Improved Algorithms for Neural Active Learning

Yikun Ban*, Yuheng Zhang*, Hanghang Tong, Arindam Banerjee, and Jingrui He

University of Illinois Urbana-Champaign
{yikunb2, yuhengz2, htong, arindamb, jingrui}@illinois.edu

Abstract

We improve the theoretical and empirical performance of neural-network(NN)-based active learning algorithms for the non-parametric streaming setting. In particular, we introduce two regret metrics by minimizing the population loss that are more suitable in active learning than the one used in state-of-the-art (SOTA) related work. Then, the proposed algorithm leverages the powerful representation of NNs for both exploitation and exploration, has the query decision-maker tailored for k-class classification problems with the performance guarantee, utilizes the full feedback, and updates parameters in a more practical and efficient manner. These careful designs lead to a better regret upper bound, improving by a multiplicative factor $\mathcal{O}(\log T)$ and removing the curse of input dimensionality. Furthermore, we show that the algorithm can achieve the same performance as the Bayes-optimal classifier in the long run under the hard-margin setting in classification problems. In the end, we use extensive experiments to evaluate the proposed algorithm and SOTA baselines, to show the improved empirical performance.

1 Introduction

The Neural Network (NN) is one of the indispensable paradigms in machine learning and is widely used in multifarious supervised-learning tasks [25]. As more and more complicated NNs are developed, the requirement of the training procedure on the labeled data grows, incurring significant cost of label annotation. Active learning investigates effective techniques on a much smaller labeled data set while attaining the comparable generalization performance to passive learning [20]. In this paper, we focus on the classification problem in the streaming setting of active learning with NN models. At every round, the learner receives an instance and is compelled to decide on-the-fly whether or not to observe the label associated with this instance. This problem seeks to maximize the generalization capability of learned NNs in a sequence of rounds, such that the model has robust performance on the unseen data from the same distribution [42].

In active learning, given access to the i.i.d. generated instances from a distribution $\mathcal{D}$, suppose there exist a class of functions $\mathcal{F}$ that formulate the mapping from instances to theirs labels. In the parametric setting, i.e., $\mathcal{F}$ has finite VC-dimension [27], existing works [26, 15, 8] have shown that the active learning algorithms can achieve the convergence rate of $\widetilde{\mathcal{O}}(1/\sqrt{N})$ to the best population loss in $\mathcal{F}$, where N is the number of label queries. In the non-parametric setting, recent works [36, 37] provide the similar convergence results while suffering from the curse of input dimensionality. Unfortunately, most of NN-based approaches to active learning do not come with the performance guarantee, despite having powerful empirical results.

The first performance guarantee for neural active learning has been established in a recent work by [50], and the analysis is for over-parameterized neural networks with the assistance of Neural Tangent Kernel (NTK). We carefully investigate the limitations of [50], which turn into the main motivations of our paper. First, [50] transforms the classification problem into a multi-armed bandit

* Both authors contribute equally.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

problem [57], to minimize a pseudo regret metric. Yet, on the grounds that they seek to minimize the *conditional* population loss on a sequence of given data, it is dubious that the pseudo regret used in [50] can explicitly measure the generalization capability of given algorithms (see Remark 2.1). Second, the proposed algorithm in [50] has an unknown prior complexity term S that reflects the complexity of function to be learned, brought by the NTK approximation (Lemma 1 in [50]). To estimate the magnitude of S, the authors introduce another complex model selection algorithm and make strong assumptions regarding the generated models. Third, the training process for NN models is not efficient, as [50] uses vanilla gradient descent and starts from randomly initialized parameters in every round. Fourth, although [50] removes the curse of input dimensionality d, the performance guarantee strongly suffers from another introduced term, the effective dimensionality $\widetilde{d}$, which can be thought of as the non-linear dimensions of Hilbert space spanned by NTK. In the worse case, the magnitude of $\widetilde{d}$ can be an unacceptably large number and thus the performance guarantee collapses.

1.1 Main contributions

In this paper, we propose a novel algorithm, I-NeurAL (**I**mproved Algorithms for **Neur**al **A**ctive **L**earning), to tackle the above limitations. Our contributions can be summarized as follows: (1) We consider the k-class classification problem, and we introduce two new regret metrics to minimize the population loss, which can directly reflect the generalization capability of NN-based algorithms. (2) I-NeurAL has a neural exploration strategy with a novel component to decide whether or not to query the label, coming with the performance guarantee. I-NeurAL does not have unknown complexity terms and the hyper-parameters are straightforward and intuitive. (3) I-NeurAL is designed to support mini-batch Stochastic Gradient Descent (SGD). In particular, at every round, I-NeurAL does mini-batch SGD starting with the parameters of the last round, i.e., with warm start, which is more efficient and practical compared to [50]. (4) Without any noise assumption on the data distribution, we provide the performance guarantee of I-NeurAL for over-parameterized neural networks. Compared to [50], we remove the curse of both the input dimensionality d and the effective dimensionality $\widetilde{d}$; Moreover, we improve the regret by a multiplicative factor $\log(T)$, where T is the number of rounds. (5) under a hard-margin assumption on the data distribution, to the best of our knowledge, we provide the first performance analysis that NN models can achieve the same generalization capability as Bayes-optimal classifier after a fixed number of label queries; (6) we conduct extensive experiments on real-world data sets to demonstrate the improved performance of I-NeurAL over state-of-the-art baselines including the closest work [50] which has not provided empirical validation of their proposed algorithms.

1.2 Related Work

Active learning has been extensively studied and applied to many essential applications [46]. Bayesian active learning methods typically use a probabilistic regression model to estimate the improvement of each query [31, 43]. In spite of effectiveness on the small or moderate data sets, the Bayesian-based approaches are difficult to scale to large-scale data sets because of the batch sampling [45]. Another important class, margin algorithms or uncertainty sampling [35], obtains considerate performance improvement over passive learning and is further developed by many practitioners [21, 30, 39, 16]. Margin algorithms are flexible and can be adapted to both streaming and pool settings. In the pool setting, a line of works utilize the neural networks in active learning to improve the empirical performance [38, 44, 6, 18, 32, 47, 49, 56, 5]. However, they do not provide performance guarantee for NN-based active learning algorithms. From the theoretical perspective, [53, 22, 7, 9, 54] provide the performance guarantee with the specific classes of functions and [28, 23] present the theoretical analysis of active learning algorithms with the surrogate loss functions for binary classification. However, their performance guarantee is restricted within hypothesis classes, i.e, the parametric setting. In contrast, our goal is to derive an NN-based algorithm in the non-parametric setting that performs well both empirically and theoretically. Neural contextual bandits[57, 55, 14, 12, 13, 41] provide the principled method to balance between the exploitation and exploration [10, 11]. [50] transforms active learning into neural contextual bandit problem and obtains a performance guarantee, of which limitations are discussed above.

As [50] is the closest related work to our paper, we emphasize the differences of our techniques from [50] throughout the paper. We introduce the problem definition and proposed algorithms in Section

2 and Section 3 respectively. Then, we provide performance guarantees in Section 4 and empirical results in Section 5, ending with the conclusion in Section 6.

2 Problem Definition

In this paper, we study the streaming setting of active learning in the k-class classification problem. Let $\mathcal{X}$ denote the input space over $\mathbb{R}^d$, $\mathcal{Y} = \{1, 2, \ldots, k\}$ represent the label space, and $\mathcal{D}$ be some unknown distribution over $\mathcal{X} \times \mathcal{Y}$. At round $t \in [T] = \{1, 2, \ldots, T\}$, an instance $\mathbf{x}_t$ is drawn from the marginal distribution $\mathcal{D}_{\mathcal{X}}$ and accordingly y_t is drawn from the conditional distribution $\mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}$. Here, y_t can be thought of as the index of the class that $\mathbf{x}_t$ belongs to. Inspired by [50], we first transform $\mathbf{x}_t$ into k context vectors representing the k classes respectively: $\mathbf{x}_{t,1} = (\mathbf{x}_t^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top)^\top, \mathbf{x}_{t,2} = (\mathbf{0}^\top, \mathbf{x}_t^\top, \ldots, \mathbf{0}^\top)^\top, \ldots, \mathbf{x}_{t,k} = (\mathbf{0}^\top, \mathbf{0}^\top, \ldots, \mathbf{x}_t^\top)^\top$ and $\mathbf{x}_{t,i} \in \mathbb{R}^{dk}, \forall i \in [k]$. In accordance with context vectors, we construct the k label vectors representing the k possible prediction: $\mathbf{y}_{t,1} = (1, 0, \ldots, 0)^\top, \mathbf{y}_{t,2} = (0, 1, \ldots, 0)^\top, \ldots, \mathbf{y}_{t,k} = (0, 0, \ldots, 1)^\top$ and $\mathbf{y}_{t,i} \in \mathbb{R}^k, \forall i \in [k]$. Thus, $\mathbf{y}_{t,y_t}$ is the ground-truth label vector for $\mathbf{x}_t$.

Under the non-parametric setting of active learning, we define an unknown function h to formulate the conditional distribution $\mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}$: $\mathcal{X}^k \to [0, 1]$, such that

$$\forall i \in [k], \mathbb{P}(\mathbf{y}_{t,y_t} = \mathbf{y}_{t,i} | \mathbf{x}_t) = h(\mathbf{x}_{t,i}) \,, \tag{2.1}$$

which is subject to $\sum_{i=1}^k h(\mathbf{x}_{t,i}) = 1$. For simplicity, we consider the k-class classification problem with 0-1 loss. Given $\mathbf{x}_t$, i.e., $\mathbf{x}_{t,i}, i \in [k]$, let $\widehat{i}$ be the index of the class predicted by some hypothesis f and thus $\mathbf{y}_{t,\widehat{i}}$ is the prediction. Then, we have the following loss:

$$L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t}) = \mathbb{1}\{\mathbf{y}_{t,\widehat{i}} \neq \mathbf{y}_{t,y_t}\} \in \{0, 1\} \,. \tag{2.2}$$

where $\mathbb{1}$ is the indicator function.

Given the number of rounds T, at each round $t \in [T]$, the learner receives an instance $\mathbf{x}_t$ drawn i.i.d. from $\mathcal{D}_{\mathcal{X}}$. Then, the learner needs to make a prediction $\mathbf{y}_{t,\widehat{i}}$, and at the same time, decide on-the-fly whether or not to query the label $\mathbf{y}_{t,y_t}$ where y_t is drawn i.i.d. from $\mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}$. As the goal of active learning tasks is often to minimize the population loss [42], we introduce the following two regret metrics.

Definition 2.1 (Latest Population Regret)**.** *Given the data distribution $\mathcal{D}$, the number of rounds T, the Latest Population Regret is defined as*

$$R_T = \mathop{\mathbb{E}}_{\mathbf{x}_T \sim D_{\mathcal{X}}} \left[\mathop{\mathbb{E}}_{y_T \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_T}} [L(\mathbf{y}_{T,\widehat{i}}, \mathbf{y}_{T,y_T}) \mid \mathbf{x}_T] \right] - \mathop{\mathbb{E}}_{\mathbf{x}_T \sim D_{\mathcal{X}}} \left[\mathop{\mathbb{E}}_{y_T \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_T}} [L(\mathbf{y}_{T,i^*}, \mathbf{y}_{T,y_T}) \mid \mathbf{x}_T] \right] \tag{2.3}$$

where $\mathbf{y}_{T,i^*}$ is the prediction the Bayes-optimal classifier would make on instance $\mathbf{x}_T$, i.e., $i^* = \arg\max_{i \in [k]} h(\mathbf{x}_{T,i})$ for $\mathbf{y}_{T,i^*}$.

Definition 2.2 (Cumulative Population Regret)**.** *Given the data distribution $\mathcal{D}$, the number of rounds T, the Cumulative Population Regret is defined as:*

$$\mathbf{R}_T = \sum_{t=1}^T \left(\mathop{\mathbb{E}}_{\mathbf{x}_t \sim D_{\mathcal{X}}} \left[\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t}) \mid \mathbf{x}_t] \right] - \mathop{\mathbb{E}}_{\mathbf{x}_t \sim D_{\mathcal{X}}} \left[\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t}) \mid \mathbf{x}_t] \right] \right) \tag{2.4}$$

where $\mathbf{y}_{t,i^*}$ is the prediction the Bayes-optimal classifier would make on instance $\mathbf{x}_t$, i.e., $i^* = \arg\max_{i \in [k]} h(\mathbf{x}_{t,i})$ for $\mathbf{y}_{t,i^*}$.

R_T measures the performance at the last round T only, and $\mathbf{R}_T$ measures the overall performance in T rounds combined. Therefore, the goal of this problem is to minimize R_T or $\mathbf{R}_T$, or both. At the same time, we also aim to minimize the following expected query cost:

$$\mathbf{N}_T = \sum_{t=1}^T \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_{\mathcal{X}}} [\mathbf{I}_t \mid \mathbf{x}_t], \tag{2.5}$$

where $\mathbf{I}_t$ is the indicator of the query decision in round t such that $\mathbf{I}_t = 1$ if y_t is observed; $\mathbf{I}_t = 0$, otherwise.

Remark 2.1. Minimizing R_T or $\mathbf{R}_T$ shows the generalization capability of the learned hypothesis on the distribution $\mathcal{D}$. However, the problem defined in [50] is to minimize the cumulative *conditional* population regret as follows:

$$\widetilde{\mathbf{R}}_T = \sum_{t=1}^{T} \left(\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})|\mathbf{x}_t] - \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})|\mathbf{x}_t] \right). \tag{2.6}$$

As $\mathbb{E}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}}[L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})|\mathbf{x}_t]$ is the population loss conditioned on $\mathbf{x}_t$, unfortunately, $\widetilde{\mathbf{R}}_T$ only measures the performance of the learned hypothesis on the collected data $\{\mathbf{x}_t\}_{t=1}^T$, and $\widetilde{\mathbf{R}}_T$ cannot directly measure the accuracy of the hypothesis on unseen data instances. Although $\widetilde{\mathbf{R}}_T$ follows the regret definition in multi-armed bandits [57], it is fair to say that $\widetilde{\mathbf{R}}_T$ may not be a good metric in active learning.

3 Proposed Algorithms

In this section, we elaborate on the proposed algorithm I-NeurAL (Algorithm 1). In contrast to the directly comparable work [50], I-NeurAL has the following novel and advantageous aspects: (1) I-NeurAL incorporates a neural-based exploration strategy (Line 6) inspired by recent advances in bandits [14] to solve the exploitation-exploration dilemma in the decision for whether or not to query labels; (2) I-NeurAL includes a novel component (Line 11) to decide whether or not to query labels in the k-class classification problem, which does not have the unknown complexity term S as in [50], thus avoiding the complicated model selection procedure (Algorithm 2 in [50]); (3) I-NeurAL infers and exploits the feedback of all the contexts (Lines 12-17), instead of only utilizing the feedback of the chosen context in [50]; (4) I-NeurAL conducts mini-batch SGD based on the parameters of the last round (Algorithm 2), which is more practical, as opposed to conducting vanilla gradient descent from the initialization at every round in [50]. Next, we will present the details of I-NeurAL.

Exploitation Network f_1. Given $\mathbf{x}_{t,i}, i \in [k]$, to learn the unknown function h (Eq. (2.1)), we use a fully-connected neural network f_1 with L-depth and m-width:

$$f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}^1) = \mathbf{W}_L^1 \sigma(\mathbf{W}_{L-1}^1 \sigma(\mathbf{W}_{L-2}^1 \dots \sigma(\mathbf{W}_1^1 \mathbf{x}_{t,i}))), \tag{3.1}$$

where $\mathbf{W}_1^1 \in \mathbb{R}^{m \times kd}, \mathbf{W}_l^1 \in \mathbb{R}^{m \times m}$, for $2 \le l \le L-1$, $\mathbf{W}_L^1 \in \mathbb{R}^{1 \times m}$, $\boldsymbol{\theta}^1 = [\text{vec}(\mathbf{W}_1^1)^\top, \dots, \text{vec}(\mathbf{W}_L^1)^\top]^\top \in \mathbb{R}^{p_1}$, and σ is the ReLU activation function $\sigma(\mathbf{x}) = \max\{0, \mathbf{x}\}$. In round t, given $\mathbf{x}_{t,i}, i \in [k]$, $f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$ is assigned to learn $h(\mathbf{x}_{t,i})$. Based on the fact $h(\mathbf{x}_{t,i}) = \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [1 - L(\mathbf{y}_{t,i}, \mathbf{y}_{t,y_t})]$, it is natural to regard $1 - L(\mathbf{y}_{t,i}, \mathbf{y}_{t,y_t})$ as the label for training f_1. Note that we take the basic fully-connected network as an example for the sake of analysis in over-parameterized networks and f_1 can be easily replaced with more complicated models depending on the tasks.

Exploration Network f_2. In addition to the network f_1, we assign another network f_2 to explore uncertain information contained in incoming instances. First, we carefully design the input of f_2 to incorporate the context vectors of the instance and the discrimination-ability of f_1, to learn the error between the Bayes-optimal probability $h(\mathbf{x}_{t,i})$ and the prediction $f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}^1)$.

Definition 3.1 (Derivative-Context (DC) Embedding)**.** *Given the exploitation network $f_1(\cdot; \boldsymbol{\theta}_{t-1}^1)$ and an input context $\mathbf{x}_{t,i}$, its DC embeding is defined as*

$$\phi(\mathbf{x}_{t,i}) = \left(\frac{vec\left(\nabla_{\mathbf{x}_{t,i}} f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)\right)^\top}{\sqrt{2}\|\nabla_{\mathbf{x}_{t,i}} f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)\|_2}, \frac{\mathbf{x}_{t,i}{}^\top}{\sqrt{2}} \right) \in \mathbb{R}^{2dk}, \tag{3.2}$$

where $\nabla_{\mathbf{x}_{t,i}} f_1$ is the partial derivative of $f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$ with respect to $\mathbf{x}_{t,i}$.

$\phi(\mathbf{x}_{t,i})$ is normalized so that $\|\phi(\mathbf{x}_{t,i})\|_2 = 1$. Note that the input for f_2 in [14] is the gradient with respect to θ_1, denoted by $\nabla_{\theta_1} f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1) \in \mathbb{R}^{p_1}$. Its dimensionality is much larger than $\nabla_{\mathbf{x}_{t,i}} f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$ in Definition 3.1, may causing significant computation cost.

