Management Science

Active Learning for Contextual Search with Binary Feedback

Xi Chen, Quanquan Liu, Yining Wang

Please scroll down for article—it is on subsequent pages

Active Learning for Contextual Search with Binary Feedback

Xi Chen,[a],* Quanquan Liu,[b] Yining Wang[b]

[a] Leonard N. Stern School of Business, New York University, New York, New York 10012; [b] Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080
*Corresponding author
Contact: xc13@stern.nyu.edu, https://orcid.org/0000-0002-9049-9452 (XC); qxl220001@utdallas.edu (QL); yxw220006@utdallas.edu, https://orcid.org/0000-0001-9410-0392 (YW)

Abstract. In this paper, we study the learning problem in contextual search, which is motivated by applications such as crowdsourcing and personalized medicine experiments. In particular, for a sequence of arriving context vectors, with each context associated with an underlying value, the decision maker either makes a query at a certain point or skips the context. The decision maker will only observe the binary feedback on the relationship between the query point and the value associated with the context. We study a probably approximately correct learning setting, where the goal is to learn the underlying mean value function in context with a minimum number of queries. To address this challenge, we propose a trisection search approach combined with a margin-based active learning method. We show that the algorithm only needs to make $\widetilde{O}(1/\varepsilon^2)$ queries to achieve an ε-estimation accuracy. This sample complexity significantly reduces the required sample complexity in the passive setting where neither sample skipping nor query selection is allowed, which is at least $\Omega(1/\varepsilon^3)$.

Keywords: active learning • binary feedback • classification • contextual search

1. Introduction

Contextual search, which extends the classical binary search problem to high dimensions, finds a wide range of applications, such as crowdsourcing and personalized medicine. In the contextual search problem, for each round $i = 1, 2, 3, \cdots$, an item (e.g., a customer or a patient) arrives sequentially, each with a contextual vector $x_i \in \mathbb{R}^d$ accessible to the decision maker. We assume that the context x_i incurs an unknown stochastic *value* $u_i = v(x_i) + \xi_i$, where $v(x_i)$ is the mean value function of x_i and ξ_i is the stochastic noise. The decision maker selects a query $b_i \in \mathbb{R}$ and then observes the *binary feedback*, that is, whether $u_i \geq b_i$ or vice versa. The true value u_i will never be revealed. To better fit our motivating applications illustrated below, the decision maker is allowed to skip making a query on certain contextual vectors to save her budget. Our goal is to learn the mean value function $v(x_i)$ with a minimum number of queries/trials. It is worth noting that we adopt the sample complexity as the objective instead of revenue/cost because we focus on the experimental phase for the learning purpose. In this phase, the number of trials is usually quite small, and thus, it is common to treat cost equally for each trial. We now briefly describe two motivating applications.

Personalized Medicine Experiment: Let us consider the example of clinical trials, where the goal of an experiment is to determine the proper dosage $v(x)$ in radiation therapy. The profile of each potential experimental unit is characterized by x (e.g., her demographics, diagnosis, medications, and genetics). Bastani and Bayati (2020) adopted a linear bandit model (i.e., the linear form of $v(x)$) to investigate the relationship between the optimal dosage and the patients' profile. In this experiment, the laboratory posts an advertisement to the public and receives nominations (e.g., someone will call the laboratory to express her interest). After the laboratory receives a nomination and conducts some prescreening to collect the profile information x, the laboratory can simply reject a nomination without further experimental procedure. For example, if the laboratory has already experimented on a similar unit (i.e., a unit with a similar profile x), the laboratory will naturally reject this potential unit. If the laboratory decides to accept the experimental unit, we assume that the laboratory will recommend a dosage b_i and receive the binary feedback on whether the recommended dosage is above or below the appropriate level. As performing a radiation therapy experiment is costly and time-consuming, a common goal is to use the minimum number of trials to learn the ideal personalized dosage level function (i.e., the $v(\cdot)$ function).

Crowdsourcing: In a crowdsourcing experiment, the decision maker hires a crowdsourced expert to help

determine the difficulty level (e.g., measured by completion time) of different tasks characterized by their context vectors x. Assuming for each task i, the underlying difficulty level is $u_i = v(x_i) + \xi_i$. Numerous psychology studies have shown one is more good at providing pairwise comparison than absolute numerical estimate (Shiffrin and Nosofsky 1994, Stewart et al. 2005). Therefore, instead of asking the expert to provide a numerical estimate of the difficulty level, the decision maker will give an estimate b_i. Then the expert provides binary feedback on whether she believes b_i is an over-estimate or under-estimate. In such a crowdsourcing experiment, it is natural that the decision maker will not bother the expert to provide feedback for some tasks (e.g., those tasks similar to previously queried jobs).

Motivated by these applications, the goal of this paper is to propose an efficient algorithm to learn $v(x)$. Following the existing literature on contextual search and feature-based pricing, we also adopt a linear model of the mean valuation function, that is, $v(x) = \langle x, w^* \rangle - \mu^*$ for some unknown coefficient vector $w^* \in \mathbb{R}^d$ and the intercept $\mu^* \in \mathbb{R}$. Compared with the existing literature, our contextual search problem has the following unique features, which calls for new algorithmic development:

1. First, the existing contextual search setup aims to minimize either the absolute loss $|b_i - v(x_i)|$ or the ϵ-ball loss $\mathbb{I}(|b_i - v(x_i)| > \epsilon)$ for some predetermined ϵ over time. Here $\mathbb{I}(\cdot)$ denotes the indicator function. In contrast, we consider a learning problem, where the goal is to learn $v(x)$ as accurately as possible. Therefore, we adopt a probably approximately correct (PAC) setting (see (2) in Section 2) instead of regret minimization setting in existing literature (Leme and Schneider 2018, Lobel et al. 2018, Cohen et al. 2020, Krishnamurthy et al. 2021). To facilitate the analysis of this learning problem, we assume the stochasticity of the contextual information x_i.
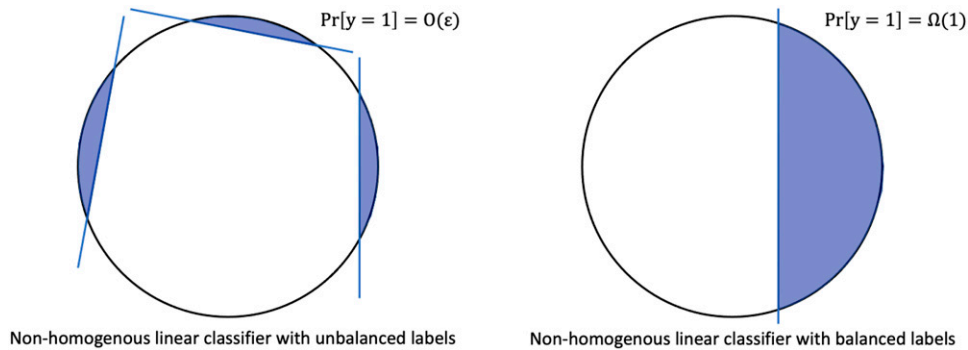
2. Second, as we are motivated by *experimental* applications, the decision maker should judge the benefit of a context x_i to the learning problem. Therefore, compared with the existing contextual search, our problem has another layer of decision, that is, whether to conduct a query or not, beyond the decision of the query point itself.

To address this problem, we adopt the active learning framework from machine learning research (Settles 2012). In particular, we adopt the margin-based active learning approach (Balcan et al. 2007). At a high level, let $\widehat{v}(\cdot)$ be the current estimate of the underlying $v(\cdot)$ function and $\widehat{b}$ be the query point. For an arriving context x, the margin-based active learning will make a query if $|\widehat{v}(x) - \widehat{b}|$ is sufficiently small, which indicates that it is difficult to determine the relationship between $\widehat{b}$ and $\widehat{v}(x)$. Although it is an intuitive approach, existing margin-based active learning approaches cannot be directly applied to address our problem because of the existence of the intercept μ^*. In fact, a famous negative result by Dasgupta (2005b) shows that active learning cannot significantly improve sample complexity over *passive learning* for linear binary classification models *with intercepts* in its most general form. It is worth noting that throughout this paper, by "passive learning," we refer to the learning paradigm in which the decision making can neither skip samples (regardless of their contextual information x) nor adaptively change actions/queries. Please refer to Figure 1 for details.

To address this challenge, we propose an active learning procedure consisting of three major stages:

1. The first stage of the algorithm is to use trisection search to locate two queries $\widehat{b}_1$ and $\widehat{b}_2$ that are close to the underlying intercept term μ^*, without consuming too many labeled (queried) samples. In this first stage sample selection (i.e., determining whether a sample is to be labeled/queried or not) is *not* carried out, but the

Figure 1. (Color online) Negative Examples of Problem Instances Constructed in Dasgupta (2005b)

Pr[y = 1] = O(ε)

Pr[y = 1] = Ω(1)

Non-homogenous linear classifier with unbalanced labels

Non-homogenous linear classifier with balanced labels

Notes. (Left) Examples of nonhomogeneous linear classifiers with *unbalanced* labels, for which Dasgupta (2005b) shows that active learning (i.e., sample selection for labeling purposes) cannot lead to significantly improved sample complexity. (Right) Nonhomogenous linear classifiers with *balanced* labels, for which improvements of sample complexity can be made via sample selection (active learning).

algorithm will actively explore different actions in order to obtain $\widehat{b}_1, \widehat{b}_2$ that are close to μ^*;

2. The second stage of the algorithm is to apply margin based active learning to learn the linear model w^* and an intercept term depending on both μ^* and $\widehat{b}_1, \widehat{b}_2$. In this second stage sample selection will be carried out, as only those users with contextual vectors x_t close to classification hyperplanes will be queried/labeled (see Algorithm 3 for details). The actions taken in this stage (on selected samples) will be fixed to either $\widehat{b}_1$ or $\widehat{b}_2$ obtained in the first stage.

Although this classification model still has nonzero intercept terms, the closeness of $\widehat{b}_1, \widehat{b}_2$ to μ^* would imply that the obtained labels under actions $\widehat{b}_1$ or $\widehat{b}_2$ are *balanced*, circumventing the negative results in the work of Dasgupta (2005b) that specifically constructed counter-examples with unbalanced labels. In Figure 1 and the following related work section, we give a detailed account of this negative example and how it presents challenges to active learning. Indeed, our theoretical analysis extends the arguments in Balcan et al. (2007) to this more general setting of linear classification *with intercepts* and balanced labels, with similar convergence rates derived.

3. The final stage of the algorithm is to reconstruct the mean utility model $\widehat{v}(\cdot)$ from the estimated linear model and intercepts. Because margin-based active learning can only estimate a linear model up to scales, we need model estimates at two different actions $\widehat{b}_1, \widehat{b}_2$ (corresponding to two different effective intercepts) in order to reconstruct w^* and μ^* in $v(\cdot)$. Details of how this reconstruction is carried out are given in the last two lines of Algorithm 1.

We establish the sample complexity bound for the proposed margin-based active learning with a trisection search scheme. We assume that with $\widetilde{O}(1/\varepsilon^3)$ total number of incoming contexts, the decision maker only needs to make $\widetilde{O}(1/\varepsilon^2)$ queries to estimate the mean value function $v(x)$ within ε-precision (with high probability). Here $\widetilde{O}$ here hides the dependence on d and other logarithmic factors. We also show that in the passive setting, where the decision maker is required to conduct queries for all arriving contexts as in the standard contextual search, the sample complexity would be at least $\Omega(1/\varepsilon^3)$ (see Remark 1).

1.1. Related Work
Our problem setting can be viewed as a variant of the contextual search problem, which is an extension of the classical binary search. In binary search, the decision maker tries to guess a fixed constant μ^* (i.e., the value $u_i \equiv \mu^*$ for all i in our problem) by iteratively making queries b_i. In the PAC learning setting, the binary search algorithm only needs $O(\log(1/\varepsilon))$ queries to estimate μ^* within ε-precision. Because of the importance of applications such as personalized

medicine and feature-based pricing, contextual search has received a lot of attention in recent years. The existing literature mainly adopts the linear model for the mean value function. For ε-ball loss $\sum_i \mathbb{I}(|b_i - v(x_i)| > \varepsilon)$, Lobel et al. (2018) established the $\Omega(d \log (1/\varepsilon\sqrt{d}))$ regret lower bound and proposed the project volume algorithm that achieves a near-optimal regret of $O(d \log(d/\varepsilon))$. For absolute loss $\sum_i |b_i - v(x_i)|$, Leme and Schneider (2018) established the regret bound of $O(\text{poly}(d))$. As we explained in the introduction, to fit the applications considered in our paper, we adopt a PAC learning setting and equip the decision maker with the ability to pass an incoming context. While most contextual search settings in the literature consider adversarial contextual information, we assume the stochasticity of the contextual information as we study a learning problem.

Active learning is an important research area in machine learning, originating from the seminal work of Cohn et al. (1994) dating back to the 1990s. The main idea behind active learning is to equip the learning algorithm with the ability to *select* samples or data points to be labeled, improving its sample complexity in applications where labels are expensive to obtain but unlabeled data are abundant. There have been many successful algorithms developed for active learning, such as bisection search for one-dimensional noiseless problems (Dasgupta 2005b), greedy method (Dasgupta 2005a), disagreement-based active learning (Hanneke 2007, Balcan et al. 2009, Zhang and Chaudhuri 2014), margin based active learning (Balcan et al. 2007, Balcan and Long 2013, Wang and Singh 2016) and active learning based on surrogate loss functions (Awasthi et al. 2017, Balcan and Zhang 2017). Because of the vast amount of literature on active learning, we cannot cite all related works here and would like to refer interested readers to the excellent review of Hanneke (2014) for an overview of this area.

Our approach in this paper resembles the margin-based active learning method (Balcan et al. 2007, Balcan and Long 2013, Wang and Singh 2016) that is developed for linear classifiers and have been popular in the active learning literature because of its intuitive nature, tight sample complexity, and relative ease of implementation. However, while linear classifiers seem simple, nonhomogeneous linear classifiers (i.e., linear classifiers with *an intercept term*) present notorious challenges to active learning algorithms. More specifically, the work of Dasgupta (2005a) shows that $d \geq 2$ and nonhomogeneous linear classifiers produce unbalanced samples, such as the example shown in the left panel of Figure 1. In this illustrative example, potential linear classifiers are within $O(\varepsilon)$ distance to the domain boundary, and thus, active learning *cannot* asymptotically improve sample complexity over passive learning as it takes $O(1/\varepsilon)$ samples to hit the

boundaries. It is easy to verify that, if a nonhomogenous linear classifier is within ϵ distance to the boundary and the underlying distribution of unlabeled samples is relatively uniform, the probability of seeing a positive sample (as indicated in the filled region in Figure 1) is also on the order of $O(\varepsilon)$. To overcome this counter-example, in this paper, we exploit the special structure in the contextual search problem to "balance" the labels, as shown in the right panel of Figure 1. Although the balanced model still possesses a nonzero intercept term, the classifier will be generally $\Omega(1)$ away from the boundary, which our theoretical analysis shows is sufficient of obtaining desired sample complexity results for active learning.

