Online Learning with Costly Features and Labels

Navid Zolghadr
Department of Computing Science
University of Alberta
zolghadr@ualberta.ca

Gábor Bartók
Department of Computer Science
ETH Zürich
bartok@inf.ethz.ch

Russell Greiner **András György** **Csaba Szepesvári**
Department of Computing Science, University of Alberta
{rgreiner,gyorgy,szepesva}@ualberta.ca

Abstract

This paper introduces the *online probing* problem: In each round, the learner is able to purchase the values of a subset of feature values. After the learner uses this information to come up with a prediction for the given round, he then has the option of paying to see the loss function that he is evaluated against. Either way, the learner pays for both the errors of his predictions and also whatever he chooses to observe, including the cost of observing the loss function for the given round and the cost of the observed features. We consider two variations of this problem, depending on whether the learner can observe the label for free or not. We provide algorithms and upper and lower bounds on the regret for both variants. We show that a positive cost for observing the label significantly increases the regret of the problem.

1 Introduction

In this paper, we study a variant of online learning, called *online probing*, which is motivated by practical problems where there is a cost to observing the features that may help one's predictions. Online probing is a class of online learning problems. Just like in standard online learning problems, the learner's goal is to produce a good predictor. In each time step t, the learner produces his prediction based on the values of some feature $x_t = (x_{t,1}, \dots, x_{t,d})^\top \in \mathcal{X} \subset \mathbb{R}^d$.[1] However, unlike in the standard online learning settings, if the learner wants to use the value of feature i to produce a prediction, he has to purchase the value at some fixed, *a priori* known cost, $c_i \geq 0$. Features whose value is not purchased in a given round remain unobserved by the learner. Once a prediction $\hat{y}_t \in \mathcal{Y}$ is produced, it is evaluated against a loss function $\ell_t : \mathcal{Y} \to \mathbb{R}$. At the end of a round, the learner has the option of purchasing the full loss function, again at a fixed prespecified cost $c_{d+1} \geq 0$ (by default, the loss function is not revealed to the learner). The learner's performance is measured by his regret as he competes against some prespecified set of predictors. Just like the learner, a competing predictor also needs to purchase the feature values needed in the prediction. If $s_t \in \{0,1\}^{d+1}$ is the indicator vector denoting what the learner purchased in round t ($s_{t,i} = 1$ if the learner purchased $x_{t,i}$ for $1 \leq i \leq d$, and purchased the label for $i = d+1$) and $c \in [0,\infty)^{d+1}$ denotes the respective costs, then the regret with respect to a class of prediction functions $\mathcal{F} \subset \{f \mid f : \mathcal{X} \to \mathcal{Y}\}$ is defined by

$$R_T = \sum_{t=1}^{T} \{\ell_t(\hat{y}_t) + \langle\, s_t, c \,\rangle\} \; - \inf_{f \in \mathcal{F}} \left\{ T \langle\, \mathrm{s}(f), c_{1:d} \,\rangle + \sum_{t=1}^{T} \ell_t(f(x_t)) \right\},$$

where $c_{1:d} \in \mathbb{R}^d$ is the vector obtained from c by dropping its last component and for a given function $f : \mathbb{R}^d \to \mathcal{Y}$, $\mathrm{s}(f) \in \{0,1\}^d$ is an indicator vector whose i^{th} component indicates whether f

[1] We use $^\top$ to denote the transpose of vectors. Throughout, all vectors $x \in \mathbb{R}^d$ will denote column vectors.

is sensitive to its i^{th} input (in particular, $\mathrm{s}_i(f) = 0$ by definition when $f(x_1, \ldots, x_i, \ldots, x_d) = f(x_1, \ldots, x'_i, \ldots, x_d)$ holds for all $(x_1, \ldots, x_i, \ldots, x_d), (x_1, \ldots, x'_i, \ldots, x_d) \in \mathcal{X}$; otherwise $\mathrm{s}_i(f) = 1$). Note that when defining the best competitor in hindsight, we did not include the cost of observing the loss function. This is because (i) the reference predictors do not need it; and (ii) if we did include the cost of observing the loss function for the reference predictors, then the loss of each predictor would just be increased by $c_{d+1}T$, and so the regret R_T would just be reduced by $c_{d+1}T$, making it substantially easier for the learner to achieve sublinear regret. Thus, we prefer the current regret definition as it promotes the study of regret when there is a price attached to observing the loss functions.

To motivate our framework, consider the problem of developing a computer-assisted diagnostic tool to determine what treatment to apply to a patient in a subpopulation of patients. When a patient arrives, the computer can order a number of tests that cost money, while other information (e.g., the medical record of the patient) is available for free. Based on the available information, the system chooses a treatment. Following-up the patient may or may not incur additional cost. In this example, there is typically a delay in obtaining the information whether the treatment was effective. However, for simplicity, in this work we have decided not to study the effect of this delay. Several works in the literature show that delays usually increase the regret in a moderate fashion (Mesterharm, 2005; Weinberger and Ordentlich, 2006; Agarwal and Duchi, 2011; Joulani et al., 2013).

As another example, consider the problem of product testing in a manufacturing process (e.g., the production of electronic consumer devices). When the product arrives, it can be subjected to a large number of diagnostic tests that differ in terms of their costs and effectiveness. The goal is to predict whether the product is defect-free. Obtaining the ground truth can also be quite expensive, especially for complex products. The challenge is that the effectiveness of the various tests is often *a priori* unknown and that different tests may provide complementary information (meaning that many tests may be required). . Hence, it might be challenging to decide what form the most cost-effective diagnostic procedure may take. Yet another example is the problem of developing a cost-effective way of instrument calibration. In this problem, the goal is to predict one or more real-valued parameters of some product. Again, various tests with different costs and reliability can be used as the input to the predictor.