Given the input $\phi(\mathbf{x}_{t,i})$, similarly, we choose the fully-connected network to build f_2:

$$f_2(\phi(\mathbf{x}_{t,i}); \boldsymbol{\theta}^2) = \mathbf{W}_L^2 \sigma(\mathbf{W}_{L-1}^2 \sigma(\mathbf{W}_{L-2}^2 \dots \sigma(\mathbf{W}_1^2 \phi(\mathbf{x}_{t,i})))), \tag{3.3}$$

Algorithm 1 I-NeurAL

Input: T (number of rounds) f_1, f_2 (neural networks), η_1, η_2 (learning rate), γ (exploration parameter), b (batch size), δ (confidence level)

1: Initialize $\boldsymbol{\theta}_0^1, \boldsymbol{\theta}_0^2$; $\widehat{\boldsymbol{\theta}}_0^1 = \boldsymbol{\theta}_0^1; \widehat{\boldsymbol{\theta}}_0^2 = \boldsymbol{\theta}_0^2$
2: $\mathcal{H}_0^1 = \emptyset; \mathcal{H}_0^2 = \emptyset$
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Observe instance $\mathbf{x}_t \in \mathbb{R}^d$ and build $\mathbf{x}_{t,i}, \forall i \in [k]$
5: **for** each $i \in [k]$ **do**
6: $f(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}) = \Big(\underbrace{f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)}_{\text{Exploitation Score}} + \underbrace{f_2(\phi(\mathbf{x}_{t,i}); \boldsymbol{\theta}_{t-1}^2)}_{\text{Exploration Score}} \Big)$
7: **end for**
8: $\widehat{i} = \arg\max_{i \in [k]} f(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1})$
9: $i^\circ = \arg\max_{i \in ([k] \setminus \{\widehat{i}\})} f(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1})$
10: Predict $\mathbf{y}_{t,\widehat{i}}$
11: $\mathbf{I}_t = \mathbb{1}\{|f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})| < 2\gamma \boldsymbol{\beta}_t\} \in \{0, 1\}$; $\boldsymbol{\beta}_t = \sqrt{\frac{2}{t}} + \left(\frac{c_1 3L}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(c_2 tk)/\delta)}{t}}$
12: **if** $\mathbf{I}_t = 1$ **then**
13: Query $\mathbf{x}_t$ and observe y_t
14: **for** $i \in [k]$ **do**
15: $r_{t,i}^1 = 1 - L(\mathbf{y}_{t,i}, \mathbf{y}_{t,y_t})$ (defined in E.q. (2.2))
16: $r_{t,i}^2 = r_{t,i}^1 - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$
17: **end for**
18: **else**
19: **for** $i \in [k]$ **do**
20: $r_{t,i}^1 = 1 - L(\mathbf{y}_{t,i}, \mathbf{y}_{t,\widehat{i}})$
21: $r_{t,i}^2 = r_{t,i}^1 - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$
22: **end for**
23: **end if**
24: $\mathcal{H}_t^1 = \mathcal{H}_{t-1}^1 \cup \{(\mathbf{x}_{t,i}, r_{t,i}^1), i \in [k]\}$
25: $\mathcal{H}_t^2 = \mathcal{H}_{t-1}^2 \cup \{(\mathbf{x}_{t,i}, r_{t,i}^2), i \in [k]\}$
26: $\boldsymbol{\theta}_t^1, \boldsymbol{\theta}_t^2$ = Mini-Batch-SGD-Warm-Start (f_1, f_2, $\mathcal{H}_t^1$, $\mathcal{H}_t^2$, b)
27: **end for**

where $\mathbf{W}_1^2 \in \mathbb{R}^{m \times 2kd}, \mathbf{W}_l^2 \in \mathbb{R}^{m \times m}$, for $2 \le l \le L-1$, $\mathbf{W}_L^2 = \mathbb{R}^{1 \times m}$ and $\boldsymbol{\theta}^2 = [\text{vec}(\mathbf{W}_1^2)^\top, \ldots, \text{vec}(\mathbf{W}_L^2)^\top]^\top \in \mathbb{R}^{p_2}$. In round t, given $\mathbf{x}_{t,i}, \forall i \in [k]$, f_2 is to predict $h(\mathbf{x}_{t,i}) - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$ for exploration. Because $h(\mathbf{x}_{t,i}) - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1) = \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [1 - L(\mathbf{y}_{t,i}, \mathbf{y}_{t,y_t}) - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)]$, we regard $1 - L(\mathbf{y}_{t,i}, \mathbf{y}_{t,y_t}) - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1)$ as the label for training f_2.

To sum up, in round t, given $\mathbf{x}_{t,i}, \forall i \in [k]$, the prediction $\widehat{i}$ ($\mathbf{y}_{t,\widehat{i}}$) is made based on the sum of exploitation and exploration scores, i.e., $f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}^1) + f_2(\phi(\mathbf{x}_{t,i}); \boldsymbol{\theta}_{t-1}^2)$ (Lines 5-10).

Query Decision-maker (Line 11). A label query is made when I-NeurAL is not confident enough to discriminate the Bayes-optimal class from other classes. $2\gamma\boldsymbol{\beta}_t$ ($\boldsymbol{\beta}_t$ is also defined in Lemma 7.3) can be thought of as a confidence interval for the distance between the optimal class and second optimal class, where γ is the hyper-parameter to tune the sensitivity of the decision-maker in practice. Given any $\gamma \ge 1, \delta \in (0, 1)$, with probability at least $1 - \delta$, based on our analysis (Lemma 7.5), $\mathbb{E}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})] = \mathbb{E}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})]$ when $\mathbf{I}_t = 0$, i.e., I-NeurAL suffers no regret. Thus, we use $\mathbf{y}_{t,\widehat{i}}$ as the pseudo-label in this case and we have the following update rules.

Utilize Full Feedback (Lines 14-25). Different from the bandit setting where the learner can only observe the reward of the selected context, we can infer the rewards of all contexts in active learning, as we know the specific class of the current instance. Thus, for each $\mathbf{x}_{t,i}, i \in [k]$, $r_{t,i}^1 = 1 - \mathcal{L}(\mathbf{y}_{t,i}, \mathbf{y}_{t,y_t})$ is regarded as the "reward" of $\mathbf{x}_{t,i}$, predicted by f_1, and $r_{t,i}^2 = r_{t,i}^1 - f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}^1)$ is regarded as the

Algorithm 2 Mini-Batch-SGD-Warm-Start (f_1, f_2, $\mathcal{H}_t^1, \mathcal{H}_t^2, b$)

1: Define $\mathcal{L}_1[(\mathbf{x}, r^1); \boldsymbol{\theta}^1] = (r^1 - f_1(\mathbf{x}; \boldsymbol{\theta}^1))^2/2$
2: Uniformly draw a set $\widehat{\mathcal{H}}_t^1 \subset \mathcal{H}_t^1, s.t., |\widehat{\mathcal{H}}_t^1| = b$
3: $\widehat{\boldsymbol{\theta}}_t^1 = \widehat{\boldsymbol{\theta}}_{t-1}^1 - \frac{\eta_1}{b} \sum_{(\mathbf{x}, r^1) \in \widehat{\mathcal{H}}_t^1} \nabla_{\boldsymbol{\theta}^1} \mathcal{L}_1[(\mathbf{x}, r^1); \widehat{\boldsymbol{\theta}}_{t-1}^1]$
4: Define $\mathcal{L}_2[(\phi(\mathbf{x}), r^2); \boldsymbol{\theta}^2] = (r^2 - f_2(\phi(\mathbf{x}); \boldsymbol{\theta}^2))^2/2$
5: Uniformly draw a set $\widehat{\mathcal{H}}_t^2 \subset \mathcal{H}_t^2, s.t., |\widehat{\mathcal{H}}_t^2| = b$
6: $\widehat{\boldsymbol{\theta}}_t^2 = \widehat{\boldsymbol{\theta}}_{t-1}^2 - \frac{\eta_2}{b} \sum_{(\phi(\mathbf{x}), r^1) \in \widehat{\mathcal{H}}_t^2} \nabla_{\boldsymbol{\theta}^2} \mathcal{L}_2[(\phi(\mathbf{x}), r^2); \widehat{\boldsymbol{\theta}}_{t-1}^2]$
7: $\Omega_t = \Omega_{t-1} \cup \{(\widehat{\boldsymbol{\theta}}_t^1, \widehat{\boldsymbol{\theta}}_t^2)\}$
8: **Return** $(\boldsymbol{\theta}_t^1, \boldsymbol{\theta}_t^2)$ uniformly from Ω_t

"residual reward" of $\mathbf{x}_{t,i}$, predicted by f_2. In summary, in round t, when $\mathbf{I}_t = 1$, $\mathbf{y}_{t,y_t}$ is observed to update $r_{t,i}^1$ and $r_{t,i}^2$; when $\mathbf{I}_t = 0$, $\mathbf{y}_{t,\widehat{i}}$ is regard as the pseudo-label to obtain $r_{t,i}^1$ and $r_{t,i}^2$, $\forall i \in [k]$. Therefore, we have the training data $\mathcal{H}_t^1$ for f_1 and $\mathcal{H}_t^2$ for f_2.

Mini-Batch SGD with Warm-Start (Algorithm 3). Unlike [50] that uses vanilla gradient descent from randomly initialized parameters in each round, causing unnecessarily expensive computation, we extend the training procedure to mini-batch SGD with warm start, i.e., we incrementally train the parameters $\boldsymbol{\theta}_t$ starting from the parameters of the last round $\boldsymbol{\theta}_{t-1}$ in each round t.

Algorithm 1 depicts the workflow of I-NeurAL. Lines 1-2 initialize the parameters where each entry of $\mathbf{W}_l$ is drawn from the normal distribution $\mathcal{N}(0, 2/m)$ and each entry of $\mathbf{W}_L$ is drawn from $\mathcal{N}(0, 1/m)$ for both f_1 and f_2. $\mathcal{H}_0^1, \mathcal{H}_0^2$ store the historical data for f_1 and f_2 respectively. In each round t, Line 4 builds the k contexts for the observed instance $\mathbf{x}_t$, and Lines 5-7 calculate the exploitation-exploration score for each context. $\widehat{i}$ (Line 8) is the index of the optimal-predicted class and thus $\mathbf{y}_{t,\widehat{i}}$ is the prediction. i° (Line 9) is the index of the second optimal-predicted class, which is used to decide whether to make a query. Line 11 is our decision component. When $\mathbf{I_t} = 1$, it shows that we are not confident enough about our prediction, so that we make a query for $\mathbf{x}_t$ and observe the rewards for each context (Lines 12-17). When $\mathbf{I_t} = 0$, based on our analysis, with high confidence, the prediction $\mathbf{y}_{t,\widehat{i}}$ matches the one predicted by the Bayes-optimal classifier. Hence, we consider $\mathbf{y}_{t,\widehat{i}}$ as the label and observe the reward for all contexts (Lines 18-23). In the end, we update the networks f_1 and f_2, based on the collected data (Lines 24-26).

4 Regret Analysis

In this section, we provide the regret analysis of I-NeurAL in the over-parameterized neural networks. First, we need the standard normalization and separateness restricted to the input instances.

Assumption 4.1. For any $t \in [T]$, $\|\mathbf{x}_t\|_2 = 1$.

This is standard assumption in the analysis of over-parameterized neural networks[2, 24, 4], neural active learning [50] and neural bandits [57, 14]. Given a constant $\nu > 0$, we define the following ν-ball of $\boldsymbol{\theta}^2$ around the random initialization: $\mathcal{B}(\boldsymbol{\theta}_0^2, \nu) = \{\widetilde{\boldsymbol{\theta}}^2 : \|\widetilde{\boldsymbol{\theta}}^2 - \boldsymbol{\theta}_0^2\|_2 \leq \mathcal{O}(\frac{\nu}{\sqrt{m}})\}$. Recall that $r_{t,\widehat{i}}^2 = r_{t,\widehat{i}}^1 - f_1(\mathbf{x}_{t,\widehat{i}}; \widehat{\boldsymbol{\theta}}_{t-1}^1)$. Let $\widehat{\boldsymbol{\theta}}_{t-1}^{1,*}$ represent the parameters trained on $\mathcal{H}_{t-1}^{1,*}$ using Algorithm 2 with the Bayes-optimal classifier, where $\mathcal{H}_{t-1}^{1,*} = \{\mathbf{x}_{\tau,i^*}, r_{\tau,i^*}^1\}_{\tau=1}^{t-1}$ are the historical Bayes-optimal pairs. We define $r_{t,i^*}^{2,*} = r_{t,i^*}^1 - f_1(\mathbf{x}_{t,i^*}; \widehat{\boldsymbol{\theta}}_{t-1}^{1,*})$. Then, we provide the following regret bound that depends on the classification ability of exploration network class induced by $\mathcal{B}(\boldsymbol{\theta}_0^2, \nu)$.

Theorem 4.1. *Given the number of rounds T, for any $\delta \in (0, 1), \gamma > 1$, suppose $m \geq \Omega(poly(T, k, L, \nu)), \eta_1 = \eta_2 = \Theta(\frac{\nu L}{\sqrt{tm}})$. Then, with probability at least $1 - \delta$ over the initialization of $\boldsymbol{\theta}_0^1, \boldsymbol{\theta}_0^2$, for any $\nu > 0$, and suppose*

$\inf_{\widetilde{\boldsymbol{\theta}}^2 \in \mathcal{B}(\boldsymbol{\theta}_0^2, \nu)} \sum_{t=1}^{T} \left(f_2(\phi(\mathbf{x}_{t,\widehat{i}}); \widetilde{\boldsymbol{\theta}}^2) - r_{t,\widehat{i}}^2 \right)^2 \& \inf_{\widetilde{\boldsymbol{\theta}}^2 \in \mathcal{B}(\boldsymbol{\theta}_0^2, \nu)} \sum_{t=1}^{T} \left(f_2(\phi(\mathbf{x}_{t,i^*}); \widetilde{\boldsymbol{\theta}}^2) - r_{t,i^*}^{2,*} \right)^2 \leq \mu$, *Algorithm 1 achieves the following regret bound:*

$$R_T \leq \mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu}}{\sqrt{2T}}\right) + 2\sqrt{\frac{2\log(\mathcal{O}(Tk)/\delta)}{T}}. \tag{4.1}$$

$$\mathbf{R}_T \leq \mathcal{O}\left(2\sqrt{T} - 1\right)\left[\frac{6L\nu + 4\sqrt{\mu}}{\sqrt{2}} + 2\sqrt{2\log(\mathcal{O}(Tk)/\delta)} + \mathcal{O}(1)\right] \tag{4.2}$$

and at the same time $\mathbf{N}_T \leq \mathcal{O}(T)$.

Theorem 4.1 provides the regret bound of I-NeurAL for R_T and $\mathbf{R}_T$ respectively, $R_T \leq \mathcal{O}(\frac{\sqrt{\log T}}{\sqrt{T}})$ and $\mathbf{R}_T \leq \mathcal{O}(\sqrt{T \log T})$. As [50] only provides the regret bound for $\widetilde{R}_T$, to show the advantages of I-NeurAL, we also provide the following lemma for fair comparison.

Lemma 4.1. *Given the number of rounds T, for any $\delta \in (0,1), \gamma > 1$, suppose $m \geq$ poly$(T, k, L, \nu, \rho^{-1}, e^{\sqrt{\log(Tk/\delta)}}], \eta_1 = \eta_2 = \Theta(\frac{\nu\rho}{\sqrt{tm}})$. Then, with probability at least $1-\delta$ over the initialization of $\boldsymbol{\theta}_0^1, \boldsymbol{\theta}_0^2$, for any $\nu > 0$ and suppose μ satisfies the conditions in Theorem 4.1, Algorithm 1 can achieve the following regret bound:*

$$\widetilde{\mathbf{R}}_T \leq \mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu}}{\sqrt{2}}\right)\sqrt{T} + 2\sqrt{2T\log(\mathcal{O}(T)/\delta)} + \mathcal{O}(1) \tag{4.3}$$

and at the same time $\mathbf{N}_T \leq \mathcal{O}(T)$.

Comparison with [50]. Lemma 4.1 shows that I-NeurAL can achieve the regret bound of same complexity for $\widetilde{\mathbf{R}}_T$ as $\mathbf{R}_T$. Under the same assumption in the over-parameterized neural networks, without any assumption on $\mathcal{D}$, Theorem 1 in [50] (i.e., the lower-noise condition with exponent $\alpha = 0$, and $k = 2$ is ignored in the binary classification) achieves the following regret bound: $\widetilde{\mathbf{R}}_T \leq \mathcal{O}(\log\det(I + \mathbf{H})\sqrt{T(\log\det(I + \mathbf{H}) + S^2)})$ where $\mathbf{H}$ is the NTK matrix [29, 4] formed by received instances of all T rounds, $S = \sqrt{\mathbf{h}^\top \mathbf{H}^{-1}\mathbf{h}}$ is a complexity term, and $\mathbf{h} = (h(\mathbf{x}_{1,\widehat{i}}), \ldots, h(\mathbf{x}_{T,\widehat{i}}))^\top \in \mathbb{R}^T$. Note that I-NeurAL and [50] have the same trivial label complexity $\mathcal{O}(T)$ in this difficult case. According to the definition of effective dimension $\widetilde{d}$ in [57], the above regret bound obtained by [50] can be represented by:

$$\widetilde{\mathbf{R}}_T \leq \mathcal{O}(\widetilde{d}\log(1+T))\sqrt{T(\widetilde{d}\log(1+T) + S^2)} \quad \text{and} \quad \widetilde{d} = \frac{\log\det(I + \mathbf{H})}{\log(1+T)} \tag{4.4}$$

Remark 4.1. The term μ reflects the possible minimal convergence error caused by the functions induced the parameters in $\mathcal{B}(\boldsymbol{\theta}_0^2, \nu)$ controlled by ν. Such complexity term is first introduced in [17]. When ν is small, the corresponding ball $\mathcal{B}(\boldsymbol{\theta}_0^2, \nu)$ is small, so μ will be large; Otherwise, when ν is large, μ will be small. In particular, when setting $\nu = \mathcal{O}(1)$, Theorem 4.1 and Lemma 4.1 suggests that if the data can be learned by a function in the function class formed by $\mathcal{B}(\boldsymbol{\theta}_0^2, \mathcal{O}(1))$ with the small training error, then I-NeurAL will have small regret. Note that [50] has the complexity term S as well, to reflect the boundary of optimal parameters specific to the data.

Remark 4.2. Theorem 4.1 and Lemma 4.1 do not depend on $\widetilde{d}$. The effective dimension $\widetilde{d}$ was first introduced in [48] and then used in [57], which can be thought of as the non-linear dimensions in the NTK kernel space. However, $\widetilde{d}$ can be $p = m + mkd + m^2(L-1)$ in the worst case, i.e., $\widetilde{d} \gg T$ (see details in Appendix 9). Eq.(4.4) has the term $\mathcal{O}(\widetilde{d})$ and thus the regret bound obtained by [50] can explode due to $\widetilde{d}$. This is because the analysis of [50] closely depends on NTK, i.e., to apply Confidence Ellipsoid bound (Theorem 2 in [1]) to the NTK approximation. This procedure inevitably bind their regret bound to the determinant of NTK that can have a very large magnitude. In contrast, Eq.(4.3) does not have the term $\widetilde{d}$, because our analysis does not depend on the NTK approximation and I-NeurAL directly utilizes the property of over-parameterized neural networks, i.e., the convergence error μ and the generalization concentration bound (Lemma 7.6). These two terms are independent of $\widetilde{d}$, which paves the way for I-NeurAL to remove the curse of $\widetilde{d}$.

Remark 4.3. Theorem 4.1 and Lemma 4.1 improve the regret by a multiplicative factor $\mathcal{O}(\log T)$ over [50]. Note that the analysis of [50] is built for binary classification and thus $k = 2$ in Theorem 4.1 and Lemma 4.1. This improvement stems from the different analysis workflow of I-NeurAL from [50]. Again, our analysis does not rely on NTK approximation and it is built on the convergence and generalization bound of wide neural networks.

Remark 4.4. Our proof workflow of Theorem 4.1 and Lemma 4.1 is inspired by [14]. Compared to [14], we provide the first regret bound supporting mini-batch SGD with warm-start and a more generic generalization bound (Lemma 7.3) that holds for every arm (class). Moreover, we carry out the performance analysis of query decision-maker (Lemma 7.5), which is a new addition.

For the label complexity, $\mathbf{N}_T$ has the trivial $\mathcal{O}(T)$ complexity which is the same as Theorem 1 in [50] (with the exponent $\alpha = 0$). Because we have to consider the worst case where the unique Bayes-optimal class does not exist, i.e., given $\mathbf{x}_{t,i}, i \in [k]$, there does not exist i^* such that $h(\mathbf{x}_{t,i^*}) > h(\mathbf{x}_{t,i}), \forall i \in [k] \setminus \{i^*\}$. Therefore, we provide the following analysis and show that $\mathbf{R}_T$ and $\mathbf{N}_T$ can be upper bounded by constants as long as there exists a unique Bayes-optimal class for the input instances, described by the following mild margin assumption.