It is also interesting to make comparisons to other margin-based active learning work. For example, the work by Awasthi et al. (2017) considers the following setting: for an underlying (unknown) model w^* and feature vector x, the algorithm observes $\text{sgn}(\langle x, w^* \rangle)$ with probability $1 - \eta$ and an adaptively chosen label with probability η. In comparison, in our problem setting the algorithm observes labels with probability related to the margin $|\langle x, w^* \rangle|$. Such a difference in the setup leads to a fundamental difference in the sample complexity: the sample complexity in Awasthi et al. (2017) is poly-logarithmic in $1/\varepsilon$, whereas in our problem setting, a poly$(1/\varepsilon)$ sample complexity is necessary. In fact, $\log(1/\varepsilon)$ sample complexity is only possible if one has deterministic labels or probabilistic labels satisfying the Massart noise condition; that is, for any $x \in \mathbb{R}^d$ with $\langle x, w^* \rangle > 0$, $\Pr[y = 1|x] > 1/2 + c$ for some constant $c > 0$ (and vice versa for all $\langle x, w^* \rangle < 0$). Such a condition clearly is not satisfied by the setting studied in this paper, in which $\Pr[y = 1|x] \to 1/2$ as $\langle x, w^* \rangle \to 0^+$. For noise distributions not satisfying the Massart condition, poly$(1/\varepsilon)$ samples are necessary (Balcan et al. 2007, Ben-David and Urner 2014, Wang and Singh 2016).

Our work is also related to the pure-exploration problem and sequential experimental design (Elfving 1952, Chernoff 1959, Albert 1961, Naghshvar and Javidi 2013, Wang and Zenios 2020, Araman and Caldentey 2021, Li et al. 2021, Wager and Xu 2021, Chen et al. 2022, Feng et al. 2022). In experimental design problems, the decision maker is capable of choosing the context vector x. However, in application settings considered in this paper (e.g., experimental units arriving sequentially), it is impractical to assume that the context vectors $\{x_t\}_{t=1}^T$ could be chosen arbitrarily. Thus, we only allow the decision maker to decide whether to skip a query.

Active learning has been an important area in machine learning. However, it has not received a lot of attention in operations management. This paper takes a preliminary step on exploring the applications of active learning, and hopefully, it will inspire more research on active learning to address challenges arising from operations management.

1.2. Paper Organization and Notations
The rest of the paper is organized as follows. Section 2 describes the problem formulation and necessary assumptions. Section 3 develops our margin-based active learning algorithm with the trisection search and establishes the sample complexity bound. The technical proofs are provided in Section 4. We provide the numerical simulation studies in Section 5, followed by the conclusion in Section 6. Proofs of some technical lemmas are relegated to the appendix.

In our paper, the asymptotic is with respect to d and ε, with all other parameters being functions of d, ε and other problem-dependent constants (e.g., $B, c_x, C_x, c_\xi, C_\xi$) that do not change with d, ε. We say that $f(x) = O(g(x))$ if there exist constants d_0, ε_0 and $C < \infty$ such that for all $d \geq d_0$ and $\varepsilon \leq \varepsilon_0$, $f(x) \leq Cg(x)$. If we omit dependency on constants $\theta = (B, c_x, C_x, c_\xi, C_\xi)$ in the big-O notation, then the constant C can be a function of θ. If we further omit poly-logarithmic dependency (by using the notation $\widetilde{O}(g(x))$), then the constant C can depend on $\log^c x$ for some constant c.

2. Problem Formulation and Assumptions
In our modeling, assuming the items (e.g., ads or experimental units) $i = 1, 2, 3, \ldots$ arrive sequentially, each with a contextual or feature vector $x_i \in \mathbb{R}^d$ accessible to the decision maker. We assume that the contextual vectors $\{x_i\}_{i \geq 1}$ are independently and identically distributed with respect to an unknown underlying distribution P_X. We also assume that $\|x_i\|_2 \leq 1$ for the ease of illustration. Given the contextual vector $x_i \in \mathbb{R}^d$, the "valuation" of the item (e.g., the appropriate dosage in personalized medical treatment) follows a linear model:

$$u_i = v(x_i) + \xi_i = \langle x_i, w^* \rangle - \mu^* + \xi_i, \quad (1)$$

where $v(\cdot) = \langle \cdot, w^* \rangle - \mu^*$ is an underlying linear model with a fixed but unknown coefficient vector $w^* \in \mathbb{R}^d$, the intercept $\mu^* \in \mathbb{R}$, and the noise $\{\xi_i\}_{i \geq 1}$, which are independently and identically distributed stochastic variations with respect to an unknown distribution P_ξ.

After observing the contextual vector $x_i \in \mathbb{R}^d$, the decision maker will do either one of the following:

1. Let the item pass without taking any actions, and thereby without obtaining any feedback/information;

2. Make a query at $b_i \in \mathbb{R}$, and observe the binary feedback $y_i = 1$ if $u_i \geq b_i$ or $y_i = -1$ if $u_i < b_i$.

Because making a query (e.g., admitting an experimental unit into a clinical trail program) incurs much higher implicit cost as compared with passing (i.e., taking no action), the main goal of the decision maker is to use as few number of queries as possible to estimate the mean valuation function $v(\cdot)$ to a certain precision. More specifically, let $\varepsilon, \delta \in (0, 1)$ be target accuracy and probability parameters. We use $n(\varepsilon, \delta)$ to

denote the number of queries a learning algorithm takes in order to produce an estimate $\widehat{v}(\cdot)$ that satisfies

$$\sup_{\|x\|_2 \leq 1} |\widehat{v}(x) - v(x)| \leq \varepsilon, \quad \text{with probability} \geq 1 - \delta. \tag{2}$$

Clearly, the smaller $n(\varepsilon, \delta)$ is the more efficient the designed learning algorithm is. The main objective of this paper is to design an active learning algorithm that minimizes $n(\varepsilon, \delta)$. Additionally, we use $m(\varepsilon, \delta)$ to denote the number of *total* samples (i.e., the number of total incoming contexts) an algorithm requires to obtain an estimate $\widehat{v}$ satisfying Equation (2). Although those incoming contexts skipped by our algorithm usually do not incur extra cost, it is desirable that $m(\varepsilon, \delta)$ is reasonable because the supply of experimental units might still be limited. In active learning literature, an $m(\varepsilon, \delta)$ is *reasonable* if it is a polynomial function in terms of $1/\varepsilon, \log(1/\delta)$ and d (Cohn et al. 1994, Cohn 1996, Balcan et al. 2007).

Throughout this paper, we impose the following assumptions.

Assumption 1. *There exists a constant $B < \infty$ such that $\|w^*\|_2 \leq B$ and $|\mu^*| \leq B$.*

Assumption 2. *The distribution P_X satisfies the following condition: it is supported on the unit ℓ_2 ball $\mathbb{B}_2(d) = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$; it admits a probability density function $f_x(\cdot)$; there exist constants $0 < c_x \leq C_x < \infty$ such that $c_x f_u(x) \leq f_x(x) \leq C_x f_u(x)$ for all $x \in \mathbb{B}_2(d)$, where f_u is the probability density function (PDF) of the uniform distribution on $\mathbb{B}_2(d)$.*

Assumption 3. *The distribution P_ξ satisfies the following condition: $\Pr[\xi \leq 0] = \Pr[\xi \geq 0] = 1/2$; it admits a probability density function $f_\xi(\cdot)$; there exist constants $0 < c_\xi \leq C_\xi < \infty$ such that $\sup_{\xi \in \mathbb{R}} f_\xi(\xi) \leq C_\xi/\|w^*\|_2$ and $\inf_{|\xi| \leq 2} f_\xi(\xi) \geq c_\xi/\|w^*\|_2$.*

Assumption 1 is a standard bounded assumption imposed on model parameters. Assumption 2 assumes that the contextual vectors are independently and identically distributed, with respect to a bounded and nondegenerate distribution P_X that is unknown. Similar "nondegenerate" or "covariate diversity" assumptions were also adopted in the contextual learning literature (Bastani and Bayati 2020, Bastani et al. 2021), and the assumption is actually weaker than some of the existing works on active learning (Balcan et al. 2007, Wang and Singh 2016), which requires P_X to be the exact uniform distribution over $\mathbb{B}_2(d)$.

Assumption 3 is a general condition imposed on the distribution P_ξ of the noise variables. Essentially, it assumes that zero is the median of the noise distribution P_ξ, which ensures that the linear classifier is the optimal Bayes classifier. The same assumption is common in the active learning literature (Balcan et al. 2007, Wang and Singh 2016). We do not assume the

noise distribution P_ξ has any specific parametric forms (e.g., logistic or Probit noises), making it generally applicable to a broad range of problems. Assumption 3 requires the noise distribution to scale together with $\|w^*\|_2$ to preserve signal-to-noise ratios. In the case of a signal-independent assumption $\sup_{\xi \in \mathbb{R}} f_\xi(\xi) \leq C'_\xi$ and $\inf_{|\xi| \leq 2} f_\xi(\xi) \geq c'_\xi$, the change-of-parameter $C'_\xi = C_\xi/\|w^*\|_2$ and $c'_\xi = c_\xi/\|w^*\|_2$ can be used to bring the signal level $\|w^*\|_2$ into the sample complexity analysis.

3. Margin-Based Active Learning with Trisection Search

Algorithm 1 (Meta-Algorithm for Actively Learning Contextual Functions)

1: **Input**: dimension d, accuracy parameters ε, δ, algorithm parameters $\kappa_m, \kappa_n, \kappa_\varepsilon, \beta_0$.
2: $\widehat{b}_1, \widehat{b}_2 \leftarrow$ TRISECTIONSEARCH$(\varepsilon_s, \delta_s)$ with $\varepsilon_s = 0.1/\sqrt{d-1}$, $\delta_s = \delta/3$;
3: Let $\varepsilon_a = \kappa_\varepsilon \varepsilon^2/\ln^2(1/\varepsilon)$, $\delta_a = \delta/3$;
4: $(\widehat{w}_1, \widehat{\beta}_1) \leftarrow$ MARGINBASEDACTIVELEARNING$(\widehat{b}_1, \varepsilon_a, \delta_a, \kappa_m, \kappa_n, \sqrt{\varepsilon_a}, \beta_0)$;
5: $(\widehat{w}_2, \widehat{\beta}_2) \leftarrow$ MARGINBASEDACTIVELEARNING$(\widehat{b}_2, \varepsilon_a, \delta_a, \kappa_m, \kappa_n, \sqrt{\varepsilon_a}, \beta_0)$;
6: Let $\widehat{\alpha} = (\widehat{b}_2 - \widehat{b}_1)/(\widehat{\beta}_2 - \widehat{\beta}_1)$;
7: **Output**: utility function estimate $\widehat{v}(\cdot) = \langle \cdot, \widehat{w} \rangle - \widehat{\mu}$, where $\widehat{w} = \widehat{\alpha} \widehat{w}_1$ and $\widehat{\mu} = \widehat{\alpha} \widehat{\beta}_1 - \widehat{b}_1$.

The main algorithm we proposed for actively learning contextual functions is given in Algorithm 1. The main idea of the proposed algorithm can be summarized as follows.

The first step is to find two actions $\widehat{b}_1, \widehat{b}_2$ that are reasonably close to the mean utility μ^*. This is to ensure that when the actions are fixed at $\widehat{b}_1$ or $\widehat{b}_2$, the labels received from user streams are relatively balanced, thereby circumventing the negative results in the work of Dasgupta (2005b). In Section 3.1, we show how $\widehat{b}_1, \widehat{b}_2$ can be found without using too many labeled samples, using a trisection search idea.

After we obtained candidate actions $\widehat{b}_1$ and $\widehat{b}_2$, we use a margin-based active learning algorithm to estimate the linear model w^* and mean utility μ^*. The margin-based active learning algorithm is similar to the work of Balcan et al. (2007), with the difference being that in our setting the active learning algorithm needs to incorporate a (relatively small) intercept term, which complicates its design and analysis.

Finally, we use the estimates $(\widehat{w}_1, \widehat{\beta}_1)$ and $(\widehat{w}_2, \widehat{\beta}_2)$ obtained from the previously mentioned active learning procedure *under two different fixed actions $\widehat{b}_1, \widehat{b}_2$* to reconstruct the linear utility parameters w^* and μ^*. The reason we need two fixed actions $\widehat{b}_1, \widehat{b}_2$ is because the active learning procedure solves a classification problem, for which we can only estimate the linear model and its intercept *up to scalings* because if one

multiplies both the linear model and its intercept by a constant the resulting classification problem is the same. Hence, we need two fixed actions $\widehat{b}_1, \widehat{b}_2$ to construct an approximate linear system of equations, the solution of which would give us consistent estimates of w^* and μ^*.

We briefly explain our intuition behind the construction of the utility function estimate $\widehat{v}(\cdot)$ in Algorithm 1. For simplicity, we will omit the learning errors that occurred in the two MARGINBASEDACTIVELEARING invocations. Because the margin based active learning algorithm learns linear classifiers up to normalization (see Algorithm 3), we have the following equivalence:

$$\langle \widehat{w}_1, x \rangle - \widehat{\beta}_1 > 0 \Longleftrightarrow \langle w^*, x \rangle - \mu^* > \widehat{b}_1;$$
$$\langle \widehat{w}_2, x \rangle - \widehat{\beta}_2 > 0 \Longleftrightarrow \langle w^*, x \rangle - \mu^* > \widehat{b}_2,$$

where $\|\widehat{w}_1\|_2 = \|\widehat{w}_2\|_2 = 1$ because of the construction of Algorithm 3. Again, we emphasize that the above equivalence only holds approximately because of learning errors of $\widehat{w}_1, \widehat{\beta}_1, \widehat{w}_2, \widehat{\beta}_2$, but we will omit these learning errors for ease of explanation. Let $\alpha = \|w^*\|_2$. We have $\widehat{\beta}_1 = (\mu^* + \widehat{b}_1)/\alpha$ and $\widehat{\beta}_2 = (\mu^* + \widehat{b}_2)/\alpha$. Therefore, we set $\widehat{\alpha} = (\widehat{b}_2 - \widehat{b}_1)/(\widehat{\mu}_2 - \widehat{\mu}_1)$ as the estimate of α, and $\widehat{\mu} = \widehat{\alpha}\widehat{\beta}_1 - \widehat{b}_1$ as the estimate of $\widehat{\mu}$. Thus, we obtain the utility function estimate $\widehat{v}(\cdot)$ in Algorithm 1.