Finally, although we pose the task as an online learning problem, it is easy to show that the procedures we develop can also be used to attack the batch learning problem, when the goal is to learn a predictor that will be cost-efficient on future data given a database of examples.

Obviously, when observing the loss is costly, the problem is related to active learning. However, to our best knowledge, the case when observing the features is costly has not been studied before in the online learning literature. Section 1.1 will discusses the relationship of our work to the existing literature in more detail.

This paper analyzes two versions of the online problem. In the first version, *free-label online probing*, there is no cost to seeing the loss function, that is, $c_{d+1} = 0$. (The loss function often compares the predicted value with some label in a known way, in which case learning the value of the label for the round means that the whole loss function becomes known; hence the choice of the name.) Thus, the learner naturally will choose to see the loss function after he provides his prediction; this provides feedback that the learner can use, to improve the predictor he produces. In the second version, *non-free-label online probing*, the cost of seeing the loss function is positive: $c_{d+1} > 0$.

In Section 2 we study the case of free-label online probing. We give an algorithm that enjoys a regret of $\mathcal{O}(\sqrt{2^d LT \ln \mathcal{N}_T(1/(TL))})$ when the losses are L-equi-Lipschitz (Theorem 2.2), where $\mathcal{N}_T(\varepsilon)$ is the ε-covering number of $\mathcal{F}$ on sequences of length T. This leads to an $\tilde{\mathcal{O}}(\sqrt{2^d LT})$ regret bound for typical function classes, such as the class of linear predictors with bounded weights and bounded inputs. We also show that, in the worst case, the exponential dependence on the dimension cannot be avoided in the bound. For the special case of linear prediction with quadratic loss, we give an algorithm whose regret scales only as $\tilde{\mathcal{O}}(\sqrt{dt})$, a vast improvement in the dependence on d.

The case of non-free-label online probing is treated in Section 3. Here, in contrast to the free-label case, we prove that the minimax growth rate of the regret is of the order $\tilde{\Theta}(T^{2/3})$. The increase of regret-rate stems from the fact that the "best competitor in hindsight" does not have to pay for the label. In contrast to the previous case, since the label is costly here, if the algorithm decides to see the

label it does not even have to reason about which features to observe, as querying the label requires paying a cost that is a constant over the cost of the best predictor in hindsight, already resulting in the $\tilde{\Theta}(T^{2/3})$ regret rate. However, in practice (for shorter horizons) it still makes sense to select the features that provide the best balance between the feature-cost and the prediction loss. Although we do not study this, we note that by combining the algorithmic ideas developed for the free-label case with the ideas developed for the non-free-label case, it is possible to derive an algorithm that reasons actively about the cost of observing the features, too.

In the part dealing with the free-label problem, we build heavily on the results of Mannor and Shamir (2011), while in the part dealing with the non-free-label problem we build on the ideas of (Cesa-Bianchi et al., 2006). Due to space limitations, all of our proofs are relegated to the appendix.

1.1 Related Work

This paper analyzes online learning when features (and perhaps labels) have to be purchased. The standard "batch learning" framework has a pure explore phase, which gives the learner a set of labeled, completely specified examples, followed by a pure exploit phase, where the learned predictor is asked to predict the label for novel instances. Notice the learner is not required (nor even allowed) to decide which information to gather. By contrast, "active (batch) learning" requires the learner to identify that information (Settles, 2009). Most such active learners begin with completely specified, but unlabeled instances; they then purchase labels for a subset of the instances. Our model, however, requires the learner to purchase feature values as well. This is similar to the "active feature-purchasing learning" framework (Lizotte et al., 2003). This is extended in Kapoor and Greiner (2005) to a version that requires the eventual predictor (as well as the learner) to pay to see feature values as well. However, these are still in the batch framework: after gathering the information, the learner produces a predictor, which is not changed afterwards.

Our problem is an online problem over multiple rounds, where at each round the learner is required to predict the label for the current example. Standard online learning algorithms typically assume that each example is given with all the features. For example, Cesa-Bianchi et al. (2005) provided upper and lower bounds on the regret where the learner is given all the features for each example, but must pay for any labels he requests. In our problem, the learner must pay to see the values of the features of each example as well as the cost to obtain its true label at each round. This cost model means there is an advantage to finding a predictor that involves few features, as long as it is sufficiently accurate. The challenge, of course, is finding these relevant features, which happens during this online learning process.

Other works, in particular Rostamizadeh et al. (2011) and Dekel et al. (2010), assume the features of different examples might be corrupted, missed, or partially observed due to various problems, such as failure in sensors gathering these features. Having such missing features is realistic in many applications. Rostamizadeh et al. (2011) provided an algorithm for this task in the online settings, with optimal $\mathcal{O}(\sqrt{T})$ regret where T is the number of rounds. Our model differs from this model as in our case the learner has the option to obtain the values of only the subset of the features that he selects.

2 Free-Label Probing

In this section we consider the case when the cost of observing the loss function is zero. Thus, we can assume without loss of generality that the learner receives the loss function at the end of each round (*i.e.,* $s_{t,d+1} = 1$). We will first consider the general setting where the only restriction is that the losses are equi-Lipschitz and the function set $\mathcal{F}$ has a finite empirical worst-case covering number. Then we consider the special case where the set of competitors are the linear predictors and the losses are quadratic.