Assumption 4.2 (ϵ-margin)**.** In round $t \in [T]$, given an instance $\mathbf{x}_t$ and the label y_t, then $\mathbf{x}_t$ has the ϵ-Unique optimal class if there exists $\epsilon > 0$ such that

$$\mathbb{P}(\mathbf{y}_{t,y_t} = \mathbf{y}_{t,i^*}|\mathbf{x}_t) - \mathbb{P}(\mathbf{y}_{t,y_t} = \mathbf{y}_{t,i^\circ}|\mathbf{x}_t) \geq \epsilon, \tag{4.5}$$

where $i^* = \arg\max_{i\in[k]} h(\mathbf{x}_{t,i})$ is the Bayes-optimal class and $i^\circ = \arg\max_{i\in([k]\setminus\{i^*\})} h(\mathbf{x}_{t,i})$ is the second Bayes-optimal class.

Given any $i \in [k]$, let i be a fixed index, i.e., suppose there exist a policy Ω_i which always select the i-th context $(\mathbf{x}_{t,i}, r^1_{t,i})$ for every round $t \in [T]$. Then, in round t, we have the collected data by Ω_i: $\mathcal{H}^{1,i}_{t-1} = \{\mathbf{x}_{\tau,i}, r^1_{\tau,i}\}_{\tau=1}^{t-1}$. Then, let $\widehat{\boldsymbol{\theta}}^{1,i}_{t-1}$ represent the parameters trained only on $\mathcal{H}^{1,i}_{t-1}$ using Algorithm 2 and $r^{2,i}_{t,i} = r^1_{t,i} - f_1(\mathbf{x}_{t,i}; \widehat{\boldsymbol{\theta}}^{1,i}_{t-1})$.

Theorem 4.2. *Suppose the instances drawn from D satisfy Assumption 4.2. Then, given the number of rounds T, for any $\delta \in (0,1), \gamma > 1, \epsilon \in (0,1)$, suppose $m \geq$ poly$(T, k, L, \nu, \rho^{-1}, e^{\sqrt{\log(Tk/\delta)}}), \eta_1 = \eta_2 = \Theta(\frac{\nu\rho}{\sqrt{tm}})$. Then, with probability at least $1-\delta$ over the initialization of $\boldsymbol{\theta}^1_0, \boldsymbol{\theta}^2_0$, for any $\nu > 0$, suppose μ satisfies the conditions in Theorem 4.1 and $\max_{i\in k} \left\{ \inf_{\widetilde{\boldsymbol{\theta}}^2 \in \mathcal{B}(\boldsymbol{\theta}^2_0, \nu)} \sum_{t=1}^T \left(f_2\left(\mathbf{x}_{t,\widehat{i}}; \widetilde{\boldsymbol{\theta}}^2\right) - r^{2,i}_{t,i} \right)^2 \right\} \leq \mu$. Then, Algorithm 1 achieves the following regret bound:*

$$\begin{cases} R_T \leq \mathcal{O}\left(\frac{6(L+1)}{\sqrt{2T}}\right) + 2\sqrt{\frac{2\log(\mathcal{O}(T)/\delta)}{T}}, \ \textit{if } T \leq \frac{48(\gamma+1)^2}{\epsilon^2}\xi; \\ R_T = 0, \textit{else}. \end{cases} \qquad \mathbf{N}_T \leq \frac{48(\gamma+1)^2}{\epsilon^2} \cdot \xi \tag{4.6}$$

$$\mathbf{R}_T \leq \mathcal{O}\left(\frac{8\sqrt{3}(\gamma+1)}{\epsilon}\sqrt{\xi} - 1\right)\left[\frac{3L}{\sqrt{2}} + \sqrt{4\log(\frac{4\sqrt{3}(\gamma+1)}{\epsilon\delta}) + 2\log(\frac{\mathcal{O}(\xi)}{\delta})}\right] + \mathcal{O}(1) \tag{4.7}$$

where $\xi = 2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon} + \log(\mathcal{O}(k)) + \frac{9}{4}(\mathcal{O}(L^2\nu^2) + \frac{4\mu}{9}) - \log\delta$.

Remark 4.5. Theorem 4.2 provides the upper bound for $\mathbf{R}_T$ that is independent of T. When other parameters are fixed, this indicates $\mathbf{R}_T$ is upper bounded by a constant when $T \to +\infty$. Moreover, the analysis of R_T indicates that I-NeurAL can achieve the same performance as Bayes-optimal classifier with high confidence after a fixed number of rounds (i.e. $T > \frac{48(\gamma+1)^2}{\epsilon^2}\xi$). To the best of our knowledge, this is the first analysis for neural active learning to show that an NN-based algorithm can achieve the performance as the Bayes-optimal classifier after a fixed number of rounds. In Theorem 1 of [50] (with the exponent $\alpha \to +\infty$ equivalent to Assumption 4.2), $\widetilde{\mathbf{R}}_T \leq \mathcal{O}(\widetilde{d}\log(1+T))\sqrt{(\widetilde{d}\log(1+T) + S^2)}$ that still is dependent on T because NTK depends on T and $\widetilde{d}$.

	Phishing	IJCNN	Letter	Fashion	MNIST	CIFAR-10
Random	1095	845	3519	444	1599	1910
Margin	704	974	3164	247	1327	1474
NeuAL-NTK-F	4898	2684	6066	6001	6192	5007
NeuAL-NTK-D	796	744	3410	742	1239	1700
ALPS	978	683	3108	379	1433	1662
I-NeurAL	$\mathbf{689}(\uparrow \mathbf{2.1}\%)$	$\mathbf{469}(\uparrow \mathbf{31.3}\%)$	$\mathbf{2571}(\uparrow \mathbf{17.3}\%)$	$\mathbf{118}(\uparrow \mathbf{52.2}\%)$	$\mathbf{674}(\uparrow \mathbf{45.6}\%)$	$\mathbf{1372}(\uparrow \mathbf{6.9}\%)$

Table 1: Total regret comparison.

5 Experiments

In this section, we evaluate I-NeurAL on public classification data sets compared with state-of-the-art (SOTA) baselines. Due to the space limit, we only report the main results here and leave the implementation details and parameter sensitivity in the Appendix 10.

We report the experimental results on the following six data sets: Phishing[1], IJCNN [40], Letter [19], Fashion [51], MNIST [34] and CIFAR-10 [33]. In each round, one instance is randomly drawn from the data set and the algorithm is compelled to make prediction on it. Then, the regret is 1 if the prediction does not match the label; the regret is 0, otherwise. At the same time, if the algorithm decides to observe the label, it costs one query budget. As the algorithm may abusively make label queries, we restrict the query budget to 3% of the total number of instances in the data set for fair comparison.

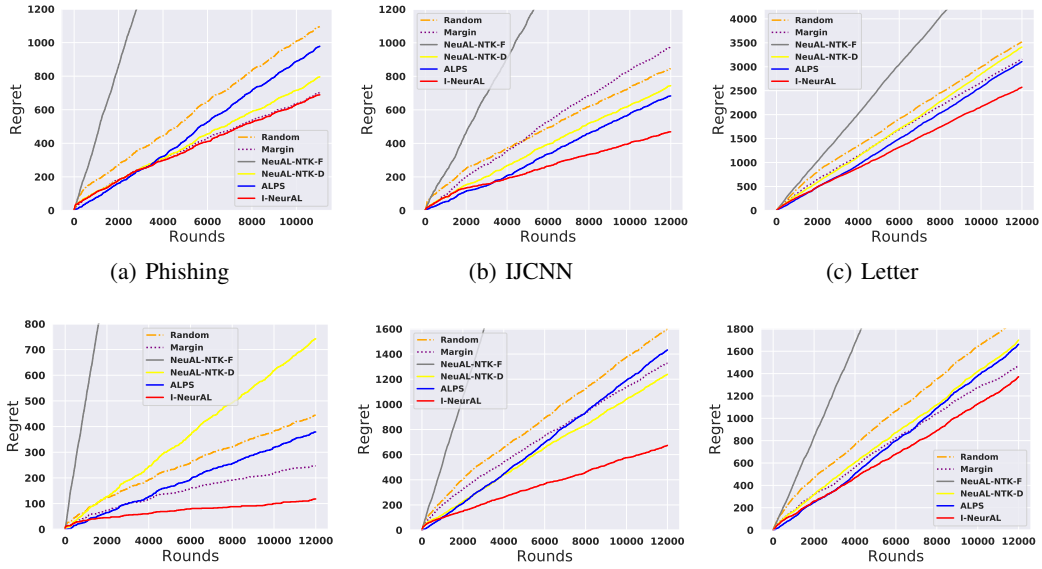
The compared baselines are described as follows. (1) **Random**: The NN classifier queries the label with a fixed probability p until the query budget is exhausted; (2) **Margin**: The NN classifier queries the label when the predicted probability is lower than a threshold. These two baselines are used in [23]. (3) **NeuAL-NTK-F (Algorithm 1 in [50]**: This model makes predictions based on the frozen NTK approximation coming with an Upper-Confidence-Bound(UCB)-based exploration strategy. (4) **NeuAL-NTK-D (Algorithm 3 in [50])**: The prediction is made based on the NN classifier with a UCB while the NTK is updated accordingly. (5) **ALPS [23]**: Given a class of pre-trained hypotheses, the hypothesis minimizing the logistic loss of labeled and pseudo-labeled data is chosen to make predictions and the label query is based on the disagreement of different hypotheses.

Results. The regret comparison on six data sets is shown in Table 1 and Figure 1. I-NeurAL consistently outperforms all baselines across all data sets. In particular, I-NeurAL surpasses the best baseline by 31.3%, 45.6%, 52.2% on IJCNN, MNIST, Fashion respectively. Since NeuAL-NTK-F uses frozen NTK approximation, the new knowledge of each round is barely utilized by the neural network and thus it turns into the worst baseline. NeuAL-NTK-D updates the network parameters with gradient descent and queries the label based on the uncertainty estimation. However, its upper confidence bound is still based on the confidence ellipsoid. Instead, I-NeurAL leverages the representation power of neural networks for both exploitation and exploration. ALPS maintains a class of pre-trained hypotheses and tries to make the best decisions based on these hypotheses. Nevertheless, the model parameters are fixed before the online active learning process. Hence, ALPS is not able to take the new knowledge obtained by queries into account and its performance is highly restricted by the hypothesis class. Although Margin algorithm is simple and straightforward, it exhibits great empirical performance in practice. This observation is consistent with other studies [52] [23]. However, Margin algorithm does not incorporate the exploitation portion and the query criterion is not adaptive to difference instances, thus still outperformed by I-NeurAL.

6 Conclusion

In this paper, we introduce two regret metrics and propose a novel neural-based algorithm (I-NeurAL) tailored for the streaming setting of non-parametric active learning. We carefully design its exploration strategy, query decision-maker, update rules, and training procedure, which lead to both the theoretical and empirical improvement compared to SOTA [50]. In the regret analysis, we remove the dependence on either the input dimensionality or effective dimensionality, and improve over [50] by a multiplicative factor of $\mathcal{O}(\log T)$. On the other hand, we empirically show that I-NeurAL consistently achieves better accuracy under the same query budget than the strong baselines including the SOTA work [50] and [23].

[1] `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html`



(a) Phishing (b) IJCNN (c) Letter

(d) Fashion (e) MNIST (f) CIFAR-10

Figure 1: Regret comparison on six data sets. I-NeurAL outperforms all baselines.

Acknowledgements

This work is supported by NSF (IIS-1947203, IIS-2117902, IIS-2137468, IIS-2002540, DMS-2134079, IIS-2131335, OAC-2130835, and DBI-2021898), DARPA (HR001121C0165), ARO (W911NF2110088), and C3.ai. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agencies or the government.

References

[1] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

[2] Z. Allen-Zhu, Y. Li, and Z. Song. A convergence theory for deep learning via over-parameterization. In *International Conference on Machine Learning*, pages 242–252. PMLR, 2019.

[3] A. Antos, V. Grover, and C. Szepesvári. Active learning in heteroscedastic noise. *Theoretical Computer Science*, 411(29-30):2712–2728, 2010.

[4] S. Arora, S. S. Du, W. Hu, Z. Li, R. R. Salakhutdinov, and R. Wang. On exact computation with an infinitely wide neural net. In *Advances in Neural Information Processing Systems*, pages 8141–8150, 2019.

[5] J. Ash, S. Goel, A. Krishnamurthy, and S. Kakade. Gone fishing: Neural active learning with fisher embeddings. *Advances in Neural Information Processing Systems*, 34, 2021.

[6] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*, 2019.

[7] P. Awasthi, M. F. Balcan, and P. M. Long. The power of localization for efficiently learning linear separators with noise. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 449–458, 2014.

[8] M.-F. Balcan, A. Beygelzimer, and J. Langford. Agnostic active learning. *Journal of Computer and System Sciences*, 75(1):78–89, 2009.

[9] M.-F. Balcan, A. Broder, and T. Zhang. Margin based active learning. In *International Conference on Computational Learning Theory*, pages 35–50. Springer, 2007.

[10] Y. Ban and J. He. Generic outlier detection in multi-armed bandit. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 913–923, 2020.

[11] Y. Ban and J. He. Local clustering in contextual multi-armed bandits. In *Proceedings of the Web Conference 2021*, pages 2335–2346, 2021.

[12] Y. Ban, J. He, and C. B. Cook. Multi-facet contextual bandits: A neural network perspective. In *The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, pages 35–45, 2021.

[13] Y. Ban, Y. Qi, T. Wei, and J. He. Neural collaborative filtering bandits via meta learning. *ArXiv abs/2201.13395*, 2022.

[14] Y. Ban, Y. Yan, A. Banerjee, and J. He. EE-net: Exploitation-exploration neural networks in contextual bandits. In *International Conference on Learning Representations*, 2022.

[15] A. Beygelzimer, S. Dasgupta, and J. Langford. Importance weighted active learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 49–56, 2009.

[16] K. Brinker. Incorporating diversity in active learning with support vector machines. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 59–66, 2003.

[17] Y. Cao and Q. Gu. Generalization bounds of stochastic gradient descent for wide and deep neural networks. *Advances in Neural Information Processing Systems*, 32:10836–10846, 2019.

[18] G. Citovsky, G. DeSalvo, C. Gentile, L. Karydas, A. Rajagopalan, A. Rostamizadeh, and S. Kumar. Batch active learning at scale. *Advances in Neural Information Processing Systems*, 34, 2021.

[19] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik. Emnist: Extending mnist to handwritten letters. In *2017 international joint conference on neural networks (IJCNN)*, pages 2921–2926. IEEE, 2017.

[20] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

[21] A. Culotta and A. McCallum. Reducing labeling effort for structured prediction tasks. In *AAAI*, volume 5, pages 746–751, 2005.

[22] S. Dasgupta, A. T. Kalai, and C. Monteleoni. Analysis of perceptron-based active learning. In *International conference on computational learning theory*, pages 249–263. Springer, 2005.

[23] G. DeSalvo, C. Gentile, and T. S. Thune. Online active learning with surrogate loss functions. *Advances in Neural Information Processing Systems*, 34, 2021.

[24] S. Du, J. Lee, H. Li, L. Wang, and X. Zhai. Gradient descent finds global minima of deep neural networks. In *International Conference on Machine Learning*, pages 1675–1685. PMLR, 2019.

[25] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[26] S. Hanneke. A bound on the label complexity of agnostic active learning. In *Proceedings of the 24th international conference on Machine learning*, pages 353–360, 2007.

[27] S. Hanneke et al. Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, 7(2-3):131–309, 2014.

[28] S. Hanneke and L. Yang. Surrogate losses in passive and active learning. *Electronic Journal of Statistics*, 13(2):4646–4708, 2019.

[29] A. Jacot, F. Gabriel, and C. Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in neural information processing systems*, pages 8571–8580, 2018.

[30] A. J. Joshi, F. Porikli, and N. Papanikolopoulos. Multi-class active learning for image classification. In *2009 ieee conference on computer vision and pattern recognition*, pages 2372–2379. IEEE, 2009.

[31] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell. Active learning with gaussian processes for object categorization. In *2007 IEEE 11th international conference on computer vision*, pages 1–8. IEEE, 2007.

[32] Y.-Y. Kim, K. Song, J. Jang, and I.-c. Moon. Lada: Look-ahead data acquisition via augmentation for deep active learning. *Advances in Neural Information Processing Systems*, 34, 2021.

[33] A. Krizhevsky, G. Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[34] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[35] D. D. Lewis and W. A. Gale. A sequential algorithm for training text classifiers. In *SIGIR'94*, pages 3–12. Springer, 1994.

[36] A. Locatelli, A. Carpentier, and S. Kpotufe. Adaptivity to noise parameters in nonparametric active learning. In *Proceedings of the 2017 Conference on Learning Theory, PMLR*, 2017.

[37] S. Minsker. Plug-in approach to active learning. *Journal of Machine Learning Research*, 13(1), 2012.

[38] J. Moon, J. Kim, Y. Shin, and S. Hwang. Confidence-aware learning for deep neural networks. In *international conference on machine learning*, pages 7034–7044. PMLR, 2020.

[39] S. Mussmann and P. S. Liang. Uncertainty sampling is preconditioned stochastic gradient descent on zero-one loss. *Advances in Neural Information Processing Systems*, 31, 2018.

[40] D. Prokhorov. Ijcnn 2001 neural network competition. *Slide presentation in IJCNN*, 1(97):38, 2001.

[41] Y. Qi, Y. Ban, and J. He. Neural bandit with arm group graph. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, page 1379–1389, New York, NY, USA, 2022. Association for Computing Machinery.

[42] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang. A survey of deep active learning. *ACM Computing Surveys (CSUR)*, 54(9):1–40, 2021.

[43] N. Roy and A. McCallum. Toward optimal active learning through monte carlo estimation of error reduction. *ICML, Williamstown*, 2:441–448, 2001.

[44] C. Schröder and A. Niekler. A survey of active learning for text classification using deep neural networks. *arXiv preprint arXiv:2008.07267*, 2020.

[45] O. Sener and S. Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

[46] B. Settles. Active learning literature survey. 2009.

[47] W. Tan, L. Du, and W. Buntine. Diversity enhanced active learning with strictly proper scoring rules. *Advances in Neural Information Processing Systems*, 34, 2021.

[48] M. Valko, N. Korda, R. Munos, I. Flaounas, and N. Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

[49] H. Wang, W. Huang, A. Margenot, H. Tong, and J. He. Deep active learning by leveraging training dynamics. *arXiv preprint arXiv:2110.08611*, 2021.

[50] Z. Wang, P. Awasthi, C. Dann, A. Sekhari, and C. Gentile. Neural active learning with performance guarantees. *Advances in Neural Information Processing Systems*, 34, 2021.

[51] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.

[52] Y. Yang and M. Loog. A benchmark and comparison of active learning for logistic regression. *Pattern Recognition*, 83:401–415, 2018.

[53] C. Zhang. Efficient active learning of sparse halfspaces. In *Conference on Learning Theory*, pages 1856–1880. PMLR, 2018.

[54] C. Zhang, J. Shen, and P. Awasthi. Efficient active learning of sparse halfspaces with arbitrary bounded noise. *Advances in Neural Information Processing Systems*, 33:7184–7197, 2020.

[55] W. Zhang, D. Zhou, L. Li, and Q. Gu. Neural thompson sampling. In *International Conference on Learning Representations*, 2021.

[56] Y. Zhang, H. Tong, Y. Xia, Y. Zhu, Y. Chi, and L. Ying. Batch active learning with graph neural networks via multi-agent deep reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36:9118–9126, 2022.