3.1. Trisection Search for Accurate Mean Utility

Now we are ready to present our trisection search algorithm for estimating the mean utility parameter μ. The graphical illustration is provided in Figure 2.

Algorithm 2 (Trisection Search Algorithm to Roughly Estimate the Mean Utility Parameter μ^*)

1: **function** TRISECTIONSEARCH(ε_s, δ_s)
2:   Initialize: $n = 0$, lower and upper bounds $\widehat{b}_1 = -B, \widehat{b}_2 = B$;
3:   **while** $\widehat{b}_2 - \widehat{b}_1 > \varepsilon_s$ **do**
4:     $\widehat{b}_3 \leftarrow \widehat{b}_1 + (\widehat{b}_2 - \widehat{b}_1)/3$, $\widehat{b}_4 \leftarrow \widehat{b}_2 - (\widehat{b}_2 - \widehat{b}_1)/3$, $\widehat{n} = \widehat{r}_3 = \widehat{r}_4 = 0$, $\underline{p}_3 = \underline{p}_4 = 0$, $\bar{p}_3 = \bar{p}_4 = 1$;
5:     **while** $\underline{p}_3 \leq 0.5 \leq \bar{p}_3$ and $\underline{p}_4 \leq 0.5 \leq \bar{p}_4$ **do**
6:       For an incoming user x, take action $\widehat{b}_3$ and observe result $y \in \{0, 1\}$;
7:       For another incoming user x', take action $\widehat{b}_4$ and observe result $y' \in \{0, 1\}$;
8:       $n \leftarrow n + 1$, $\widehat{n} \leftarrow \widehat{n} + 1$, $\widehat{r}_3 = \widehat{r}_3 + \mathbf{1}\{y_3 = 1\}$, $\widehat{r}_4 = \widehat{r}_4 + \mathbf{1}\{y_4 = 1\}$;
9:       Update: $[\underline{p}_3, \bar{p}_3] \leftarrow \frac{\widehat{r}_3}{\widehat{n}} \pm \sqrt{\frac{\ln(8n^2/\delta_s)}{2\widehat{n}}}$ and $[\underline{p}_4, \bar{p}_4] \leftarrow \frac{\widehat{r}_4}{\widehat{n}} \pm \sqrt{\frac{\ln(8n^2/\delta_s)}{2\widehat{n}}}$;
10:    **end while**
11:    Set $\widehat{b}_1 \leftarrow \widehat{b}_3$ if $\underline{p}_3 > 0.5$ or $\underline{p}_4 > 0.5$ and $\widehat{b}_2 \leftarrow \widehat{b}_4$ otherwise;
12: **end while**
13: **return** $(\widehat{b}_1, \widehat{b}_2)$.
14: **end function**

Let $b^* \in \mathbb{R}$ be the unique value such that $\Pr_{x \sim P_X}[v(x) \geq b^*] = 1/2$. Because P_X and P_ξ have PDFs, such a value of b^* exists and is unique. Intuitively, if one commits to the fixed action b^*, then the labels received by the algorithm should be balanced. Algorithm 2 shows how to find actions $\widehat{b}_1, \widehat{b}_2$ that are reasonably close to b^*, without consuming too many labeled samples.

The main idea behind Algorithm 2 is a trisection search approach, motivated by the fact that the probability $\Pr_{x \sim P_X}[v(x) \geq b]$ is a monotonically decreasing function of b, and furthermore as $|b - b^*|$ increases the gap between $\Pr_{x \sim P_X}[v(x) \geq b]$ and $\Pr_{x \sim P_X}[v(x) \geq b^*] = 1/2$ will also increase (see Lemma 4 in the proof). This allows us to use a trisection search procedure to localize the value of b^*, by simply comparing an empirical estimate of $\Pr_{x \in P_X}[v(x) \geq b]$ at the current value of b. More specifically, at an iteration $\widehat{b}_3, \widehat{b}_4$ are the two midpoints and $[\underline{p}_3, \bar{p}_3]$ are lower and upper estimates of $\Pr_{x \sim P_X}[v(x) \geq \widehat{b}_3]$ and similarly $[\underline{p}_4, \bar{p}_4]$ are lower and upper estimates for $\Pr_{x \sim P_X}[v(x) \geq \widehat{b}_4]$. With either probability being separated from $1/2$, the algorithm could move $\widehat{b}_1$ or $\widehat{b}_2$ to $\widehat{b}_3$ or $\widehat{b}_4$. The algorithm is guaranteed to maintain that $b^* \in [\widehat{p}_1, \widehat{p}_2]$, thanks to the monotonicity of $\Pr_{x \sim P_X}[v(x) \geq b]$ with respect to b.

The following technical lemmas are the main results explaining the objective and guarantee of Algorithm 2, which are proved in Section 4.1.

Lemma 1. *Suppose $d \geq 2$ and let $\beta^* = b^* + \mu^*$. Then $\frac{|\beta^*|}{\|w^*\|_2} \leq \sqrt{\frac{2\ln(100 C_x C_\xi / c_x c_\xi)}{d-1}} = O(1/\sqrt{d})$.*

Lemma 2. *Suppose $d \geq 2$ and let $(\widehat{b}_1, \widehat{b}_2)$ be the values returned by BISECTIONSEARCH(ε_s, δ_s). With probability $1 - \delta_s$ the following hold: $\widehat{b}_1 \leq b^* \leq \widehat{b}_2$, and at most $O(\varepsilon_s^{-2} \log(1/\delta_s \varepsilon_s)) = O(d \log(d/\delta))$ queried samples are consumed.*

Intuitively, Lemma 1 establishes that the "balancing" intercept b^* is $O(1/\sqrt{d})$ close to the intercept μ^* in the utility model, which is helpful for our later analysis. Lemma 2 further establishes that the returned two actions $\widehat{b}_1, \widehat{b}_2$ sandwich the "label-balancing" action b^* and also upper bound the total number of labeled (queried) samples consumed in the algorithmic procedure.

3.2. Margin-Based Active Learning

In Algorithm 3, we provide the pseudocode description of the margin based active learning algorithm we use in this problem to actively learn a linear model with intercepts.

Algorithm 3 (Margin-Based Active Learning Nonhomogeneous Linear Classifiers)

1: **function** MARGINBASEDACTIVELEARNING($b, \varepsilon_a, \delta_a, \kappa_m, \kappa_n, \epsilon_0, \beta_0$)

2:   Collect $n_0 = \lceil \kappa_n / \epsilon_0^2 \rceil$ samples with action b and let $\mathcal{D}_0 = \{(x,y)\} \subseteq \mathbb{B}_d(2) \times \{\pm 1\}, |\mathcal{D}_0| = n_0$ be the queried samples;

3:   Let $\widehat{w}_0, \widehat{\beta}_0 \leftarrow \arg\min_{\|w\|_2 = 1, |\beta| \le \beta_0} \sum_{(x,y) \in \mathcal{D}_0} \mathbf{1}\{y \ne \text{sgn} (\langle x, w \rangle - \beta)\}$;

4:   Let $k_0 = \min\{k \in \mathbb{N} : 2^{-k} \epsilon_0 \le \epsilon_a\}$;

5:   **for** $k = 1, 2, \ldots, k_0$ **do**

6:       $\epsilon_k \leftarrow 2^{-k} \epsilon_0, m_k \leftarrow \kappa_m \sqrt{\epsilon_k}, n_k \leftarrow \lceil \kappa_n d / \epsilon_k \rceil, \mathcal{D}_k = \emptyset$;

7:       **while** $|\mathcal{D}_k| < n_k$ **do**

8:           Observe context vector $x \in \mathbb{R}^d$ for the next object;

9:           **if** $|\langle x, \widehat{w}_{k-1} \rangle - \widehat{\beta}_{k-1}| \le m_k$ **then**

10:             Invoke action b and let $y \in \{\pm 1\}$ be the collected binary feedback;

11:             Update $\mathcal{D}_k \leftarrow \mathcal{D}_k \cup \{x, y\}$;

12:         **end if**

13:     **end while**

14:       $\widehat{w}_k, \widehat{\beta}_k \leftarrow \arg\min_{\|w\|_2 = 1, |\beta| \le \beta_0} \sum_{(x,y) \in \mathcal{D}_k} \mathbf{1}\{y \ne \text{sgn} (\langle x, w \rangle - \beta)\}$;

15:   **end for**

16:   **return** $\widehat{w}_{k_0}, \widehat{\beta}_{k_0}$.

17: **end function**

In Algorithm 3, the query point b is fixed, with the algorithm only able to select which sample/contextual vector to act on. Because the query point b is fixed, we can consider linear models with intercepts as $\widehat{v}(\cdot) = \langle \cdot, \widehat{w} \rangle - \widehat{\beta}$. For such a model, we define the *error* of $\widehat{v}$ under the query point b as

$$\text{err}_b(\widehat{v}) := \Pr_{x \sim P_X, \xi \sim P_\xi} [\text{sgn}(\underbrace{v(x) + \xi}_{u(x)} - b) \ne \text{sgn}(\widehat{v}(x))], \quad (3)$$

where $v(x) = \langle x, w^* \rangle - \mu^*$. For any $b \in \mathbb{R}$, the model $v_b^*(\cdot) := \langle \cdot, w^* \rangle - \mu^* - b$ has the smallest error defined in Equation (3). This is because $v_b^*(\cdot)$ is the Bayes classifier; that is, $v_b^*(x) \ge 0$ if and only if $\Pr[v(x) + \xi \ge b \mid x] \ge 1/2$. Hence, we can also define the *excess error* of a model $\widehat{v}(\cdot)$ as

$$\Delta\text{err}_b(\widehat{v}) := \text{err}_b(\widehat{v}) - \text{err}_b(v_b^*). \quad (4)$$

Figure 3 illustrates the principles of Algorithm 3. The main idea of Algorithm 3 is simple: The algorithm first uses a "warm-up" epoch consisting of n_0 queried samples to construct a preliminary model estimate $\widehat{w}_0$ and $\widehat{\beta}_0$. There is no sample selection or active learning in this warm-up procedure, and the analysis of excess errors of $\widehat{w}_0, \widehat{\beta}_0$ follows the standard Vapnik–Chervonenkis (VC) theory analyzing empirical risk minimizers of binary classifiers (see Lemma A.3 in the proof; Balcan et al. 2007, Vapnik 2013, Vapnik and Chervonenkis 2015). Next, in each epoch, the algorithm only takes action b for those users with contextual vectors that are *close* to the current classification hyperplane (i.e., those users with small "margin" $|\langle x, \widehat{w}_{k-1} \rangle - \widehat{\beta}_{k-1}|$). This concentrates our labeled/queried samples to the region that are close to the classification hyperplane, which helps reduce the number of queried samples as the queried samples are collected on regions that are the most uncertain from a binary classification perspective.
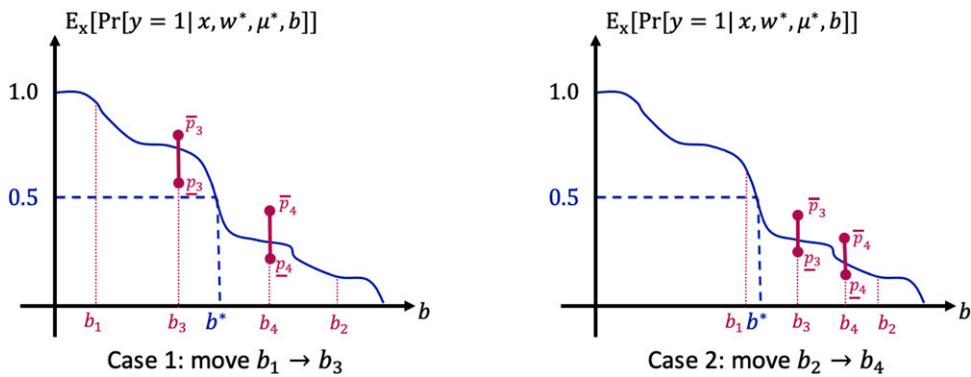
The following lemma is the main result of this section, which is proved in Section 4.2.

Lemma 3. *Let* $(\widehat{w}, \widehat{\beta})$ *be returned by Algorithm 3 with parameters satisfying* $|\mu^* - b| / \|w^*\|_2 \le \beta_0 = O(1/\sqrt{d})$, $\kappa_m = \Omega(1)$, $\kappa_n = \Omega(d + \log\log(1/\epsilon_a) + \log(1/\delta_a))$ *and* $\epsilon_0 = \sqrt{\epsilon_a}$. *Let* $\widehat{v}(\cdot) = \langle \cdot, \widehat{w} \rangle - \widehat{\beta}$. *Then for sufficiently large d and sufficiently small ϵ_a, with probability $1 - \delta_a$ the following hold:*
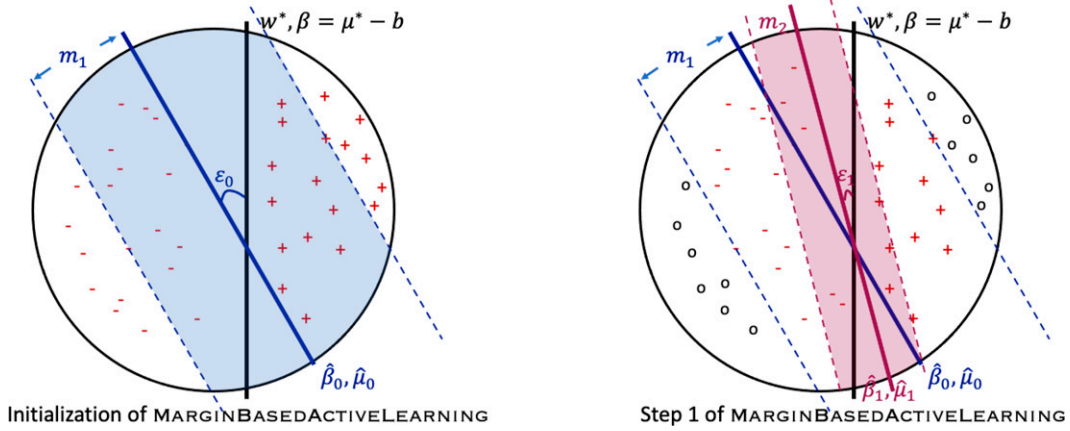
1. *The excess error satisfies* $\Delta\text{err}(\widehat{v}) \le \epsilon_a$;

2. *Algorithm 3 consumes* $O(\kappa_n d / \epsilon_a)$ *queried samples and* $\widetilde{O}(\kappa_n \sqrt{d} e^{d\epsilon_a} / \epsilon_a^{3/2})$ *total samples.*

Essentially, Lemma 3 shows that the estimated linear model $\widehat{v}(\cdot)$ produced by Algorithm 3 has the target excess risk ϵ_a with high probability. The lemma also upper bounds the number of queried and total samples

Figure 2. (Color online) Graphical Illustration of the Main Idea Behind Algorithm 2

Notes. (Left) First case of the trisection search, in which $\wp(b_3) = \mathbb{E}_{x \sim P_X}[y \mid x, w^*, \mu^*, b_3] > 1/2$. Once $\underline{p}_3$ exceeds $1/2$, the algorithm will move b_1 to b_3. (Right) Second case of the trisection search, in which $\wp(b_4) = \mathbb{E}_{x \sim P_X}[y \mid x, w^*, \mu^*, b_4] < 1/2$. As both $\overline{p}_3, \overline{p}_4$ are below $1/2$, the algorithm will move b_2 to b_4. The strict monotonicity of $\wp(\cdot)$ as a function of b ensures that the trisection search will never exclude b^* from $[b_1, b_4]$, and that the search will terminate in $O(d \log(d/\delta))$ iterations (see Lemma 2).