2.1 The Case of Lipschitz losses

In this section we assume that the loss functions, ℓ_t, are Lipschitz with a known, common Lipschitz constant L over $\mathcal{Y}$ w.r.t. to some semi-metric $d_{\mathcal{Y}}$ of $\mathcal{Y}$: for all $t \geq 1$

$$\sup_{y,y' \in \mathcal{Y}} |\ell_t(y) - \ell_t(y')| \leq L \, d_{\mathcal{Y}}(y, y'). \tag{1}$$

Clearly, the problem is an instance of prediction with expert advice under partial information feedback (Auer et al., 2002), where each expert corresponds to an element of $\mathcal{F}$. Note that, if the learner chooses to observe the values of some features, then he will also be able to evaluate the losses of all the predictors $f \in \mathcal{F}$ that use only these selected features. This can be formalized as follows: By a slight abuse of notation let $s_t \in \{0,1\}^d$ be the indicator showing the features selected by the learner at time t (here we drop the last element of s_t as s_{t,d_1} is always 1); similarly, we will drop the last coordinate of the cost vector c throughout this section. Then, the learner can compute the loss of any predictor $f \in \mathcal{F}$ such that $\mathrm{s}(f) \leq s_t$, where $\leq$ denotes the conjunction of the component-wise comparison. However, for some loss functions, it may be possible to estimate the losses of other predictors, too. We will exploit this when we study some interesting special cases of the general problem. However, in general, it is not possible to infer the losses for functions such that $s_{t,i} < \mathrm{s}(f)_i$ for some i (cf. Theorem 2.3).

The idea is to study first the case when $\mathcal{F}$ is finite and then reduce the general case to the finite case by considering appropriate finite coverings of the space $\mathcal{F}$. The regret will then depend on how the covering numbers of the space $\mathcal{F}$ behave.

Mannor and Shamir (2011) studied problems similar to this in a general framework, where in addition to the loss of the selected predictor (expert), the losses of some other predictors are also communicated to the learner in every round. The connection between the predictors is represented by a directed graph whose nodes are labeled as elements of $\mathcal{F}$ (i.e., as the experts) and there is an edge from $f \in \mathcal{F}$ to $g \in \mathcal{F}$ if, when choosing f, the loss of g is also revealed to the learner. It is assumed that the graph of any round t, $G_t = (\mathcal{F}, E_t)$ becomes known to the learner at the beginning of the round. Further, it is also assumed that $(f,f) \in E_t$ for every $t \geq 1$ and $f \in \mathcal{F}$. Mannor and Shamir (2011) gave an algorithm, called ELP (exponential weights with linear programming), to solve this problem, which calls the Exponential Weights algorithm, but modifies it to explore less, exploiting the information structure of the problem. The exploration distribution is found by solving a linear program, explaining the name of the algorithm. The regret of ELP is analyzed in the following theorem.

Theorem 2.1 (Mannor and Shamir 2011)**.** *Consider a prediction with expert advice problem over $\mathcal{F}$ where in round t, $G_t = (\mathcal{F}, E_t)$ is the directed graph that encodes which losses become available to the learner. Assume that for any $t \geq 1$, at most $\chi(G_t)$ cliques of G_t can cover all vertices of G_t. Let B be a bound on the non-negative losses ℓ_t: $\max_{t\geq 1, f\in\mathcal{F}} \ell_t(f(x_t)) \leq B$. Then, there exists a constant $C_{\mathrm{ELP}} > 0$ such that for any $T > 0$, the regret of Algorithm 2 (shown in the Appendix) when competing against the best predictor using* ELP *satisfies*

$$\mathbb{E}[R_T] \quad \leq \quad C_{\mathrm{ELP}} B \sqrt{(\ln|\mathcal{F}|) \sum_{t=1}^{T} \chi(G_t)}\,. \tag{2}$$

The algorithm's computational cost in any given round is $\mathrm{poly}(|\mathcal{F}|)$.

For a finite $\mathcal{F}$, define $E_t \equiv E \doteq \{(f,g) \,|\, \mathrm{s}(g) \leq \mathrm{s}(f)\}$. Then clearly, $\chi(G_t) \leq 2^d$. Further, $B = \|c_{1:d}\|_1 + \max_{t\geq 1, y\in\mathcal{Y}} \ell_t(y) \doteq C_1 + \ell_{\max}$ (i.e., $C_1 = \|c_{1:d}\|_1$). Plugging these into (2) gives

$$\mathbb{E}[R_T] \quad \leq \quad C_{\mathrm{ELP}}(C_1 + \ell_{\max})\sqrt{2^d T \ln|\mathcal{F}|}\,. \tag{3}$$

