote $\Delta := \mathrm{O}\big(\sqrt{\frac{\log T}{T^z}}\big)$. Therefore we will be having (using convexity and smoothness of σ)

$$\begin{aligned} &\sigma(e-\Delta) \ge \sigma(e) - \sigma'(e)(-\Delta) \ge \sigma(e) + L\Delta \\ &\Rightarrow \sigma(e-\Delta) - \sigma(e) \ge \underline{L}\Delta. \end{aligned} \tag{8.7}$$

This creates $b\underline{L}\Delta$ difference in $\mathbb{E}[\sum_{k=2}^5 S_j^k(t)]$ based on Eqn.(8.3) (exerting e^* and e). Suppose worker i is deviating, then we prove:

Lemma 8. *After $\mathrm{O}\big(T^{z'}\big)$ selections, the number of selection of worker i can be bounded as follows*

$$\mathbb{E}[n_i(T; t \ge T^{z'})] \le \mathrm{O}\big(\frac{\log T}{\Delta^2}\big) = \mathrm{O}\big(T^z\big)\ .$$

Then by deviating the number of selection of worker i is bounded by $\max\{\mathrm{O}\big(T^{z'}\big), \mathrm{O}\big(T^z\big)\} \le \mathrm{O}\big(T^{z'}\big)$. Following which we know the collected reward for worker i is then upper bounded by $(\gamma+\Delta)\cdot\mathrm{O}\big(T^{z'}\big) < \gamma\cdot\mathrm{O}\big(T\big)$, when $\gamma = \Omega(\sqrt{\frac{\log T}{T}})$, and T is large, and z' is selected such that $z < z' < \frac{z+1}{2}$. On the other hand, when the deviation is no more than $\mathrm{O}\big(\sqrt{\frac{\log T}{T}}\big)$, the per round gain is bounded by

$$\underbrace{\mathrm{O}\big(\sqrt{\frac{\log T}{T}}\big)+\gamma}_{\text{after deviation}} - \underbrace{\frac{\gamma\cdot(T-\text{const.})}{T}}_{\text{before deviation}} \le \mathrm{O}\big(\sqrt{\frac{\log T}{T}}\big)\ ,$$

Thus the above argument establishes that a consistent deviation will result in at most $\sqrt{\frac{\log T}{T}}$ more net-payment per task.

We now prove the case when workers may deviate differently at different step. Take worker i as an example, denote its effort level at step t as $e_i(t)$. Denote $\Delta_i(t) = e^* - e_i(t) \ge 0$ as a per-step deviation. First we have by convexity $\sigma(e_i(t)) - \sigma(e^*) \ge \sigma'(e^*)\Delta_i(t)$. Sum over all period of time we have

$$\frac{\sum_{t=1}^T \sigma(e_i(t))}{T} - \sigma(e^*) \ge \sigma'(e^*)\frac{\sum_{t=1}^T \Delta_i(t)}{T}\ . \tag{8.8}$$

If $\frac{\sum_{t=1}^T \Delta_i(t)}{T} \le 0$, we know that the total cost is higher than Te^*. Then

$$\begin{aligned} &\frac{\sum_{t=1}^T 1(i\in d(t))(p_i - e_i(t)) - E[n_i(t,e^*)]\gamma}{T} \\ \le &\frac{\sum_{t=1}^T 1(i\in d(t))(e_i+\gamma-e_i(t)) - (T-\text{const.})\gamma}{T} \\ \le &\frac{\text{const.}}{T}\gamma + \frac{\sum_{t=1}^T 1(i\in d(t))\Delta_i(t)}{T} \le \frac{\text{const.}}{T}\gamma\,. \end{aligned}$$

So the per-round profit is upper bounded by $\frac{\text{const.}}{T}\gamma$ by such a deviation. Now consider the case $\frac{\sum_{t=1}^{T}\Delta_i(t)}{T} > 0$. We then have

$$\frac{\sum_{t=1}^{T}\sigma(e_i(t))}{T} - \sigma(e^*) \geq \underline{L}\frac{\sum_{t=1}^{T}\Delta_i(t)}{T} \ . \tag{8.9}$$

Denote by $\Delta := \underline{L}\frac{\sum_{t=1}^{T}\Delta_i(t)}{T} > 0$. Denote by $t'-1$ the last time such that

$$\underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'-1} < \Delta/2.$$

If there does not exist such a t', that is for all t' $\underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'-1} \geq \Delta/2$, we simply set $t'=1$. Then starting from t', we have

$$\underline{L}\frac{\sum_{t=1}^{t'}\Delta_i(t)}{t'} \geq \Delta/2.$$

When $\Delta = \mathrm{O}\big(\sqrt{\log T/T^z}\big),\ z \to 1$, we discuss in three cases.

- **Case 1:** When $t' \geq T - \mathrm{O}\big(\sqrt{T}\big)$. We must have $\underline{L}\frac{\sum_{t=1}^{t'}\Delta_i(t)}{t'} \leq 2\Delta/3$, as otherwise
$$\underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'-1} \geq \underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'} \geq \underline{L}\frac{\sum_{t=1}^{t'}\Delta_i(t)-\bar{e}}{t'} \geq 2\Delta/3 - \frac{\bar{e}}{t'} \geq \Delta/2,$$
which contradicts the definition of t'. Then for this case, the average utility gain is upper bounded by the following case (being selected for all the rest of $\mathrm{O}\big(\sqrt{T}\big)$ steps): $2\Delta/3 + \mathrm{O}\big(\sqrt{T}/T\big)$. So this establishes the $\mathrm{O}\big(\sqrt{\log T/T}\big)$-BNE.

- **Case 2:** For the second case that $t' = o(T)$, specifically say $t' = \mathrm{O}\big(T^z\big), 0 < z < 1$. We can prove a result that is similar to Lemma 8 stating that
$$\mathbb{E}[n_i(T; t \geq T^{z'})] \leq \mathrm{O}\big(\frac{\log T}{\Delta^2}\big) \ .$$
All previous analysis establishes themselves directly except for the convergence of the fifth term $S_i^5(t)$, as now it consists of non-identical noise terms. Nevertheless using Hoeffding bound, we can establish the convergence of the sum of sequence of non-identical but independent samples $S_i^5(t) \to \frac{\sum_{t=1}^{T}\sigma(e_i(t))}{T}$. If $\frac{\sum_{t=1}^{T}\Delta_i(t)}{T} = \Omega(\sqrt{\frac{\log T}{T^z}})$, we will again have
$$\frac{\sum_{t=1}^{T}\sigma(e_i(t))}{T} - \sigma(e^*) = \Omega(\sqrt{\frac{\log T}{T^z}}) \ , \tag{8.10}$$
from which we can prove a contradiction on profitable deviations, via similarly proving the bound on the number of selection (Lemma (8)), i.e., by deviating the number of selection of worker i is bounded by $\max\{\mathrm{O}\big(T^{z'}\big), \mathrm{O}\big(T^z\big)\} = \mathrm{O}\big(T^{z'}\big)$; and the rest analysis follows.

- **Case 3:** For the third case that $\mathrm{O}\big(T^z\big) \leq t' \leq T - \mathrm{O}\big(\sqrt{T}\big)$. Again we must have $\underline{L}\frac{\sum_{t=1}^{t'}\Delta_i(t)}{t'} \leq 2\Delta/3$, as otherwise
$$\underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'-1} \geq \underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'} \geq \underline{L}\frac{\sum_{t=1}^{t'}\Delta_i(t)-\bar{e}}{t'} \geq 2\Delta/3 - \mathrm{O}\big(1/T^z\big) \geq \Delta/2,$$
as $\mathrm{O}\big(1/T^z\big) \leq \sqrt{\log T/T^z}$. Then we can repeat the argument for **Case 2**, but with a deviation analysis on the interval of $[\mathrm{O}\big(T^z\big), T]$, with the starting time being $\mathrm{O}\big(T^z\big)$ or larger. Then similar to the case with $t' = o(T)$ (as now $\mathrm{O}\big(T^z\big)$ is as if $t'=1$), we can prove that $\mathbb{E}[n_i(T; t \geq T^{z'})] \leq \mathrm{O}\big(\frac{\log T}{\Delta^2}\big)$. Yet the average gain per step before t' is bounded by $2\Delta/3 = \mathrm{O}\big(\sqrt{\frac{\log T}{T^z}}\big)$.

When $\Delta > \sqrt{\frac{\log T}{T}}$, we will take t' as the last time that $\underline{L}\frac{\sum_{t=1}^{t'-1}\Delta_i(t)}{t'-1} < \Delta/4$ instead. This argument repeats by above halfing procedure until Δ reduces to the order of $\sqrt{\frac{\log T}{T}}$, and t' will remain $o(T)$. Then the above argument can be applied. Combine all above we proved the theorem. □

8.5 Proof of Lemma 6

Proof. We follow the notations and definitions in Section 8.2 and Lemma 4 therein (S_i^k etc). Suppose at a certain time t we have $n_i(t) = (1+\delta_U)n_j(t), \delta_U > 0$. We would like to bound the following probability $\Pr[I_i(t) \geq I_j(t)]$. This is equivalent with proving the following:

$$\begin{aligned}\Pr[I_i(t) \geq I_j(t)] &= \Pr\left[S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} \geq S_j(t) + c\sqrt{\frac{\log t}{n_i(t)}}\right] \\ &= \Pr\left[S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} \geq S_j(t) + c\sqrt{\frac{(1+\delta_U)\log t}{n_j(t)}}\right] \\ &= \Pr\left[S_i(t) - S_j(t) \geq (\sqrt{1+\delta_U} - 1)c\sqrt{\frac{\log t}{n_i(t)}}\right].\end{aligned}$$

Using Lemma 5, and denote $C_1 := bZ^4(d+2)^6\frac{(1+\xi)^2}{(1-\xi)^4}$. Then we know with probability at least $1 - e^{-K\sum_{k\neq i} n_k(t)}$, and $1 - e^{-K\sum_{k\neq j} n_k(t)}$ (with $K > 0$ being a constant) respectively (when worker i, j exert effort levels e_i, e_j respectively),

$$|S_i^1(t)| \leq \frac{C_1}{(\sum_{k\neq i} n_k(t))^2}, \;\; |S_j^1(t)| \leq \frac{C_1}{(\sum_{k\neq j} n_k(t))^2},$$