Figure 3. (Color online) Graphical Illustration of the Main Idea of Algorithm 3

Notes. (Left) Initialization step (Lines 2 and 4) of Algorithm 3. In the initialization step, sample selection is not carried out and therefore the obtained model estimates $\widehat{w}_0, \widehat{\beta}_0$ have error upper bounded by ε_0. (Right) First iteration of Algorithm 3. As shown in the figure, only those samples that are within an m_1 margin around $\widehat{w}_0, \widehat{\beta}_0$ (those within the blue dashed lines) are labeled/queried. After the first iteration, a more refined estimate $\widehat{w}_1, \widehat{\beta}_1$ is obtained and a shrunk margin m_2 is imposed (the shadowed region) for the next iteration.

consumed in the estimation procedure. As we can see, the number of labeled samples required is on the order of $O(1/\varepsilon_a)$, which is an order of magnitude fewer than the total number of samples consumed (on the order of $\widetilde{O}(1/\varepsilon^{3/2})$). This shows that the active learning procedure is capable of drastically reducing the number of queried samples required to attain an accurate model estimate $\widehat{v}$, by being selective in the user context vectors.

3.3. Sample Complexity Analysis of Algorithm 1

In this section, we establish the following theorem, which analyzes the sample complexity (both samples that are queried on and samples that are passed) of Algorithm 1 and provides guidance on the selection of the algorithm input parameters.

Theorem 1. *Suppose Algorithm 1 is executed with $\kappa_m \asymp 1$, $\kappa_n \asymp d + \log\log(1/\varepsilon) + \log(1/\delta)$, $\kappa_\varepsilon \asymp 1/d$, and $\beta_0 \asymp 1/\sqrt{d}$. Then for sufficiently small $\varepsilon > 0$ and sufficiently large d, with probability $1 - \delta$ it holds that $|\widehat{v}(x) - v(x)| \le \varepsilon$ for all $x \in \mathbb{B}_2(d)$. Furthermore, the algorithm makes $n(\delta, \varepsilon)$ queries among $m(\delta, \varepsilon)$ total samples/contexts, with*

$$n(\delta, \varepsilon) = O\left(\frac{d^3 \log(d \log(d\varepsilon^{-1})/\delta) \log^2(1/\varepsilon)}{\varepsilon^2}\right),$$

$$m(\delta, \varepsilon) = O\left(\frac{d^3 \log(d \log(d\varepsilon^{-1})/\delta) \log^3(1/\varepsilon)}{\varepsilon^3}\right).$$

Theorem 1 shows that, by using more unlabeled/unqueried samples than those that are labeled (more specifically, $1/\varepsilon^3$ total samples and $1/\varepsilon^2$ labeled ones), the utility function estimate $\widehat{v}(\cdot)$ produced by our active learning algorithm is within ε estimation error with high probability. In Section 5 of numerical studies, we will see that the availability of unlabeled samples

will greatly improve the estimation accuracy of an active learning algorithm compared with a passive learning baseline which cannot skip or select samples to query.

Remark 1. If the decision maker needs to make queries to all incoming contexts/samples (i.e., skipping uninformative samples is not allowed), then at least $\Omega(d/\varepsilon^3)$ samples are required. To see this, the standard classification theory establishes that $\Omega(d/\varepsilon^{3/2})$ samples are needed to obtain a linear classifier $\widehat{w}$ such that $\Pr[\text{sgn}(\widehat{w}^\top x) \ne \text{sgn}((w^*)^\top x)] \le \varepsilon$ (Mammen and Tsybakov 1999; Ben-David and Urner 2014, table 1) probabilistic labels with Bayes classifier in H and the Tsybakov noise conditions (TNC) parameter $\alpha = 1/2$. Conversely, it can be shown via an integration argument as follows. Let $\angle(\widehat{w}, w^*)$ denote the angle between $\widehat{w}$ and w^*. If both $\widehat{w}, w^*$ are normalized (i.e., $\|\widehat{w}\|_2 = \|w^*\|_2 = 1$) and $\|\widehat{w} - w^*\|_2 \approx \angle(\widehat{w}, w^*) \approx \varepsilon$ then $\Pr_x[\text{sgn}(\widehat{w}^\top x) \ne \text{sgn}((w^*)^\top x)] \approx \varepsilon^2$. This shows that to achieve $|\widehat{v}(\cdot) - v(\cdot)| \le \varepsilon$, we must have $\Pr_x[\text{sgn}(\widehat{w}^\top x) \ne \text{sgn}((w^*)^\top x)] \lesssim \varepsilon^2$, indicating a sample complexity lower bound of $\Omega(d/(\varepsilon^{3/2})^2) = \Omega(d/\varepsilon^3)$.

Remark 2. When each skipped sample has a cost of $\rho \in [0, 1)$ compared with a labeled sample, the combined sample complexity of our proposed active learning algorithm is on the order of $\widetilde{O}(\varepsilon^{-2} + \rho\varepsilon^{-3})$, omitting polynomial dependency on other problem parameters. Conversely, an algorithm incapable of skipping samples requires $\Omega(\varepsilon^{-3})$ samples as indicated in the previous remark, significantly higher than $\widetilde{O}(\varepsilon^{-2} + \rho\varepsilon^{-3})$ especially when ρ is small (indicating that skipped samples are much less costly compared with labeled samples).

Remark 3. In this remark, we discuss an "intermediate setting" in which the algorithm can select action levels

but not skip samples. This intermediate setting is stronger than the passive learning setting but weaker than the active learning setting. We remark that the intermediate setting is likely to have similar sample complexity compared with passive learning.

Consider the uniform distribution on the unit ℓ_2 ball in $\mathbb{R}^d$ and let $w \in \mathbb{R}^d$ be an arbitrary (unknown) classifier. It is easy to observe that, up to polynomial constants in d, for every small $\epsilon > 0$ the probability of $x \in \{x : \theta(x, w) \leq \epsilon\}$ is $O(\epsilon)$, where $\theta(\cdot, \cdot)$ denotes the angle between two vectors in $\mathbb{R}^d$. This means that, without the ability to skip samples, for a batch of n samples only $O(n\epsilon)$ of them are sufficiently close to the decision boundary w to offer a good amount of information. On the other hand, active learning allows the algorithm to only collect labels/responses on the $O(n\epsilon)$ samples that are sufficiently close to the boundary, thus leading to more efficient usage of information from labeled samples. Although the intermediate setting can still adaptively change the action levels (corresponding to changing the intercept in a nonhomogeneous linear classification model), such ability is unlikely to achieve the "sample concentration" effect because only changing one parameter in a multivariate linear model cannot bring a uniformly sampled data point arbitrarily close to the (unknown) decision boundary.

4. Technical Proofs

In this section, we state the proofs of the main results in this paper. There are also some technical lemmas that are either easy to prove or cited/rephrased from existing works, which will be presented in the appendix. For simplicity, let P_U be the uniform distribution on $\mathbb{B}_2(d) = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$ for all proofs in this section.

4.1. Proof of Results in Section 3.1

4.1.1. Proof of Lemma 1. First, $\langle w^*, x \rangle - \mu^* \geq b^*$ is equivalent to $\langle w^*, x \rangle - \beta^* \geq 0$, with $\beta^* = \mu^* + b^*$. Also, we may assume $\|w^*\|_2 = 1$ because $|\beta^*|/\|w^*\|_2$ is invariant to $\|w^*\|_2$. In this proof, we shall use the lower and upper bounds of f_x by connecting it with the uniform distribution on $\mathbb{B}_2(d)$, P_U. Because P_U is isotropic, we may assume without loss of generality that $w^* = (1, 0, \ldots, 0)$ and $\beta^* \geq 0$. We will also abbreviate $\eta = \eta_{b^*}$ and $\Delta = \Delta_{b^*}$ because all margins in this proof are with respect to b^*. Then for all $x \in \mathbb{B}_2(d)$ with $x_1 \geq \beta^*$, $\eta(x) \geq 1/2$, and furthermore,

$$\eta(x) - \frac{1}{2} = \phi(x_1 - \beta^*) = \int_0^{x_1 - \beta^*} \phi'(u)\mathrm{d}u \leq C_\xi(x_1 - \beta^*).$$

Subsequently, by Assumption 2 and Lemma A.1, it holds that

$$\int_{x_1 \geq \beta^*} \left(\eta(x) - \frac{1}{2}\right)\mathrm{d}P_x(x)$$

$$\leq C_x \int_{x_1 \geq \beta^*} \left(\eta(x) - \frac{1}{2}\right)\mathrm{d}P_U(x)$$

$$\leq C_x C_\xi \int_{x_1 \geq \beta^*} (x_1 - \beta^*)\mathrm{d}P_U(x)$$

$$\leq C_x C_\xi \int_0^1 \sqrt{\frac{d+1}{2\pi}} e^{-(d-1)(\beta^* + \gamma)^2/2} \gamma \mathrm{d}\gamma$$

$$\leq C_x C_\xi \sqrt{d} e^{-(d-1)(\beta^*)^2/2} \int_0^1 \gamma e^{-(d-1)\gamma^2/2}\mathrm{d}\gamma. \tag{5}$$

With $\gamma \mapsto \gamma/\sqrt{d-1}$, we have $\int_0^1 \gamma e^{-(d-1)\gamma^2/2}\mathrm{d}\gamma \leq \sqrt{\frac{2\pi}{d-1}}\mathbb{E}_{z\sim\mathcal{N}(0, 1/(d-1))}[|z|]/2 \leq 1/(d-1)$. Noting that $\sqrt{d} \leq \sqrt{2(d-1)}$ for $d \geq 2$, Equation (5) can then be simplified to

$$\int_{x_1 \geq \beta^*} \left(\eta(x) - \frac{1}{2}\right)\mathrm{d}P_x(x) \leq \frac{\sqrt{2}C_x C_\xi}{\sqrt{d-1}} e^{-(d-1)(\beta^*)^2/2}. \tag{6}$$

Conversely, for all $x \in \mathbb{B}_2(d)$ with $x_1 \leq \beta^*$, $\eta(x) \leq 1/2$, and furthermore,

$$\frac{1}{2} - \eta(x) = -\phi(x_1 - \beta^*) = \int_0^{\beta^* - x_1} \phi'(u)\mathrm{d}u \geq c_\xi(\beta^* - x_1).$$

Subsequently, by Assumption 2 and Lemma A.1, it holds that

$$\int_{x_1 \leq \beta^*} \left(\frac{1}{2} - \eta(x)\right)\mathrm{d}P_x(x) \geq c_x \int_{x_1 \leq 0} \left(\frac{1}{2} - \eta(x)\right)\mathrm{d}P_U(x)$$

$$\geq c_x c_\xi \int_{x_1 \leq 0} (\beta^* - x_1)\mathrm{d}P_U(x)$$

$$\geq c_x c_\xi \int_{x_1 \leq 0} -x_1 \mathrm{d}P_U(x) \geq c_x c_\xi \int_0^1 \sqrt{\frac{d+1}{16\pi}} e^{-(d-1)\gamma^2/2} \gamma \mathrm{d}\gamma$$

$$\geq \frac{c_x c_\xi \sqrt{d}}{4\sqrt{\pi}} \int_{1/\sqrt{d-1}}^{\sqrt{2}/\sqrt{d-1}} e^{-(d-1)\gamma^2/2} \gamma \mathrm{d}\gamma$$

$$\geq \frac{c_x c_\xi \sqrt{d}}{4\sqrt{\pi}} \times \frac{1}{e\sqrt{d-1}} \times \frac{\sqrt{2}-1}{\sqrt{d-1}} \geq \frac{(\sqrt{2}-1)c_x c_\xi}{4e\sqrt{\pi(d-1)}}. \tag{7}$$

Combining Equations (5) and (7), we obtain

$$\frac{1}{2} = \Pr_{x \sim P_x}[y = 1 \mid b^*]$$

$$\leq \frac{1}{2} + \frac{\sqrt{2}C_x C_\xi}{\sqrt{d-1}} e^{-(d-1)(\beta^*)^2/2} - \frac{(\sqrt{2}-1)c_x c_\xi}{4e\sqrt{\pi(d-1)}}.$$

To satisfy the previous inequality, $\beta^* \geq 0$ must satisfy

$$\beta^* \leq \sqrt{\frac{2\ln(100 C_x C_\xi / c_x c_\xi)}{d-1}} = O(1/\sqrt{d}),$$

which proves Lemma 1.

4.1.2. Proof of Lemma 2. For notational simplicity define $\wp(\widehat{b}) := \Pr_{x \sim P_x}[v(x) \geq \widehat{b}]$. Clearly, $\wp(b^*) = 1/2$ and $\wp(\cdot)$ is a monotonically decreasing function. By Hoeffding's inequality, at sample n, we have $\Pr[\wp(\widehat{b}_3) \in [\underline{p}_3, \bar{p}_3]] \geq 1 - 2e^{-2\hat{n} \times \ln(8n^2/\delta_s)/(2\hat{n})} \geq 1 - \frac{\delta_s}{4n^2}$. The same inequality holds for $\Pr[\wp(\widehat{b}_4) \in [\underline{p}_4, \bar{p}_4]]$ as well. By the union bound, the probability that $\wp(\widehat{b}_3) \in [\underline{p}_3, \bar{p}_3]$ and $\wp(\widehat{b}_4) \in [\underline{p}_4, \bar{p}_4]$ throughout the entire Algorithm 2 is lower bounded by

$$1 - \sum_{n \geq 1} 2 \times \frac{\delta_s}{4n^2} = 1 - \frac{\delta_s}{2} \sum_{n \geq 1} \frac{1}{n^2} \geq 1 - \frac{\delta_s}{2} \frac{\pi^2}{6} \geq 1 - \delta_s.$$

This shows that with probability $1 - \delta_s$ the $(\widehat{b}_1, \widehat{b}_2)$ pair returned by Algorithm 2 satisfies $\widehat{b}_1 \leq b^* \leq \widehat{b}_2$ because of the monotonicity of the $\wp(\widehat{b})$ function.