[57] D. Zhou, L. Li, and Q. Gu. Neural contextual bandits with ucb-based exploration. In *International Conference on Machine Learning*, pages 11492–11502. PMLR, 2020.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [No]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] The random seed is fixed to 42 in all the experiments.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

In this **Appendix**, we first present the proof of Theorem 4.1 and 4.2 in Section 7; second, show the proof of Lemma 4.1 in Section 8; third, provide an upper bound for the effective dimension $\widetilde{d}$ in Section 9; in the end, present the more experiment details in Section 10.

7 Proofs of Theorem 4.1 and 4.2

7.1 Proof of Theorem 4.1

Proof. Let $f(\mathbf{x};\boldsymbol{\theta}) = f_1(\mathbf{x};\boldsymbol{\theta}^1) + f_2(\phi(\mathbf{x});\boldsymbol{\theta}^2)$ and we use $\mathbb{E}_{\mathbf{x}_t,y_t}$ to denote $\mathbb{E}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}$ for brevity. For any $t \in [T] \wedge (\mathbf{I}_t = 1)$, we have

$$
\begin{aligned}
R_t|(\mathbf{I}_t = 1) &= \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}[L(\mathbf{y}_{t,\widehat{i}},\mathbf{y}_{t,y_t}) - L(\mathbf{y}_{t,i^*},\mathbf{y}_{t,y_t})] \\
&= \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[1 - L(\mathbf{y}_{t,i^*},\mathbf{y}_{t,y_t}) - \left(1 - L(\mathbf{y}_{t,\widehat{i}},\mathbf{y}_{t,y_t})\right)\right] \\
&= \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}[r^1_{t,i^*} - r^1_{t,\widehat{i}}] \\
&\overset{E_1}{=} \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}[\min\{r^1_{t,i^*} - r^1_{t,\widehat{i}}, 1\}] \\
&= \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) + f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}\right] \\
&\overset{E_2}{\leq} \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1}) + f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}\right] \\
&\overset{E_3}{=} \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\Big[\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1}) + f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1}) \\
&\quad + f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}\Big] \\
&\leq \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1}), 1\}\right] + \mathop{\mathbb{E}}_{\mathbf{x}_t}[\min\left\{f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1}), 1\right\}] \\
&\quad + \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}\right] \\
&\leq \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{|r^1_{t,i^*} - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1})|, 1\}\right] + \mathop{\mathbb{E}}_{\mathbf{x}_t}[\min\left\{|f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})|, 1\right\}] \\
&\quad + \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{|f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}|, 1\}\right]
\end{aligned}
\tag{7.1}
$$

where E_1 is based on the fact $r_{t,i} \in [0,1], \forall i \in [k]$, E_2 is due to $f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1}) \leq f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1})$ according to our selection criterion, and $\boldsymbol{\theta}^*_{t-1}$ in E_3 are intermediate parameters to bound errors.

For any $t \in [T] \wedge (\mathbf{I}_t = 0)$, we have $R_t|(\mathbf{I}_t = 0) = \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}[L(\mathbf{y}_{t,\widehat{i}},\mathbf{y}_{t,y_t}) - L(\mathbf{y}_{t,i^*},\mathbf{y}_{t,y_t})] = 0$ based on Lemma 7.5.

Therefore, for any $t \in [T]$, we have

$$
\begin{aligned}
R_t \leq &\mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{|r^1_{t,i^*} - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1})|, 1\}\right] + \mathop{\mathbb{E}}_{\mathbf{x}_t}[\min\left\{|f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})|, 1\right\}] \\
&+ \mathop{\mathbb{E}}_{\mathbf{x}_t,y_t}\left[\min\{|f(\mathbf{x}_{t,i_t};\boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}|, 1\}\right]
\end{aligned}
\tag{7.2}
$$

Based on Lemma 7.6, Lemma 7.7, and Lemma 7.14, with probability at least $1-\delta$, we have

$$
R_t \leq 2\left(\mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2t}}\right) + \mathcal{O}\left(\sqrt{\frac{2\log(\mathcal{O}(k)/\delta)}{t}}\right) + \xi_t\right), \tag{7.3}
$$

where

$$
\xi_t = \left(1 + \mathcal{O}\left(\frac{tL^3\log^{5/6} m}{\rho^{1/3}m^{1/6}}\right)\right)\mathcal{O}\left(\frac{Lt^3}{\rho\sqrt{m}}\log m\right) + \mathcal{O}\left(\frac{t^4L^2\log^{11/6} m}{\rho^{4/3}m^{1/6}}\right). \tag{7.4}
$$

Applying the union bound over all the rounds, with probability at least $1-\delta$, we have

$$\forall t \in [T], \quad R_t \leq 2\left(\mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2t}}\right) + \mathcal{O}\left(\sqrt{\frac{2\log(\mathcal{O}(tk)/\delta)}{t}})\right) + \xi_t.\right) \tag{7.5}$$

When $m \geq \widetilde{\Omega}(T^{27})$, we have $\xi_t = \mathcal{O}(\frac{1}{\sqrt{T}})$. Therefore, in round T, we have

$$R_T \leq \mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu})}{\sqrt{2T}}\right) + \mathcal{O}\left(\sqrt{\frac{2\log(\mathcal{O}(Tk)/\delta)}{T}}\right). \tag{7.6}$$

Finally, the regret of T rounds is

$$\begin{aligned}
\mathbf{R}_T =& \sum_{t=1}^{T} R_t \\
\leq& \sum_{t=1}^{T} 2\left(\underbrace{\mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(\mathcal{O}(Tk)/\delta)}{t}}}_{I_1} + \underbrace{\xi_t}_{I_2}\right) \\
\leq& 2\left(\underbrace{(2\sqrt{T}-1)\left[\mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2}}\right) + \sqrt{2\log(\mathcal{O}(Tk)/\delta)} + \underbrace{\mathcal{O}(1)}_{I_2}\right]}_{I_1}\right)
\end{aligned} \tag{7.7}$$

where I_1 is due to $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \int_1^T \frac{1}{\sqrt{t}}\,dx + 1 = 2\sqrt{T} - 1$ and I_2 is because of the choice of m ($m \geq \widetilde{\Omega}(T^{27})$). The proof is complete. □

7.2 Proof of Theorem 4.2

Proof. Given $\bar{\mathcal{T}} = \frac{48(\gamma+1)^2}{\epsilon^2}\left[2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon} + \log C_2 k + \frac{9}{4}(C_1^2L^2\nu^2 + \frac{4\mu}{9}) - \log\delta\right]$, suppose $T > \bar{\mathcal{T}}$, we have

$$\begin{aligned}
\mathbf{R}_T =& \sum_{t=1}^{T} R_t \\
=& \sum_{t=1}^{\bar{\mathcal{T}}} R_t + \sum_{t=\bar{\mathcal{T}}+1}^{T} R_t \\
=& \underbrace{\sum_{t=1}^{\bar{\mathcal{T}}} \mathop{\mathbb{E}}_{\mathbf{x}_t, y_t}[L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t}) - L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})]}_{I_1} + \underbrace{\sum_{t=\bar{\mathcal{T}}+1}^{T} \mathop{\mathbb{E}}_{\mathbf{x}_t, y_t}[L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t}) - L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})]}_{I_2}
\end{aligned} \tag{7.8}$$

For I_1, based on Eq.(7.5), for any $\mu \in (0,1), t \in [\bar{\mathcal{T}}]$, we have

$$\begin{aligned}
I_1 \leq& \sum_{t=1}^{\bar{\mathcal{T}}} 2\left(\mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(\mathcal{O}(tk)/\delta)}{t}} + \xi_t.\right) \\
\overset{E_1}{\leq}& (2\sqrt{\bar{\mathcal{T}}} - 1)\left[\mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu})}{\sqrt{2}}\right) + \sqrt{2\log(\mathcal{O}(\bar{\mathcal{T}}k)/\delta)}\right]
\end{aligned} \tag{7.9}$$

where E_1 is because of $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \int_1^T \frac{1}{\sqrt{t}}\,dx + 1 = 2\sqrt{T} - 1$ and the choice of m. It is straight forward to show that R_T also satisfies this upper bound when $T \leq \bar{\mathcal{T}}$. For I_2, based on the Lemma 7.1, we have $\mathbb{E}_{\mathbf{x}_t \sim \mathcal{D}_{\mathcal{X}}}[h(\mathbf{x}_{t,\widehat{i}}) - h(\mathbf{x}_{t,i^*}))] = 0$ when $t \geq \bar{\mathcal{T}}$. This implies

$$\mathbb{E}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t}) - L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})] = 0. \tag{7.10}$$

Therefore, we have $I_2 = 0$. Putting them together, we have

$$\mathbf{R}_T \leq (2\sqrt{\bar{\mathcal{T}}} - 1)\left[\mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu}}{\sqrt{2}}\right) + \sqrt{2\log(\mathcal{O}(\bar{\mathcal{T}})/\delta)} + \mathcal{O}(\frac{1}{\bar{\mathcal{T}}})\right]. \tag{7.11}$$

According to Eq.(7.5) and Eq.(7.10), we have

$$\begin{cases} R_T \leq \mathcal{O}\left(\frac{6L\nu+4\sqrt{\mu}}{\sqrt{2T}}\right) + 2\sqrt{\frac{2\log(\mathcal{O}(T)/\delta)}{T}}, \text{ if } T \leq \bar{\mathcal{T}}; \\ R_T = 0, \text{else.} \end{cases} \tag{7.12}$$

Then, replace $\bar{\mathcal{T}}$ and the proof is complete. □

7.3 Main Lemmas

Lemma 7.1. *For any $\delta \in (0,1)$, $\gamma \geq 1$, suppose $T \geq \bar{\mathcal{T}}$. Then, with probability at least $1-\delta$, these exist constants C_1, C_2, such that the following two event $\mathcal{E}_1, \mathcal{E}_2$ happens*

$$\mathcal{E}_1 = \left\{t \geq \bar{\mathcal{T}}, \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*}) - h(\mathbf{x}_{t,\widehat{i}})] = 0\right\}, \tag{7.13}$$

$$\mathcal{E}_2 = \left\{t \geq \bar{\mathcal{T}}, \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1})] = 0\right\}, \tag{7.14}$$

where

$$\bar{\mathcal{T}} = \frac{48(\gamma+1)^2}{\epsilon^2}\left[2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon} + \log C_2 k + \frac{9}{4}(C_1^2 L^2 \nu^2 + \frac{4\mu}{9}) - \log\delta\right]. \tag{7.15}$$

Proof. According to Lemma 7.3 and Jensen's inequality, for any $i \in [k]$, with probability at least $1-\delta$, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[\min\{|f(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}) - h(\mathbf{x}_{t,i})|, 1\}] &\leq \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[\min\{|f(\mathbf{x}_{t,i}; \boldsymbol{\theta}_{t-1}) - r_{t,j}|, 1\}] \\ &\leq \sqrt{\frac{2\mu}{t}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}} + 2\xi_t \end{aligned} \tag{7.16}$$

In round t, define the event

$$\widehat{\mathcal{E}}_0 = \left\{\tau \in [t], i \in [k], \mathop{\mathbb{E}}_{\mathbf{x}_\tau \sim \mathcal{D}_\mathcal{X}}[\min\{|f(\mathbf{x}_{\tau,i}; \boldsymbol{\theta}_{\tau-1}) - h(\mathbf{x}_{\tau,i})|, 1\}] \leq \boldsymbol{\beta}_\tau\right\} \tag{7.17}$$

Then, applying the union bound over t and k, then, with probability at least $1-\delta$, $\mathcal{E}$ happens, where

$$\boldsymbol{\beta}_\tau = \sqrt{\frac{2\mu}{\tau}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2\tau}}\right) + \sqrt{\frac{2\log(\mathcal{O}(tk)/\delta)}{\tau}}. \tag{7.18}$$

where we merge ξ_τ into $\mathcal{O}\left(\frac{3L}{\sqrt{2\tau}}\right)$ as a result of choice of m. Next, define the event

$$\widehat{\mathcal{E}}_1 = \left\{t \geq \bar{\mathcal{T}}, \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] < 2\gamma\boldsymbol{\beta}_t\right\}. \tag{7.19}$$

When $\widehat{\mathcal{E}}_0$ happens with probability at least $1-\delta$, based on the fact $h(\cdot) \in [0,1]$, we have

$$\begin{cases} \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1})] - \min\{\boldsymbol{\beta}_t, 1\} \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1})] + \min\{\boldsymbol{\beta}_t, 1\} \\ \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] - \min\{\boldsymbol{\beta}_t, 1\} \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] + \min\{\boldsymbol{\beta}_t, 1\} \end{cases} \tag{7.20}$$

Then, based on Lemma 7.2, with probability at least $1-\delta$, when $t > \bar{\mathcal{T}}$, $2(\gamma+1)\boldsymbol{\beta}_t \leq \epsilon \Rightarrow \boldsymbol{\beta}_t < 1$. This implies

$$\begin{cases} \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})] - \boldsymbol{\beta}_t \leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})] + \boldsymbol{\beta}_t \\ \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] - \boldsymbol{\beta}_t \leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] + \boldsymbol{\beta}_t \end{cases} \tag{7.21}$$

Therefore, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*}) - h(\mathbf{x}_{t,i^\circ})] &\leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})] + \boldsymbol{\beta}_t - \left(\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] - \boldsymbol{\beta}_t\right) \\ &\leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] + 2\boldsymbol{\beta}_t. \end{aligned} \tag{7.22}$$

Suppose $\widehat{\mathcal{E}}_1$ happens, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*}) - h(\mathbf{x}_{t,i^\circ})] \leq 2(\gamma+1)\boldsymbol{\beta}_t. \tag{7.23}$$

Then, based on Lemma 7.2, when $t > \bar{\mathcal{T}}$, $2(\gamma+1)\boldsymbol{\beta}_t \leq \epsilon$. Therefore, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*}) - h(\mathbf{x}_{t,i^\circ})] \leq 2(\gamma+1)\boldsymbol{\beta}_t \leq \epsilon. \tag{7.24}$$

This contradicts Assumption 4.2, i.e., $h(\mathbf{x}_{t,i^*}) - h(\mathbf{x}_{t,i^\circ}) \geq \epsilon$. Hence, $\widehat{\mathcal{E}}_1$ will not happen. Accordingly, with probability at least $1-\delta$, the following event will happen

$$\widehat{\mathcal{E}}_2 = \left\{ t \geq \bar{\mathcal{T}}, \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] \geq 2\gamma\boldsymbol{\beta}_t \right\}. \tag{7.25}$$

Therefore, we have $\mathbb{E}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})] > \mathbb{E}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})]$. Recall that $i^* = \arg\max_{i\in[k]} h(\mathbf{x}_{t,i})$ and $\widehat{i} = \arg\max_{i\in[k]} f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})$. As

$$\begin{aligned} &\forall i \in ([k] \setminus \{\widehat{i}\}), f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}) \leq f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1}) \\ \Rightarrow\ &\forall i \in ([k] \setminus \{\widehat{i}\}), \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] \end{aligned} \tag{7.26}$$

we have

$$\forall i \in ([k] \setminus \{\widehat{i}\}), \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})] > \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})]. \tag{7.27}$$

Based on the definition of $\widehat{i}$, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*};\boldsymbol{\theta}_{t-1})] = \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1})] = \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[\max_{i\in[k]} f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})]. \tag{7.28}$$

This indicates $\mathcal{E}_2$ happens with probability at least $1-\delta$.

Therefore, based on $\widehat{\mathcal{E}}_2$, the following inferred event $\widehat{\mathcal{E}}_3$ happens with probability at least $1-\delta$:

$$\widehat{\mathcal{E}}_3 = \left\{ t \geq \bar{\mathcal{T}}, \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] \geq 2\gamma\boldsymbol{\beta}_t \right\}. \tag{7.29}$$

Then, based on Eq. 7.21, we have

$$\begin{aligned} \mathbb{E}[h(\mathbf{x}_{t,\widehat{i}}) - h(\mathbf{x}_{t,i^\circ})] &\geq \mathbb{E}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1})] - \boldsymbol{\beta}_t - (\mathbb{E}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] + \boldsymbol{\beta}_t) \\ &= \mathbb{E}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})] - 2\boldsymbol{\beta}_t \\ &\overset{E_1}{\geq} 2(\gamma-1)\boldsymbol{\beta}_t \\ &\geq 0 \end{aligned} \tag{7.30}$$

where E_1 is because $\widehat{\mathcal{E}}_3$ happened with probability at least $1-\delta$. Therefore, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] - \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})] > 0. \tag{7.31}$$

Similarly, we can prove that

$$\Rightarrow \forall i \in ([k] \setminus \{\widehat{i}\}), \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] - \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i})] > 0. \tag{7.32}$$

Then, based on the definition of $\mathbf{x}_{t,i^*}$, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] = \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*})] = \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[\max_{i\in[k]} h(\mathbf{x}_{t,i})]. \tag{7.33}$$

Thus, the event $\mathcal{E}_1$ happens with probability at least $1-\delta$. □

Lemma 7.2. *When* $t > \mathcal{T} = \frac{48(\gamma+1)^2}{\epsilon^2}\left[2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon} + \log C_2 k + \frac{9}{4}(C_1^2L^2\nu^2 + \frac{4\mu}{9}) - \log\delta\right]$*, it has* $2(\gamma+1)\boldsymbol{\beta}_t \leq \epsilon$.