For $\sum_{k\neq i} n_k(t)$ we discuss two cases. For the first case, if there exists a constant ν such that $n_i(t) \leq (1-\nu)t$, then $\sum_{k\neq i} n_k(t) \geq \nu t$. Otherwise if $n_i(t) > (1-\nu)t$ we will also have

$$\sum_{k\neq i} n_k(t) \geq n_j(t) \geq \frac{n_i(t)}{1+\delta_U} \geq \frac{1-\nu}{1+\delta_U}t$$

so to summarize

$$\sum_{k\neq i} n_k(t) \geq \min\{\nu, \frac{1-\nu}{1+\delta_U}\}t \; .$$

Similarly we can prove that

$$\sum_{k\neq j} n_k(t) \geq \min\{\nu, (1-\nu)(1+\delta_U)\}t \; .$$

Denote as $C_2 = \min\{\nu, \frac{1-\nu}{1+\delta_U}, (1-\nu)(1+\delta_U)\}$. We will have with probability at least $1 - e^{-KC_2 t}$

$$\max\{|S_i^1(t)|, |S_j^1(t)|\} \leq \frac{C_1}{C_2^2 t^2}.$$

Then

$$\begin{aligned}&\Pr\left[S_i(t) - S_j(t) \geq (\sqrt{1+\delta_U} - 1)c\sqrt{\frac{\log t}{n_i(t)}}\right] \\ \leq &\Pr\left[\sum_{k=2}^5 S_i^k(t) - \sum_{k=2}^5 S_j^k(t) \geq (\sqrt{1+\delta_U} - 1)c\sqrt{\frac{\log t}{n_i(t)}} - \frac{2C_1}{C_2^2 t^2}\right].\end{aligned}$$

Since $\mathbb{E}[\sum_{k=2}^5 S_i^k(t)] = \mathbb{E}[\sum_{k=2}^5 S_j^k(t)]$ (at equilibria, and worker i is also exerting the same amount of effort), using union bound, the above implies that

$$\begin{aligned}&\Pr\left[\sum_{k=2}^5 S_i^k(t) - \sum_{k=2}^5 S_i^j(t) \geq (\sqrt{1+\delta_U} - 1)c\sqrt{\frac{\log t}{n_i(t)}} - \frac{2C_1}{C_2^2 t^2}\right] \\ \leq &\Pr\left[\sum_{k=2}^5 S_i^k(t) - \mathbb{E}[\sum_{k=2}^5 S_i^k(t)] \geq \frac{\sqrt{1+\delta_U} - 1}{2}c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{C_2^2 t^2}\right] \\ + &\Pr\left[\sum_{k=2}^5 S_j^k(t) - \mathbb{E}[\sum_{k=2}^5 S_j^k(t)] \leq \frac{\sqrt{1+\delta_U} - 1}{2}c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{C_2^2 t^2}\right].\end{aligned} \tag{8.11}$$

We bound each of above two terms. (Due to symmetry we only show the bound for one of them.) For worker i, via union bound:

$$\Pr\left[\sum_{k=2}^{5} S_i^k(t) - \mathbb{E}[\sum_{k=2}^{5} S_i^k(t)] \geq \frac{\sqrt{1+\delta_U}-1}{2} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{C_2^2 t^2}\right]$$
$$\leq \sum_{k=2}^{5} \Pr\left[S_i^k(t) - \mathbb{E}[S_i^k(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t^2}\right].$$

Since $n_i(t) \leq t$ we know when t is large $\sqrt{\frac{\log t}{n_i(t)}} \geq \sqrt{\frac{\log t}{t}} \geq \frac{1}{t^2}$. So when c is large enough, e.g. $\frac{\sqrt{1+\delta_U}-1}{8} c > \frac{C_1}{4C_2}$, we will be having

$$\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t^2} > 0 .$$

$S_i^2(t), S_i^3(t)$ can be bounded similarly, while $S_i^4(t), S_i^5(t)$ share similar concentration bound. W.l.o.g., we show the derivation for one of each pair. For $S_i^2(t)$, first of all notice

$$\begin{aligned}
|b \cdot l_{i,t}^2(n)| &= |b(\tilde{\theta}_{-i}(t) - \theta)^T x_i(n) \cdot z(n)| \leq b||\tilde{\theta}_{-i}(t) - \theta||_2 ||x_i(n)||_2 |z(n)| \\
&\leq b(||\tilde{\theta}_{-i}(t)||_2 + ||\theta||_2) Z \leq 2MZ \\
&\Rightarrow -2bMZ \leq l_{i,t}^2(n) \leq 2bMZ .
\end{aligned}$$

Then via Hoeffding inequality we know

$$\begin{aligned}
&\Pr\left[S_i^2(t) - E[S_i^2(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t^2}\right] \\
\leq &\exp\left(-\frac{2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t^2})^2 n_i^2(t)}{16b^2M^2Z^2 n_i(t)}\right) \\
\leq &\exp\left(-\frac{2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t^2})^2 n_i(t)}{16b^2M^2Z^2}\right) \\
\leq &\exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1}{4C_2^2 t^2}\sqrt{\log t \cdot n_i(t)}}{16b^2M^2Z^2}\right) \\
\leq &\exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1}{4C_2^2 t}}{16b^2M^2Z^2}\right) \\
\leq &\frac{1}{t^2} \cdot \exp(2/t) \leq \frac{2}{t^2} ,
\end{aligned}$$

when δ_U and c are selected to be large enough, and t large enough: for example

$$\frac{\sqrt{1+\delta_U}-1}{8} c \geq 4bMZ \cdot \max\{1, \frac{4C_2^2}{C_1}\}, \text{ and } t \geq 4.$$

Similarly we can bound $S_i^3(t)$. Now consider $S_i^4(t)$. We use Hoeffding bound via first observing the boundedness of each term $bl_{i,t}^4(n) = |bz^2(n)| \Rightarrow 0 \le b \cdot l_{i,t}^4(n) \le bZ^2$. Then

$$\begin{aligned}
&\Pr\Big[S_i^4(t) - \mathbb{E}[S_i^4(t)] \ge \frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t^2}\Big] \\
\le &\exp(-\frac{2((\frac{\sqrt{1+\delta_U}-1}{8})c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C}{4C_2^2t^2})n_i^2(t)}{b^2Z^4 n_i(t)}) \\
\le &\exp(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2c^2}{b^2Z^4}\log t)\cdot\exp(2\frac{\frac{\sqrt{1+\delta_U}-1}{8}c\cdot\frac{C_1}{4C_2^2t^2}\sqrt{\log t\cdot n_i(t)}}{b^2Z^4}) \\
\le &\exp(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2c^2}{b^2Z^4}\log t)\cdot\exp(2\frac{\frac{\sqrt{1+\delta_U}-1}{8}c\cdot\frac{C_1}{4C_2^2t}}{b^2Z^4}) \\
\le &\frac{1}{t^2}\cdot e^{2/t} \le \frac{2}{t^2}\ ,
\end{aligned}$$

again when δ_U and c are selected to be large enough, and t large enough: for example

$$\frac{\sqrt{1+\delta_U}-1}{8}c \ge bZ^2\cdot\max\{1,\frac{4C_2^2}{C_1}\},\ \text{and}\ t\ge 4.$$

Similarly we can bound the term invoking $S_i^4(t)$. Also similarly we can bound

$$\Pr\Big[S_j(t) - \mathbb{E}[S_j(t)] \le (\frac{\sqrt{1+\delta_U}-1}{2})c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C}{C_2^2t^2}\Big] \le \mathrm{O}\big(\frac{1}{t^2}\big)\ .$$

And in all summarize we proved $\Pr[I_i(t)\ge I_j(t)] \le \mathrm{O}\big(\frac{1}{t^2}\big)$.

Now at time t, if $n_i(t) > (1+\delta_U)n_j(t)$, we must have a time point t' that $n_i(t')$ changes from $\le (1+\delta_U)n_j(t')$ to $> (1+\delta_U)n_j(t')$, where we must have $n_i(t') \ge (1+\delta_U)n_j(t') - 1 \ge (1+\delta_U-1)n_j(t')$. Choose δ_U large enough so $\delta_U - 1$ also satisfies the above claim that $\Pr[I_i(t)\ge I_j(t)] \le \mathrm{O}\big(\frac{1}{t^2}\big)$.. We know at time t', it must be i is selected but not j, otherwise the ratio between them can only go down (both being selected will not increase a >1 ratio), i.e., it must be $I_i(t') \ge I_j(t')$. We discuss two cases. When $t' \in [t/2,t]$, we know this is upper bounded by $\mathrm{O}\big(\frac{1}{(t/2)^2}\big) = \mathrm{O}\big(\frac{1}{t^2}\big)$.