To apply this algorithm in the case when $\mathcal{F}$ is infinite, we have to approximate $\mathcal{F}$ with a finite set $\mathcal{F}' \subset \{f \,|\, f : X \to \mathcal{Y}\}$. The worst-case maximum approximation error of $\mathcal{F}$ using $\mathcal{F}'$ over sequences of length T can be defined as

$$A_T(\mathcal{F}', \mathcal{F}) = \max_{x\in\mathcal{X}^T} \sup_{f\in\mathcal{F}} \inf_{f'\in\mathcal{F}'} \frac{1}{T}\sum_{t=1}^{T} d_{\mathcal{Y}}(f(x_t), f'(x_t)) + \langle\, (\mathrm{s}(f') - \mathrm{s}(f))^+, c_{1:d}\,\rangle\,,$$

where $(\mathrm{s}(f') - \mathrm{s}(f))^+$ denotes the coordinate-wise positive part of $\mathrm{s}(f') - \mathrm{s}(f)$, that is, the indicator vector of the features used by f' and not used by f. The average error can also be viewed as a (normalized) $d_{\mathcal{Y}}$-"distance" between the vectors $(f(x_t))_{1\leq t\leq T}$ and $(f'(x_t))_{1\leq t\leq T}$ penalized with the extra feature costs. For a given positive number α, define the *worst-case empirical covering number* of $\mathcal{F}$ at level α and horizon $T > 0$ by

$$\mathcal{N}_T(\mathcal{F}, \alpha) = \min\{\ |\mathcal{F}'| \ \mid\ \mathcal{F}' \subset \{f \,|\, f : X \to \mathcal{Y}\},\ A_T(\mathcal{F}', \mathcal{F}) \leq \alpha\,\}.$$

We are going to apply the ELP algorithm to $\mathcal{F}'$ and apply (3) to obtain a regret bound. If f' uses more features than f then the cost-penalized distance between f' and f is bounded from below by the cost of observing the extra features. This means that unless the problem is very special, $\mathcal{F}'$ has to contain, for all $s \in \{\mathrm{s}(f) \,|\, f \in \mathcal{F}\}$, some f' with $\mathrm{s}(f') = s$. Thus, if $\mathcal{F}$ contains a function for all $s \in \{0,1\}^d$, $\chi(G_t) = 2^d$. Selecting a covering $\mathcal{F}'$ that achieves accuracy α, the approximation error becomes $TL\alpha$ (using equation 1), giving the following bound:

Theorem 2.2. *Assume that the losses $(\ell_t)_{t\geq 1}$ are L-Lipschitz (cf. (1)) and $\alpha > 0$. Then, there exists an algorithm such that for any $T > 0$, knowing T, the regret satisfies*

$$\mathbb{E}[R_T] \leq C_{\mathrm{ELP}}(C_1 + \ell_{\max})\sqrt{2^d T \ln \mathcal{N}_T(\mathcal{F}, \alpha)} + TL\alpha\,.$$

In particular, by choosing $\alpha = 1/(TL)$, we have

$$\mathbb{E}[R_T] \leq C_{\mathrm{ELP}}(C_1 + \ell_{\max})\sqrt{2^d T \ln \mathcal{N}_T(\mathcal{F}, 1/(TL))} + 1\,.$$

We note in passing that the the dependence of the algorithm on the time horizon T can be alleviated, using, for example, the doubling trick.

In order to turn the above bound into a concrete bound, one must investigate the behavior of the *metric entropy*, $\ln \mathcal{N}_T(\mathcal{F}, \alpha)$. In many cases, the metric entropy can be bounded independently of T. In fact, often, $\ln \mathcal{N}_T(\mathcal{F}, \alpha) = D \ln(1 + c/\alpha)$ for some $c, D > 0$. When this holds, D is often called the "dimension" of $\mathcal{F}$ and we get that

$$\mathbb{E}\left[R_T\right] \leq C_{\mathrm{ELP}}(C_1 + \ell_{\max})\sqrt{2^d T D \ln(1 + cTL)} + 1\,.$$

As a specific example, we will consider the case of real-valued linear functions over a ball in a Euclidean space with weights belonging to some other ball. For a normed vector space V with norm $\|\cdot\|$ and dual norm $\|\cdot\|_*$, $x \in V$, $r \geq 0$, let $B_{\|\cdot\|}(x, r) = \{v \in V \,|\, \|v\| \leq r\}$ denote the ball in V centered at x that has radius r. For $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{W} \subset \mathbb{R}^d$, let

$$\mathcal{F} \subset \mathrm{Lin}(\mathcal{X}, \mathcal{W}) \doteq \{g : \mathcal{X} \to \mathbb{R} \,|\, g(\cdot) = \langle\, w, \cdot\,\rangle\,, w \in \mathcal{W}\} \tag{4}$$

be the space of linear mappings from $\mathcal{X}$ to reals with weights belonging to $\mathcal{W}$. We have the following lemma:

Lemma 2.1. *Let $X, W > 0$, $d_Y(y, y') = |y - y'|$, $\mathcal{X} \subset B_{\|\cdot\|}(0, X)$ and $\mathcal{W} \subset B_{\|\cdot\|_*}(0, W)$. Consider a set of real-valued linear predictors $\mathcal{F} \subset \mathrm{Lin}(\mathcal{X}, \mathcal{W})$. Then, for any $\alpha > 0$,*

$$\ln \mathcal{N}_T(\mathcal{F}, \alpha) \leq d \ln(1 + 2WX/\alpha).$$

The previous lemma, together with Theorem 2.2 immediately gives the following result:

Corollary 2.1. *Assume that $\mathcal{F} \subset \mathrm{Lin}(\mathcal{X}, \mathcal{W})$, $\mathcal{X} \subset B_{\|\cdot\|}(0, X)$, $\mathcal{W} \subset B_{\|\cdot\|_*}(0, W)$ for some $X, W > 0$. Further, assume that the losses $(\ell_t)_{t\geq 1}$ are L-Lipschitz. Then, there exists an algorithm such that for any $T > 0$, the regret of the algorithm satisfies,*

$$\mathbb{E}\left[R_T\right] \leq C_{\mathrm{ELP}}(C_1 + \ell_{\max})\sqrt{d 2^d T \ln(1 + 2TLWX)} + 1\,.$$