Proof. As the choice of m, we have $m \leq \mathcal{O}(1/\sqrt{t})$. Thus, to achieve $2(\gamma+1)\boldsymbol{\beta}_t \leq \epsilon$, there exist constants C_1, C_2, such that

$$\begin{aligned}
\sqrt{\frac{2\mu}{t}} + \left(\frac{3C_1L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(C_2Tk/\delta)}{t}} &\leq \frac{\epsilon}{2(\gamma+1)} \\
\left(\sqrt{\frac{2\mu}{t}} + \left(\frac{3C_1L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(C_2Tk/\delta)}{t}}\right)^2 &\leq \left(\frac{\epsilon}{2(\gamma+1)}\right)^2 \\
3\left(\left(\sqrt{\frac{2\mu}{t}}\right)^2 + \left(\frac{3C_1L\nu}{\sqrt{2t}}\right)^2 + \left(\sqrt{\frac{2\log(\mathcal{O}(Tk)/\delta)}{t}}\right)^2\right) &\leq \left(\frac{\epsilon}{2(\gamma+1)}\right)^2
\end{aligned} \tag{7.34}$$

By calculations, we have

$$\log t \leq \frac{t\epsilon^2}{24(\gamma+1)^2} - \log C_2 k. + \log\delta - \frac{9}{4}(C_1^2L^2\nu^2 + \frac{4\mu}{9}) \tag{7.35}$$

Then, based on Lemme 8 in [3] and Lemma 8.1 in [11], we have, when

$$t \geq \frac{48(\gamma+1)^2}{\epsilon^2}\left[2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon} + \log C_2 k + \frac{9}{4}(C_1^2L^2\nu^2 + \frac{4\mu}{9}) - \log\delta\right], \tag{7.36}$$

$2(\gamma+1)\boldsymbol{\beta}_t \leq \epsilon$. □

Lemma 7.3. *For any* $\delta \in (0,1)$*, suppose* m *satisfies the conditions in Theorem 4.1. Then, with probability at least* $1-\delta$*, given any fixed index* $i \in [k]$*, it holds that*

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t)\sim\mathcal{D}}\left[\min\left\{\left|f_1(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}^1) + f_2(\phi(\mathbf{x}_{t,i});\boldsymbol{\theta}_{t-1}^2) - r_{t,i}^1\right|, 1\right\}\right] \\
\leq&\sqrt{\frac{2\mu}{t}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}} + 2\xi_t.
\end{aligned} \tag{7.37}$$

Proof. Given any $i \in [k]$, let i be a fixed index, i.e., suppose there exist a policy Ω_i which always select the i-th context $(\mathbf{x}_{t,i}, r_{t,i}^1)$ for every round $t \in [T]$. Then, in round t, we have the collected data by Ω_i: $\mathcal{H}_{t-1}^{1,i} = \{\mathbf{x}_{\tau,i}, r_{\tau,i}^1\}_{\tau=1}^{t-1}$. Then, let $\boldsymbol{\theta}_{t-1}^{1,i}, \boldsymbol{\theta}_{t-1}^{2,i}$ represent the parameters trained only on $\mathcal{H}_{t-1}^i$ using Algorithm 2, satisfying $\|\boldsymbol{\theta}_{t-1}^{1,i} - \boldsymbol{\theta}_0^1\|_2 \leq \mathcal{O}(\frac{t^3}{\rho\sqrt{m}}\log m)$ and $\|\boldsymbol{\theta}_{t-1}^{2,i} - \boldsymbol{\theta}_0^2\|_2 \leq \mathcal{O}(\frac{t^3}{\rho\sqrt{m}}\log m)$. Note that $\boldsymbol{\theta}_{t-1}^{1,i}, \boldsymbol{\theta}_{t-1}^{2,i}$ are uniformly drawn from $\{\widehat{\boldsymbol{\theta}}_{\tau-1}^{1,i}, \widehat{\boldsymbol{\theta}}_{\tau-1}^{2,i}\}_{\tau=0}^{t-1}$ and these parameters are unknown but introduced for the sake of analysis. Then, for $\tau \in [t]$, we define

$$\begin{aligned}
V_\tau =& \mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau)\sim\mathcal{D}}\left[\min\{|f_1(\mathbf{x}_{\tau,i};\widehat{\boldsymbol{\theta}}_{\tau-1}^{1,i}) + f_2(\phi(\mathbf{x}_{\tau,i});\widehat{\boldsymbol{\theta}}_{\tau-1}^{2,i}) - r_{\tau,i}^1|, 1\}\right] \\
&- \min\{|f_1(\mathbf{x}_{\tau,i};\widehat{\boldsymbol{\theta}}_{\tau-1}^{1,i}) + f_2(\phi(\mathbf{x}_{\tau,i});\widehat{\boldsymbol{\theta}}_{\tau-1}^{2,i}) - r_{\tau,i}^1|, 1\}
\end{aligned} \tag{7.38}$$

Then, we have

$$\begin{aligned}
\mathbb{E}[V_\tau|F_{\tau-1}] =& \mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau)\sim\mathcal{D}}\left[\min\{|f_1(\mathbf{x}_{\tau,i};\widehat{\boldsymbol{\theta}}_{\tau-1}^{1,i}) + f_2(\phi(\mathbf{x}_{\tau,i});\widehat{\boldsymbol{\theta}}_{\tau-1}^{2,i}) - r_{\tau,i}^1|, 1\}\right] \\
&- \mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau)\sim\mathcal{D}}\left[\min\{|f_1(\mathbf{x}_{\tau,i};\widehat{\boldsymbol{\theta}}_{\tau-1}^{1,i}) + f_2(\phi(\mathbf{x}_{\tau,i});\widehat{\boldsymbol{\theta}}_{\tau-1}^{2,i}) - r_{\tau,i}^1|, 1\}\right] \\
=&0
\end{aligned} \tag{7.39}$$

where $F_{\tau-1}$ denotes the σ-algebra generated by the history $\mathcal{H}_{\tau-1}^i$. Therefore, $\{V_\tau\}_{\tau=1}^t$ are the martingale difference sequence.

Then, using the similar proof method in Lemma 7.6, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{\left|f_1(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}^{1,i})+f_2(\phi(\mathbf{x}_{t,i});\boldsymbol{\theta}_{t-1}^{2,i})-r_{t,i}^1\right|,1\right\}\mid\mathcal{H}_{t-1}^i\right]\\ \leq&\sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}}.\end{aligned}\tag{7.40}$$

Let $f(\mathbf{x};\boldsymbol{\theta}_{t-1})=f_1(\mathbf{x};\boldsymbol{\theta}_{t-1}^1)+f_2(\phi(\mathbf{x});\boldsymbol{\theta}_{t-1}^2)$ and $f(\mathbf{x};\boldsymbol{\theta}_{t-1}^i)=f_1(\mathbf{x};\boldsymbol{\theta}_{t-1}^{1,i})+f_2(\phi(\mathbf{x});\boldsymbol{\theta}_{t-1}^{2,i})$. And we have $\|\boldsymbol{\theta}_{t-1}^1-\boldsymbol{\theta}_{t-1}^{1,i}\|_2\leq\mathcal{O}(\frac{t^3}{\rho\sqrt{m}}\log m)$, $\|\boldsymbol{\theta}_{t-1}^2-\boldsymbol{\theta}_{t-1}^{2,i}\|_2\leq\mathcal{O}(\frac{t^3}{\rho\sqrt{m}}\log m)$, based on Lemma 7.13 (2).

Then, given $i\in[k]$, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{|f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})-r_{t,i}^1|,1\right\}\right]\\ =&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{|f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})-f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}^i)+f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}^i)-r_{t,i}^1|,1\right\}\right]\\ \leq&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}[|f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})-f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}^i)|]+\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}[\min\left\{|f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1}^i)-r_{t,i}^1|,1\right\}]\\ \leq&\sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}}+2\xi_t\end{aligned}\tag{7.41}$$

where the last inequality is the application of Lemma 7.14 and Eq. (7.40). The proof is complete.

Lemma 7.4 (Label Complexity Analysis)**.** *For any $\delta\in(0,1),\gamma\geq1$, suppose m satisfies the conditions in Theorem 4.1. Then, with probability at least $1-\delta$, we have*

$$\mathbf{N}_T\leq\frac{48(\gamma+1)^2}{\epsilon^2}\left[2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon}+\log C_2k+\frac{9}{4}(C_1^2L^2\nu^2+\frac{4\mu}{9})-\log\delta\right].\tag{7.42}$$

Proof. Recall that $\mathbf{x}_{t,\widehat{i}}=\max_{\mathbf{x}_{t,i},i\in[k]}f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})$, and $\mathbf{x}_{t,i^\circ}=\max_{\mathbf{x}_{t,i},i\in([k]/\{\mathbf{x}_{t,\widehat{i}}\})}f(\mathbf{x}_{t,i};\boldsymbol{\theta}_{t-1})$. With probability at least $1-\delta$, according to Eq. (7.17) the event

$$\widehat{\mathcal{E}}_0=\left\{\tau\in[t],i\in[k],\mathop{\mathbb{E}}_{\mathbf{x}_\tau\sim\mathcal{D}_\mathcal{X}}[\min\{|f(\mathbf{x}_{\tau,i};\boldsymbol{\theta}_{\tau-1})-h(\mathbf{x}_{\tau,i})|,1\}]\leq\boldsymbol{\beta}_\tau\right\}$$

happens. Therefore, we have

$$\begin{cases}\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})]-\min\{\boldsymbol{\beta}_t,1\}\leq\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1})]\leq\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})]+\min\{\boldsymbol{\beta}_t,1\}\\ \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})]-\min\{\boldsymbol{\beta}_t,1\}\leq\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})]\leq\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})]+\min\{\boldsymbol{\beta}_t,1\}.\end{cases}\tag{7.43}$$

Then, we have

$$\begin{cases}\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1})-f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})]\leq\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})]-\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})]+2\min\{\boldsymbol{\beta}_t,1\}\\ \mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1})-f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})]\geq\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})]-\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})]-2\min\{\boldsymbol{\beta}_t,1\}.\end{cases}\tag{7.44}$$

Let $\epsilon_t=|\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})]-\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})]|$. Then, based on Lemma 7.2, when $t\geq\bar{\mathcal{T}}_t$, we have

$$2(\gamma+1)\boldsymbol{\beta}_t\leq\epsilon_t\leq1,\tag{7.45}$$

where

$$\bar{\mathcal{T}}_t=\frac{48(\gamma+1)^2}{\epsilon_t^2}\left[2\log\frac{2\sqrt{6}(\gamma+1)}{\epsilon_t}+\log C_2k+\frac{9}{4}(C_1^2L^2\nu^2+\frac{4\mu}{9})-\log\delta\right]\tag{7.46}$$

For any $t\in[T]$ and $t<\bar{\mathcal{T}}_t$, we have $\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[\mathbf{I}_t]\leq1$. For the round $t>\bar{\mathcal{T}}_t$, suppose $\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})]-\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})]=-\epsilon_t$, then, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t\sim\mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_{t-1})-f(\mathbf{x}_{t,i^\circ};\boldsymbol{\theta}_{t-1})]\leq-\epsilon_t+2\boldsymbol{\beta}_t\overset{E_1}{\leq}-\epsilon_t+\frac{\epsilon_t}{2}\leq0,\tag{7.47}$$

where E_1 is because of Eq. (7.45) since $\gamma \geq 1$. This contradicts the fact $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] \geq 0$. Therefore, $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] - \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})] = \epsilon_t$. Then, based on Eq.(7.44), we have

$$\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] \geq \epsilon_t - 2\boldsymbol{\beta}_t \overset{E_2}{\geq} 2\gamma\boldsymbol{\beta}_t, \tag{7.48}$$

where E_2 is because of Eq. (7.45).

According to Lemma 7.1, when $t > \bar{\mathcal{T}}$, $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1})] = \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1})]$. Then, applying Eq. (7.48), for the round $t > \bar{\mathcal{T}}$, we have $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[\mathbf{I}_t] = 0$.

Then, assume $T > \bar{\mathcal{T}}$, we have

$$\begin{aligned}
\mathbf{N}_T &= \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}} \left[\mathbb{1}\{f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1}) < 2\gamma\boldsymbol{\beta}_t\} \right] \\
&\leq \sum_{t=1}^{\bar{\mathcal{T}}} 1 + \sum_{t=\bar{\mathcal{T}}+1}^{T} \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}} \left[\mathbb{1}\{f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1}) < 2\gamma\boldsymbol{\beta}_t\} \right] \\
&= \bar{\mathcal{T}} + 0.
\end{aligned} \tag{7.49}$$

Therefore, we have $\mathbf{N}_T \leq \bar{\mathcal{T}}$. □

Lemma 7.5. *For any $\delta \in (0,1), \gamma \geq 1$, suppose m satisfies the conditions in Theorem 4.1. Then, with probability at least $1-\delta$, when $\mathbf{I}_t = 0$, we have*

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] &= \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*})], \\
\mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})] &= \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})].
\end{aligned}$$

Proof. As $\mathbf{I}_t = 0$, we have

$$|f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})| = f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1}) \geq 2\gamma\boldsymbol{\beta}_t$$

When $\widehat{\mathcal{E}}_0$ (Eq. (7.17)) happens with probability at least $1-\delta$, based on the fact $h(\cdot) \in [0,1]$, we have

$$\begin{cases}
\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1})] - \min\{\boldsymbol{\beta}_t, 1\} \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1})] + \min\{\boldsymbol{\beta}_t, 1\} \\
\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] - \min\{\boldsymbol{\beta}_t, 1\} \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^\circ})] \leq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] + \min\{\boldsymbol{\beta}_t, 1\}
\end{cases} \tag{7.50}$$

Then, with probability at least $1-\delta$,we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}}) - h(\mathbf{x}_{t,i^\circ})] &\geq \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - f(\mathbf{x}_{t,i^\circ}; \boldsymbol{\theta}_{t-1})] - 2\min\{\boldsymbol{\beta}_t, 1\} \\
&\geq 2\gamma\boldsymbol{\beta}_t - 2\min\{\boldsymbol{\beta}_t, 1\} \\
&\geq 0
\end{aligned} \tag{7.51}$$

where the last inequality is because of $\gamma \geq 1$. Then, similarly, for any $i' \in ([k] \setminus \{\widehat{i}, i^\circ\})$, we have $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}}) - h(\mathbf{x}_{t,i'})] \geq 0$. Thus, based on the definition of $h(\mathbf{x}_{t,i^*})$, we have $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] = \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,i^*})]$. Because $\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{D}_\mathcal{X}}[h(\mathbf{x}_{t,\widehat{i}})] = \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[1 - L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})]$, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[1 - L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})] = \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[1 - L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})] \\
&\Rightarrow \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})] = \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}}[L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})].
\end{aligned}$$

The proof is complete. □

Lemma 7.6. *For any $\delta \in (0,1), \nu > 0$, suppose m satisfies the conditions in Theorem 4.1. In round $t \in [T]$, given $(\mathbf{x}_t, y_t) \sim \mathcal{D}$, let*

$$\widehat{i} = \arg\max_{i \in [k]} \left(f_1(\mathbf{x}_{t,i}; \boldsymbol{\theta}^1_{t-1}) + f_2(\phi(\mathbf{x}_{t,i}); \boldsymbol{\theta}^2_{t-1}) \right).$$

Then, with probability at least $1-\delta$, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}} \left[\min\left\{ \left| f_1(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}^1_{t-1}) + f_2(\phi(\mathbf{x}_{t,\widehat{i}}); \boldsymbol{\theta}^2_{t-1}) - r^1_{t,\widehat{i}} \right|, 1 \right\} | \mathcal{H}^1_{t-1} \right] \\ &\qquad \leq \mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2t}} \right) + 2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}}, \end{aligned} \tag{7.52}$$

where $\mathcal{H}^1_{t-1} = \{\mathbf{x}_{\tau,\widehat{i}}, r^1_{\tau,\widehat{i}}\}_{\tau=1}^{t-1}$ is historical data and the expectation is taken over $(\boldsymbol{\theta}^1_{t-1}, \boldsymbol{\theta}^2_{t-1})$.

Proof. This lemma is inspired by Lemma 5.1 in [14]. For any round $\tau \in [t]$, define

$$\begin{aligned} V_\tau = &\mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau) \sim \mathcal{D}} \left[\min\{ |f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,\widehat{i}}); \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}}|, 1 \} \right] \\ &- \min\{ |f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,\widehat{i}}); \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}}|, 1 \} \end{aligned} \tag{7.53}$$

Then, we have

$$\begin{aligned} \mathbb{E}[V_\tau | F_{\tau-1}] = &\mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau) \sim \mathcal{D}} \left[\min\{ |f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,\widehat{i}}); \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}}|, 1 \} \right] \\ &- \mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau) \sim \mathcal{D}} \left[\min\{ |f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,\widehat{i}}); \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}}|, 1 \} \right] \\ = &0 \end{aligned} \tag{7.54}$$

where $F_{\tau-1}$ denotes the σ-algebra generated by the history $\mathcal{H}^1_{\tau-1}$. Therefore, $\{V_\tau\}_{\tau=1}^t$ are the martingale difference sequence.

Then, applying the Hoeffding-Azuma inequality, with probability at least $1-\delta$, we have

$$\mathbb{P}\left[\frac{1}{t}\sum_{\tau=1}^t V_\tau - \underbrace{\frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[V_\tau | \mathbf{F}_{\tau-1}]}_{I_1} > \sqrt{\frac{2\log(1/\delta)}{t}} \right] \leq \delta \tag{7.55}$$

As I_1 is equal to 0, we have

$$\begin{aligned} &\frac{1}{t}\sum_{\tau=1}^t \mathop{\mathbb{E}}_{(\mathbf{x}_\tau, y_\tau) \sim \mathcal{D}} \left[\min\{ |f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,\widehat{i}}); \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}}|, 1 \} \right] \\ \leq &\min\{ |f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,\widehat{i}}); \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}}|, 1 \} + \sqrt{\frac{2\log(1/\delta)}{t}} \\ \leq &\frac{1}{t}\sum_{\tau=1}^t \left| f_2\left(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^2_{\tau-1}\right) - \left(r^1_{\tau,\widehat{i}} - f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) \right) \right| + \sqrt{\frac{2\log(1/\delta)}{t}}. \end{aligned} \tag{7.56}$$

Based on the the definition of $\boldsymbol{\theta}^1_{t-1}, \boldsymbol{\theta}^2_{t-1}$, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}(\boldsymbol{\theta}^1, \boldsymbol{\theta}^2)} \left[\min\{ |f_1(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}^1_{t-1}) + f_2(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}^2_{t-1}) - r^1_{t,\widehat{i}}|, 1 \} \right] \\ = &\frac{1}{t}\sum_{\tau=1}^t \mathop{\mathbb{E}}_{(\mathbf{x}_t, y_t) \sim \mathcal{D}} \left[\min\left\{ \left| f_1(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^1_{\tau-1}) + f_2(\mathbf{x}_{\tau,\widehat{i}}; \widehat{\boldsymbol{\theta}}^2_{\tau-1}) - r^1_{\tau,\widehat{i}} \right|, 1 \right\} \right]. \end{aligned} \tag{7.57}$$

Therefore, putting them together, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}(\boldsymbol{\theta}^1,\boldsymbol{\theta}^2)}\mathop{\mathbb{E}}\left[\left|f_1(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}^1_{t-1})+f_2(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}^2_{t-1})-r^1_{t,\widehat{i}}\right|\right]\\&\leq\underbrace{\frac{1}{t}\sum_{\tau=1}^{t}\left|f_2\left(\mathbf{x}_{\tau,\widehat{i}};\widehat{\boldsymbol{\theta}}^2_{\tau-1}\right)-\left(r^1_{\tau,\widehat{i}}-f_1(\mathbf{x}_{\tau,\widehat{i}};\widehat{\boldsymbol{\theta}}^1_{\tau-1})\right)\right|}_{I_2}+\sqrt{\frac{2\log(1/\delta)}{t}}\,.\end{aligned}\tag{7.58}$$

For I_2, based on Lemma 7.8, we have

$$\begin{aligned}I_2&\leq\frac{1}{t}\sum_{\tau=1}^{t}\left|f_2\left(\mathbf{x}_{\tau,\widehat{i}};\widetilde{\boldsymbol{\theta}}^2\right)-\left(r^1_{\tau,\widehat{i}}-f_1(\mathbf{x}_{\tau,\widehat{i}};\widehat{\boldsymbol{\theta}}^1_{\tau-1})\right)\right|+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(1/\delta)}{t}}\\&\leq\frac{1}{t}\sqrt{t}\sqrt{\underbrace{\sum_{\tau=1}^{t}\left(f_2\left(\mathbf{x}_{\tau,\widehat{i}};\widetilde{\boldsymbol{\theta}}^2\right)-\left(r^1_{\tau,\widehat{i}}-f_1(\mathbf{x}_{\tau,\widehat{i}};\widehat{\boldsymbol{\theta}}^1_{\tau-1})\right)\right)^2}_{I_3}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(1/\delta)}{t}}\\&\leq\sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(1/\delta)}{t}}\end{aligned}\tag{7.59}$$

where I_3 is based on $\mathcal{L}(\widetilde{\boldsymbol{\theta}}^2)=\frac{1}{2}\sum_{\tau=1}^{t}\left(f_2\left(\mathbf{x}_{\tau,\widehat{i}};\widetilde{\boldsymbol{\theta}}^2\right)-\left(r^1_{\tau,\widehat{i}}-f_1(\mathbf{x}_{\tau,\widehat{i}};\widehat{\boldsymbol{\theta}}^1_{\tau-1})\right)\right)^2\leq\mu$, the definition of μ.

Combining above Eq. (7.58) and (7.59) together, with probability at least $1-\delta$, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{\left|f_1(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}^1_{t-1})+f_2(\phi(\mathbf{x}_{t,\widehat{i}});\boldsymbol{\theta}^2_{t-1})-r^1_{t,\widehat{i}}\right|,1\right\}\right]\\&\qquad\leq\mathcal{O}\left(\frac{3L\nu+2\sqrt{\mu}}{\sqrt{2t}}\right)+2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}}.\end{aligned}\tag{7.60}$$

where we apply union bound over δ to make above events occur concurrently.