If not, consider the worker who has been selected most of the times between $[t/2,t]$. Denote it as k. Then we must have $n_k(t) \ge \frac{t}{N2}$. If $n_k(t) \le (1+\delta_U)n_j(t)$, we will have $n_i(t) \ge \frac{t}{2\delta_U N}$, so $n_i(t)/n_j(t) \le \frac{t/2}{\frac{t}{2\delta_U N}} = \delta_U\cdot N$. Otherwise if $n_k(t) > (1+\delta_U)n_j(t)$. We must have there exists a t' such that $t' \ge t/2$ and

$$n_k(t') \ge (1+\delta_U)n_j(t') - 1 \ge (1+\delta_U-1)n_j(t'),$$

and such that k is selected but not j. However we know the probability for this event is also upper bounded by $\mathrm{O}\big(1/t^2\big)$. Reset $\delta_U := \delta_U\cdot N$ we finished the proof. □

8.6 Proof for Lemma 7

Proof. Following Lemma 6 we know w.h.p. ($\ge 1-\mathrm{O}\big(\frac{1}{t^2}\big)$)

$$n_i(t) \ge \frac{t}{(1+\delta_U)(N-1)+1}.$$

Following proof for Lemma 6 we know that w.h.p. ($\ge 1-\mathrm{O}\big(\frac{1}{t^2}\big)$),

$$|S_i(t) - \mathbb{E}[S_i(t)]| \le \frac{\sqrt{1+\delta_U}-1}{2}c\sqrt{\frac{\log t}{n_i(t)}} + \frac{2C_1}{C_2^2t^2}.$$

Plug in $n_i(t) \geq \frac{t}{(1+\delta_U)(N-1)+1}$ we have

$$|I_i(t) - (a - b(\sigma_z + \sigma(e^*)))| = |S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} - (a - b(\sigma_z + \sigma(e^*)))|$$

$$\leq |S_i(t) - \mathbb{E}[S_i(t)]| + |\mathbb{E}[S_i(t)] - (a - b(\sigma_z + \sigma(e^*)))| + c\sqrt{\frac{\log t}{n_i(t)}}$$

$$\leq \frac{\sqrt{1+\delta_U}+1}{2} c\sqrt{(1+\delta_U)(N-1)+1}\sqrt{\frac{\log t}{t}} + \frac{3C_1}{C_2^2 t^2}$$

$$\leq \frac{\sqrt{1+\delta_U}+1}{2} c\sqrt{(1+\delta_U)N}\sqrt{\frac{\log t}{t}} + \frac{3C_1}{C_2^2 t^2} \ .$$

Then if we set $\tau(t)$ to be two times of above bound:

$$\tau(t) := 2\left(\frac{\sqrt{1+\delta_U}+1}{2} c\sqrt{(1+\delta_U)N}\sqrt{\frac{\log t}{t}} + \frac{3C_1}{C_2^2 t^2}\right)$$

$$= (\sqrt{1+\delta_U}+1) c\sqrt{(1+\delta_U)N}\sqrt{\frac{\log t}{t}} + \frac{6C_1}{C_2^2 t^2}.$$

we will have

$$\Pr\left[I_j(t) \geq \max_i I_i(t) - \tau(t)\right] \leq \mathrm{O}\big(\frac{1}{t^2}\big), \text{ i.e., } \Pr[j \in d(t)] \geq 1 - \mathrm{O}\big(\frac{1}{t^2}\big), \ \forall j,t \ .$$

Therefore we know

$$\mathbb{E}[n_i(T)] = \mathbb{E}[\sum_{n=1}^{T} 1(i \in d(n))] = \sum_{n=1}^{T} \Pr[i \in d(n)] \geq T - \mathrm{O}\big(\sum_{n=1}^{T} \frac{1}{n^2}\big) \geq T - \text{const.} \ .$$

□

8.7 Proof for Lemma 8

Proof. To bound the number of selections of worker i we need to bound $\Pr[I_i(t) \geq \max_j I_j(t) - \tau(t)]$. We further bound this term by the following term $\forall j \neq i, j \in \{1,2\}$(top 2 workers):

$$\Pr[I_i(t) \geq I_j(t) - \tau(t)] = \Pr\left[S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} \geq S_j(t) + c\sqrt{\frac{\log t}{n_j(t)}} - \tau(t)\right].$$

Notice the event $\{S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} \geq S_j(t) + c\sqrt{\frac{\log t}{n_j(t)}} - \tau(t)\}$ implies at least one of the following should hold

$$\sum_{k=2}^{5} S_i^k(t) - \mathbb{E}[\sum_{k=2}^{5} S_i^k(t)] \geq c\sqrt{\frac{\log t}{n_i(t)}}, \quad \sum_{k=2}^{5} S_j^k(t) - \mathbb{E}[\sum_{k=2}^{5} S_j^k(t)] \leq -c\sqrt{\frac{\log t}{n_i(t)}}$$

$$b\underline{L}\Delta \leq 2c\sqrt{\frac{\log t}{n_i(t)}} + \tau(t) + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2} + \frac{C_1}{(\sum_{k\neq j} n_k(t))^2} \ .$$

As otherwise we will have

$$S_j(t) + c\sqrt{\frac{\log t}{n_j(t)}} - \tau(t) > \sum_{k=2}^{5} S_j^k(t) + c\sqrt{\frac{\log t}{n_j(t)}} - \tau(t) - \frac{C_1}{(\sum_{k\neq j} n_k(t))^2}$$

$$> \mathbb{E}[\sum_{k=2}^{5} S_j^k(t)] - \tau(t) - \frac{C_1}{(\sum_{k\neq j} n_k(t))^2} \geq \mathbb{E}[\sum_{k=2}^{5} S_i^k(t)] + b\overline{L}\Delta - \tau(t) - \frac{C_1}{(\sum_{k\neq j} n_k(t))^2}$$

$$> \mathbb{E}[\sum_{k=2}^{5} S_i^k(t)] + 2c\sqrt{\frac{\log t}{n_i(t)}} + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2} > \sum_{k=2}^{5} S_i^k(t) - c\sqrt{\frac{\log t}{n_i(t)}} + 2c\sqrt{\frac{\log t}{n_i(t)}} + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2}$$

$$= \sum_{k=2}^{5} S_i^k(t) + c\sqrt{\frac{\log t}{n_i(t)}} + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2} \geq S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} \ ,$$

which is a contradiction. Similarly as in the proof for Lemma 6 we can prove

$$\Pr\Big[|\sum_{k=2}^{5} S_i^k(t) - \mathbb{E}[\sum_{k=2}^{5} S_i^k(t)]| \geq c\sqrt{\frac{\log t}{n_i(t)}}\Big] = \mathrm{O}\big(\frac{1}{t^4}\big),$$

Then when $n_i(t) \geq \mathrm{O}\big(T^{z'}\big)$, we will be having

$$\sum_{k\neq j} n_k(t) \geq n_i(t) \geq \mathrm{O}\big(T^{z'}\big),$$

also repeating argument in the proof for Lemma 6 to establish that $n_i(t) \leq (1+\delta_U) n_j(t), j \neq i$ (intuitively, S_i converges to a smaller quantity, due to lack of effort. So this side of inequality holds; particularly Eqn.(8.11) holds. We omit the details). So

$$\sum_{k\neq i} n_k(t) \geq n_j(t) \geq \mathrm{O}\big(T^{z'}\big),$$

Thus w.p. at least $1 - e^{K\,\mathrm{O}\left(T^{z'}\right)} \geq 1 - \mathrm{O}\big(\frac{1}{t^4}\big)$ (as $T^{z'} \geq \log T \geq \log t$ when T, z' are large) ,

$$2c\sqrt{\frac{\log t}{n_i(t)}} + \tau(t) + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2} + \frac{C_1}{(\sum_{k\neq j} n_k(t))^2} \leq 2c\sqrt{\frac{\log t}{n_i(t)}} + \tau(t) + \mathrm{O}\big(\frac{1}{T^{2z'}}\big)\ .$$

Note this is a much smaller quantity compared with $\mathrm{O}\big(\sqrt{\frac{\log T}{T^z}}\big)$ (since $z' > z$, and this is the amount of deviation). When t, T are larger than certain constants such that $\tau(t) + \mathrm{O}\big(\frac{1}{T^{2z'}}\big) < \frac{bL\Delta}{2}$, and when $n_i(t) \geq \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2}$:

$$2c\sqrt{\frac{\log t}{n_i(t)}} + \tau(t) + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2} + \frac{C_1}{(\sum_{k\neq j} n_k(t))^2} < bL\Delta\ .$$

Combined above we know

$$\Pr[I_i(t) \geq \max_j I_j(t) - \tau(t), n_i(t) \geq \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2}] = \mathrm{O}\big(\frac{1}{t^4}\big)\ .$$

That is after $n_i(t) \geq \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2}$ number of selections, worker i will not be selected, except for the $\mathrm{O}\big(\frac{1}{t^4}\big)$ fraction of probability. Then following the classical method detailed in [1] for UCB1 (the three way arguments), we know the expected number of selection $\mathbb{E}[n_i(T)]$ bounds as follows: for some $\zeta > 0$:

$$\begin{aligned}
n_i(t) &\leq \zeta + \sum_{s=\zeta+1}^{t} 1\Big(S_i(t) + c\sqrt{\frac{\log t}{n_i(t)}} \geq S_j(t) + c\sqrt{\frac{\log t}{n_j(t)}} - \tau(t)\Big) \\
&\leq \zeta + \sum_{s=\zeta+1}^{t} 1\Big(\min_{0<n^*<s} S_j(n^*) + c\sqrt{\frac{\log s}{n^*}} - \tau(n^*) \leq \max_{\zeta<n<s} S_i(n) + c\sqrt{\frac{\log s}{n}}, j = 1 \text{ or } 2\Big) \\
&\leq \zeta + \sum_{j\in\{1,2\}} \sum_{s=1}^{\infty} \sum_{n^*=1}^{s-1} \sum_{n=\zeta}^{s-1} 1\Big(S_j(n^*) + c\sqrt{\frac{\log s}{n^*}} - \tau(s) \leq S_i(n) + c\sqrt{\frac{\log s}{n}}, j = 1 \text{ or } 2\Big)\ .
\end{aligned}$$

Take expectation and set $\zeta = \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2}$ we know

$$\begin{aligned}
\mathbb{E}[n_i(T)] &\leq \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2} + \sum_{j\in\{1,2\}} \sum_{s=1}^{\infty} \sum_{n^*=1}^{s-1} \sum_{n=\zeta}^{s-1} \mathrm{O}\big(\frac{1}{s^4}\big) \\
&\leq \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2} + \sum_{j\in\{1,2\}} \sum_{s=1}^{\infty} \mathrm{O}\big(\frac{1}{s^2}\big) \\
&\leq \frac{(2c)^2 \log t}{(\frac{bL\Delta}{2})^2} + \text{const.} = \mathrm{O}\big(\frac{\log T}{\Delta^2}\big)\ .
\end{aligned}$$

□

8.8 Proof for Theorem 2

Proof. We first prove that regardless of workers' decision on efforts exertion we will be having:

Lemma 9. *Under SR-UCB for linear least square, we have when t is large* $n_i(t) = \Omega(\log t), a.s.$

Note the classical bandit argument cannot be applied directly to establish a $\mathrm{O}(\log t)$ lower bound since the underlying distribution for the index terms can be different for different arms, as now $S_i(t)$ depends not only on each worker's parameter e_i, but will also depend on other workers e_i and their labeled data. With the help of this lemma we have the following results:

Lemma 10. *At any time t, the number of selection of workers* $i > 2$ *with* $e_i(t) \leq e_1^* + \gamma, \forall t$ *satisfies* $\mathbb{E}[n_i(t)] = \mathrm{O}(\frac{\log t}{\Delta^2})$ *. And moreover if* $e_1(t) \equiv e_1^*, e_2(t) \equiv e_2^*$*, we will be having* $\mathbb{E}[n_1(t)], \mathbb{E}[n_2(t)] = T - \mathrm{O}(\log T)$ *.*

Also since $\sigma_1(e_1^*) = \sigma_2(e_2^*)$, following previous argument for Lemma 6 we can similarly establish that there exists a constant $\delta_U > 0$ s.t. with probability at least $1 - \mathrm{O}(\frac{1}{t^2})$, $\frac{1}{1+\delta_U} n_2(t) \leq n_1(t) \leq (1+\delta_U) n_2(t)$, following which we know $\mathbb{E}[n_1(t)], \mathbb{E}[n_2(t)] \geq T - \mathrm{O}(\log T)$. Therefore further deviating to $e_{1(2)} > e_{1(2)}^*$ will give the corresponding worker at most $O(\frac{\log T}{T}) < \mathrm{O}(\sqrt{\frac{\log T}{T}})$ additional profit per task. For deviation to $e_i < e_i^*, i = 1, 2$, similar to the symmetric case we can again show such a deviation can bring in at most $\mathrm{O}(\sqrt{\frac{\log T}{T}})$ additional payment: what we need to establish is similar to Lemma 8 that

$$\mathbb{E}[n_2(T; t \geq T^z)] \leq \mathrm{O}(\frac{\log T}{\Delta^2}) = \mathrm{O}(T^z) \ .$$

With above we establish the fact that exerting efforts e_1^*, e_2^* is $\mathrm{O}(\sqrt{\frac{\log T}{T}})-$BNE for worker 1 & 2.

For worker $i > 2$, since we already proved that for any effort level $e_i \leq e_1^* + \gamma$, the expected number of selection is bounded up by $\mathrm{O}(\log T)$, as γ is set to be small enough such that $\overline{L}\gamma \leq \frac{\Delta_1}{2}$, and we will then be having $\sigma_i(e_i) - \sigma_1(e_1^*) \geq \frac{\Delta_1}{2}$. Therefore any profitable deviation will lead to at most $\mathrm{O}(\frac{\log T}{T}) < \mathrm{O}(\sqrt{\frac{\log T}{T}})$ additional profit per task. Apparently deviating to $e_i > e_1^* + \gamma$ is not profitable at all (negative marginal gain).

Again consider the dynamic case, where workers can choose to exert different level of efforts at each different steps. When exerting efforts to reach the same effort level as worker 1 & 2 $\frac{\sum_{t=1}^T \sigma_i(e_i(t))}{T} = \sigma_1(e_1^*)$, [9] suppose $\sigma_1(e_1^*) = \sigma_i(e_i^*)$ and we know $e_i^* > e_1^*$; and further

$$\sigma_i(e_i^*) - \sigma_i(e_1^*) \leq \overline{L}(e_i^* - e_1^*) \Rightarrow e_i^* \geq \frac{\Delta}{\overline{L}} + e_1^*.$$

Also we have (by convexity)

$$\sigma_i(e_i^*) = \frac{\sum_{t=1}^T \sigma_i(e_i(t))}{T} \geq \sigma_i(\frac{\sum_{t=1}^T e_i(t)}{T}) \Rightarrow \frac{\sum_{t=1}^T e_i(t)}{T} \geq e_i^* \geq \frac{\Delta}{\overline{L}} + e_1^* \geq e_1^* + \gamma$$

However the average payment is only $e_1^* + \gamma$, we know such a deviation is not profitable for workers $i > 2$. □

8.9 Proof for Lemma 9

Proof. Suppose there is i, such that $n_i(t) = o(\log t)$. The basic intuition of a contradiction is as follows: denote the worker with maximum number of selection as $j \neq i$, and we know $n_j(t) = \mathrm{O}(t)$. Then we will have

$$I_i(t) \geq a - 4M^2 b + c\sqrt{\frac{\log t}{n_i(t)}} > a + c\sqrt{\frac{\log t}{\mathrm{O}(t)}} - \tau(t) \geq I_j(t) - \tau(t) \ , \tag{8.12}$$

[9] This is really a relaxed argument, in fact we need to prove for a $o(1)$-close to $\sigma_1(e_1^*)$.

thus i will be selected when t is large. More rigorously consider t' as the earliest time such that $n_j(t') \geq t^z$. We know $t^z \leq t' \leq t - t^z$, where $0 < z < 1$ is a constant, and the second inequality comes as otherwise $n_i(t) \leq n_j(t') + t^z \leq t^z + t^z + 1 < \mathrm{O}(t)$. Then

$$\begin{aligned} n_i(t) &\geq \sum_{n=t'}^{t} 1(j \in d(n)) \cdot 1\left(c\sqrt{\frac{\log n}{n_i(n)}} \geq 4M^2 b + c\sqrt{\frac{\log n}{n_j(n)}}\right) \\ &\geq \sum_{n=t'}^{t} 1(j \in d(n)) \cdot 1\left(c\sqrt{\frac{z\log t}{n_i(t)}} \geq 4M^2 b + c\sqrt{\frac{\log t}{t^z}}\right), \end{aligned}$$

where the second inequality comes from the facts that

$$\frac{\log n}{n_i(n)} \geq \frac{\log t'}{n_i(t)},\ \frac{\log n}{n_j(n)} \leq \frac{\log t}{n_j(t')} \leq \frac{\log t}{t^z}\ .$$

If $n_i(t) = o(\log t)$, when t is large

$$1\left(c\sqrt{\frac{z\log t}{n_i(t)}} \geq 4M^2 b + c\sqrt{\frac{\log t}{t^z}}\right) = 1 \Rightarrow n_i(t) \geq \sum_{n=t'}^{t} 1(j \in d(n)) = \mathrm{O}(t),$$

which is a contradiction. □

8.10 Proof for Lemma 10

Proof. With an appropriately selected γ, for $i > 2$, worker i will only be willing to pay $e_i \leq e_1^* + \gamma$, as otherwise no matter how many times they got selected, they always receive negative payment. Therefore we will be having $\sigma_i(e_i) \geq \sigma_i(e_1^* + \gamma) \geq \sigma_i(e_1^*) - \bar{L}\gamma$. If we make γ small enough such that $\bar{L}\gamma \leq \frac{\Delta_1}{2}$, we will then be having $\sigma_i(e_i) - \sigma_1(e_1^*) \geq \frac{\Delta_1}{2}$, which leads to $b\frac{\Delta_1}{2}$ difference in the expected value of index. With this, following the proof for Lemma (8), we upper bound the number of selections on the order of $\mathrm{O}(\frac{\log T}{\Delta_1^2})$: the only difference is in bounding the following event:

$$\{b\frac{\Delta_1}{2} \leq 2c\sqrt{\frac{\log t}{n_i(t)}} + \tau(t) + \frac{C_1}{(\sum_{k\neq i} n_k(t))^2} + \frac{C_1}{(\sum_{k\neq j} n_k(t))^2})\}\ .$$

By Lemma 9, we claim, a.s.,

$$(\sum_{k\neq i} n_k(t))^2 \geq \mathrm{O}\big((\log t)^2\big),\ (\sum_{k\neq j} n_k(t))^2 \geq \mathrm{O}\big((\log t)^2\big),$$

and thus since Δ is a positive constant, we know when t is large enough, there exists a $\Delta' = \alpha\Delta_1$ where $0 < \alpha < 1$ such that

$$\begin{aligned} \{b\frac{\Delta_1}{2} \leq& 2c\sqrt{\frac{\log t}{n_i(t)}} - \tau(t) - \mathrm{O}\big(\frac{1}{(\sum_{k\neq i} n_k(t))^2}\big) - \mathrm{O}\big(\frac{1}{(\sum_{k\neq j} n_k(t))^2}\big)\} \\ &\subseteq \{b\frac{\Delta'}{2} \leq 2c\sqrt{\frac{\log t}{n_i(t)}}\}, \end{aligned}$$

from where we can follow the reasoning for Lemma (8) to finish the proof. □

8.11 Removing bad equilibira

Proof. When following the equilibria e^*, the average utility for each worker is $p_s\gamma$. With adding this independent randomization device, most of the key parts of the proof will go through. For example, the proof of Lemma 6 and 7 will go through directly, except for the small change that, the number of selections up to time t is now lower bounded (instead of being lower bounded by t) by the following random variable that satisfies: denote the event of a selection as $s(t) \in \{1(\text{selection}), 0(\text{no selection})\}$

$$\Pr\left[\frac{\sum_n s(n)}{t} \geq p_s - \sqrt{\frac{\log t}{t}}\right] \leq \exp(-2\frac{\log t}{t} \cdot t) = 1/t^2.$$

Then based on Lemma 7, we know with $\mathrm{O}(T)$ number of times, the worker will be jointly selected with others. This finishes the $p_s \cdot \gamma$ argument.

Now if worker i deviates, his utility will be upper bounded by the following case (1) he becomes the monopoly for $\mathrm{O}(T)$ number of times. (2) His marginal gain is upper bounded by $\gamma + \mathrm{O}(\sqrt{\frac{\log T}{T}})$. (as otherwise if the deviation of effort is higher than $\mathrm{O}(\sqrt{\frac{\log T}{T}})$, the number of selection will be upper bounded at the order of sublinear). Then the utility gain for such a deviation is

$$\gamma + \mathrm{O}(\sqrt{\frac{\log T}{T}}) - p_s \cdot \gamma = \mathrm{O}(\sqrt{\frac{\log T}{T}}/\gamma) \cdot \gamma + \mathrm{O}(\sqrt{\frac{\log T}{T}}) = \mathrm{O}(\sqrt{\frac{\log T}{T}}).$$

This establishes the $\mathrm{O}(\sqrt{\frac{\log T}{T}})$-BNE.