To analyze the number of queried samples/objects by Algorithm 2, we require some additional technical results. The following lemma connects the deviation $|\wp(\widehat{b}) - 1/2|$ with $|\widehat{b} - b^*|$.

Lemma 4. *Recall the definition that* $\wp(\widehat{b}) = \Pr_{x \sim P_x}[w(x) \geq \widehat{b}]$ *and* b^* *such that* $\wp(b^*) = 1/2$. *Then* $0.07 c_x c_\xi |\widehat{b} - b^*| \leq |\wp(\widehat{b}) - 1/2| \leq C_\xi |\widehat{b} - b^*|$.

Proof of Lemma 4. Define $\beta^* = b^* + \mu^*$ and $\widehat{\beta} = \widehat{b} + \mu^*$. Define also $s := \widehat{\beta} - \beta^*$, so that $\widehat{\beta} = \beta^* + s$. Recall the definition of margin that $\Delta_{b^*}(x) = v(x) - \beta^*$, and $\Pr[y = 1 | x, b^*] = \phi(\Delta_{b^*}(x))$. Under $\widehat{b}$, we have $\Delta_{\hat{b}}(x) = v(x) - \widehat{\beta} = \Delta_{b^*}(x) - s$ and $\Pr[y = 1 | x, \widehat{b}] = \phi(\Delta_{b^*}(x) - s)$. Subsequently,

$$|\wp(\widehat{b}) - \wp(b^*)| \leq \mathbb{E}_{x \sim P_x}\left[|\phi(\Delta_{b^*}(x) - s) - \phi(\Delta_{b^*}(x))|\right]$$

$$\leq \sup_{|\gamma| \leq 1} |\phi(\gamma - s) - \phi(\gamma)| \leq C_\xi |s|.$$

This proves the upper bound on $|\wp(\widehat{b}) - 1/2|$.

We next consider the lower bound of $|\wp(\widehat{b}) - 1/2|$. Without loss of generality, assume $\beta^* \geq 0$, $w^* = (1, 0, \ldots, 0)$ and $s \geq 0$. We will lower bound $|\wp(\widehat{b}) - 1/2|$ by studying the decrease of $\Pr[y = 1 | x]$ on the ball segment $\mathbb{B}_2(d) \cap \{x \in \mathbb{R}^d : -r \leq x_1 \leq 0\}$ with $r = 1/\sqrt{2(d-1)} \leq 1/\sqrt{2}$ for $d \geq 2$. More specifically,

$$|\wp(\widehat{b}) - 1/2| \geq \int_{r \leq x_1 \leq 0} [\phi(\beta^* - x_1) - \phi(\beta^* - x_1 - s)] dP_X(x)$$

$$\geq c_x \int_{r \leq x_1 \leq 0} [\phi(\beta^* - x_1) - \phi(\beta^* - x_1 - s)] dP_U(x) \quad (8)$$

$$\geq c_x \int_0^r \sqrt{\frac{d+1}{16\pi}} e^{-(d-1)\gamma^2/2} [\phi(\beta^* + \gamma) - \phi(\beta^* + \gamma - s)] d\gamma \quad (9)$$

$$\geq c_x \int_0^r \sqrt{\frac{d+1}{16\pi}} e^{-(d-1)\gamma^2/2} c_\xi s \, d\gamma \quad (10)$$

$$\geq c_x r \times \sqrt{\frac{d+1}{16\pi}} \times e^{-(d-1)r^2/2} \times c_\xi s = \frac{c_x c_\xi s}{4\sqrt{2\pi}\sqrt{e}} \geq 0.07 c_x c_\xi s. \quad (11)$$

Here Equation (8) is because of Assumption 2, Equation (9) is because of Lemma A.2, and Equation (10) is because of Assumption 3. □

We are now ready to analyze the number of queried samples in Algorithm 2. Fix an arbitrary pair of $(\widehat{b}_1, \widehat{b}_2)$ at outer iteration τ such that $\widehat{b}_2 - \widehat{b}_1 = \varepsilon_\tau = 2(2/3)^\tau B \geq \varepsilon_a$. Then either $|\widehat{b}_3 - b^*| \geq \varepsilon_\tau/6$ or $|\widehat{b}_4 - b^*| \geq \varepsilon_\tau/6$. Let $\widehat{n}_\tau$ be the final count when outer iteration τ ends. The condition $0.5 \in [\underline{p}_3, \bar{p}_3] \wedge 0.5 \in [\underline{p}_4, \bar{p}_4]$ in the inside while loop will be violated if $\sqrt{\ln(8n_\tau^2/\delta_s)/2\widehat{n}_\tau} < 0.07 c_x c_\xi \varepsilon_\tau/6$, which translates to

$$\widehat{n}_\tau \leq 1 + \frac{7,500 \ln(8n_\tau/\delta_s)}{c_x^2 c_\xi^2 \varepsilon_\tau^2},$$

where $n_\tau = \sum_{\tau' \leq \tau} \widehat{n}_{\tau'}$. Let τ_0 be the largest integer such that $\varepsilon_{\tau_0} \geq \varepsilon_a$. Then the total number of queried samples is upper bounded by

$$2 \sum_{\tau \leq \tau_0} \widehat{n}_\tau = O\left(\sum_{\tau \leq \tau_0} \frac{1}{\varepsilon_\tau^2} \log\left(\frac{1}{\delta_s \varepsilon_\tau}\right)\right) = O(\varepsilon_{\tau_0}^{-2} \log(1/(\delta_s \varepsilon_{\tau_0})))$$

$$= O(\varepsilon_a^{-2} \log(1/(\delta_s \varepsilon_a))).$$
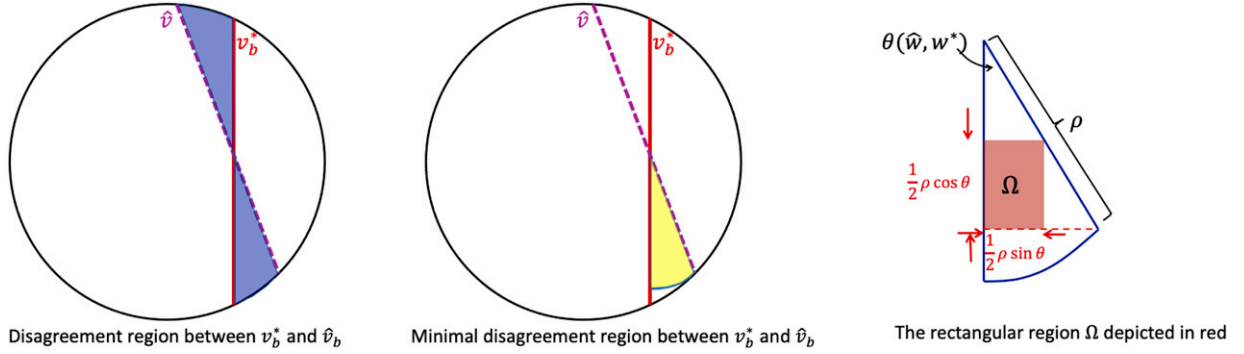
4.2. Proof of Results in Section 3.2

The objective of this section is to prove the key Lemma 3. Throughout this proof, we assume that d is sufficiently large, and $\varepsilon_a > 0$ is sufficiently small. We also define $\theta(w, w')$ as the smallest angle between $w, w' \in \mathbb{R}^d$.

Recall the definition that $v_b^*(\cdot) = \langle \cdot, w^* \rangle - \beta$ with $\beta = b + \mu^*$ is the nonhomogeneous linear classifier with the smallest classification error. For presentation simplicity, we shall normalize $v_b^*(\cdot)$ (because only the *signs* of $v_b^*(\cdot)$ matter in a binary classification problem) as $v_b^*(\cdot) = \langle \cdot, \widetilde{w}^* \rangle - \widetilde{\beta}$, where $\widetilde{w}^* = w^*/\|w^*\|_2$ and $\widetilde{\beta} = \beta/\|w^*\|_2$. Our first technical lemma shows that if another classifier $\widehat{v}(\cdot) = \langle \cdot, \widehat{w} \rangle - \widehat{\beta}$ has small excess error, then the angle between $\widehat{w}$ and $\widetilde{w}^*$ must be small.

Lemma 5. *Let* $\widehat{v}(\cdot) = \langle \cdot, \widehat{w} \rangle - \widehat{\beta}$, $\|\widehat{w}\|_2 = 1$ *be a learnt classifier such that* $\Delta \text{err}(\widehat{v}) = \text{err}(\widehat{v}) - \text{err}(v_b^*) \leq \epsilon$. *Then for sufficiently small* ϵ, *it holds that* $\tan \theta(\widehat{w}, \widetilde{w}^*) \leq 23 e^{(d-2)\beta_0^2/2} \sqrt{\epsilon} = O(\sqrt{\epsilon})$.

Proof of Lemma 5. Abbreviate $\theta = \theta(\widehat{w}, \widetilde{w}^*)$. Without loss of generality, assume $w^* = (1, 0, \ldots, 0)$, $\widehat{w} = (1 - \cos \theta, \sin \theta, 0, \ldots, 0)$ and $\widetilde{\beta} \geq 0$. For sufficiently small ϵ, we have $\tan \theta \leq 1$, and the disagreement region between $\widehat{v}$ and v_b^* is shadowed in solid color in the left panel of Figure 4. Also, as one adjusts the intercept $\widehat{\beta}$ in $\widehat{v}$, one disagreement region will enlarge and the other one will shrink. As a result, the minimal

Figure 4. (Color online) Graphical Illustration of the Proof of Lemma 5

Disagreement region between v_b^* and $\widehat{v}_b$ Minimal disagreement region between v_b^* and $\widehat{v}_b$ The rectangular region Ω depicted in red

disagreement region is shadowed in solid color in the middle panel of Figure 4, with the radius ρ to be at least $1/2$ for sufficiently large d because $|\beta| \le \beta_0 = O(1/\sqrt{d})$. To further simplify, we take only the upper triangle of the disagreement region with $r = 1/2\sqrt{d} \le \rho$ and study the rectangular region designated as Ω in the right panel of Figure 4, whose size is $\frac{r}{2} \times h$ where $h = \frac{r}{2}\tan\theta$.

The excess error of $\widehat{v}$ can be lower bounded by the deviation of $\eta(x) = \Pr[y = 1\,|\,x, b]$ from $1/2$ on Ω. More specifically,

$$\Delta\mathrm{err}(\widehat{v}) \ge \int_\Omega \left(\eta(x) - \frac{1}{2}\right)\mathrm{d}P_X(x) \ge c_x \int_\Omega \left(\eta(x) - \frac{1}{2}\right)\mathrm{d}P_U(x)$$

$$= c_x \int_\Omega \phi(x_1 - \beta)\mathrm{d}P_U(x) \tag{12}$$

$$\ge c_x c_\xi \int_\Omega (x_1 - \beta)\mathrm{d}P_U(x) \ge c_x c_\xi \int_\Omega \frac{d}{4\pi}e^{-(d-2)(x_1^2 + x_2^2)/2}$$

$$(x_1 - \beta)\mathrm{d}x_1\mathrm{d}x_2 \tag{13}$$

$$\ge \frac{c_x c_\xi d}{4\pi}\int_0^h \gamma e^{-(d-2)((\beta+h)^2 + r^2)/2}\mathrm{d}\gamma$$

$$\ge \frac{c_x c_\xi d}{4\pi}e^{-(d-2)(\beta^2 + r^2)}\frac{1}{2}h^2 \ge \frac{e^{-(d-2)\beta^2}}{128\pi\sqrt[4]{e}}\tan^2\theta. \tag{14}$$

Here, Equation (12) is because of Assumption 2 and the definition of ϕ; Equation (13) is because of

Assumption 3 and Lemma A.2. Taking the square root on both sides of Equation (14) and noting that $\Delta\mathrm{err}(\widehat{v}) = \epsilon$, we complete the proof of Lemma 5. □
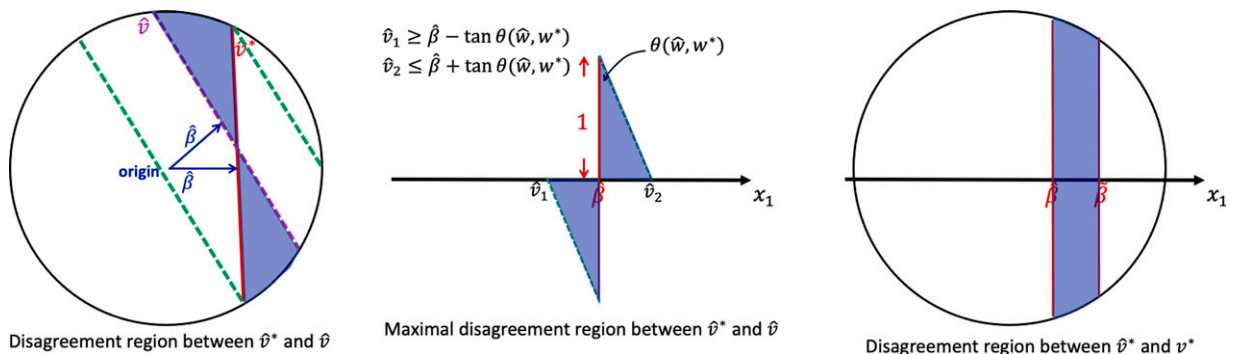
The next lemma shows that if $\Delta\mathrm{err}(\widehat{v})$ is small, then the intercept $\widehat{\beta}$ cannot be too far away from β either.

Lemma 6. *Let* $\widehat{v}(\cdot) = \langle\,\cdot\,, \widehat{w}\rangle - \widehat{\beta}$, $\|\widehat{w}\|_2 = 1$, $|\widehat{\beta}| \le \beta_0$ *be a learnt classifier such that* $\Delta\mathrm{err}(\widehat{v}) = \mathrm{err}(\widehat{v}) - \mathrm{err}(v_b^*) = \epsilon$. *Then for sufficiently small* ϵ, $|\widehat{\beta} - \widetilde{\beta}| \le 3601\sqrt{\pi}C_x C_\xi c_x^{-1} c_\xi^{-1}$ $\max\{e^{(d-1)\beta_0^2}, 1\}\sqrt{\epsilon} = O(\sqrt{\epsilon})$.