Then, based on Lemma 7.13 (2), it is sufficient to show that $\boldsymbol{\theta}^1_{t-1},\boldsymbol{\theta}^2_{t-1}$ are close to initialization for any $t\in[T]$. The proof is complete. □

Lemma 7.7. *In round $t\in[T]$, given $(\mathbf{x}_t,y_t)\sim\mathcal{D}$, let $i^*=\arg\max_{i\in[k]}h(\mathbf{x}_{\tau,i})$. Let $\boldsymbol{\theta}^{1,*}_{t-1},\boldsymbol{\theta}^{2,*}_{t-1}$ are the parameters trained on $\mathcal{H}^*_{t-1}$ using Algorithm 2. For any $\nu>0$, suppose $\inf_{\widetilde{\boldsymbol{\theta}}^{2,*}\in\mathcal{B}(\theta_0^2,\nu)}\frac{1}{2}\sum_{\tau=1}^{t}\left(f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1})+f_2\left(\phi(\mathbf{x}_{\tau,i^*});\widetilde{\boldsymbol{\theta}}^{2,*}\right)-r^1_{\tau,i^*}\right)^2\leq\mu$. Then, with probability at least $1-\delta$, we have*

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{\left|f_1(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^{1,*}_{t-1})+f_2(\phi(\mathbf{x}_{t,i^*});\boldsymbol{\theta}^{2,*}_{t-1})-r^1_{t,i^*}\right|,1\right\}|\mathcal{H}^*_{t-1}\right]\\&\qquad\leq\mathcal{O}\left(\frac{3L\nu+2\sqrt{\mu}}{\sqrt{2t}}\right)+2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}},\end{aligned}\tag{7.61}$$

where $\mathcal{H}^*_{t-1}=\{\mathbf{x}_{\tau,i^*},r^1_{\tau,i^*}\}_{\tau=1}^{t-1}$ is optimal data of past rounds the expectation is taken over $\boldsymbol{\theta}^{1,*}_{t-1}$, $\boldsymbol{\theta}^{2,*}_{t-1}$.

Proof. This lemma is a direct corollary of Lemma 7.6. For any $\tau\in[t]$, define

$$\begin{aligned}V_\tau=&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\{|f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1})+f_2(\phi(\mathbf{x}_{\tau,i^*});\widehat{\boldsymbol{\theta}}^{2,*}_{\tau-1})-r^1_{\tau,i^*}|,1\}\right]\\&-\min\{|f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1})+f_2(\phi(\mathbf{x}_{\tau,i^*});\widehat{\boldsymbol{\theta}}^{2,*}_{\tau-1})-r^1_{\tau,i^*}|,1\}\end{aligned}\tag{7.62}$$

Then, we have

$$\begin{aligned}\mathbb{E}[V_\tau|F_{\tau-1}] =& \mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\{|f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,i^*});\widehat{\boldsymbol{\theta}}^{2,*}_{\tau-1}) - r^1_{\tau,i^*}|,1\}\right] \\ & - \mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\{|f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,i^*});\widehat{\boldsymbol{\theta}}^{2,*}_{\tau-1}) - r^1_{\tau,i^*}|,1\}\right] \\ =&0\end{aligned} \tag{7.63}$$

where $F_{\tau-1}$ denotes the σ-algebra generated by the history $\mathcal{H}_{\tau-1}$. Therefore, $\{V_\tau\}_{\tau=1}^t$ are the martingale difference sequence.

Then, applying the Hoeffding-Azuma inequality, with probability at least $1-\delta$, we have

$$\mathbb{P}\left[\frac{1}{t}\sum_{\tau=1}^t V_\tau - \underbrace{\frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[V_\tau|\mathbf{F}_\tau]}_{I_1} > \sqrt{\frac{2\log(1/\delta)}{t}}\right] \le \delta \tag{7.64}$$

As I_1 is equal to 0, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}(\boldsymbol{\theta}^1,\boldsymbol{\theta}^2)}\mathop{\mathbb{E}}\left[\min\{|f_1(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^1_{t-1}) + f_2(\phi(\mathbf{x}_{t,i^*});\boldsymbol{\theta}^2_{t-1}) - r^1_{t,i^*}|,1\}\right] \\ =&\frac{1}{t}\sum_{\tau=1}^t \mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{\left|f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1}) + f_2(\phi(\mathbf{x}_{\tau,i^*});\widehat{\boldsymbol{\theta}}^{2,*}_{\tau-1}) - r^1_{\tau,i^*},1\right\}\right|\right] \\ \le&\underbrace{\frac{1}{t}\sum_{\tau=1}^t\left|f_2\left(\phi(\mathbf{x}_{\tau,i^*});\widehat{\boldsymbol{\theta}}^{2,*}_{\tau-1}\right) - \left(r^1_{\tau,i^*} - f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1})\right)\right|}_{I_2} + \sqrt{\frac{2\log(1/\delta)}{t}}.\end{aligned} \tag{7.65}$$

For I_2, applying Lemma 7.8, for any $\widetilde{\boldsymbol{\theta}}^{2,*}$ satisfying $\|\widetilde{\boldsymbol{\theta}}^{2,*} - \boldsymbol{\theta}^2_0\|_2 \le \mathcal{O}(\frac{\nu}{\sqrt{m}})$, with probability at least $1-3\delta$, we have

$$\begin{aligned}I_2 \le& \frac{1}{t}\sum_{\tau=1}^t |f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1}) + f_2\left(\phi(\mathbf{x}_{\tau,i^*});\widetilde{\boldsymbol{\theta}}^{2,*}\right) - r^1_{\tau,i^*}| + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(1/\delta)}{t}} \\ \le& \frac{1}{t}\sqrt{t}\sqrt{\underbrace{\sum_{\tau=1}^t\left(f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1}) + f_2\left(\phi(\mathbf{x}_{\tau,i^*});\widetilde{\boldsymbol{\theta}}^{2,*}\right) - r^1_{\tau,i^*}\right)^2}_{I_3}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(1/\delta)}{t}} \\ \le& \sqrt{\frac{2\mu}{t}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(1/\delta)}{t}}\end{aligned} \tag{7.66}$$

where I_3 is because: $\mathcal{L}(\widetilde{\boldsymbol{\theta}}^{2,*}) = \frac{1}{2}\sum_{\tau=1}^t\left(f_1(\mathbf{x}_{\tau,i^*};\widehat{\boldsymbol{\theta}}^{1,*}_{\tau-1}) + f_2\left(\phi(\mathbf{x}_{\tau,i^*});\widetilde{\boldsymbol{\theta}}^{2,*}\right) - r^1_{\tau,i^*}\right)^2 \le \mu$.

Combining above inequalities together, as $\mu \in (0,1]$, with probability at least $1-\delta$, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x}_t,y_t)\sim\mathcal{D}}\left[\min\left\{\left|f_1(\mathbf{x}_{t,i^*};\boldsymbol{\theta}^{1,*}_{t-1}) + f_2(\phi(\mathbf{x}_{t,i^*});\boldsymbol{\theta}^{2,*}_{t-1}) - r^1_{t,i^*}\right|,1\right\}\right] \\ &\qquad\le \mathcal{O}\left(\frac{3L\nu + 2\sqrt{\mu}}{\sqrt{2t}}\right) + 2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}},\end{aligned} \tag{7.67}$$

where we apply union bound over δ to make above events occur concurrently. Then, based on Lemma 7.13 (2), it is sufficient to show that $\boldsymbol{\theta}^{1,*}_{t-1}, \boldsymbol{\theta}^{2,*}_{t-1}$ are close to initialization for any $t \in [T]$. □

Lemma 7.8. *For any $\delta \in (0,1)$, suppose m satisfies the condition in Theorem 4.1. Then, with probability at least $1-\delta$, setting $\eta_2 = \frac{\nu'\nu}{m\sqrt{t}}$ for algorithm 1, for $\nu > 0$ and any $\widetilde{\boldsymbol{\theta}}^2$ satisfying*

$\|\widetilde{\boldsymbol{\theta}}^2 - \boldsymbol{\theta}_0^2\|_2 \leq \mathcal{O}(\frac{\nu}{\sqrt{m}})$*, and* $i \in [k]$*, there exists a small enough constant* v'*, such that*

$$\sum_{\tau=1}^{t} \left| f_2\left(\phi(\mathbf{x}_{\tau,i}); \widehat{\boldsymbol{\theta}}_{\tau-1}^2\right) - \left(r_{\tau,i}^1 - f_1(\mathbf{x}_{\tau,i}; \widehat{\boldsymbol{\theta}}_{\tau-1}^1)\right)\right|$$
$$\leq \sum_{\tau=1}^{t} \left| f_2\left(\phi(\mathbf{x}_{t,i}); \widetilde{\boldsymbol{\theta}}^2\right) - \left(r_{\tau,i}^1 - f_1(\mathbf{x}_{\tau,i}; \widehat{\boldsymbol{\theta}}_{\tau-1}^1)\right)\right| + \mathcal{O}\left(\frac{3L\nu\sqrt{t}}{\sqrt{2}}\right) + \sqrt{2t\log(1/\delta)}.$$

Proof. This is a direct application of Lemma 7.9 by setting $\hat{\epsilon} = \frac{L\nu}{\sqrt{2\nu' t}}$, and, where ν' is some small enough absolute constant. We set $L_\tau(\widehat{\boldsymbol{\theta}}_{\tau-1}^2) = \left| f_2(\phi(\mathbf{x}_{t,i}); \widehat{\boldsymbol{\theta}}_{\tau-1}^2) - \left(r_{\tau,i}^1 - f_1(\mathbf{x}_{\tau,i}; \widehat{\boldsymbol{\theta}}_{\tau-1}^1)\right)\right|$. Then, for any $\widetilde{\boldsymbol{\theta}}^2$ satisfying $\|\widetilde{\boldsymbol{\theta}}^2 - \boldsymbol{\theta}_0^2\|_2 \leq \mathcal{O}(\frac{\nu}{\sqrt{m}})$, there exist a small enough absolute constant ν', such that

$$\sum_{\tau=1}^{t} L_\tau(\widehat{\boldsymbol{\theta}}_{\tau-1}^2) \leq \sum_{\tau=1}^{t} L_\tau(\widetilde{\boldsymbol{\theta}}^2) + \mathcal{O}(3t\hat{\epsilon}) + \sqrt{2t\log(1/\delta)}. \tag{7.68}$$

Then, replacing $\hat{\epsilon}$ completes the proof. □

Lemma 7.9. *With probability at least* $1-\delta$ *over the randomness of* $\boldsymbol{\theta}_0$*, given the convex loss* L *satisfying* $\nabla_f L \leq \mathcal{O}(1)$*, for any* $\hat{\epsilon}, \nu > 0$ *and* $\widetilde{\boldsymbol{\theta}}$ *satisfying* $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 \leq \mathcal{O}(\frac{\nu}{\sqrt{m}})$*, Algorithm 1 with* $\eta = \frac{\nu'\hat{\epsilon}}{Lm}$ *and* $t = \frac{L^2\nu^2}{2\nu'\hat{\epsilon}^2}$ *for some small enough constant* ν' *has the following bound:*

$$\sum_{\tau=1}^{t} \min\{L_{(\mathbf{x}_\tau, r_\tau)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x}_\tau, r)}(\widetilde{\boldsymbol{\theta}}), 1\} \leq \mathcal{O}(3t\hat{\epsilon}) + \sqrt{2t\log(1/\delta)}. \tag{7.69}$$

Proof. Define $\mathcal{B}(\boldsymbol{\theta}_0, w) = \{\boldsymbol{\theta} \in \mathbb{R}^p : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \leq w\}$ and $L_{(\mathbf{x};r)}(\boldsymbol{\theta}) = |r - f(\mathbf{x};\boldsymbol{\theta})|$. First, we need to show $\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_t$ also are in $\mathcal{B}(\boldsymbol{\theta}_0, w)$, where. According to Lemma 7.14, when $\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0, w)$, we have

$$\|\nabla_{\boldsymbol{\theta}} f(\mathbf{x};\boldsymbol{\theta})\|_2 \leq \mathcal{O}(L), \;\; \|\nabla_{\boldsymbol{\theta}} L_{(\mathbf{x};r)}(\boldsymbol{\theta})\|_2 \;\; \leq \sqrt{\sum_{l=1}^{L} \|\mathcal{O}(\nabla_{\mathbf{W}_l} f(\mathbf{x};\boldsymbol{\theta}))\|_2^2} \leq \mathcal{O}(L). \tag{7.70}$$

The proof follows a simple induction. Suppose that $\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_t \in \mathcal{B}(\boldsymbol{\theta}_0, w)$, by triangle inequality, we have

$$\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\|_2 \leq \sum_{\tau=0}^{t-1} \|\widehat{\boldsymbol{\theta}}_{\tau+1} - \widehat{\boldsymbol{\theta}}_\tau\|_2 \leq \frac{1}{|\widehat{\mathcal{H}}_t|} \sum_{\tau=0}^{t-1} \sum_{(\mathbf{x},r)\in\widehat{\mathcal{H}}_t} \|\nabla_{\boldsymbol{\theta}} L(\mathbf{x};r)\|_2 \leq \mathcal{O}(L\eta t). \tag{7.71}$$

Because $\eta = \mathcal{O}(\frac{1}{m})$, we have $\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\|_2 \leq w$. In round $\tau \in [t]$, recall that $|\widehat{\mathcal{H}}_\tau| = b$ and $\widehat{\mathcal{H}}_\tau \subset \mathcal{H}_\tau$. Given the context $\mathbf{x}_\tau$ and its reward r, we have the fact

$$\begin{aligned}
\tau \mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}\left[\frac{1}{b}\sum_{(\mathbf{x},r)\in\widehat{\mathcal{H}}_\tau} \nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1})\right] &\overset{E_1}{=} \tau \mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}\left[\nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1})\right] \\
&= \tau \sum_{(\mathbf{x},r)\in\mathcal{H}_\tau} \frac{1}{|\mathcal{H}_\tau|} \nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1}) \\
&\overset{E_2}{=} \sum_{(\mathbf{x},r)\in\mathcal{H}_\tau} \nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1}),
\end{aligned} \tag{7.72}$$

where E_1 is because $\widehat{\mathcal{H}}_\tau$ is uniformly drawn from $\mathcal{H}_\tau$ and E_2 is duo to $|\mathcal{H}_\tau| = \tau$. Then, based on Lemma 7.12, for any $\epsilon > 0$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}})] &\leq \langle \mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[\nabla_{\widehat{\boldsymbol{\theta}}} L_\tau(\widehat{\boldsymbol{\theta}}_{\tau-1})], \widehat{\boldsymbol{\theta}}_{\tau-1} - \widetilde{\boldsymbol{\theta}}\rangle + \hat{\epsilon} \\
&\leq \left\langle \mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}\left[\frac{1}{b}\sum_{(\mathbf{x},r)\in\widehat{\mathcal{H}}_\tau} \nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1})\right], \widehat{\boldsymbol{\theta}}_{\tau-1} - \widetilde{\boldsymbol{\theta}}\right\rangle + \hat{\epsilon} \\
&= \frac{\langle \widehat{\boldsymbol{\theta}}_{\tau-1} - \mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}[\widehat{\boldsymbol{\theta}}_\tau], \widehat{\boldsymbol{\theta}}_{\tau-1} - \widetilde{\boldsymbol{\theta}}\rangle}{\eta} + \hat{\epsilon}.
\end{aligned} \tag{7.73}$$

Based on the fact $2\langle \mathbf{A}, \mathbf{B}\rangle = \|\mathbf{A}\|_2^2 + \|\mathbf{B}\|_2^2 - \|\mathbf{A}, \mathbf{B}\|_2^2$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}})] &\leq \frac{\|\widehat{\boldsymbol{\theta}}_{\tau-1} - \mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}[\widehat{\boldsymbol{\theta}}_\tau]\|_2^2 + \|\widehat{\boldsymbol{\theta}}_{\tau-1} - \widetilde{\boldsymbol{\theta}}\|_2^2 - \|\mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}[\widehat{\boldsymbol{\theta}}_\tau] - \widetilde{\boldsymbol{\theta}}\|_2^2}{2\eta} + \hat{\epsilon} \\
&\overset{E_3}{\leq} \frac{\|\widehat{\boldsymbol{\theta}}_{\tau-1} - \widetilde{\boldsymbol{\theta}}\|_2^2 - \|\mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}[\widehat{\boldsymbol{\theta}}_\tau] - \widetilde{\boldsymbol{\theta}}\|_2^2}{2\eta} + \mathcal{O}(L^2\eta) + \hat{\epsilon}
\end{aligned} \tag{7.74}$$

where E_3 is because of 7.72:

$$\begin{aligned}
\|\widehat{\boldsymbol{\theta}}_{\tau-1} - \mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}[\widehat{\boldsymbol{\theta}}_\tau]\|_2 &= \eta\left\|\mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}\left[\frac{1}{b}\sum_{(\mathbf{x},r)\in\widehat{\mathcal{H}}_\tau} \nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1})\right]\right\|_2 \\
&= \eta\frac{1}{\tau}\|\sum_{(\mathbf{x},r)\in\mathcal{H}_\tau} \nabla_{\widehat{\boldsymbol{\theta}}} L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}^{t-1})\|_2 \leq \mathcal{O}(\eta L).
\end{aligned} \tag{7.75}$$

Therefore, we have

$$\begin{aligned}
\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}})] &\leq \frac{\|\widehat{\boldsymbol{\theta}}_0 - \widetilde{\boldsymbol{\theta}}\|_2^2 - \|\mathop{\mathbb{E}}_{\widehat{\mathcal{H}}_\tau}[\widehat{\boldsymbol{\theta}}_t] - \widetilde{\boldsymbol{\theta}}\|_2^2}{2\eta} + \mathcal{O}(tL^2\eta) + t\hat{\epsilon} \\
&\leq \frac{\|\widehat{\boldsymbol{\theta}}_0 - \widetilde{\boldsymbol{\theta}}\|_2^2}{2\eta} + \mathcal{O}(tL^2\eta) + t\hat{\epsilon} \\
&\leq \frac{LR^2}{2\eta m} + \mathcal{O}(tL^2\eta) + t\hat{\epsilon}.
\end{aligned} \tag{7.76}$$

Then, for $\tau \in [t]$, define

$$V_\tau = \min\{L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}}), 1\} - \mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[\min\{L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}}), 1\}]. \tag{7.77}$$

Then, we have

$$\begin{aligned}
\mathbb{E}[V_\tau|\mathcal{F}_{\tau-1}] =& \mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[\min\{L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}), -L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}}), 1\}] \\
&- \mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau}[\min\{L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}}), 1\}] = 0,
\end{aligned} \tag{7.78}$$

where where $F_{\tau-1}$ denotes the σ-algebra generated by the history $\mathcal{H}_{\tau-1}$. Therefore, $\{V_0, \ldots, V_t\}$ is the martingale difference sequence. Then, applying the Hoeffding-Azuma inequality, with probability

at least $1-\delta$, we have

$$\begin{aligned}
&\sum_{\tau=1}^{t} \min\{L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x}_\tau,r)}(\widetilde{\boldsymbol{\theta}}), 1\} \\
\leq &\sum_{\tau=1}^{t} \mathop{\mathbb{E}}_{(\mathbf{x},r)\sim\mathcal{H}_\tau} [\min\{L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}}), 1\}] + t\sqrt{\frac{2\log(1/\delta)}{t}} \\
\overset{E_4}{\leq} &\frac{LR^2}{2\eta m} + \mathcal{O}(tL^2\eta) + t\hat{\epsilon} + \sqrt{2t\log(1/\delta)} \\
\overset{E_5}{\leq} &\mathcal{O}(3t\hat{\epsilon}) + \sqrt{2t\log(1/\delta)}
\end{aligned} \tag{7.79}$$

where E_4 be because of 7.76 and E_5 is by placing the parameter choice $\eta = \frac{\nu'\hat{\epsilon}}{Lm}$ and $t = \frac{L^2\nu^2}{2\nu'\hat{\epsilon}}$. The proof is completed. □