When others are exerting $e = e^* - \Delta e$ ($\Delta e > \mathrm{O}(\sqrt{\frac{\log T}{T}})$). If a particular worker i is also exerting the same level of effort, the average payoff is $p_s(\gamma + \Delta e)$. However if the worker deviates by exerting $e = e^* - \Delta e + \tilde{\Delta e}$, where $\tilde{\Delta e} > \mathrm{O}(\sqrt{\frac{\log T}{T}})$, we have the number of times the other workers being selected bounded by (by Lemma 8): $\mathrm{O}(\frac{\log T}{\tilde{\Delta e}^2}) = o(T)$. This fact helps establish that the number of unique selection for worker i is $\mathrm{O}(T)$. Then his marginal payment will become $\gamma + \Delta e - \tilde{\Delta e}$. The gain of such a deviation is

$$\mathrm{O}(\frac{\sqrt{\log T/T}}{\gamma})(\gamma + \Delta e - \tilde{\Delta e}) > \mathrm{O}(\sqrt{\frac{\log T}{T}})$$

when $\Delta e > \mathrm{O}(\gamma)$ and $\tilde{\Delta e} < \sqrt{\log T/T} \cdot \frac{\Delta e}{\gamma}$. □

8.12 With unknown σ

Proof. Within our sequential learning setting, we now show we can even afford to assign tasks and induce efforts without knowing the exact σ values. The idea is as follows: we fix an arbitrary effort level for a certain period of time, and at any time t, we can use collected data from past with this particular effort level to learn a regression model $\theta(t)$. Using this estimated regression model, we are able to estimate $\sigma(e)$. When the space of effort level is finite, we can further impose a bandit selection procedure over effort(i.e., the effort levels are bandits). When the effort level is continuous, using the assumption we made earlier that σ is continuous in e, and suppose the effort level has a bounded support $[0, \overline{e}]$, we can then separate $[0, \overline{e}]$ into $T^z, 0 < z < 1$ intervals uniformly, with each interval having length $1/T^z$. For each of the interval we assign T^κ number of data. Both $0 < z, \kappa < 1$ are constant parameters by design. We choose that $z + \kappa < 1$. For each interval $k = 1, ..., T^z$, we assign $e(k) = \frac{k}{T^z}\overline{e}$. Then use the T^κ samples to estimate $\sigma(e(k))$ in the following way (denote the samples assigned as $(x(n), \tilde{y}(x(n)))$):

$$\tilde{\sigma}(e(k)) = \frac{\sum_{n=1}^{\kappa} (\theta^T(T^\kappa)x(n) - \tilde{y}(x(n)))^2}{T^\kappa} .$$

Now we have the following lemma,

Lemma 11. *With SR-UCB for linear least square, but adaptive effort selection mechanism detailed above, with probability at least* $1 - \mathrm{O}(\frac{1}{T^2})$, $|\tilde{\sigma}(e) - \sigma(e)| \leq \mathrm{O}(\sqrt{\frac{\kappa \log T}{T^\kappa}}) + \mathrm{O}(\frac{1}{T^z}), \forall e.$

Proof. Using Chernoff bound, we can prove the following concentration results for the estimation when T^κ samples are available: with probability at least $1 - \mathrm{O}(\frac{1}{T^2})$,

$$|\tilde{\sigma}(e(k)) - \sigma(e(k))| \leq \mathrm{O}(\sqrt{\frac{\kappa \log T}{T^\kappa}}) .$$

Then using Lipschitz condition we know

$$|\tilde{\sigma}(e) - \sigma(e)| \leq |\tilde{\sigma}(e) - \sigma(e(k))| + |\sigma(e(k)) - \sigma(e)| \leq \mathrm{O}(\sqrt{\frac{\kappa \log T}{T^\kappa}}) + \mathrm{O}(\frac{1}{T^z}) .$$

Note in order to use such an estimation, we need to make sure that during each interval each worker will exert $e(k)$. We can similarly establish a $\mathrm{O}\big(\sqrt{\frac{\log T}{T^{\kappa}}}\big)$-BNE for worker to contribute the corresponding effort for each interval. The reason that we can decouple the above argument for each interval that worker will exert effort $e(k)$ is due to the fact that net payment $p_i - e_i = \gamma$ is independent of the effort level, so the workers have no incentives to mislead the learner into believing a wrong mapping between σ and e; and within each assignment block, workers again try to maximize total payment. Therefore for any e, suppose $\frac{k-1}{T^z} \le e \le \frac{k}{T^z}$, use $\tilde{\sigma}(e(k))$ to serve as an approximation we will have

$$|\tilde{\sigma}(e) - \sigma(e)| \le |\tilde{\sigma}(e) - \sigma(e(k))| + |\sigma(e(k)) - \sigma(e)| \le \mathrm{O}\big(\sqrt{\frac{\kappa \log T}{T^{\kappa}}}\big) + \mathrm{O}\big(\frac{1}{T^z}\big) \ .$$

□

The above error bound reaches the optimal order when $\kappa/2 = z$, and since $\kappa + z < 1$, we have the error decays roughly on the order of $\mathrm{O}\big(T^{-1/3}\big)$. □

8.13 Performance with contributed data

Proof. For outputting the final regression model, we will use the data from only the top 2 most selected workers. First since $\mathbb{E}[n_i(T)] \ge T - \mathrm{O}\big(\log T\big)$ for $i = 1,2$ we know

$$\Pr[T - n_i(T) \ge T/2] \le \frac{\mathbb{E}[T - n_i(T)]}{T/2} \le \mathrm{O}\big(\frac{\log T}{T}\big), i = 1,2.$$

So w.h.p., $n_1(T), n_2(T) \ge T/2$, and then w.p. being at least $1 - e^{-\mathrm{O}\big(T\big)}$ we know the square error loss of the trained regression model is bounded as follows:

$$\begin{aligned}
&\mathbb{E}[\sigma_1(e_1^*)/(\sum_{i=1,2} n_i(T))^2 - \sigma_1(e_1^*)/(2T)^2] \\
\le &\mathbb{E}[\max_{\sum_{i=1,2} n_i(T)} \frac{2\sigma_1(e_1^*)}{(\sum_{i=1,2} n_i(T))^3}(2T - \sum_{i=1,2} n_i(T))] \\
\le &\frac{2\sigma_1(e_1^*)}{T^3}(2T - \mathbb{E}[\sum_{i=1,2} n_i(T)]) \\
\le &\frac{2\sigma_1(e_1^*)}{T^3}(2T - 2T + 2\,\mathrm{O}\big(\log T\big)) \\
= &\mathrm{O}\big(\frac{\sigma_1(e_1^*)\log T}{T^3}\big) \ ,
\end{aligned}$$

where the first inequality is due to mean value theorem, and the second is due to $\sum_i n_i(T) \ge T$, as there is at least one selection at a time. □

8.14 Ridge regression: Proof for Lemma 1

Proof. Again denote the stacked data in a matrix form as $\mathbf{X} \in \mathbb{R}^{n \times d}$, and the corresponding labeling outcome $\tilde{y} \in \mathbb{R}^n$. Following classical results from linear regression we know

$$\begin{aligned}
||\tilde{\theta}_{-i}(t) - \mathbb{E}[\tilde{\theta}_{-i}(t)]||_2^2 &= \mathrm{trace}((\rho I + \mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\tilde{y} - y)(\tilde{y} - y)^T\mathbf{X}(\rho I + \mathbf{X}^T\mathbf{X})^{-1}) \ . \\
||\mathbb{E}[\tilde{\theta}_{-i}(t)] - \theta||_2^2 &= ||-\rho(\rho I + \mathbf{X}^T\mathbf{X})^{-1}\theta||_2^2 \ .
\end{aligned}$$

The variance term is independent of the ground-truth regression model θ and will converge similarly with our previous arguments. The bias term $||\mathbb{E}[\tilde{\theta}_{-i}(t)] - \theta||_2^2$ is depending on θ which is unknown. Therefore without knowing such a quantity[10], it is hard for both the workers and requester to evaluate the one step payment rule. Within our dynamic setting, workers do not need to calculate the specific form of θ; instead workers only need to form the belief that under the same effort level, they will have comparable indexes. Further with more and more data being collected, the bias term will

[10] And we cannot assume we know it as we are learning it.

be decreasing and its effects will diminish – this is by observing the following fact that [4] w.h.p. $\geq 1-e^{-C_1 n}$, when there is n sample being available (following the notations in Lemma 5)

$$||\rho I+\mathbf{X}^T\mathbf{X}|| \leq \rho+(1+\xi)\frac{n}{d+2},\ ||(\rho I+\mathbf{X}^T\mathbf{X})^{-1}|| \leq \frac{1}{\rho+(1-\xi)\frac{n}{d+2}}\ .$$

As in the proof for Lemma 5, we also know

$$||\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}||_2^2 \leq Z^2\frac{1+\xi}{2+d}n\ ,$$

with which we will be able to prove

$$\begin{aligned}
&\text{trace}\left((\rho I+\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\tilde{y}-y)(\tilde{y}-y)^T\mathbf{X}(\rho I+\mathbf{X}^T\mathbf{X})^{-1}\right)\\
\leq&\left((\frac{(1+\xi)\frac{n}{d+2}}{(\rho+(1-\xi)\frac{n}{d+2})^2})^2\cdot Z^2\frac{1+\xi}{2+d}n\right)\\
\leq&\left(\frac{Z^2(1+\xi)^3(d+2)}{(1-\xi)^4}\right)^2\frac{1}{n^2}\ ,
\end{aligned}$$

and

$$\begin{aligned}
&||-\rho(\rho I+\mathbf{X}^T\mathbf{X})^{-1}\theta||_2^2 \leq \left(\frac{\rho M}{\rho+(1-\xi)\frac{n}{d+2}}\right)^2\\
=&(\frac{\rho M(d+2)}{1-\xi})^2\cdot\left(\frac{1}{\rho(d+2)/(1-\xi)+n}\right)^2\\
\leq&(\frac{\rho M(d+2)}{1-\xi})^2\frac{1}{n^2}\ .
\end{aligned}$$

To summarize

$$||\tilde{\theta}_{-i}(t)-\theta||_2^2 \leq ((\frac{Z^2(1+\xi)^3(d+2)}{(1-\xi)^4})^2+(\frac{\rho M(d+2)}{1-\xi})^2)\frac{1}{n^2}\ .$$

□

8.15 (Sketch)-Proof for π-BNE for Non-linear estimator

Proof. Again we evaluate the score for each worker:

$$(\tilde{f}_{-i,t}(x)-\tilde{y}(x))^2=(\tilde{f}_{-i,t}(x)-y(x))^2-2(\tilde{f}_{-i,t}(x)-y(x))\cdot(z+z_i(e_i))+(z+z_i(e-i))^2\ .$$

We want to bound $(\tilde{f}_{-i,t}(x)-y(x))^2$. More specifically according to the results from [20], for non-linear regression model we can establish:

Lemma 12. *With n i.i.d. samples, w.h.p.* $||\tilde{\theta}_i(t)-\theta||_2 \leq O(\frac{1}{\sqrt{n}})$.