Proof of Lemma 6. Let $\theta = \theta(\widehat{w}, \widetilde{w}^*)$, and assume without loss of generality that $\widetilde{w}^* = (1, 0, \dots, 0)$ and $\widetilde{\beta} \ge 0$. Let also $\Delta_\beta = \widehat{\beta} - \widetilde{\beta}$.

First we compare the two models of $\widehat{v}(\cdot) = \langle\,\cdot\,, \widehat{w}\rangle - \widehat{\beta}$ and $\widehat{v}^*(\cdot) = \langle\,\cdot\,, \widetilde{w}^*\rangle - \widehat{\beta}$. When d is sufficiently large and ϵ is sufficiently small, the disagreement region between $\widehat{v}$ and $\widehat{v}^*$ is depicted in the left panel of Figure 5. Because the two line segments intersect when $|\widehat{\beta}| \le \beta_0 \to 0$ as $d \to \infty$ and $\theta(\widehat{w}, \widetilde{w}^*) \to 0$ as $\epsilon \to 0$, the maximal disagreement region between $\widehat{v}$ and $\widehat{v}^*$ is reached by the two green dashed lines in the left panel of Figure 5, an upper bound of which is depicted in the middle panel of Figure 5 by projecting onto the one-dimensional space along the direction of $\widetilde{w}^*$. Subsequently, the disagreement between $\widehat{v}$ and $\widehat{v}^*$ can be upper bounded by

Figure 5. (Color online) Graphical Illustration of Proof of Lemma 6

Disagreement region between $\widehat{v}^*$ and $\widehat{v}$ Maximal disagreement region between $\widehat{v}^*$ and $\widehat{v}$ Disagreement region between $\widehat{v}^*$ and v^*

$$\int_{\text{sgn}(\hat{v}(x))\neq\text{sgn}(\hat{v}^*(x))}\left|\eta(x)-\frac{1}{2}\right|dP_X(x)$$

$$\leq \int_{x_1\in[\hat{\beta}\pm\tan\theta]}\left|\eta(x)-\frac{1}{2}\right|dP_X(x)$$

$$\leq C_x C_\xi \int_{x_1\in[\hat{\beta}\pm\tan\theta]}|x_1-\widetilde{\beta}|dP_U(x)$$

$$\leq C_x C_\xi \int_{-\tan\theta}^{\tan\theta}\sqrt{\frac{d+1}{4\pi}}e^{-(d-1)(\hat{\beta}+\gamma)^2/2}|\gamma-\Delta_\beta|d\gamma \quad (15)$$

$$\leq C_x C_\xi \sqrt{\frac{d+1}{4\pi}}\int_{-\tan\theta}^{\tan\theta}|\gamma-\Delta_\beta|d\gamma$$

$$= C_x C_\xi \sqrt{\frac{d+1}{4\pi}}\left(\tan^2\theta+2\Delta_\beta\tan\theta\right)$$

$$\leq 150 C_x C_\xi \sqrt{d+1}\left(e^{(d-2)\beta_0^2}\epsilon+2e^{(d-2)\beta_0^2/2}\sqrt{\epsilon}\Delta_\beta\right)$$

$$= O(\sqrt{d}\epsilon+\sqrt{d\epsilon}\Delta_\beta). \quad (16)$$

Here, Equation (15) is because of Lemma A.1, and the last inequality of Equation (17) holds by Lemma 5.

Next, consider the disagreement between the two models of $\widehat{v}^*(\cdot)=\langle\cdot,\widetilde{w}^*\rangle-\widehat{\beta}$ and $v^*(\cdot)=\langle\cdot,\widetilde{w}^*\rangle-\widetilde{\beta}$. First consider the case of $\widehat{\beta}\geq\widetilde{\beta}$, and let $\Delta_\beta=\widehat{\beta}-\widetilde{\beta}$. The disagreement region between $\widehat{v}^*$ and v^* in this case is depicted in the right panel of Figure 5. The disagreement between $\widehat{v}^*$ and v^* can then be lower bounded by

$$\int_{\text{sgn}(\hat{v}^*(x))\neq\text{sgn}(v^*(x))}\left(\eta(x)-\frac{1}{2}\right)dP_X(x)$$

$$\geq \int_{x_1\in[\widetilde{\beta},\hat{\beta}]}\left(\eta(x)-\frac{1}{2}\right)dP_X(x)$$

$$\geq c_x c_\xi \int_{x_1\in[\widetilde{\beta},\hat{\beta}]}(x_1-\widetilde{\beta})dP_U(x)$$

$$\geq c_x c_\xi \int_0^{\Delta_\beta}\sqrt{\frac{d+1}{16\pi}}e^{-(d-1)(\widetilde{\beta}+\gamma)^2/2}\gamma d\gamma$$

$$\geq c_x c_\xi \sqrt{\frac{d+1}{16\pi}}e^{-(d-1)\beta_0^2/2}\int_0^{\Delta_\beta}\gamma d\gamma \quad (17)$$

$$\geq \frac{c_x c_\xi\sqrt{d+1}}{8\sqrt{\pi}}e^{-(d-1)\beta_0^2/2}\Delta_\beta^2. \quad (18)$$

Here the second inequality in Equation (17) holds because $\widehat{\beta}=\beta+\Delta_\beta\leq\beta_0$ by optimization constraint. If $\widehat{\beta}<\widetilde{\beta}$, the disagreement region has more density because the region (the $[\hat{\beta},\widetilde{\beta}]$ strip) is closer to the origin than the perimeter of the $\mathbb{B}_2(d)$ ball.

Combining Equations (16) and (18) and noting that v^* is the Bayes classifier (i.e., the classifier that minimizes classification error), we have that

$$\Delta\text{err}(\widehat{v})\geq\frac{c_x c_\xi\sqrt{d+1}}{8\sqrt{\pi}}e^{-(d-1)\beta_0^2/2}\Delta_\beta^2-150 C_x C_\xi\sqrt{d+1}$$

$$\left(e^{(d-2)\beta_0^2}\epsilon+2e^{(d-2)\beta_0^2/2}\sqrt{\epsilon}\Delta_\beta\right). \quad (19)$$

Because $\Delta\text{err}(\widehat{v})=\epsilon$ and $|\Delta_\beta|\leq 2\beta_0=O(1/\sqrt{d})$, for sufficiently large d, the previous inequality solves to

$$|\Delta_\beta|\leq 3601\sqrt{\pi}\frac{C_x C_\xi}{c_x c_\xi}\max\left\{e^{(d-1)\beta_0^2},1\right\}\times\sqrt{\epsilon}=O(\sqrt{\epsilon}),$$

which is to be proved. □

We are now ready to prove the key Lemma 3 in Section 3.2.

Proof of Lemma 3. Recall the definition that $\epsilon_k=2^{-k}\epsilon_0$. We use mathematical induction to prove that, at the end of each outer iteration $k\in\{0,1,2,\ldots,k_0\}$, with probability $1-\delta_a/(k_0+1)$ it holds that $\Delta\text{err}(\widehat{w}_k,\widehat{\beta}_k)\leq\epsilon_k$.

4.2.1. Base of Induction. For $k=0$, invoke Lemma A.3 with $n=n_0$ and $\delta=\delta_a/(k_0+1)$, we have with probability $1-\delta$ that $\Delta\text{err}(\widehat{w}_0,\widehat{\mu}_0)\leq O\left(\sqrt{\frac{d+\ln(k_0/\delta_a)}{n_0}}\right)$. Also, $k_0\asymp\log(1/\varepsilon_a)$. Hence, with $n_0=\Omega(\varepsilon_a^{-1}(d+\log\log(1/\varepsilon_a)+\log(1/\delta_a)))$, we have with probability $1-\delta_a/(k_0+1)$ that $\Delta\text{err}(\widehat{w}_0,\widehat{\mu}_0)\leq\epsilon_0$.

4.2.2. Inductive Steps. We assume the inductive hypothesis is true for $k-1$, that is, $\Delta\text{err}(\widehat{w}_{k-1},\widehat{\mu}_{k-1})\leq\epsilon_{k-1}=2^{-(k-1)}\epsilon_0$. We will prove in this step that $\Delta\text{err}(\widehat{w}_k,\widehat{\mu}_k)\leq\epsilon_k=2^{-k}\epsilon_0$ with probability $1-\delta_a/(k_0+1)$.

Denote $S_1=\{x\in\mathbb{B}_2(d):|\langle x,\widehat{w}_{k-1}\rangle-\widehat{\beta}_{k-1}|\leq m_k\}$ and $S_2=\mathbb{B}_2(d)\backslash S_1$. Because $\Delta\text{err}(\widehat{w}_{k-1},\widehat{\beta}_{k-1})\leq\epsilon_{k-1}$, by Lemmas 5 and 6, we have that $\tan\theta(\widehat{w}_{k-1},w^*)\leq C\sqrt{\epsilon_{k-1}}$ and $|\widehat{\beta}_{k-1}-\beta|\leq C\sqrt{\epsilon_{k-1}}$ for some constant C depending only on C_x,c_x,C_ξ,c_ξ. Hence, with m_k selected as $m_k=\Omega(\sqrt{\epsilon_{k-1}})$, we have that $\text{sgn}(x^\top\widehat{w}_{k-1}-\widehat{\beta}_{k-1})=\text{sgn}(x^\top w^*-\beta)$ for all $x\in S_2$. Subsequently,

$$\Delta\text{err}(\widehat{w}_k,\widehat{\beta}_k)=[\text{err}(\widehat{w}_k,\widehat{\beta}_k\mid S_1)-\text{err}(w^*,\beta\mid S_1)]\Pr[x\in S_1]$$

$$=:\Delta\text{err}(\widehat{w}_k,\widehat{\beta}_k\mid S_1)\Pr[x\in S_1], \quad (20)$$

where $\text{err}(w,\beta\mid S_1)=\Pr_{(x,y)}[y\neq\text{sgn}(w^\top x-\beta)\mid x\in S_1]$.

Invoking Lemma A.3, if $n_k\geq\Omega(C_x^2 m_k^2 d^2\epsilon_k^{-2}\ln(k_0/\delta_a))=\Omega(d^2\epsilon_k^{-1}\ln(k_0/\delta_a))$ then it holds with probability $1-\delta_a/(k_0+1)$ that

$$\Delta\text{err}(\widehat{w}_k,\widehat{\beta}_k\mid S_1)\leq\frac{\epsilon_k}{1.3 C_x m_k\sqrt{d}}. \quad (21)$$

Conversely, we have that

$$\Pr_{x\sim P_X}[x\in S_1]\leq C_x\Pr_{x\sim P_U}[x\in S_1]\leq C_x\Pr_{x\sim P_U}[|x_1|\leq m_k]$$

$$\leq C_x\sqrt{\frac{2(d+1)}{\pi}}\int_0^{m_k}e^{-(d-1)u^2/2}du \quad (22)$$

$$\leq 1.3 C_x m_k\sqrt{d}. \quad (23)$$

Here, the last inequality in Equation (22) holds by invoking Lemma A.1. Plug Equations (21) and (23) into Equation (20). We proved that $\Delta\mathrm{err}(\widehat{w}_k, \widehat{\beta}_k) \leq \epsilon_k$, which completes the induction step.

In the final part of the proof, we upper bound the total number of labeled (queried) and unlabeled samples used in Algorithm 3. The number of labeled samples is simply $n_0 + \sum_{k=1}^{k_0} n_k$. It can be upper bounded by

$$n_0 + \sum_{k=1}^{k_0} n_k \leq O(\kappa_n \epsilon_0^{-2}) + \sum_{k=1}^{k_0} O(\kappa_n d \epsilon_k^{-1}) \leq O(\kappa_n)$$
$$\times \left(\frac{1}{\epsilon_a} + \sum_{k=1}^{k_0} \frac{2^k d}{\sqrt{\epsilon_a}} \right) \leq O\left(\frac{\kappa_n d}{\epsilon_a} \right),$$

where the last inequality holds because $k_0 = \min\{k \in \mathbb{M} : 2^{-k}\epsilon_0 \leq \epsilon_a\}$ and $\epsilon_0 = \sqrt{\epsilon_a}$. This shows that the total number of labeled samples consumed is on the order of $O(\kappa_n d / \epsilon_a)$.

To upper bound the total number of samples (labeled/queried or unlabeled/not queried), note that at epoch k the number of total samples is upper bounded by $\widetilde{O}(n_k / \Pr[x \in S_1(k)])$, where $S_1(k) = \{x \in \mathbb{B}_2(d) : |x^\top \widehat{w}_{k-1} - \widehat{\beta}_{k-1}| \leq m_k\}$. Because $|\widehat{\beta}_{k-1}| \leq \beta_0$, we can lower bound $\Pr[x \in S_1(k)]$ as

$$\Pr_{x \sim P_X} [x \in S_1(k)]$$
$$\geq c_x \Pr_{x \sim P_U} [x \in S_1(k)]$$
$$\geq c_x \Pr_{x \sim P_U} [|x - \widehat{\beta}_{k-1}| \leq m_k]$$
$$\geq c_x \sqrt{\frac{d+1}{16\pi}} \int_0^{m_k} e^{-(d-1)(|\widehat{\beta}_{k-1}|+u)^2/2} \mathrm{d}u$$
$$\geq c_x \sqrt{\frac{d+1}{16\pi}} e^{-(d-1)\beta_0^2} \times m_k e^{-(d-1)m_k^2}$$
$$\geq \Omega(\sqrt{d}) \times m_k e^{-(d-1)m_k^2}.$$

Hence, the total number of samples consumed can be upper bounded by

$$n_0 + \sum_{k=1}^{k_0} \widetilde{O}\left(\frac{n_k}{\Pr[x \in S_1(k)]} \right)$$
$$\leq \widetilde{O}(\kappa_n \epsilon_0^{-2}) + \sum_{k=1}^{k_0} \widetilde{O}\left(\frac{\kappa_n \sqrt{d} e^{(d-1)m_k^2}}{m_k \epsilon_k} \right)$$
$$\leq \widetilde{O}(\kappa_n \epsilon_0^{-2}) + \sum_{k=1}^{k_0} \widetilde{O}\left(\frac{\kappa_n \sqrt{d} e^{(d-1) \times \widetilde{O}(\epsilon_k)}}{\epsilon_k^{3/2}} \right)$$
$$\leq \widetilde{O}\left(\frac{\kappa_n \sqrt{d} e^{d\epsilon_a}}{\epsilon_a^{3/2}} \right).$$