Note that if one is given an *a priori bound* p on the maximum number of features that can be used in a single round (allowing the algorithm to use fewer than p, but not more features) then 2^d in the above bound could be replaced by $\sum_{1\leq i\leq p} \binom{d}{i} \approx d^p$, where the approximation assumes that $p < d/2$. Such a bound on the number of features available per round may arise from strict budgetary considerations. When d^p is small, this makes the bound non-vacuous even for small horizons T. In addition, in such cases the algorithm also becomes computationally feasible. It remains an interesting open question to study the computational complexity when there is no restriction on the number of features used. In the next theorem, however, we show that the worst-case exponential dependence of the regret on the number of features cannot be improved (while keeping the root-T dependence on the horizon). The bound is based on the lower bound construction of Mannor and Shamir (2011), which reduces the problem to known lower bounds in the multi-armed bandit case.

Theorem 2.3. *There exist an instance of* free-label online probing *such that the minimax regret of any algorithm is $\Omega\left(\sqrt{\binom{d}{d/2} T}\right)$.*

2.2 Linear Prediction with Quadratic Losses

In this section, we study the problem under the assumption that the predictors have a linear form and the loss functions are quadratic. That is, $\mathcal{F} \subset \mathrm{Lin}(\mathcal{X}, \mathcal{W})$ where $\mathcal{W} = \{w \in \mathbb{R}^d \,|\, \|w\|_* \le w_{\lim}\}$ and $\mathcal{X} = \{x \in \mathbb{R}^d \,|\, \|x\| \le x_{\lim}\}$ for some given constants $w_{\lim}, x_{\lim} > 0$, while $\ell_t(y) = (y - y_t)^2$, where $|y_t| \le x_{\lim} w_{\lim}$. Thus, choosing a predictor is akin to selecting a weight vector $w_t \in \mathcal{W}$, as well as a binary vector $s_t \in \mathcal{G} \subset \{0,1\}^d$ that encodes the features to be used in round t. The prediction for round t is then $\hat{y}_t = \langle\, w_t, s_t \odot x_t \,\rangle$, where $\odot$ denotes coordinate-wise product, while the loss suffered is $(\hat{y}_t - y_t)^2$. The set $\mathcal{G}$ is an arbitrary non-empty, a priori specified subset of $\{0,1\}^d$ that allows the user of the algorithm to encode extra constraints on what subsets of features can be selected.

In this section we show that in this case a regret bound of size $\tilde{\mathcal{O}}(\sqrt{\mathrm{poly}(d)T})$ is possible. The key idea that permits the improvement of the regret bound is that a randomized choice of a weight vector W_t (and thus, of a subset) helps one construct unbiased estimates of the losses $\ell_t(\langle\, w, s \odot x_t \,\rangle)$ for all weight vectors w and all subsets $s \in \mathcal{G}$ under some mild conditions on the distribution of W_t. That the construction of such unbiased estimates is possible, despite that some feature values are unobserved, is because of the special algebraic structure of the prediction and loss functions. A similar construction has appeared in a different context, e.g., in the paper of Cesa-Bianchi et al. (2010).

The construction works as follows. Define the $d \times d$ matrix, X_t by $(X_t)_{i,j} = x_{t,i} x_{t,j}$ $(1 \le i, j \le d)$. Expanding the loss of the prediction $\hat{y}_t = \langle\, w, x_t \,\rangle$, we get that the loss of using $w \in \mathcal{W}$ is

$$\ell_t(w) \doteq \ell_t(\langle\, w, x_t \,\rangle) = w^\top X_t\, w - 2\, w^\top x_t y_t + y_t^2,$$

where with a slight abuse of notation we have introduced the loss function $\ell_t : \mathcal{W} \to \mathbb{R}$ (we'll keep abusing the use of ℓ_t by overloading it based on the type of its argument). Clearly, it suffices to construct unbiased estimates of $\ell_t(w)$ for any $w \in \mathcal{W}$.

We will use a discretization approach. Therefore, assume that we are given a finite subset $\mathcal{W}'$ of $\mathcal{W}$ that will be constructed later. In each step t, our algorithm will choose a random weight vector W_t from a probability distribution supported on $\mathcal{W}'$. Let $p_t(w)$ be the probability of selecting the weight vector, $w \in \mathcal{W}'$. For $1 \le i \le d$, let

$$q_t(i) = \sum_{w \in \mathcal{W}' : i \in \mathrm{s}(w)} p_t(w)\,,$$

be the probability that $\mathrm{s}(W_t)$ will contain i, while for $1 \le i, j \le d$, let

$$q_t(i,j) = \sum_{w \in \mathcal{W}' : i,j \in \mathrm{s}(w)} p_t(w)\,,$$

be the probability that both $i, j \in \mathrm{s}(W_t)$.[2] Assume that $p_t(\cdot)$ is constructed such that $q_t(i,j) > 0$ holds for any time t and indices $1 \le i, j \le d$. This also implies that $q_t(i) > 0$ for all $1 \le i \le d$. Define the vector $\tilde{x}_t \in \mathbb{R}^d$ and matrix $\tilde{X}_t \in \mathbb{R}^{d \times d}$ using the following equations:

$$\tilde{x}_{t,i} = \frac{\mathbb{1}_{\{i \in \mathrm{s}(W_t)\}} x_{t,i}}{q_t(i)}\,, \qquad (\tilde{X}_t)_{i,j} = \frac{\mathbb{1}_{\{i,j \in \mathrm{s}(W_t)\}} x_{t,i} x_{t,j}}{q_t(i,j)}\,. \tag{5}$$