The key difference that is going to affect establishing the π-BNE is in proving Lemma 6. With Lipschitz condition we know now w.h.p.$|S_i^1(t)| \leq \frac{C_1}{n}$, where C_1 is re-defined as the constant proved in Lemma 12. Again we know w.h.p., $n \geq C_2 t$ which gives us $|S_i^1(t)| \leq \frac{C_1}{C_2 t}$. So what we need to prove is to bound for each $k=2,3,4,5$

$$\Pr\left[S_i^k(t)-\mathbb{E}[S_i^k(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8}c\sqrt{\frac{\log t}{n_i(t)}}-\frac{C_1}{4C_2 t}\right]\ .$$

Take $S_i^2(t)$ for example, and the rest follow the same logic.

$$\begin{aligned}
&\Pr\left[S_i^2(t) - \mathbb{E}[S_i^2(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2 t}\right] \\
&\leq \exp\left(-2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2 t})^2 n_i(t)\right) \\
&\leq \exp\left(-\frac{2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1}{4C_2^2 t})^2 n_i(t)}{16b^2M^2Z^2}\right) \\
&\leq \exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right)\cdot\exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c\cdot\frac{C_1}{4C_2^2 t}\sqrt{\log t\cdot n_i(t)}}{16b^2M^2Z^2}\right) \\
&\leq \exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right)\cdot\exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c\cdot\frac{C_1}{4C_2^2}}{16b^2M^2Z^2}\right) \leq \frac{\exp(2)}{t^2} \ .
\end{aligned}$$

□

8.16 Example: logistic regression

Proof. To see this, denote $\mu := \theta^T x$ and $\tilde{\mu} := \tilde{\theta}_i^T(t)x$ and apply mean value theorem to $\frac{1}{1+e^{-y}}$ we have

$$\begin{aligned}
&|\frac{1}{1+e^{-\mu}} - \frac{1}{1+e^{-\tilde{\mu}}}| \\
\leq &\max(\frac{1}{1+e^{-y}})'|\mu-\tilde{\mu}| \\
= &\max\frac{1}{e^y+e^{-y}+2}|(\theta-\tilde{\theta}_i(t))^T x| \\
\leq &\frac{1}{4}|(\theta-\tilde{\theta}_i(t))^T x| \ ,
\end{aligned}$$

where we used the fact $e^y + e^{-y} \geq 2$. Since

$$|(\theta-\tilde{\theta}_i(t))^T x| \leq ||\theta-\tilde{\theta}_i(t)||_2||x||_2 \leq ||\theta-\tilde{\theta}_i(t)||_2,$$

we proved the claim. □

9 Proofs for Section 6

9.1 $\mathrm{O}\big(\sqrt{\log T/T}\big)$-BNE for OSR1-UCB

Proof. Following the results detailed in [19] for online learning algorithm for strongly convex function (ρ-ridge regularized loss function is 2ρ-strongly convex), set $\eta_t = 1/(2\rho t)$ we have

Lemma 13. *$\forall t$ of OSR1-UCB, w.p. $\geq 1-\delta$, $||\tilde{\theta}_{-i}^{online}(t) - \tilde{\theta}_{-i}(t)||_2^2 \leq \mathrm{O}\big(\log(\log t/\delta)/\rho t\big)$.*

We can similarly establish the $\mathrm{O}\big(\sqrt{\log T/T}\big)$-BNE for effort exertion – the only difference is compared to what we established earlier that $||\tilde{\theta}_{-i}(t) - \theta||_2^2 \leq \mathrm{O}\big(1/t^2\big)$, here we will have a $\mathrm{O}\big(\log t/\rho t\big)$ (by setting $\delta = 1/t^2$) convergence rate which is much slower in the order. Nevertheless we show this is enough – the intuition is $\mathrm{O}\big(\log t/\rho t\big) < \mathrm{O}\big(\sqrt{\log t/t}\big)$ which is the order of the bias term in our SR-UCB index, such small converging term will not affect the analysis by much. The argument is similar to the proof in Section 8.15, in that we only need to prove bound on $S_i^k(t) - \mathbb{E}[S_i^k(t)], k = 1,2,...,5$

with a different confidence/bias term. Take $S_i^2(t)$ for example:

$$\begin{aligned}
&\Pr\left[S_i^2(t) - \mathbb{E}[S_i^2(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1 \log t}{4C_2 t}\right] \\
&\leq \exp\left(-2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1 \log t}{4C_2 t})^2 n_i(t)\right) \\
&\leq \exp\left(-\frac{2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1 \log t}{4C_2 t})^2 n_i(t)}{16b^2M^2Z^2}\right) \\
&\leq \exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1 \log t}{4C_2 t}\sqrt{\log t \cdot n_i(t)}}{16b^2M^2Z^2}\right) \\
&\leq \exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1}{4C_2^2}\log t\sqrt{\frac{\log t}{t}}}{16b^2M^2Z^2}\right).
\end{aligned}$$

If we choose $\frac{\sqrt{1+\delta_U}-1}{8} c \geq 4bMZ \cdot \max\{1, \frac{4C_2^2}{C_1}\}$, and $t \geq 100$, we know

$$\begin{aligned}
&\exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1}{4C_2^2}\log t\sqrt{\frac{\log t}{t}}}{16b^2M^2Z^2}\right) \\
&\leq \exp(-2\log t)\exp(2) \leq \frac{\exp(2)}{t^2}\ .
\end{aligned}$$

□

9.2 $\mathrm{O}\big(\log T/\sqrt{T}\big)$-BNE for OSR2-UCB

Proof. First we prove

Lemma 14. *With $S_i^{online}(t)$, $\forall t$, w.p. $1-\mathrm{O}\big(1/t^2\big)$:*

$$\frac{1}{t}\sum_{n=1}^{t} 1(i \in d(n))\left((\tilde{\theta}_{-i}^{online}(n) - \theta)^T x_i(n)\right)^2 \leq \mathrm{O}\big(\log t/\sqrt{n_i(t)}\big)\ .$$

Then the argument is similar to the proof in Section 8.15, in that we again need to prove bound on $S_i^k(t) - \mathbb{E}[S_i^k(t)], k = 1,2,...,5$ with a different confidence/bias bound. Take $S_i^2(t)$ for example:

$$\begin{aligned}
&\Pr\left[S_i^2(t) - \mathbb{E}[S_i^2(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log t}{n_i(t)}} - \frac{C_1 \log t}{4C_2\sqrt{n_i(t)}}\right] \\
&\leq \exp\left(-2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log^2 t}{n_i(t)}} - \frac{C_1 \log t}{4C_2\sqrt{n_i(t)}})^2 n_i(t)\right) \\
&\leq \exp\left(-\frac{2(\frac{\sqrt{1+\delta_U}-1}{8} c\sqrt{\frac{\log^2 t}{n_i(t)}} - \frac{C_1 \log t}{4C_2\sqrt{n_i(t)}})^2 n_i(t)}{16b^2M^2Z^2}\right) \\
&\leq \exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log^2 t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1 \log t}{4C_2\sqrt{n_i(t)}}\log t\sqrt{n_i(t)}}{16b^2M^2Z^2}\right) \\
&\leq \exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2 c^2}{16b^2M^2Z^2}\log^2 t\right) \cdot \exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8} c \cdot \frac{C_1}{4C_2^2}\log^2 t}{16b^2M^2Z^2}\right).
\end{aligned}$$

If we choose $\frac{\sqrt{1+\delta_U}-1}{8}c \geq \max\{2\frac{C_1}{4C_2^2}, 4bMZ\}$, we know

$$\exp\left(-2\frac{(\frac{\sqrt{1+\delta_U}-1}{8})^2c^2}{16b^2M^2Z^2}\log^2 t\right)\cdot\exp\left(2\frac{\frac{\sqrt{1+\delta_U}-1}{8}c\cdot\frac{C_1}{4C_2^2}\log^2 t}{16b^2M^2Z^2}\right)\leq \exp(-\log^2 t)\leq \frac{1}{t^2}\ , \text{ if } t\geq e^2.$$

□

9.3 Proof for Lemma 14

Proof. As in Lemma 13, let $\delta=\frac{1}{t^3}$, we know with probability at least $1-\frac{1}{t^3}$ we have

$$((\tilde{\theta}_{-i}^{\text{online}}(n)-\theta)^T x_i(n))^2 \leq ||\tilde{\theta}_{-i}^{\text{online}}(n)-\theta||_2^2 \leq \mathrm{O}\big(\frac{\log t}{n}\big)\ .$$

Via union bounds with probability at least $1-\frac{1}{t^2}$,

$$\sum_{n=1}^{n_i(t)}((\tilde{\theta}_{-i}^{\text{online}}(n)-\theta)^T x_i(n))^2 = \sqrt{n_i(t)}2M^2+\sum_{n=\sqrt{n_i(t)}}^{n_i(t)}\mathrm{O}\big(\frac{\log t}{n}\big)=\mathrm{O}\big(\log t\log n_i(t)\big)\ .$$

which leads to the average error

$$\frac{\sqrt{n_i(t)}2M^2+\log t\log n_i(t)}{n_i(t)}\leq\frac{\log t}{\sqrt{n_i(t)}}\ ,$$

which is due to the fact shown in Lemma 9 that when t is large, $n_i(t)\geq\mathrm{O}\big(\log t\big)$ a.s. and when $n_i(t)=\Omega(\log t)$ we have $\sqrt{n_i(t)}\geq\log n_i(t)$ when t is large. □