7.4 Ancillary Lemmas

Lemma 7.10 (Theorem 5, [2])**.** *For any $\delta \in (0,1)$, if w satisfies that*

$$\mathcal{O}(m^{-3/2}L^{-3/2}\max\{\log^{-3/2} m, \log^{3/2}(Tn/\delta)\}) \leq w \leq \mathcal{O}(L^{-9/2}\log^{-3} m), \tag{7.80}$$

then, with probability at least $1-\delta$, for all $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2 \leq w$, we have

$$\|\nabla_{\boldsymbol{\theta}} f(\mathbf{x};\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}_0} f(\mathbf{x};\boldsymbol{\theta}_0)\|_2 \leq \mathcal{O}(\sqrt{\log m} w^{1/3} L^3)\|\nabla_{\boldsymbol{\theta}_0} f(\mathbf{x};\boldsymbol{\theta}_0)\|_2. \tag{7.81}$$

Lemma 7.11 (Lemma 4.1, [17])**.** *For any $\delta \in (0,1)$, if w satisfies*

$$\mathcal{O}(m^{-3/2}L^{-3/2}[\log(tnL^2/\delta)]^{3/2}) \leq w \leq \mathcal{O}(L^{-6}[\log m]^{-3/2}),$$

then, with probability at least $1-\delta$ over randomness of $\boldsymbol{\theta}_0$, for any $t \in [T]$, $\|\mathbf{x}\|_2 = 1$, and $\boldsymbol{\theta}, \boldsymbol{\theta}'$ satisfying $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2 \leq w$ and $\|\boldsymbol{\theta}'-\boldsymbol{\theta}_0\|_2 \leq w$, it holds uniformly that

$$|f(\mathbf{x}_i;\boldsymbol{\theta}) - f(\mathbf{x}_i;\boldsymbol{\theta}') - \langle \nabla_{\boldsymbol{\theta}'} f(\mathbf{x}_i;\boldsymbol{\theta}'), \boldsymbol{\theta}-\boldsymbol{\theta}'\rangle| \leq \mathcal{O}(w^{1/3}L^2\sqrt{m\log(m)})\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2. \tag{7.82}$$

Lemma 7.12 (Lemma 4.2, [17])**.** *For any $\delta \in (0,1), \hat{\epsilon} > 0$, if w satisfies*

$$\mathcal{O}(m^{-3/2}L^{-3/2}[\log(tnL^2/\delta)]^{3/2}) \leq w \leq \kappa L^{-6}m^{-3/8}[\log m]^{-3/2}\hat{\epsilon}^{3/4},$$

then, with probability at least $1-\delta$ over randomness of $\boldsymbol{\theta}^{(0)}$, for any $\hat{\epsilon} > 0, i \in [n]$, and $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}$ satisfying $\|\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)}\|_2 \leq w$ and $\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^{(0)}\|_2 \leq w$, it holds uniformly that

$$L_{(\mathbf{x},r)}(\widehat{\boldsymbol{\theta}}_{\tau-1}) - L_{(\mathbf{x},r)}(\widetilde{\boldsymbol{\theta}}) \leq \langle \nabla_{\widehat{\boldsymbol{\theta}}} L_{\mathbf{x},r}(\widehat{\boldsymbol{\theta}}_{\tau-1}), \widehat{\boldsymbol{\theta}}_{\tau-1} - \widetilde{\boldsymbol{\theta}}\rangle + \hat{\epsilon} \tag{7.83}$$

Lemma 7.13. *Given a constant $0 < \hat{\epsilon} < 1$, suppose m satisfies the conditions in Lemma 4.1, the learning rate $\eta = \Omega(\frac{\rho}{poly(t,n,L)m})$, the number of iterations $K = \Omega(\frac{poly(t,n,L)}{\rho^2} \cdot \log\hat{\epsilon}^{-1})$. Then, with probability at least $1-\delta$, starting from random initialization $\boldsymbol{\theta}_0$,*

(1) (Theorem 1 in [2]) In round $t \in [T]$, given the collected data $\{\mathbf{x}_\tau, r_\tau\}_{i=\tau}^{t}$, the loss function is defined as: $\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{\tau=1}^{t}(f(\mathbf{x}_\tau;\boldsymbol{\theta}) - r_\tau)^2$. Then, there exists $\widetilde{\boldsymbol{\theta}}$ satisfying $\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2 \leq \mathcal{O}\left(\frac{t^3}{\rho\sqrt{m}}\log m\right)$, such that $\mathcal{L}(\widetilde{\boldsymbol{\theta}}) \leq \hat{\epsilon}$ in $K = \Omega(\frac{poly(t,n,L)}{\rho^2}\cdot\log\hat{\epsilon}^{-1})$ iterations;

(2) For any $t \in [T]$, it holds uniformly that $\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_0\|_2 \leq \mathcal{O}\left(\frac{t^3}{\rho\sqrt{m}}\log m\right)$;

(3) (Lemma C.4 in [14]) Following the initialization, given $\|\mathbf{x}\|_2 = 1$, it holds that

$$\|\nabla_{\boldsymbol{\theta}_0} f(\mathbf{x};\boldsymbol{\theta}_0)\|_2 \leq \mathcal{O}(L), \quad |f(\mathbf{x};\boldsymbol{\theta}_0)| \leq \mathcal{O}(1).$$

Proof. (2) is a corollary of Theorem 1 in [2]. Suppose $\boldsymbol{\theta}_\tau = \boldsymbol{\theta}_{\tau-1} - \frac{\eta}{b}\sum_{(\mathbf{x},r)\in\widehat{\mathcal{H}}_\tau} \nabla_{\boldsymbol{\theta}}\mathcal{L}[(\mathbf{x},r);\boldsymbol{\theta}_{\tau-1}]$. The proof is based on the following induction. Let $w = \mathcal{O}\left(\frac{t^3}{\delta\sqrt{m}}\log m\right)$. Then, based on the Theorem 1 in [2], we have

$$\mathcal{L}[(\mathbf{x},r);\boldsymbol{\theta}_\tau] = \left(1 - \Omega\left(\frac{\eta\rho m}{t^2}\right)\right)\mathcal{L}[(\mathbf{x},r);\boldsymbol{\theta}_{\tau-1}].$$

Then, we have

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|_2 \leq \sum_{\tau=1} \|\frac{\eta}{b} \sum_{(\mathbf{x},r)\in\widehat{\mathcal{H}}_\tau} \nabla_{\boldsymbol{\theta}} \mathcal{L}[(\mathbf{x},r);\boldsymbol{\theta}_{\tau-1}]\| \leq \mathcal{O}(\eta\sqrt{tm}) \sum_{\tau=1}^{t} \sqrt{\mathcal{L}[(\mathbf{x},r);\boldsymbol{\theta}_{\tau-1}]}$$
$$\leq \mathcal{O}(\eta\sqrt{tm}) \cdot \Omega(\frac{t^2}{\eta\rho m}) \cdot \mathcal{O}(\sqrt{t\log^2 m}) \leq \mathcal{O}\left(\frac{t^3}{\rho\sqrt{m}}\log m\right).$$

□

Lemma 7.14 (Lemma C.2 [14]). *For any $\delta \in (0,1), \rho \in (0, \mathcal{O}(\frac{1}{L})]$, suppose m satisfies the conditions in Theorem 4.1. Then, with probability at least $1-\delta$, in each round $t \in [T]$, for any $\mathbf{x}$ satisfying $\|\mathbf{x}\|_2 = 1$, $\boldsymbol{\theta}_{t-1}^{1,*}, \boldsymbol{\theta}_{t-1}^1$ satisfying $\|\boldsymbol{\theta}_{t-1}^{1,*} - \boldsymbol{\theta}_{t-1}^1\|_2 \leq \mathcal{O}\left(\frac{t^3}{\rho\sqrt{m}}\log m\right)$, and $\boldsymbol{\theta}_{t-1}^{2,*}, \boldsymbol{\theta}_{t-1}^2$ satisfying $\|\boldsymbol{\theta}_{t-1}^{2,*} - \boldsymbol{\theta}_{t-1}^2\|_2 \leq \mathcal{O}\left(\frac{t^3}{\rho\sqrt{m}}\log m\right)$, we have*

$$(1) \quad |f_1(\mathbf{x};\boldsymbol{\theta}_{t-1}^{1,*}) - f_1(\mathbf{x};\boldsymbol{\theta}_{t-1}^1)|$$
$$\leq \left(1 + \mathcal{O}\left(\frac{tL^3\log^{5/6} m}{\rho^{1/3}m^{1/6}}\right)\right)\mathcal{O}\left(\frac{Lt^3}{\rho\sqrt{m}}\log m\right) + \mathcal{O}\left(\frac{t^4L^2\log^{11/6} m}{\rho^{4/3}m^{1/6}}\right); \tag{7.84}$$

$$(2) \quad \left|f_2\left(\phi(\mathbf{x});\boldsymbol{\theta}_{t-1}^{2,*}\right) - f_2\left(\phi(\mathbf{x});\boldsymbol{\theta}_{t-1}^2\right)\right|$$
$$\leq \left(1 + \mathcal{O}\left(\frac{tL^3\log^{5/6} m}{\rho^{1/3}m^{1/6}}\right)\right)\mathcal{O}\left(\frac{Lt^3}{\rho\sqrt{m}}\log m\right) + \mathcal{O}\left(\frac{t^4L^2\log^{11/6} m}{\rho^{4/3}m^{1/6}}\right); \tag{7.85}$$

$$(3) \quad \|\nabla_{\boldsymbol{\theta}_{t-1}^1} f_1(\mathbf{x};\boldsymbol{\theta}_{t-1}^1)\|_2, \|\nabla_{\boldsymbol{\theta}_{t-1}^2} f_2\left(\phi(\mathbf{x});\boldsymbol{\theta}_{t-1}^2\right)\|_2$$
$$\leq \left(1 + \mathcal{O}\left(\frac{tL^3\log^{5/6} m}{\rho^{1/3}m^{1/6}}\right)\right)\mathcal{O}(L). \tag{7.86}$$

Proof. $\boldsymbol{\theta}_{t-1}^{1,*}, \boldsymbol{\theta}_{t-1}^1$ stay in the same ball is by the application of Lemma 7.13.

□

8 Proof of Lemma 4.1

Algorithm 3 Batch-GD-Warm-Start (f_1, f_2, $\mathcal{H}_t^1, \mathcal{H}_t^2$)

1: Define $\mathcal{L}_1[(\mathbf{x}, r^1); \boldsymbol{\theta}^1] = (r^1 - f_1(\mathbf{x};\boldsymbol{\theta}^1))^2/2$
2: Uniformly draw a set $\widehat{\mathcal{H}}_t^1 \subset \mathcal{H}_t^1, s.t., |\widehat{\mathcal{H}}_t^1| = b$
3: $\widehat{\boldsymbol{\theta}}_t^1 = \widehat{\boldsymbol{\theta}}_{t-1}^1 - \frac{\eta_1}{b} \sum_{(\mathbf{x},r^1)\in\widehat{\mathcal{H}}_t^1} \nabla_{\boldsymbol{\theta}^1} \mathcal{L}_1[(\mathbf{x}, r^1); \widehat{\boldsymbol{\theta}}_{t-1}^1]$
4: Define $\mathcal{L}_2[(\phi(\mathbf{x}), r^2); \boldsymbol{\theta}^2] = (r^2 - f_2(\phi(\mathbf{x});\boldsymbol{\theta}^2))^2/2$
5: Uniformly draw a set $\widehat{\mathcal{H}}_t^2 \subset \mathcal{H}_t^2, s.t., |\widehat{\mathcal{H}}_t^2| = b$
6: $\widehat{\boldsymbol{\theta}}_t^2 = \widehat{\boldsymbol{\theta}}_{t-1}^2 - \frac{\eta_2}{b} \sum_{(\phi(\mathbf{x}),r^1)\in\widehat{\mathcal{H}}_t^2} \nabla_{\boldsymbol{\theta}^2} \mathcal{L}_2[(\phi(\mathbf{x}), r^2); \widehat{\boldsymbol{\theta}}_{t-1}^2]$
7: **Return** $(\widehat{\boldsymbol{\theta}}_t^1, \widehat{\boldsymbol{\theta}}_t^2)$

Lemma 8.1 (Lemma 4.1 Restate). *For any $\delta \in (0,1), \rho \in (0, \mathcal{O}(\frac{1}{\rho}))$, suppose $m \geq$ poly$(T, k, L, \rho^{-1}, e^{\sqrt{\log(Tn/\delta)}})$, $\eta_1 = \eta_2 = \frac{T^5}{\delta^2 m}$. Then, with probability at least $1-\delta$ over the initialization of $\boldsymbol{\theta}_0^1, \boldsymbol{\theta}_0^2$, Algorithm 1 with Algorithm 3 achieves the following regret bound:*

$$\widetilde{\mathbf{R}}_T \leq \mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu}}{\sqrt{2}}\right)\sqrt{T} + 2\sqrt{2T\log(\mathcal{O}(T)/\delta)} + \mathcal{O}(1) \tag{8.1}$$

and at the same time $\mathbf{N}_T \leq \mathcal{O}(T)$.

Proof. For any $t \in [T] \wedge (\mathbf{I}_t = 1)$, the regret of one round can be bounded as:

$$
\begin{aligned}
&R_t|(\mathbf{I}_t = 1) \\
=&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\mathcal{L}(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})|\mathbf{x}_t] - \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\mathcal{L}(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})|\mathbf{x}_t] \\
=&1 - \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\mathcal{L}(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})|\mathbf{x}_t] - (1 - \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\mathcal{L}(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t})|\mathbf{x}_t]) \\
=&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [r^1_{t,i^*} - r^1_{t,\widehat{i}}] \\
=&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{r^1_{t,i^*} - r^1_{t,\widehat{i}}, 1\}] \\
=&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) + f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}] \\
\overset{E_1}{\leq}&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) + f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}] \\
=&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) + f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}] \\
=&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1}) + f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) + f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}] \\
\leq&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{r^1_{t,i^*} - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1}), 1\}] + f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) \\
&+ \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}, 1\}] \\
\leq&\mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{|r^1_{t,i^*} - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1})|, 1\}] + |f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1})| \\
&+ \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{|f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}|, 1\}]
\end{aligned}
\tag{8.2}
$$

where E_1 is because of $f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1}) \leq f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1})$. For any $t \in [T] \wedge (\mathbf{I}_t = 0)$, we have $R_t|(\mathbf{I}_t = 0) = \mathop{\mathbb{E}}_{\mathbf{x}_t, y_t} [L(\mathbf{y}_{t,\widehat{i}}, \mathbf{y}_{t,y_t}) - L(\mathbf{y}_{t,i^*}, \mathbf{y}_{t,y_t})] = 0$ based on Lemma 7.5. Therefore, we have

$$
\begin{aligned}
\mathbf{R}_T =& \sum_{t=1}^{T} R_t \\
\leq& \underbrace{\sum_{t=1}^{T} \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{|r^1_{t,i^*} - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1})|, 1\}]}_{I_1} + \underbrace{\sum_{t=1}^{T} |f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}^*_{t-1}) - f(\mathbf{x}_{t,i^*}; \boldsymbol{\theta}_{t-1})|}_{I_2} \\
&+ \underbrace{\sum_{t=1}^{T} \mathop{\mathbb{E}}_{y_t \sim \mathcal{D}_{\mathcal{Y}|\mathbf{x}_t}} [\min\{|f(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}_{t-1}) - r^1_{t,\widehat{i}}|, 1\}]}_{I_3} \\
\leq& 2\left(2\sqrt{t\mu} + \mathcal{O}\left(\frac{3}{\sqrt{2}} L\nu\sqrt{T}\right) + \sqrt{2T\log(\mathcal{O}(T)/\delta)}\right) + T\xi_t \\
\overset{E_2}{\leq}& \mathcal{O}(1) + \mathcal{O}\left(\frac{6L\nu + 4\sqrt{\mu}}{\sqrt{2}}\right)\sqrt{T} + 2\sqrt{2T\log(\mathcal{O}(T)/\delta)}
\end{aligned}
\tag{8.3}
$$

where I_1 is because of Lemma 8.3, I_3 is due to Lemma 8.2, I_2 is the application of Lemma 7.14 and E_2 is the result of choice of m ($m \geq \widetilde{\mathcal{O}}(T^{30})$). □

Lemma 8.2. *For any $\delta \in (0,1)$, $\rho \in (0, \mathcal{O}(\frac{1}{L})]$, $\gamma \geq 1$, suppose m satisfies the conditions in Lemma 4.1. In round $t \in [T]$, given $(\mathbf{x}_t, y_t) \sim \mathcal{D}$, let*

$$
\widehat{i} = \arg\max_{i \in [k]} \left(f_1(\mathbf{x}_{t,\widehat{i}}; \boldsymbol{\theta}^1_{t-1}) + f_2(\phi(\mathbf{x}_{t,\widehat{i}}); \boldsymbol{\theta}^2_{t-1}) \right).
$$

Then, with probability at least $1-\delta$, we have

$$\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\left\{\left|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\phi(\mathbf{x}_{\tau,\widehat{i}});\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}\right|,1\right\}\right] \leq \sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}}. \tag{8.4}$$

Proof. For any $\tau\in[t]$, define

$$\begin{aligned}V_\tau =& \mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}|,1\}\right]\\ &-\min\{|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}|,1\}\end{aligned} \tag{8.5}$$

Then, we have

$$\begin{aligned}\mathbb{E}[V_\tau|F_{\tau-1}] =& \mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}|,1\}\right]\\ &-\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}|,1\}\right]\\ =&0\end{aligned} \tag{8.6}$$

where $F_{\tau-1}$ denotes the σ-algebra generated by the history $\mathcal{H}_{\tau-1}$. Therefore, $\{V_\tau\}_{\tau=1}^t$ are the martingale difference sequence.

Applying the Hoeffding-Azuma inequality, with probability at least $1-\delta$, we have

$$\mathbb{P}\left[\frac{1}{t}\sum_{\tau=1}^{t}V_\tau-\underbrace{\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}[V_\tau|\mathbf{F}_\tau]}_{I_1}>\sqrt{\frac{2\log(1/\delta)}{t}}\right]\leq\delta \tag{8.7}$$

As I_1 is equal to 0, we have

$$\begin{aligned}&\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{\left|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}\right|,1\}\right]\\ \leq&\underbrace{\frac{1}{t}\sum_{\tau=1}^{t}\min\{\left|f_2\left(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^2_{\tau-1}\right)-\left(r_{\tau,\widehat{i}}-f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})\right)\right|,1\}}_{I_3}+\sqrt{\frac{2\log(1/\delta)}{t}}\,.\end{aligned} \tag{8.8}$$

For I_3, based on Lemma 7.8, for any $\widetilde{\boldsymbol{\theta}}^2$ satisfying $\|\widetilde{\boldsymbol{\theta}}^2-\boldsymbol{\theta}^2_0\|_2\leq\mathcal{O}(\frac{\nu}{\sqrt{m}})$, with probability at least $1-\delta$, we have

$$\begin{aligned}I_3\leq&\frac{1}{t}\sum_{\tau=1}^{t}\min\{|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2\left(\mathbf{x}_{\tau,\widehat{i}};\widetilde{\boldsymbol{\theta}}^2\right)-r^1_{\tau,\widehat{i}}|,1\}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)\\ \leq&\frac{1}{t}\sqrt{t}\sqrt{\underbrace{\sum_{\tau=1}^{t}\left(f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2\left(\mathbf{x}_{\tau,\widehat{i}};\widetilde{\boldsymbol{\theta}}^2\right)-r^1_{\tau,\widehat{i}}\right)^2}_{I_4}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(1/\delta)}{t}}\\ \leq&\sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+\sqrt{\frac{2\log(1/\delta)}{t}}.\end{aligned} \tag{8.9}$$

where I_4 is by the definition of μ.