It can be readily verified that $\mathbb{E}\left[\tilde{x}_t \,|\, p_t\right] = x_t$ and $\mathbb{E}\left[\tilde{X}_t \,|\, p_t\right] = X_t$. Further, notice that both $\tilde{x}_t$ and $\tilde{X}_t$ can be computed based on the information available at the end of round t, *i.e.,* based on the feature values $(x_{t,i})_{i \in \mathrm{s}(W_t)}$. Now, define the estimate of prediction loss

$$\tilde{\ell}_t(w) = w^\top \tilde{X}_t\, w - 2\, w^\top \tilde{x}_t y_t + y_t^2\,. \tag{6}$$

Note that y_t can be readily computed from $\ell_t(\cdot)$, which is available to the algorithm (equivalently, we may assume that the algorithm observed y_t). Due to the linearity of expectation, we have $\mathbb{E}\left[\tilde{\ell}_t(w) | p_t\right] = \ell_t(w)$. That is, $\tilde{\ell}_t(w)$ provides an unbiased estimate of the loss $\ell_t(w)$ for any $w \in \mathcal{W}$. Hence, by adding a feature cost term we get $\tilde{\ell}_t(w) + \langle\, \mathrm{s}(w), c \,\rangle$ as an estimate of the loss that the learner would have suffered at round t had he chosen the weight vector w.

[2] Note that, following our earlier suggestion, we view the d-dimensional binary vectors as subsets of $\{1, \ldots, d\}$.

Algorithm 1 The LQDEXP3 Algorithm

Parameters: Real numbers $0 \le \eta$, $0 < \gamma \le 1$, $\mathcal{W}' \subset \mathcal{W}$ finite set, a distribution μ over $\mathcal{W}'$, horizon $T > 0$.
Initialization: $u_1(w) = 1$ ($w \in \mathcal{W}'$).
for $t = 1$ **to** T **do**
Draw $W_t \in \mathcal{W}'$ from the probability mass function

$$p_t(w) = (1-\gamma)\frac{u_t(w)}{U_t} + \gamma\mu(w), \qquad w \in \mathcal{W}'\,.$$

Obtain the features values, $(x_{t,i})_{i \in \mathrm{s}(W_t)}$.
Predict $\hat{y}_t = \sum_{i \in \mathrm{s}(W_t)} w_{t,i} x_{t,i}$.
for $w \in \mathcal{W}'$ **do**
Update the weights using (6) for the definitions of $\tilde{\ell}_t(w)$:

$$u_{t+1}(w) = u_t(w) e^{-\eta(\tilde{\ell}_t(w) + \langle\, c, \mathrm{s}(w)\,\rangle)}, \quad w \in \mathcal{W}'\,.$$

end for
end for

2.2.1 LQDExp3 – A Discretization-based Algorithm

Next we show that the standard EXP3 Algorithm applied to a discretization of the weight space $\mathcal{W}$ achieves $\mathcal{O}(\sqrt{dT})$ regret. The algorithm, called LQDEXP3 is given as Algorithm 1. In the name of the algorithm, LQ stands for linear prediction with quadratic losses and D denotes discretization. Note that if the exploration distribution μ in the algorithm is such that for any $1 \le i, j \le d$, $\sum_{w \in W' : i,j \in \mathrm{s}(w)} \mu(w) > 0$ then $q_t(i,j) > 0$ will be guaranteed for all time steps. Using the notation $y_{\lim} = w_{\lim} x_{\lim}$ and $E_{\mathcal{G}} = \max_{s \in \mathcal{G}} \sup_{w \in \mathcal{W} : \|w\|_* = 1} \|w \odot s\|_*$, we can state the following regret bound on the algorithm

Theorem 2.4. *Let $w_{\lim}, x_{\lim} > 0$, $c \in [0,\infty)^d$ be given, $\mathcal{W} \subset B_{\|\cdot\|_*}(0, w_{\lim})$ convex, $\mathcal{X} \subset B_{\|\cdot\|}(0, x_{\lim})$ and fix $T \ge 1$. Then, there exist a parameter setting for* LQDEXP3 *such that the following holds: Let R_T denote the regret of* LQDEXP3 *against the best linear predictor from $\mathrm{Lin}(\mathcal{W}, \mathcal{X})$ when* LQDEXP3 *is used in an online free-label probing problem defined with the sequence $((x_t, y_t))_{1 \le t \le T}$ ($\|x_t\| \le x_{\lim}$, $|y_t| \le y_{\lim}$, $1 \le t \le T$), quadratic losses $\ell_t(y) = (y - y_t)^2$, and feature-costs given by the vector c. Then,*

$$\mathbb{E}\left[R_T\right] \le C \sqrt{Td\,(4y_{\lim}^2 + \|c\|_1)(w_{\lim}^2 x_{\lim}^2 + 2y_{\lim} w_{\lim} x_{\lim} + 4y_{\lim}^2 + \|c\|_1)\ln(E_{\mathcal{G}} T)}\,,$$

where $C > 0$ is a universal constant (i.e., the value of C does not depend on the problem parameters).