10 Proof for Section 7

10.1 Proof for Lemma 2

Proof. This can be proved by induction. At time $t=2$, $n_i(2)$ is a function of $\{S_j(1)\}_j$, the initial value. Assume this is true for t. Consider time $t+1$. $n_i(t+1)$ is an outcome from an ordering function based on inputs of $\{S_j(t)\}_j$ and $\{n_j(t)\}_j$. Based on induction hypothesis, $\{n_j(t)\}_j$ can be written as functions of $\{S_j(n), n<t\}_j$. With this we proved that $\{n_j(t+1)\}_j$ can also be written as functions of $\{S_j(n), n<t+1\}_j$. Proved. □

10.2 Proof for Lemma 3

Proof. We first prove that a finite deviation from worker i creates at most $\mathrm{O}\big(1/T\big)$ differences in $\tilde{\theta}(T)$ (sensitivity) with high probability. Denote this event as $\mathcal{E}(T)$, we know $\Pr[\mathcal{E}(T)]\leq e^{-KT}$. Shorthand the contributed data as $\tilde{y}_i(n):=\tilde{y}_i(n,e_i(n))$, $\tilde{y}_i'(n):=\tilde{y}_i'(n,e_i'(n))$. And denote the regression model trained with $\tilde{y}_i(n),\tilde{y}_i'(n)$ as $\tilde{\theta}(T),\tilde{\theta}'(T)$ (differ only in one data point). Then we have

$$\begin{aligned}\Pr[\tilde{\theta}^p(T)|\tilde{y}_i'(n)] &= \Pr[\tilde{\theta}^p(T)|\tilde{y}_i'(n),\overline{\mathcal{E}}(T)]\cdot\Pr[\overline{\mathcal{E}}(T)]+\Pr[\tilde{\theta}^p(T)|\tilde{y}_i'(n),\mathcal{E}(T)]\Pr[\mathcal{E}(T)]\\ &\leq \Pr[\tilde{\theta}^p(T)|\tilde{y}_i'(n),\overline{\mathcal{E}}(T)]+O(e^{-KT})\ .\end{aligned}$$

Consider the first term above $\Pr[\tilde{\theta}^p(T)|\tilde{y}_i'(n),\overline{\mathcal{E}}(T)]$:

$$\begin{aligned}&\Pr[\tilde{\theta}^p(T)|\tilde{y}_i'(n),\overline{\mathcal{E}}(T)]=\Pr[v_\theta=\tilde{\theta}^p(T)-\tilde{\theta}'(T)]\\ =&\Pr[v_\theta=\tilde{\theta}^p(T)-\tilde{\theta}(T)+\tilde{\theta}'(T)-\tilde{\theta}(T)]\\ =&C\cdot\exp(-\varepsilon_\theta||\tilde{\theta}^p(T)-\tilde{\theta}(T)+\tilde{\theta}'(T)-\tilde{\theta}(T)||_2)\\ \leq&C\cdot\exp(-\varepsilon_\theta||\tilde{\theta}^p(T)-\tilde{\theta}(T)||_2)\cdot\exp(\varepsilon_\theta||\tilde{\theta}'(T)-\tilde{\theta}(T)||_2)\\ =&\Pr[\tilde{\theta}^p(T)|\tilde{y}_i(n)]\cdot\exp(\varepsilon_\theta\cdot\mathrm{O}(\frac{1}{T}))\ .\end{aligned}$$

What is left to prove is that $\tilde{\theta}(T)$'s sensitivity is $\mathrm{O}\big(\frac{1}{T}\big)$ w.h.p., that is we want to bound the difference in the regression model: $||\tilde{\theta}(T)-\tilde{\theta}'(T)||_2$. First

$$\begin{aligned}&||\tilde{\theta}(T)-\tilde{\theta}'(T)||_2\\ =&||\tilde{\theta}(T)-\theta+\theta-\tilde{\theta}'(T)||_2\\ \leq&||\tilde{\theta}(T)-\theta||_2+||\theta-\tilde{\theta}'(T)||_2\ .\end{aligned}$$

Since $\sum_i n_i(T)\geq T$, by results from Theorem 4 we know with probability at least $1-e^{-KT}$,

$$||\tilde{\theta}(T)-\theta||_2\leq \mathrm{O}\big(\frac{1}{T}\big),\ ||\theta-\tilde{\theta}'(T)||_2\leq \mathrm{O}\big(\frac{1}{T}\big)\ .$$

Thus we know (via union bound) w.p. being at leat $1-2e^{-KT}$,

$$||\tilde{\theta}(T)-\tilde{\theta}'(T)||_2\leq \mathrm{O}\big(\frac{1}{T}\big)\ .$$

Also notice that by the CDF of Laplacian distribution,

$$\Pr\left[||\mathbf{v}_\theta||_2\geq\frac{\log T}{\sqrt{T}}\right]=\exp(-\varepsilon_\theta\cdot\frac{\log T}{\sqrt{T}})=\exp(-2\sqrt{T}\frac{\log T}{\sqrt{T}})=\frac{1}{T^2}\ .$$

□

10.3 Proof for Theorem 3

Proof. First for $S_i^{\text{online}}(t)$, each $\tilde{y}_i(\cdot)$ appears in at most $\log T+1$ partial sums. The reason is that a noisy partial sum is discarded only when the size of the partial sum doubles (combine two partial sums). So if the number of such partial sum is greater than $\log T+1$ we will have the total number of data being greater than $2^{\log T}=T$ which is a contradiction. Then by composition theory we know the privacy leakage in $S_i^{\text{online}}(t)$ is bounded by $\mathrm{O}\big((\log T+1)\varepsilon\big)=\mathrm{O}\big(\frac{1}{\log^2 T}\big)$.

Now consider the privacy leakage in $\tilde{\tilde{\theta}}_{-j}^{\text{online}}(t)$. First we prove the following:

Lemma 15. *The sensitivity of $\tilde{\theta}_{-j}(t)$ for each $\tilde{y}_i(n,e_i(n)), n\leq t, i\neq j$ is $||\tilde{\theta}_{-j}(t)-\tilde{\theta}'_{-j}(t)||_2\leq \mathrm{O}\big(1/t\big)$ with probability at least $1-\mathrm{O}\big(1/t^3\big)$.*

The reasoning is similar to a combination of proof for Lemma 7 and Lemma 3. First similar to Lemma 7, we can prove the number of samples that come from $j\neq i$ is at the order of $\mathrm{O}\big(t\big)$ with probability at least $1-\mathrm{O}\big(1/t^3\big)$. Then with $O(t)$ samples, similar to Lemma 3, $||\tilde{\theta}_{-j}(t)-\tilde{\theta}'_{-j}(t)||_2\leq \mathrm{O}\big(1/t\big)$ with probability at least $1-e^{-\mathrm{O}\big(t\big)}$. Combine above we proved the Lemma.

Again due to the decoupling procedure of the partial sum, each $\tilde{\theta}_{-j}(t)$ appears in at most $\log T+1$ partial sums ([3]). Then the privacy leakage of $\tilde{y}_i(n,e_i(n))$ from $\tilde{\theta}_{-j}(t)$ is bounded as $(\log T+1)\varepsilon\cdot \mathrm{O}\big(\frac{1}{t}\big)$, with probability at least $1-\mathrm{O}\big(1/t^3\big)$ (similar to the argument made in the proof for Section 7). Based on this we know for $t\geq \mathrm{O}\big(\log T\big)$, with an appropriately selected constant we know w.p. at least $1-\mathrm{O}\big(\frac{1}{\log T}\cdot\frac{1}{t^2}\big)$, we have the sensitivity of $\tilde{\theta}_{-j}(t)$ is at the order of $\mathrm{O}\big(1/t\big)$. Via union bound we know with probability at least

$$1-\sum_t \mathrm{O}\big(\frac{1}{\log T}\cdot\frac{1}{t^2}\big)=1-\mathrm{O}\big(\frac{1}{\log T}\big),$$

it satisfies for all $t\geq \mathrm{O}\big(\log T\big)$, the sensitivity results hold. Sum over all $\tilde{\theta}_{-j}(t)$, we have by composition theory the total privacy leakage is

$$\sum_{t=1}^{\mathrm{O}\left(\log T\right)}(\log T+1)\varepsilon\cdot\mathrm{O}\big(1\big)+\sum_{t=\mathrm{O}\left(\log T\right)}^{T}(\log T+1)\varepsilon\cdot\mathrm{O}\big(\frac{1}{t}\big)=\mathrm{O}\big((\log T)^2\big)\cdot\varepsilon\ .$$

Then we set $\varepsilon=\frac{1}{\log^4 T}$ we have the preserved privacy level is at the order of $\mathrm{O}\big(\frac{1}{\log T}\big)$. Combined with Lemma 3, composited with the privacy preserving level in $\tilde{\theta}^p(T)$ we proved the theorem.

□

10.4 $\mathrm{O}\big(\log^6 T/\sqrt{T}\big)$-BNE for PSR-UCB

Theorem 5. *With PSR-UCB for linear regression, set fixed payment p_i for all workers as follows: $p_i = e^* + \gamma$, $\gamma = \Omega(\log^6 T/\sqrt{T})$, and set c to be large enough $c \geq Const.(M,Z,N,b)$. Then exerting effort e^* is $\mathrm{O}\big(\log^6 T/\sqrt{T}\big)$-BNE.*

Proof. The main challenge of the proof is to re-establish the convergence of indexes $I_i(t)$ with newly added noises, so that the noise exertion process will not make the indexes useless. There are three sources of noises:

- 1. Noise in $\tilde{\theta}_{-i}^{\text{online}}(t)$, due to the change to a batch summation: $\tilde{\tilde{\theta}}_{-j}^{\text{online}}(t) := \sum_{n=1}^{t} \tilde{\theta}_{-j}(n)/t$.
- 2. Noise in $S_i^{\text{online}}(t)$, due to added noise v_S to partial sums for privacy preserving.
- 3. Noise in $\tilde{\tilde{\theta}}_{-i}^{\text{online}}(t)$, due to added noise $\mathbf{v}_{\tilde{\theta}}$ to partial sums for privacy preserving.