Combining above inequalities together, with probability at least $1-\delta$, we have

$$\begin{aligned}\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\left\{\left|f_1(\mathbf{x}_{\tau,\widehat{i}};\boldsymbol{\theta}^1_{\tau-1})+f_2(\phi(\mathbf{x}_{\tau,\widehat{i}});\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,\widehat{i}}\right|,1\right\}\right]\\ \leq\sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}},\end{aligned}\tag{8.10}$$

where we applied union bound over δ to make above events occur concurrently. □

Lemma 8.3. *For any $\delta\in(0,1)$, $\rho\in(0,\mathcal{O}(\frac{1}{L})]$, $\gamma\geq 1$, suppose m satisfies the conditions in Lemma 4.1. In round $t\in[T]$, given $(\mathbf{x}_t,y_t)\sim\mathcal{D}$, let $i^*=\arg\max_{i\in[k]}h(\mathbf{x}_{\tau,i})$. Then, with probability at least $1-\delta$, there exists $\boldsymbol{\theta}^{1,*}_{t-1},\boldsymbol{\theta}^{2,*}_{t-1}$, such that*

$$\begin{aligned}\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\left\{\left|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})+f_2(\phi(\mathbf{x}_{\tau,i^*});\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,i^*}\right|,1\right\}\right]\\ \leq\sqrt{\frac{2\mu}{t}}+\mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right)+2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}},\end{aligned}\tag{8.11}$$

where $\mathcal{H}_{t-1}=\{\mathbf{x}_{\tau,i^*},r^1_{\tau,i^*}\}_{\tau=1}^{t-1}$ is historical data.

Proof. For any $\tau\in[t]$, define

$$\begin{aligned}V_\tau=&\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,i^*})|,1\}\right]\\ &-\min\{|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,i^*}|,1\}\end{aligned}\tag{8.12}$$

Then, we have

$$\begin{aligned}\mathbb{E}[V_\tau|F_{\tau-1}]=&\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,i^*})|,1\}\right]\\ &-\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,i^*}|,1\}\right]\\ =&0\end{aligned}\tag{8.13}$$

where $F_{\tau-1}$ denotes the σ-algebra generated by the history $\mathcal{H}_{\tau-1}$. Therefore, $\{V_\tau\}_{\tau=1}^t$ are the martingale difference sequence.

Applying the Hoeffding-Azuma inequality, with probability at least $1-\delta$, we have

$$\mathbb{P}\left[\frac{1}{t}\sum_{\tau=1}^{t}V_\tau-\underbrace{\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}[V_\tau|\mathbf{F}_\tau]}_{I_1}>\sqrt{\frac{2\log(1/\delta)}{t}}\right]\leq\delta\tag{8.14}$$

As I_1 is equal to 0, we have

$$\begin{aligned}&\frac{1}{t}\sum_{\tau=1}^{t}\mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}}\left[\min\{\left|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})+f_2(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^2_{\tau-1})-r^1_{\tau,i^*})\right|,1\}\right]\\ \leq&\underbrace{\frac{1}{t}\sum_{\tau=1}^{t}\min\{\left|f_2\left(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^2_{\tau-1}\right)-\left(r_{\tau,i^*}-f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}^1_{\tau-1})\right)\right|,1\}}_{I_3}+\sqrt{\frac{2\log(1/\delta)}{t}}\,.\end{aligned}\tag{8.15}$$

For I_3, based on Lemma7.8, for any $\widetilde{\boldsymbol{\theta}}^2$ satisfying $\|\widetilde{\boldsymbol{\theta}}^2 - \boldsymbol{\theta}_0^2\|_2 \leq \mathcal{O}(\frac{\nu}{\sqrt{m}})$, with probability at least $1-3\delta$, we have

$$\begin{aligned}
I_3 &\leq \frac{1}{t}\sum_{\tau=1}^{t} \min\{|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}_{\tau-1}^1) + f_2\left(\mathbf{x}_{\tau,i^*};\widetilde{\boldsymbol{\theta}}^2\right) - r_{\tau,i^*}^1|, 1\} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) \\
&\leq \frac{1}{t}\sqrt{t}\sqrt{\underbrace{\sum_{\tau=1}^{t}\left(f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}_{\tau-1}^1) + f_2\left(\mathbf{x}_{\tau,i^*};\widetilde{\boldsymbol{\theta}}^2\right) - r_{\tau,i^*}^1\right)^2}_{I_4}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(1/\delta)}{t}} \\
&\leq \sqrt{\frac{2\mu}{t}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + \sqrt{\frac{2\log(1/\delta)}{t}}.
\end{aligned} \tag{8.16}$$

where I_4 is by the definition of μ.

Combining above inequalities together, with probability at least $1-\delta$, we have

$$\begin{aligned}
\frac{1}{t}\sum_{\tau=1}^{t} \mathop{\mathbb{E}}_{y_\tau\sim\mathcal{D}_{\mathcal{Y}|\mathbf{x}_\tau}} &\left[\min\left\{\left|f_1(\mathbf{x}_{\tau,i^*};\boldsymbol{\theta}_{\tau-1}^1) + f_2(\phi(\mathbf{x}_{\tau,i^*});\boldsymbol{\theta}_{\tau-1}^2) - r_{\tau,i^*}^1\right|, 1\right\}\right] \\
&\leq \sqrt{\frac{2\mu}{t}} + \mathcal{O}\left(\frac{3L\nu}{\sqrt{2t}}\right) + 2\sqrt{\frac{2\log(\mathcal{O}(1)/\delta)}{t}},
\end{aligned} \tag{8.17}$$

where we applied union bound over δ to make above events occur concurrently. □

9 Bounds for Effective Dimension $\widetilde{d}$

Let $\{\mathbf{x}_{t,\widehat{i}}\}_{t=1}^T$ be the selected contexts in T rounds, then we have the following definition of NTK.

Definition 9.1 (NTK [29, 4]). *Let $\mathcal{N}$ denote the normal distribution. Define*

$$\begin{aligned}
&\mathbf{H}_{i,j}^0 = \boldsymbol{\Sigma}_{i,j}^0 = \langle \mathbf{x}_i, \mathbf{x}_j\rangle, \;\; \mathbf{N}_{i,j}^l = \begin{pmatrix} \boldsymbol{\Sigma}_{i,i}^l & \boldsymbol{\Sigma}_{i,j}^l \\ \boldsymbol{\Sigma}_{j,i}^l & \boldsymbol{\Sigma}_{j,j}^l \end{pmatrix} \\
&\boldsymbol{\Sigma}_{i,j}^l = 2\mathbb{E}_{a,b\sim\mathcal{N}(\mathbf{0},\mathbf{N}_{i,j}^{l-1})}[\sigma(a)\sigma(b)] \\
&\mathbf{H}_{i,j}^l = 2\mathbf{H}_{i,j}^{l-1}\mathbb{E}_{a,b\sim\mathcal{N}(\mathbf{0},\mathbf{N}_{i,j}^{l-1})}[\sigma'(a)\sigma'(b)] + \boldsymbol{\Sigma}_{i,j}^l.
\end{aligned}$$

Then, over the contexts $\{\mathbf{x}_{t,\widehat{i}}\}_{t=1}^T$, the Neural Tangent Kernel (NTK) is defined as $\mathbf{H} = (\mathbf{H}^L + \boldsymbol{\Sigma}^L)/2$.

Then, we define the following gram matrix $\mathbf{G}$. Let $g(x;\boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\theta}} f(x;\boldsymbol{\theta}_0) \in \mathbb{R}^p$ and $G = [g(\mathbf{x}_{1,\widehat{i}};\boldsymbol{\theta}_0)/\sqrt{m}, \ldots, g(\mathbf{x}_{T,\widehat{i}};\boldsymbol{\theta}_0)/\sqrt{m}] \in \mathbb{R}^{p\times T}$ where $p = m + mkd + m^2(L-1)$. Therefore, we have $\mathbf{G} = G^\top G$. Based on Theorem 3.1 in [4], when $m \geq \mathcal{O}(T^4k^6\log(2Tk/\delta)/\lambda_0^4)$ where λ_0 is the smallest eigenvalue of $\mathbf{H}$, with probability at least $1-\delta$, we have

$$\|\mathbf{G} - \mathbf{H}\| \leq \frac{\lambda_0}{2}. \tag{9.1}$$

Then, we have the following bound:

$$\begin{aligned}
\log\det(\mathbf{I}+\mathbf{H}) &= \log\det\left(\mathbf{I} + \mathbf{G} + (\mathbf{H}-\mathbf{G})\right) \\
&\leq \log\det(\mathbf{I}+\mathbf{G}) + \langle(\mathbf{I}+\mathbf{G})^{-1}, (\mathbf{H}-\mathbf{G})\rangle \\
&\leq \log\det(\mathbf{I}+\mathbf{G}) + \|(\mathbf{I}+\mathbf{G})^{-1}\|_F\|\mathbf{H}-\mathbf{G}\|_F \\
&\leq \log\det(\mathbf{I}+\mathbf{G}) + \sqrt{T}\|\mathbf{H}-\mathbf{G}\|_F \\
&\leq \log\det(\mathbf{I}+\mathbf{G}) + 1
\end{aligned} \tag{9.2}$$

where the first inequality is because of the concavity of $\log\det(\cdot)$ and the third inequality is by Lemma B.1 in [57] with the choice of m. Then, the effective dimension $\widetilde{d}$ can be bounded by:

$$\begin{aligned}\widetilde{d} =& \frac{\log\det(\mathbf{I}+\mathbf{H})}{\log(1+T)} \\ \leq& \frac{\log\det(\mathbf{I}+\mathbf{G})+1}{\log(1+T)} \\ \overset{E_1}{=}& \frac{\log\det(\mathbf{I}+GG^\top)+1}{\log(1+T)} \\ \overset{E_2}{\leq}& p\cdot\frac{\log\|\mathbf{I}+GG^\top\|_2}{\log(1+T)}+\frac{1}{\log(1+T)} \\ \overset{E_3}{\leq}& p+\frac{1}{\log(1+T)}\end{aligned} \tag{9.3}$$

where E_1 is because of $\det(\mathbf{I}+G^\top G)=\det(\mathbf{I}+GG^\top)$ and E_2 is due to $\det(GG^\top)=\|GG^\top\|_2^p$ ($GG^\top\in\mathbb{R}^{p\times p}$) and E_3 is according to

$$\|\mathbf{I}+GG^\top\|_2\leq 1+\|GG^\top\|_2\leq 1+\sum_{t=1}^{T}\|g(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_0)g(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_0)^\top/m\|_2\leq 1+T,$$

where the last inequality is as the result of $\|g(\mathbf{x}_{t,\widehat{i}};\boldsymbol{\theta}_0)/\sqrt{m}\|_2\leq 1$ (Lemma B.3 in [17]). Therefore, we have

$$\widetilde{d}\leq p+\frac{1}{\log(1+T)}\ \text{ and }\ p=m+mkd+m^2(L-1). \tag{9.4}$$

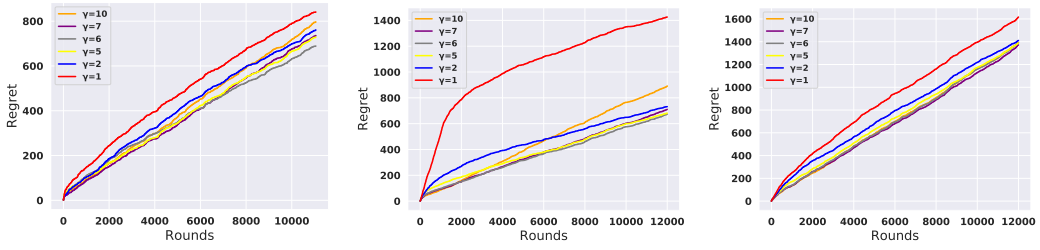
10 Further Details in Experiments

In this section, we report the specific configurations in the experiments, the sensitivity study of the core hyperparameter γ for I-NeurAL, and the ablation study for label budget. Table 2 exhibits the details of using datasets.

Dataset	Features	Samples	Classes
Phishing	68	11,055	2
IJCNN	22	12,000	2
Letter	784	12,000	26
Fashion	784	12,000	10
MNIST	784	12,000	10
CIFAR-10	3,072	12,000	10

Table 2: Statistics of the datasets used in our experiments. We conduct experiments on binary classification tasks. For Letter, the binary task is to separate 'A-M' versus 'N-Z'. For Fashion, the binary task is to separate 'T-shirt' versus 'Trouser' images. For MNIST, the binary task is to separate odd and even digits. For CIFAR-10, the binary task is to separate 'horse' and 'ship' images.

Implementation Details. We use PyTorch as our backend, and all experiments were conducted on a server with NVIDIA Tesla V100 SXM2 GPU. The classification model in all methods is the same 2-layer fully-connected network with 100-width for the fair comparison. We use Adam optimizer to train the classification model with the fixed learning rate is 0.001, and the batch size is 64, since these are model-agnostic hyperparameters. As NeuAL-NTK-F and NeuAL-NTK-D only work on the binary classification problem, we transformed the k-class classification problem into the binary classification problem when $k > 2$. In detail, given k class, we regard the first $k/2$ classes as one class and remaining classes as another class. For Random algorithm, the query probability p is set as 0.1. To find the best performance of each method, we conduct the grid search over all hyperparameters. In Margin algorithm, the query threshold is searched over $\{0.3, 0.5, 0.7, 0.9, 0.95\}$ for all datasets. For NeuAL-NTK-F and NeuAL-NTK-D, there is also an exploration parameter γ to determine the query aggressiveness and we conduct the grid search over $\{0.1, 0.3, 0.5, 0.8, 1.0\}$ for it. For ALPS, following the method in [23], we form the hypothesis class by generating 20 hypotheses on the 3% of total data samples (as same as the query budget) with different random seeds, and we conduct the grid search $\{0.1, 0.25, 0.5, 0.75, 0.9\}$ over the two slack terms in ALPS. We have tried to generate more hypotheses in the experiments, but the performance of ALPS does not improve accordingly. For I-NeurAL, the only hyperparameter γ is searched over $\{1, 2, 5, 6, 7, 10\}$ for all datasets (c_1 and c_2 in $\boldsymbol{\beta}_t$ is set as 1). The confidence level δ is set as 0.1 for all the needed methods. In the end, we report the average results of 5 runs for all methods.



(a) Phishing (b) MNIST (c) CIFAR-10

Figure 2: Parameter sensitivity on three datasets.

Sensitivity study for γ. As γ is closely related to the query threshold of I-NeurAL, we test the sensitivity of I-NeurAL with regard to γ. Based on our analysis, it is required that $\gamma \geq 1$. When γ is the smallest number (e.g., $\gamma = 1$), I-NeurAL queries the labels only if the difference between the top two classes is very small (i.e., the confidence level is very low). In this manner, I-NeurAL will save more query budget but take more risks on many instances, incurring more regret. This explains why the red line ($\gamma = 1$) is above all other lines. In contrast, if γ is a large number, I-NeurAL will be more aggressive in making queries and thus obtain satisfactory performance. However, if γ is too large,

I-NeurAL tends to query on these instances even when our model is very confident to the predictions, wasting the query budget that could have been used on these uncertain instances. Therefore, we expect that γ is neither too small nor too large. The experiments verify our assumption, when γ is 6 or 7, I-NeurAL almost achieves the best performance throughout all datasets and configurations.

Ablation study for label budget. To examine the final performance of each algorithm, we conduct new experiments with different percentages of label budget: 3%, 10%, 20%, 50%. After T rounds, we evaluate the latest model on the test (unseen) data to calculate the accuracy, which evaluates the population accuracy. For all the datasets, T is set as 10000, except that $T = 2000$ for Phishing because Phishing has fewer data instances. Table 3 - 6 reports the results. To sum up, I-NeurAL still achieves the best accuracy with different label budget. With a small amount of label budget (3%, 10%), I-NeurAL can make smart decisions to query labels on these instances with big uncertainty and leverage the full feedback to exploit the past knowledge, which enable I-NeurAL to outperform all the baselines. With the larger label budget (20%, 50 %), all methods have enough labels to train. Thus, the advantages of I-NeurAL is less significant and the gap between I-NeurAL and baselines is decreasing. Nevertheless, I-NeurAl still has the best performance benefiting from smart query choices.

	Phishing	IJCNN	Letter	Fashion	MNIST	CIFAR-10
Random	91.75%	93.80%	71.60%	95.70%	87.90%	86.40%
Margin	93.46%	92.95%	73.55%	98.15%	90.25%	88.25%
NeuAL-NTK-F	54.69%	75.15%	48.05%	51.30%	51.10%	71.00%
NeuAL-NTK-D	92.89%	93.65%	73.80%	97.70%	90.15%	84.05%
ALPS	91.47%	93.25%	71.45%	95.70%	86.95%	85.40%
I-NeurAL	**94.22%**	**95.75%**	**77.45%**	**99.15%**	**94.45%**	**89.00%**

Table 3: Test Accuracy with **3%** budget.

	Phishing	IJCNN	Letter	Fashion	MNIST	CIFAR-10
Random	93.93%	96.70%	79.70%	97.30%	90.90%	89.00%
Margin	94.98%	97.10%	81.50%	98.60%	94.40%	89.45%
NeuAL-NTK-F	54.69%	87.65%	48.05%	51.30%	51.10%	70.95%
NeuAL-NTK-D	92.99%	96.90%	80.55%	98.70%	94.85%	89.05%
ALPS	92.89%	96.20%	78.05%	97.50%	93.00%	89.35%
I-NeurAL	**95.64%**	**97.90%**	**83.95%**	**99.30%**	**97.20%**	**91.75%**

Table 4: Test Accuracy with **10%** label budget.

	Phishing	IJCNN	Letter	Fashion	MNIST	CIFAR-10
Random	93.93%	96.70%	81.90%	98.15%	93.00%	89.50%
Margin	95.17%	98.15%	82.45%	98.90%	95.05%	89.75%
NeuAL-NTK-F	54.69%	89.95%	48.05%	51.30%	51.10%	72.15%
NeuAL-NTK-D	94.98%	97.75%	82.35%	99.30%	96.15%	90.90%
ALPS	94.41%	97.05%	83.20%	98.30%	94.45%	88.95%
I-NeurAL	**95.64%**	**98.35%**	**84.65%**	**99.30%**	**97.95%**	**91.80%**

Table 5: Test Accuracy with **20%** label budget.

	Phishing	IJCNN	Letter	Fashion	MNIST	CIFAR-10
Random	94.98%	97.75%	86.05%	98.95%	96.40%	90.75%
Margin	95.73%	98.40%	86.35%	99.05%	96.20%	91.25%
NeuAL-NTK-F	54.69%	90.85%	48.05%	51.30%	51.10%	72.35%
NeuAL-NTK-D	95.83%	98.00%	83.35%	99.30%	97.15%	90.55%
ALPS	94.50%	97.80%	87.10%	99.05%	96.70%	90.85%
I-NeurAL	**96.02%**	**98.75%**	86.05%	**99.35%**	**97.80%**	**92.30%**

Table 6: Test Accuracy with **50%** label budget.