The actual parameter setting to be used with the algorithm is constructed in the proof. The computational complexity of LQDEXP3 is exponential in the dimension d due to the discretization step, hence quickly becomes impractical when the number of features is large. On the other hand, one can easily modify the algorithm to run without discretization by replacing EXP3 with its continuous version. The resulting algorithm enjoys essentially the same regret bound, and can be implemented efficiently whenever efficient sampling is possible from the resulting distribution. This approach seems to be appealing, since, from a first look, it seems to involve sampling from truncated Gaussian distributions, which can be done efficiently. However, it is easy to see that when the sampling probabilities of some feature are small, the estimated loss will not be convex as $\tilde{X}_t$ may not be positive semi-definite, and therefore the resulting distributions will not always be truncated Gaussians. Finding an efficient sampling procedure for such situations is an interesting open problem.

The optimality of LQDEXP3 can be seen by the following lower bound on the regret:

Theorem 2.5. *Let $d > 0$, and consider the online free label probing problem with linear predictors, where $\mathcal{W} = \{w \in \mathbb{R}^d \,|\, \|w\|_1 \le w_{\lim}\}$ and $\mathcal{X} = \{x \in \mathbb{R}^d \,|\, \|x\|_\infty \le 1\}$. Assume, for all $t \ge 1$, that the loss functions are of the form $\ell_t(w) = (w^\top x_t - y_t)^2 + \langle\, \mathrm{s}(w), c\,\rangle$, where $|y_t| \le 1$ and $c = 1/2 \times \mathbf{1} \in \mathbb{R}^d$. Then, for any prediction algorithm and for any $T \ge \frac{4d}{8\ln(4/3)}$, there exists a*

sequence $((x_t, y_t))_{1 \le t \le T} \in (\mathcal{X} \times [-1, 1])^T$ *such that the regret of the algorithm can be bounded from below as*

$$\mathbb{E}[R_T] \ge \frac{\sqrt{2}-1}{\sqrt{32 \ln(4/3)}} \sqrt{Td}\,.$$

3 Non-Free-Label Probing

If $c_{d+1} > 0$, the learner has to pay for observing the true label. This scenario is very similar to the well-known label-efficient prediction case in online learning (Cesa-Bianchi et al., 2006). In fact, the latter problem is a special case of this problem, immediately giving us that the regret of any algorithm is at least of order $T^{2/3}$. It turns out that if one observes the (costly) label in a given round then it does not effect the regret rate if one observes all the features at the same time. The resulting "revealing action algorithm", given in Algorithm 3 in the Appendix, achieves the following regret bound for finite expert classes:

Lemma 3.1. *Given any non-free-label online probing with finitely many experts, Algorithm 3 with appropriately set parameters achieves*

$$\mathbb{E}[R_T] \le C \max\left(T^{2/3}(\ell_{\max}^2 \|c\|_1 \ln|\mathcal{F}|)^{1/3}, \ell_{\max}\sqrt{T \ln|\mathcal{F}|}\right)$$

for some constant $C > 0$.

Using the fact that, in the linear prediction case, approximately $(2TLWX+1)^d$ experts are needed to approximate each expert in $\mathcal{W}$ with precision $\alpha = \frac{1}{LT}$ in worst-case empirical covering, we obtain the following theorem (note, however, that the complexity of the algorithm is again exponential in the dimension d, as we need to keep a weight for each expert):

Theorem 3.1. *Given any non-free-label online probing with linear predictor experts and Lipschitz prediction loss function with constant L, Algorithm 3 with appropriately set parameters running on a sufficiently discretized predictor set achieves*

$$\mathbb{E}[R_T] \le C \max\left(T^{2/3}\left[\ell_{\max}^2 \|c\|_1\, d\, \ln(TLWX)\right]^{1/3}, \ell_{\max}\sqrt{Td\ln(TLWX)}\right)$$

for some universal constant $C > 0$.

That Algorithm 3 is essentially optimal for linear predictions and quadratic losses is a consequence of the following almost matching lower bound:

Theorem 3.2. *There exists a constant* C *such that, for any non-free-label probing with linear predictors, quadratic loss, and* $c_j > (1/d)\sum_{i=1}^d c_i - 1/2d$ *for every* $j = 1, \ldots, d$*, the expected regret of any algorithm can be lower bounded by*

$$\mathbb{E}[R_T] \ge C(c_{d+1}d)^{1/3}T^{2/3}\,.$$

4 Conclusions

We introduced a new problem called *online probing*. In this problem, the learner has the option of choosing the subset of features he wants to observe as well as the option of observing the true label, but has to pay for this information. This setup produced new challenges in solving the online problem. We showed that when the labels are free, it is possible to devise algorithms with optimal regret rate $\Theta(\sqrt{T})$ (up to logarithmic factors), while in the non-free-label case we showed that only $\Theta(T^{2/3})$ is achievable. We gave algorithms that achieve the optimal regret rate (up to logarithmic factors) when the number of experts is finite or in the case of linear prediction. Unfortunately either our bounds or the computational complexity of the corresponding algorithms are exponential in the problem dimension, and it is an open problem whether these disadvantages can be eliminated simultaneously.

Acknowledgements

The authors thank Yevgeny Seldin for finding a bug in an earlier version of the paper. This work was supported in part by DARPA grant MSEE FA8650-11-1-7156, the Alberta Innovates Technology Futures, AICML, and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Agarwal, A. and Duchi, J. C. (2011). Distributed delayed stochastic optimization. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. C. N., and Weinberger, K. Q., editors, *NIPS*, pages 873–881.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77.