1. First of all we show with the averaging $\tilde{\theta}_{-i}^{\text{online}}(t)$ we do not loss too much performance in converging. Denote $n(t)$ as the number of updates on $\tilde{\theta}_{-j}^{\text{online}}(t)$ up to time t. Notice

$$||\tilde{\theta}_{-j}^{\text{online}}(t) - \theta||_2 = ||\frac{\sum_{n=1}^{n_i(t)} \tilde{\theta}_{-j}(n)}{t} - \theta||_2 \leq \frac{\sum_{n=1}^{n_i(t)} ||\tilde{\theta}_{-j}(n) - \theta||_2}{t} .$$

Consider the summation from $n = 1$ to $n_i(t)$. Select a constant D. For $n < D\sqrt{n_i(t)}$, we have

$$\sum_{n=1}^{D\sqrt{n_i(t)}} ((\tilde{\theta}_{-i}^{\text{online}}(n) - \theta)^T x_i(n))^2 \leq 2M^2 D\sqrt{n_i(t)} .$$

For $n \geq D\sqrt{n_i(t)}$, since we know $n_i(t) \geq (\log T)^6 \log^6 t$ a.s. (similarly argued as in Lemma 9, but with different bias order), we know $\sqrt{n_i(t)} \geq \mathrm{O}\big(\log T\big), a.s.$ Therefore for such n, with probability at least $\frac{1}{T^3}$ we will be having

$$||\tilde{\theta}_{-j}^{\text{online}}(n) - \theta||_2 \leq \mathrm{O}\big(\frac{1}{\sqrt{n}}\big). \tag{10.1}$$

And sum over

$$\sum_{n \geq D\sqrt{n_i(t)}} \mathrm{O}\big(\frac{1}{\sqrt{n}}\big) = \mathrm{O}\big(\sqrt{n_i(t)}\big).$$

Using union bound we have w.p. being at least $1 - \frac{1}{T^2}$

$$\frac{2M^2 D\sqrt{n_i(t)} + \mathrm{O}\big(\sqrt{n_i(t)}\big)}{n_i(t)} = \mathrm{O}\big(\frac{1}{\sqrt{n_i(t)}}\big) .$$

2. Now we analyze how these noises affect the accuracy of our indexes. First consider the added noise in $\sum(\tilde{\theta}_i^T(n)x_i(n) - \tilde{y}_i(n))^2$ (as in $S_i^{\text{online}}(t)$). At any time t we have added at most $\lceil \log t \rceil$ number of Laplacian noise with parameter ε. Denote the sum of $E(t) := \sum_{k=1}^{\lceil \log t \rceil} v_S(k)$. From Lemma 2.8 in [3], we know

$$\Pr[|E(t)| > \lambda] \leq 2\exp(-\frac{\lambda^2}{8\lceil \log t \rceil \frac{1}{\varepsilon^2}}) \leq \exp\bigg(-\frac{\lambda^2}{8}\frac{1}{(\log t + 1)(\log T + 1)^6}\bigg) . \tag{10.2}$$

Let $\lambda := 4(\log t+1)(\log T+1)^3$ we have

$$\begin{aligned}&\Pr\Big[|E(t)| > 4(\log t+1)(\log T+1)^3\Big]\\ \leq& 2\exp\Big(-\frac{16(\log t+1)^2(\log T+1)^6}{8}\frac{1}{(\log t+1)(\log T+1)^6}\Big)\\ \leq& 2\exp(-2\log t) = 2/t^2\,.\end{aligned}$$

Note this additional error term is creating a larger than the index bias by order: $\mathrm{O}\big(\frac{(\log t+1)(\log T+1)^3}{n_i(t)}\big)$.

3. Now consider the noise $\mathbf{v}_{\tilde{\theta}}$ inserted in $\tilde{\tilde{\theta}}^{\text{online}}_{-i}(t)$. Denote $\mathbf{E}(t) := \sum_{k=1}^{\lceil \log t\rceil}\mathbf{v}_{\tilde{\theta}}(k)$. Consider the following fact: for any sample $(\mathbf{x}, y)$

$$((\theta+\mathbf{E}(t)/t)^T\cdot\mathbf{x}-y)^2 = (\theta^T\cdot\mathbf{x}-y)^2+(\mathbf{E}^T(t)/t\cdot\mathbf{x})^2+2\mathbf{E}^T(t)/t\cdot\mathbf{x}\cdot(\theta^T\cdot\mathbf{x}-y)$$

The additional noises appear in two terms:

$$|2\mathbf{E}^T(t)\mathbf{x}\cdot(\theta^T\cdot\mathbf{x}-y)| \leq \mathrm{O}\big(||\mathbf{E}^T(t)||_2\big),$$

due to boundedness of $\mathbf{x}$ and $\theta^T\cdot\mathbf{x}-y$. For the other quadratic term: $(\mathbf{E}^T(t)\mathbf{x})^2 \leq ||\mathbf{E}(t)||_2^2$. For $||\mathbf{E}(t)||_2^2$ we know

$$||\mathbf{E}(t)||_2^2 \leq (\sum_{k=1}^{\lceil \log t\rceil}||\mathbf{v}_{\tilde{\theta}}(k)||_2)^2 := (E(t))^2.$$

Note each $||\mathbf{v}_{\tilde{\theta}}(k)||_2$ is an exponential random variable with parameter ε (mean ε^{-1}). And $E(t)$ is a summation of i.i.d. exponential random variables. Therefore from Theorem 5.1 of [12] we know

$$\Pr[|E(t)| \geq \lambda'/\varepsilon\cdot(\log t+1)] \leq e^{1-\lambda'}.$$

Take $\lambda' := 4(\log t+1)$, we know

$$\Pr[|E(t)| \geq 4(\log t+1)^2\log^3 T \leq \mathrm{O}\big(1/t^2\big).$$

Denote by $\lambda(t) := 4(\log t+1)^2\cdot\log^3 T$. Then the total noise added up to time t is bounded as follows:

$$\lambda^2(t)\cdot\sum_{n=1}^{n_i(t)}\frac{1}{n^2} = \mathrm{O}\big(\lambda^2(t)\big).$$

Since $\mathrm{O}\big(||\mathbf{E}^T(t)||_2\big)$ is on a much smaller order, the average error bounds as: $\mathrm{O}\big(\frac{\lambda^2(t)}{n_i(t)}\big) = \mathrm{O}\big(\frac{\log^4 t\cdot\log^6 T}{n_i(t)}\big)$.

To summarize the total error induced is at the order of

$$\mathrm{O}\big(\frac{1}{\sqrt{n_i(t)}}+\frac{\log^4 t\cdot\log^6 T}{n_i(t)}\big)$$

The rest of the proof is then similar to the reasoning in the proofs in Section 8.15, we need to bound the following term

$$\Pr\Big[S_i^2(t)-\mathbb{E}[S_i^2(t)] \geq \frac{\sqrt{1+\delta_U}-1}{8}c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}-\mathrm{O}\big(\frac{1}{\sqrt{n_i(t)}}\big)-\mathrm{O}\big(\frac{\log^4 t\cdot\log^6 T}{n_i(t)}\big)\Big].$$

After applying the Hoeffding bound, the exponent term is proportional to :

$$-\left(\frac{\sqrt{1+\delta_U}-1}{8}c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}-\mathrm{O}\big(\frac{1}{\sqrt{n_i(t)}}\big)-\mathrm{O}\big(\frac{\log^4 t\cdot\log^6 T}{n_i(t)}\big)\right)^2 n_i(t).$$

Expand it we will have the positive components coming from the inter-product term: first consider the inter-product term

$$\begin{aligned}&\frac{\sqrt{1+\delta_U}-1}{8}c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}\cdot\mathrm{O}\big(\frac{1}{\sqrt{n_i(t)}}\big)\cdot n_i(t)\\ =&\mathrm{O}\big(\log^3 t\cdot\log^3 T\big)<\mathrm{O}\big((\log^3 t\cdot\log^3 T)^2\big),\end{aligned}$$

For the other inter-product term:

$$\begin{aligned}&\frac{\sqrt{1+\delta_U}-1}{8}c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}\cdot\mathrm{O}\big(\frac{\log^4 t\cdot\log^6 T}{n_i(t)}\big)\cdot n_i(t)\\ =&\mathrm{O}\big(\frac{\sqrt{1+\delta_U}-1}{8}c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}\cdot(\log^4 t\cdot\log^6 T)\big)\\ \le&\mathrm{O}\big(\log^4 t\cdot\log^6 T\big)\\ <&\mathrm{O}\big((\log^3 t\cdot\log^3 T)^2\big),\end{aligned}$$

where the first inequality is due to the fact that after we change the bias term we can proved $\sqrt{n_i(t)}=\Omega(\log^3 t\log^3 T)$ a.s., when t is large.

The inner-product term (lower bounded by the first inner-product term):

$$\mathrm{O}\big(\left(\frac{\sqrt{1+\delta_U}-1}{8}c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}\right)^2\cdot n_i(t)\big)=\mathrm{O}\big((\log^3 t\cdot\log^3 T)^2\big).$$

Therefore the inter-products (positive exponents) is on a smaller order compared to inner-product terms (negative exponents), and thus can be ignored. We can similarly prove the convergence results.

Meanwhile with changing the bias term from $\sqrt{\log t/n_i(t)}$ to $\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}$, workers have stronger incentives to deviate. The difference we need to bound lies in changing the bounding of the following events (in Lemma 8)

$$2c\sqrt{\frac{\log t}{n_i(t)}}+\tau(t)+\frac{C_1}{(\sum_{k\neq i}n_k(t))^2}+\frac{C_1}{(\sum_{k\neq j}n_k(t))^2}<b\underline{L}\Delta\,,$$

to the following one

$$2c\frac{\log^3 t\log^3 T}{\sqrt{n_i(t)}}+\tau(t)+\frac{C_1}{(\sum_{k\neq i}n_k(t))^2}+\frac{C_1}{(\sum_{k\neq j}n_k(t))^2}<b\underline{L}\Delta\,,$$

from which we know the number of selection after deviating by Δ is bounded as follows

$$\mathbb{E}[n_i(t)]\le\mathrm{O}\big(\frac{(\log^3 t\cdot\log^3 T)^2}{\Delta^2}\big).$$

Let $t=T$, and when

$$\Delta>\mathrm{O}\big(\sqrt{\frac{(\log^3 T\cdot\log^3 T)^2}{T}}\big)=\frac{\log^6 T}{\sqrt{T}}\,,$$

we will have $\mathbb{E}[n_i(t)]=o(T)$, from where we can prove a contradiction on non-profitable deviation. □