This completes the proof of Lemma 3. □

4.3. Proof of Theorem 1

Recall the definition that $v(\cdot) = \langle \cdot, w^* \rangle - \mu^*$. Define $\alpha := \|w^*\|_2 \leq B$, $\widetilde{w}^* = w^*/\alpha$, and for $j \in \{1, 2\}$ define $\widetilde{\beta}_j = $

$(\mu^* + \widehat{b}_j)/\alpha$. By Lemma 3, we have $\Delta\mathrm{err}(\widehat{w}_j, \widehat{\beta}_j) \leq \epsilon_a$, which by Lemmas 5 and 6 implies $\tan\theta(\widehat{w}_j, \widetilde{w}^*) = O(\sqrt{\epsilon_a})$ and $|\widehat{\beta}_j - \widetilde{\beta}_j| = O(\sqrt{\epsilon_a})$. This implies that $|\alpha\widehat{\beta}_j - \mu^* - \widehat{b}_j| \leq \alpha \times O(\sqrt{\epsilon_a}) = O(B\sqrt{\epsilon_a})$. On the other hand, the stopping condition in Algorithm 3 implies $|\widehat{b}_2 - \widehat{b}_1| = \Omega(\epsilon_s) = \Omega(1/\sqrt{d})$, which yields $|\widehat{\beta}_2 - \widehat{\beta}_1| = \Omega(1/(\alpha\sqrt{d}))$ for sufficiently small ϵ because $\widehat{\beta}_1 \to (\mu^* + \widehat{b}_1)/\alpha$ and $\widehat{\beta}_2 \to (\mu^* + \widehat{b}_2)/\alpha$ as $\epsilon \to 0$. Subsequently,

$$|\widehat{\alpha} - \alpha| = \left| \alpha - \frac{\alpha\widehat{\beta}_2 - \mu^* \pm O(B\sqrt{\epsilon_a}) - \alpha\widehat{\beta}_1 + \mu^* \pm O(B\sqrt{\epsilon_a})}{\widehat{\beta}_2 - \widehat{\beta}_1} \right|$$
$$= \frac{O(B\sqrt{\epsilon_a})}{|\widehat{\beta}_2 - \widehat{\beta}_1|} = O(B^2 \sqrt{d\epsilon_a}). \quad (24)$$

We now upper bound $|\widehat{\mu} - \mu^*|$ and $\|\widehat{w} - w^*\|_2$. By definition, $\widehat{\mu} = \widehat{\alpha}\widehat{\beta}_1 - \widehat{b}_1$ and $\mu^* = \alpha\widetilde{\beta}_1 - \widehat{b}_1$. Subsequently,

$$|\widehat{\mu} - \mu^*| \leq |\widehat{\alpha} - \alpha| \cdot |\widehat{\beta}_1| + \alpha|\widehat{\beta}_1 - \widetilde{\beta}_1|$$
$$\leq O(B^2 \sqrt{d\epsilon_a}\beta_0) + O(B\sqrt{\epsilon_a}) \leq O(B^2\sqrt{d\epsilon_a}). \quad (25)$$

Similarly, $\widehat{w} = \widehat{\alpha}\widehat{w}_1$ and $w^* = \alpha\widetilde{w}^*$. Therefore,

$$\|\widehat{w} - w^*\| \leq |\widehat{\alpha} - \alpha| \cdot \|\widehat{w}\|_2 + \alpha\|\widehat{w}_1 - \widetilde{w}^*\|_2$$
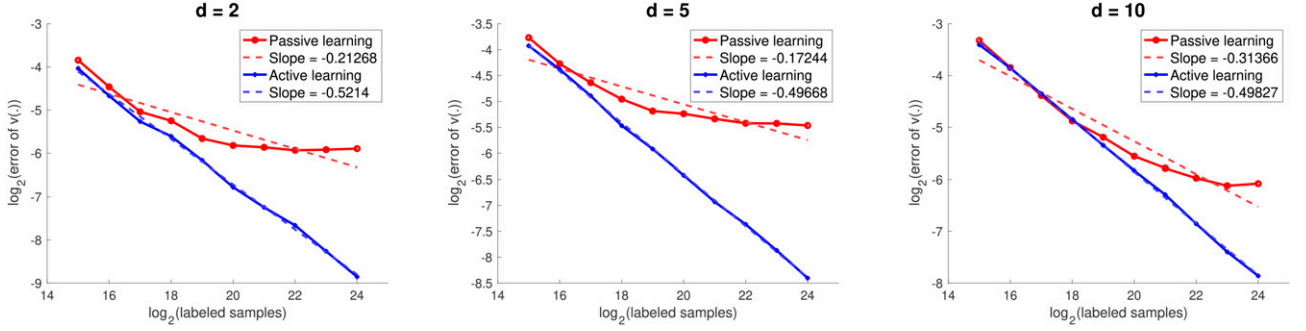$$\leq O(B^2\sqrt{d\epsilon_a}) + O(B\sqrt{\epsilon_a}) = O(B^2\sqrt{d\epsilon_a}). \quad (26)$$

With the choice of $\epsilon_a = \kappa_\epsilon \epsilon^2 / \ln^2(1/\epsilon)$ and $\kappa_\epsilon \asymp 1/d$, and with $\epsilon \to 0$ being sufficiently small, Equations (25) and (26) yield that $\sup_{x \in \mathbb{B}_2(d)} |\widehat{v}(x) - v^*(x)| \leq \epsilon$. Finally, plugging in the expression of $\epsilon_a = \kappa_\epsilon \epsilon^2 / \ln^2(1/\epsilon)$ and invoking Lemma 2 and 3, we obtain the upper bounds on $n(\epsilon, \delta)$ and $m(\epsilon, \delta)$.

5. Numerical Results

We use synthetic data to study the numerical performance of our proposed active learning methods and compare it with baseline methods. The main baseline method we are comparing against is a passive learning method:

• **The baseline method** will first invoke the TRISECTIONSEARCH routine in Algorithm 2 to obtain actions $\widehat{b}_1, \widehat{b}_2$. The method then divides the remaining number of samples into two halves and use Logistic regression to form two model estimates $\widehat{w}_1, \widehat{\beta}_1$ and $\widehat{w}_2, \widehat{\beta}_2$ under actions $\widehat{b}_1$ and $\widehat{b}_2$ respectively, *without* sample selection. The method finally uses Lines 6 and 7 of Algorithm 2 to produce an estimate $\widehat{v}(\cdot)$ of the utility function $v(\cdot)$.

Theoretically, a passive learning baseline algorithm cannot adaptively change actions in queries. However, we observe in our simulations that if the default actions for passive learning are too far away from optimal, very little information is gained and the accuracy of passive learning is very low. Therefore, we use the actions estimated by the TRISECTIONSEARCH routine

Figure 6. (Color online) Log-Log Plot of the Estimation Errors of $v(\cdot)$ as a Function of the Number of Labeled (Queried) Samples n, for $d \in \{2,3,5\}$

Note. The dotted lines are fitted linear regression of the log-log plots.

as the default actions of a passive learning algorithm in our experiments to form a more reasonable comparison.

We also mention details of the implementation of our proposed active learning algorithm. The implementation slightly deviates from the descriptions of the algorithms and the selection of parameter values in the theoretical results, because of computational efficiency issues and other factors we observe could impact the algorithm's numerical performances. In Line 14, the 0/1-error empirical risk minimization step is replaced with logistic regression as the former formulation is computationally expensive. We also remove the $\|w\|_2 = 1, |\beta| \le \beta_0$ constraints in the optimization but normalize the estimator after optimization. The parameters of Algorithm 2 are set as $\varepsilon_s = 0.5$ and $\delta_s = 0.1$. The parameters of Algorithm 3 are set as $\varepsilon_0 = 0.2$, $\kappa_m = 1.0$ and $\kappa_n = d + \ln(n)$. We no longer need the β_0 parameter with the logistic regression formulation.

For the problem settings, we adopt $P_X = P_U$ being the uniform distribution on the d-dimensional ℓ_2 ball $\mathbb{B}_2(d)$. We set the mean utility model $v(\cdot)$ as $v(\cdot) = \langle \cdot, \theta^* \rangle - \mu^*$ with $\theta^* = (2/\sqrt{d}, \ldots, 2/\sqrt{d})$, and $\mu^* = -2.5$. The noise distribution P_ξ is set as the uniform distribution on interval $[-1, 1]$.

5.1. Convergence of Utility Estimates
In the first set of reports, we report how fast the utility estimates $\widehat{v}(\cdot)$ of our proposed algorithm (and the passive learning baseline) converge to the ground truth $v(\cdot)$ as the number of labeled (queried) samples n increases. The estimation errors between $\widehat{v}(\cdot) = \langle \cdot, \widehat{w} \rangle - \widehat{\mu}$ and $v(\cdot) = \langle \cdot, w^* \rangle - \mu^*$ are reported as $\|\widehat{w} - w^*\|_2 + |\widehat{\mu} - \mu^*|$.

Figure 6 reports the estimation errors of the active learning algorithm and the passive learning baseline for dimension settings of $d \in \{2, 5, 10\}$. Each reported error statistic is averaged over 200 independent trials because both the labels and algorithm decisions contain randomness. As we can see, our proposed active learning algorithm (the lower curves with solid circle

marks) outperforms significantly the estimates of the baseline passive learning algorithm (the middle curves with hollow circle marks), demonstrating the sample efficiency of active learning.

We further fit linear regression models on the log-log plots for both algorithms. For the active learning algorithm, the slopes of the fitted linear models are very close to –0.5, suggesting an asymptotic convergence rate of $n(\varepsilon, \delta) \asymp 1/\varepsilon^2$. This matches our theoretical results established in Theorem 1. Conversely, the slopes of fitted models for the passive learning baseline range from –0.17 to –0.31, which are orders of magnitudes slower convergence rates compared with the $1/\varepsilon^2$ rates for active learning methods.

5.2. Sensitivity of Model Dimensions
We use numerical results to evaluate the sensitivity of estimation errors with respect to the dimensions of the underlying linear model d. In Figure 7, we report the estimation errors of the active learning algorithm and the passive learning baseline for dimensions d ranging from 3 to 30.

As we can see in Figure 7, the estimation errors of our proposed active learning approach scale near linearly with the dimension d of the underlying linear model. The active learning algorithm also consistently outperforms the passive learning baseline, especially in large n or d settings.
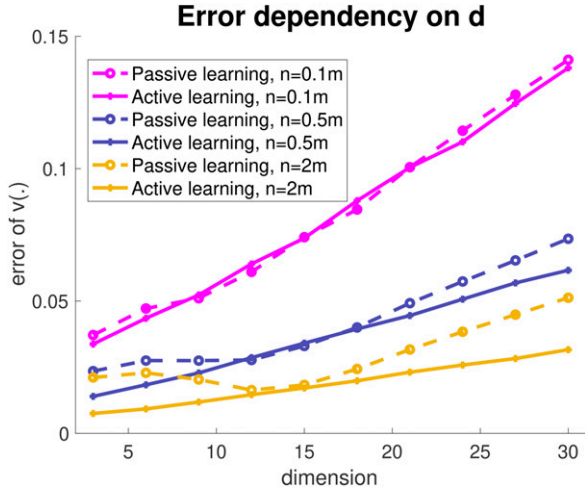
5.3. Sensitivity of Unlabeled Samples
In this section, we report numerical results showing how the estimation errors of the proposed active learning algorithm decrease as the algorithm has access to more unlabeled samples. For this purpose, we define the *ratio* between unlabeled and labeled samples as

$$\rho := \frac{m(\varepsilon, \delta) - n(\varepsilon, \delta)}{n(\varepsilon, \delta)},$$

where $m(\varepsilon, \delta)$ is the total number of samples consumed and $n(\varepsilon, \delta)$ is the number of samples that are

Figure 7. (Color online) Plot of the Estimation Errors of $v(\cdot)$ for Different Labeled (Queried) Samples n and $d \in [3, 30]$ Settings

Note. Here m in the legend stands for million (e.g., $n = 2m$ means the sample size is two million).

labeled/queried. Thus, the numerator $m(\varepsilon, \delta) - n(\varepsilon, \delta)$ is the number of skipped samples. However, our theoretical results in Theorem 1 indicate that ρ has to scale as large as $O(1/\varepsilon)$ for the margin-based active learning algorithm. In practice, however, it is possible to achieve significant estimation accuracy improvements with smaller values of ρ: when the unlabeled sample budget is completely consumed, the margin-based active learning algorithm will revert back to passive learning without any additional sample selection being carried out.

In Figure 8, we plot the estimation errors of the utility $v(\cdot)$ as a function of $\log_2(\rho)$, with larger values of $\log_2(\rho)$ indicating more unlabeled samples involved.

We also report the estimation errors of the passive learning algorithm as a benchmark, which can be regarded as an instance of $\rho = 0$ (i.e., all the samples are labeled). As we can see, the estimation errors of our proposed active learning algorithm decrease rapidly with increasing ρ, and the performance increase is significant when ρ is as small as 0.5 or 1.0. This shows that even with a modest amount of unlabeled samples, the active learning procedure can already significantly increase the accuracy of the estimated utility function $\widehat{v}(\cdot)$.
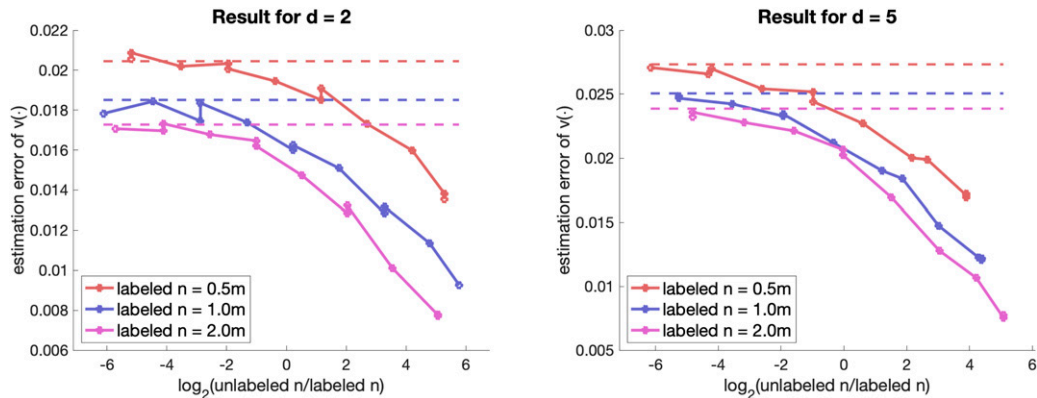
6. Conclusions

In this paper, we study a learning problem in contextual search, where the goal is to use as fewer queries as possible to accurately estimate the mean value function. To this end, we propose a margin-based active learning algorithm with trisection search scheme and establish the corresponding PAC learning sample complexity bound. Our bound shows a significant improvement over the passive setting.

There are several interesting future directions. First, we assume a linear model in this paper. It would be interesting to extend the linear model to more general parametric and nonparametric models. Second, in general, establishing lower bound result in active learning for binary feedback is very challenging. Despite that, it is worth to explore the optimality of our algorithm. Third, we hope the proposed active learning algorithm would inspire more research on adaption of active learning to solve important operations problems.

Figure 8. (Color online) Plots of Estimation Errors of $v(\cdot)$ by the Active Learning Algorithm with Different Budgets of Unlabeled Samples (i.e., Number of Labeled Samples $n(\varepsilon, \delta)$ Varies from 0.5 and 1 to 2 Million)

Notes. The x axis is the log of the ratio between the number of unlabeled samples and the number of labeled samples. The y axis is the estimation error. Dotted lines are errors of the passive learning algorithm.

Appendix A. Some Technical Lemmas

Lemma A.1. *Suppose $x \sim P_U$. Then for any measurable set $A \subseteq [-1,1]$, it holds that $\Pr[x_1 \in A] \leq \sqrt{\frac{d+1}{2\pi}} \int_{u \in A} e^{-(d-1)u^2/2} \mathrm{d}u$. If $A \subseteq [-1/\sqrt{2}, 1/\sqrt{2}]$ then $\Pr[x_1 \in A] \geq \sqrt{\frac{d+1}{16\pi}} \int_{u \in A} e^{-(d-1)u^2/2} \mathrm{d}u$.*

Proof of Lemma A.1. Let $V_d = \pi^{d/2}/\Gamma(1+d/2)$ be the volume of $\mathbb{B}_2(d)$, where $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \mathrm{d}x$ is the Gamma function. Using change-of-variable in multivariate integration, it is easy to verify that $\Pr[x_1 \in A] = \frac{V_{d-1}}{V_d} \int_{u \in A} (1-u^2)^{(d-1)/2} \mathrm{d}u = \frac{1}{\sqrt{\pi}} \frac{\Gamma(d/2+1)}{\Gamma(d/2+1/2)} \int_{u \in A} (1-u^2)^{(d-1)/2} \mathrm{d}u$. By Kershaw's inequality (Kershaw 1983), for any $m > 0$ it holds that $\frac{m}{\sqrt{m+1}} < \frac{\Gamma(m+1/2)}{\Gamma(m)} < \sqrt{m}$. Subsequently,

$$\Pr[x_1 \in A] \leq \frac{\sqrt{d/2 + 1/2}}{\sqrt{\pi}} \int_{u \in A} (1-u^2)^{(d-1)/2} \mathrm{d}u$$
$$\leq \sqrt{\frac{d+1}{2\pi}} \int_{u \in A} e^{-(d-1)u^2/2} \mathrm{d}u,$$

where the last inequality holds because $1 - z \leq e^{-z}$ for all $u \geq 0$. For the other direction, note that $m/\sqrt{m+1} \geq 1/\sqrt{2}$ for $m \geq 1$ and $1 - z \geq 0.5 e^{-z}$ for all $0 \leq z \leq 1/2$. We have

$$\Pr[x_1 \in A] \geq \frac{\sqrt{d/2+1/2}}{\sqrt{2\pi}} \int_{u \in A} (1-u^2)^{(d-1)/2} \mathrm{d}u$$
$$\geq \sqrt{\frac{d+1}{16\pi}} \int_{u \in A} e^{-(d-1)u^2} \mathrm{d}u,$$

where the last inequality holds because $u^2 \leq 1/2$ for all $u \in A$ as assumed. $\square$

Lemma A.2. *Suppose $d \geq 2$ and $x \sim P_U$. Then for any measurable set $A \subseteq [-1,1]^2 \cap \mathbb{B}_2(d)$, it holds that $\Pr[(x_1, x_2) \in A] \leq \frac{d}{2\pi} \int_{(u_1,u_2) \in A} e^{-(d-2)(u_1^2+u_2^2)/2} \mathrm{d}u_1 \mathrm{d}u_2$. If $x_1^2 + x_2^2 \leq 1/2$ for all $(x_1, x_2) \in A$, then $\Pr[(x_1,x_2) \in A] \geq \frac{d}{4\pi} \int_{(u_1,u_2) \in A} e^{-(d-2)(u_1^2+u_2^2)/2} \mathrm{d}u_1 \mathrm{d}u_2$.*

Proof of Lemma A.2. By the change-of-variable formula, $\Pr[(x_1,x_2) \in A] = \frac{V_{d-2}}{V_d} \int_{(u_1,u_2) \in A} (1-u_1^2-u_2^2)^{(d-2)/2} \mathrm{d}u_1 \mathrm{d}u_2 = \frac{d}{2\pi} \int_{(u_1,u_2) \in A} (1-u_1^2-u_2^2)^{(d-2)/2} \mathrm{d}u_1 \mathrm{d}u_2$. The rest of the proof is identical to the proof of Lemma A.1. $\square$

We next define some useful notations that will make our proof similar. For any $x \in \mathbb{B}_2(d)$ and $b \in \mathbb{R}$, define

$$\eta_b(x) := \Pr[y = 1 \mid x, b], \quad \Delta_b(x) := v(x) - b = \langle x, w^* \rangle + \mu^* - b. \tag{A.1}$$

Because $y = 1$ if and only if $v(x) + \xi \geq b$ and $\xi \sim P_\xi$ with $\int_{-\infty}^0 f_\xi(u) \mathrm{d}u = \int_0^\infty f_\xi(u) \mathrm{d}u = 1/2$ (see Assumption 3), we have that $\eta_b(x) - \frac{1}{2} = \int_{-\Delta_b(x)}^0 f_\xi(u) \mathrm{d}u = F_\xi(0) - F_\xi(-\Delta_b(x))$ if $\Delta_b(x) \geq 0$, and $\eta_b(x) - \frac{1}{2} = -\int_0^{-\Delta_b(x)} f_\xi(u) \mathrm{d}u = F_\xi(0) - F_\xi(-\Delta_b(x))$ if $\Delta_b(x) < 0$, where $F_\xi(\cdot)$ is the cumulative distribution function (CDF) of P_ξ. Because $\eta_b(x)$ only depends on $\Delta_b(x)$, we can define

$$\phi(\Delta) := F_\xi(0) - F_\xi(-\Delta). \tag{A.2}$$

It then holds that $\eta_b(x) - \frac{1}{2} = \phi(\Delta_b(x))$. Furthermore, by definition, we have that $\phi(\Delta) \leq 0$ for all $\Delta \leq 0$, $\phi(\Delta) \geq 0$ for all $\Delta \geq 0$, $\phi(0) = 0$ and $\phi'(\Delta) = f_\xi(-\Delta) \in [c_\xi, C_\xi]$ for all $|\Delta| \leq 2$, thanks to Assumption 3.

Now let $\mathcal{V} = \{v(\cdot) : v(\cdot) = \langle \cdot, w \rangle - \beta, w \in \mathbb{R}^d, \beta \in \mathbb{R}\}$ be a hypothesis class of nonhomogeneous d-dimensional linear classifiers. The following lemma is a consequence of the classical VC theory of classification (Balcan et al. 2007, theorem 8).

Lemma A.3. *Fix a distribution P supported on $\mathbb{B}_2(d)$ and a joint distribution Q supported on $\mathbb{B}_2(d) \times \{0,1\}$, such that the marginal of Q on $\mathbb{B}_2(d)$ is P. Let $v^* = \arg\min_{v \in \mathcal{V}} \mathrm{err}(v \mid Q)$, where $\mathrm{err}(v \mid Q) = \Pr_{(x,y) \sim Q}[y \neq \mathrm{sgn}(v(x))]$. Let $\{(x_i, y_i)\}_{i=1}^n \overset{i.i.d.}{\sim} Q$ be n i.i.d. samples, and $\hat{v} = \arg\min_{v \in \mathcal{V}} \sum_{i=1}^n \mathbf{1}\{y_i \neq v(x_i)\}$ be the empirical risk minimizer. Then there exists a universal constant $C > 0$ such that for any $\epsilon, \delta \in (0,1)$, if $n \geq C\epsilon^{-2}(d + 1 + \ln(1/\delta))$ then it holds with probability $1 - \delta$ that $\mathrm{err}(\hat{v} \mid Q) - \mathrm{err}(v^* \mid Q) \leq 2\epsilon$.*

Proof of Lemma A.3. The VC dimension of $\mathcal{V}$ is $d + 1$. By Balcan et al. (2007, theorem 8), it holds with probability $1 - \delta$ that

$$\Pr\left[\forall v \in \mathcal{V}, \left|\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{y_i \neq \mathrm{sgn}(v(x_i))\} - \mathrm{err}(v \mid Q)\right| \leq \epsilon\right] \geq 1 - \delta.$$

The lemma is proved by using triangle inequality. $\square$

References

Albert AE (1961) The sequential design of experiments for infinitely many states of nature. *Ann. Math. Statist.* 32(3):774–799.

Araman VF, Caldentey R (2021) Diffusion approximations for a class of sequential experimentation problems. *Management Sci.*, ePub ahead of print December 28, https://doi.org/10.1287/mnsc.2021.4195.

Awasthi P, Balcan MF, Long PM (2017) The power of localization for efficiently learning linear separators with noise. *J. ACM* 63(6):1–27.

Balcan M-F, Long P (2013) Active and passive learning of linear separators under log-concave distributions. Shalev-Shwartz S, Steinwart I, eds. *Proc. 26th Annual Conf. on Learn. Theory*, vol. 30, 288–316.

Balcan M-F, Zhang H (2017) Sample and computationally efficient learning algorithms under s-concave distributions. *Proc. 31st Inter. Conf. Neural Inform. Processing Systems* (Curran Associates Inc., New York), 4799–4808.

Balcan M-F, Beygelzimer A, Langford J (2009) Agnostic active learning. *J. Comput. System Sci.* 75(1):78–89.

Balcan M-F, Broder A, Zhang T (2007) Margin based active learning. Bshouty NH, Gentile C, eds. *Learning Theory,* (Springer Berlin Heidelberg), 35–50.

Bastani H, Bayati M (2020) Online decision making with high-dimensional covariates. *Oper. Res.* 68(1):276–294.

Bastani H, Bayati M, Khosravi K (2021) Mostly exploration-free algorithms for contextual bandits. *Management Sci.* 67(3):1329–1349.

Ben-David S, Urner R (2014) The sample complexity of agnostic learning under deterministic labels. Balcan MF, Feldman V, Szepesvári C, eds. *Proc. 27th Annual Conf. on Learn. Theory*, vol. 35, 527–542.

Chen X, Chen Y, Li X (2022) Asymptotically optimal sequential design for rank aggregation. *Math. Oper. Res.* Forthcoming.

Chernoff H (1959) Sequential design of experiments. *Ann. Math. Statist.* 30(3):755–770.

Cohen MC, Lobel I, Leme RP (2020) Feature-based dynamic pricing. *Management Sci.* 66(11):4921–4943.

Cohn DA (1996) Neural network exploration using optimal experiment design. *Neural Networks* 9(6):1071–1083.

Cohn D, Atlas L, Ladner R (1994) Improving generalization with active learning. *Machine Learn.* 15(2):201–221.

Dasgupta S (2005a) Analysis of a greedy active learning strategy. *Adv. Neural Inform. Processing Systems 18 (NIPS 2018).*

Dasgupta S (2005b) Coarse sample complexity bounds for active learning. *Adv. Neural Inform. Processing Systems 18 (NIPS 2018).*

Elfving G (1952) Optimum allocation in linear regression theory. *Ann. Math. Statist.* 23(2):255–262.

Feng Y, Caldentey R, Ryan CT (2022) Robust learning of consumer preferences. *Oper. Res.* 70(2):918–962.

Hanneke S (2007) A bound on the label complexity of agnostic active learning. *Proc. 24th Internat. Conf. on Machine Learn.,* 353–360.

Hanneke S (2014) Theory of disagreement-based active learning. *Foundations Trends Machine Learn.* 7(2–3):131–309.

Kershaw D (1983) Some extensions of W. Gautschi's inequalities for the gamma function. *Math. Comput.* 41(164):607–611.

Krishnamurthy A, Lykouris T, Podimata C, Schapire RE (2021) Contextual search in the presence of irrational agents. *Proc. 53rd Annual ACM SIGACT Sympos. on Theory of Comput.* (ACM, New York), 910–918.

Leme RP, Schneider J (2018) Contextual search via intrinsic volumes. *Proc. IEEE 59th Annual Sympos. on Foundations of Comput. Sci.* (IEEE Los Alamitos, CA), 268–282.

Li X, Chen Y, Chen X, Liu J, Ying Z (2021) Optimal stopping and worker selection in crowdsourcing: An adaptive sequential probability ratio test framework. *Statist. Sinica* 31:519–546.

Lobel I, Leme RP, Vladu A (2018) Multidimensional binary search for contextual decision-making. *Oper. Res.* 66(5):1346–1361.

Mammen E, Tsybakov AB (1999) Smooth discrimination analysis. *Ann. Statist.* 27(6):1808–1829.

Naghshvar M, Javidi T (2013) Active sequential hypothesis testing. *Ann. Statist.* 41(6):2703–2738.

Settles B (2012) *Active Learning* (Morgan & Claypool, Williston, VT).

Shiffrin RM, Nosofsky RM (1994) Seven plus or minus two: A commentary on capacity limitations. *Psych. Rev.* 101:357–361.

Stewart N, Brown GDA, Chater N (2005) Absolute identification by relative judgment. *Psych. Rev.* 112(4):881–911.

Vapnik V (2013) *The Nature of Statistical Learning Theory* (Springer Science & Business Media, New York).

Vapnik VN, Chervonenkis AY (2015) On the uniform convergence of relative frequencies of events to their probabilities. *Measures of Complexity* (Springer, Berlin), 11–30.

Wager S, Xu K (2021) Diffusion asymptotics for sequential experiments. Preprint, submitted January, https://arxiv.org/abs/2101.09855.

Wang Y, Singh A (2016) Noise-adaptive margin-based active learning and lower bounds under Tsybakov noise condition. *Proc. 30th AAAI Conf. on Artificial Intelligence* (AAAI, Menlo Park, CA), 2180–2186.

Wang Z, Zenios S (2020) Adaptive design of clinical trials: A sequential learning approach. Preprint, submitted October, https://dx.doi.org/10.2139/ssrn.3713924.

Zhang C, Chaudhuri K (2014) Beyond disagreement-based agnostic active learning. *Adv. Neural Inform. Processing Systems 27 (NIPS 2014).*