Bartók, G. (2012). *The role of information in online learning*. PhD thesis, Department of Computing Science, University of Alberta.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge Univ Pr.

Cesa-Bianchi, N., Lugosi, G., and Stoltz, G. (2005). Minimizing regret with label efficient prediction. *IEEE Transactions on Information Theory*, 51(6):2152–2162.

Cesa-Bianchi, N., Lugosi, G., and Stoltz, G. (2006). Regret minimization under partial monitoring. *Math. Oper. Res.*, 31(3):562–580.

Cesa-Bianchi, N., Shalev-Shwartz, S., and Shamir, O. (2010). Efficient learning with partially observed attributes. *CoRR*, abs/1004.4421.

Dekel, O., Shamir, O., and Xiao, L. (2010). Learning to classify with missing and corrupted features. *Machine Learning*, 81(2):149–178.

Joulani, P., György, A., and Szepesvári, C. (2013). Online learning under delayed feedback. In *30th International Conference on Machine Learning*, Atlanta, GA, USA.

Kapoor, A. and Greiner, R. (2005). Learning and classifying under hard budgets. In *European Conference on Machine Learning (ECML)*, pages 166–173.

Lizotte, D., Madani, O., and Greiner, R. (2003). Budgeted learning of naive-Bayes classifiers. In *Conference on Uncertainty in Artificial Intelligence (UAI)*.

Mannor, S. and Shamir, O. (2011). From bandits to experts: On the value of side-observations. *CoRR*, abs/1106.2436.

Mesterharm, C. (2005). On-line learning with delayed label feedback. In *Proceedings of the 16th international conference on Algorithmic Learning Theory*, ALT'05, pages 399–413, Berlin, Heidelberg. Springer-Verlag.

Rostamizadeh, A., Agarwal, A., and Bartlett, P. L. (2011). Learning with missing features. In *UAI*, pages 635–642.

Settles, B. (2009). Active learning literature survey. Technical report.

Weinberger, M. J. and Ordentlich, E. (2006). On delayed prediction of individual sequences. *IEEE Trans. Inf. Theor.*, 48(7):1959–1976.

APPENDIX—SUPPLEMENTARY MATERIAL

A.1 Free-Label Probing: Lipschitz losses

A.1.1 The ELP algorithm of Mannor and Shamir (2011)

Algorithm 2 The ELP Algorithm. In the pseudocode, Δ_N denotes the N-dimensional simplex: $\Delta_N = \{s \in [0,1]^N \mid \sum_{i=1}^N s_i = 1\}$.

Parameters: Neighborhood graphs $G_t = (\mathcal{F}, E_t)$, $1 \le t \le T$, a bound B on the losses.
Initialization: $N = |\mathcal{F}|$, $\beta = \sqrt{(\ln N)/(3B^2 \sum_t \chi(G_t))}$, $w_{0,j} = 1/N$, $1 \le j \le N$.
for $t = 1$ **to** T **do**
  Let $s_t = \arg\max_{q \in \Delta_N} \min_{1 \le i \le N} \sum_{(i,k) \in E_t} q_k$.
  Let $s_t^* = \min_{1 \le i \le N} \sum_{(i,k) \in E_t} s_{t,i}$.
  Let $\gamma_t = \beta B / s_t^*$.
  Choose action i_t randomly from probability mass function

$$p_{t,i} = (1 - \gamma_t) \frac{w_{t,i}}{\sum_{j=1}^N w_{t,j}} + \gamma_t s_{t,i} \qquad (1 \le i \le N).$$

  Receive loss $(\ell_{t,k})_{(i_t,k) \in E_t}$.
  Compute $\tilde{g}_{t,j} = \frac{B - \ell_{t,j}}{\sum_{(l,j) \in E_t} p_{t,l}}$ if $(j, i_t) \in E_t$, and $\tilde{g}_j(t) = 0$ otherwise.
  $w_{t+1,j} = w_{t,j} \exp(\beta \tilde{g}_{t,j})$, $1 \le j \le N$.
end for

A.1.2 Proofs

Lemma 2.1. *Let $X, W > 0$, $d_{\mathcal{Y}}(y, y') = |y - y'|$, $\mathcal{X} \subset B_{\|\cdot\|}(0, X)$ and $\mathcal{W} \subset B_{\|\cdot\|_*}(0, W)$. Consider a set of real-valued linear predictors $\mathcal{F} \subset \mathrm{Lin}(\mathcal{X}, \mathcal{W})$. Then, for any $\alpha > 0$,*

$$\ln \mathcal{N}_T(\mathcal{F}, \alpha) \le d \ln(1 + 2WX/\alpha).$$

Proof. An appropriate covering of $\mathcal{F}$ can be constructed as follows: Consider an ε-covering $\mathcal{W}'$ of the ball $\mathcal{W}$ with respect to $\|\cdot\|_*$ for some $\varepsilon > 0$ (*i.e.,* for any $w \in \mathcal{W}$ there exists $w' \in \mathcal{W}'$ such that $\|w - w'\|_* \le \varepsilon$). Then,

$$\mathcal{F}' = \{g : \mathcal{X} \to \mathbb{R} \mid g(x) = \langle\, x, w \,\rangle, w \in \mathcal{W}'\} \tag{7}$$

is an εX-covering of $\mathcal{F}$. To see this pick any $f \in \mathcal{F}$. Thus, $f(x) = \langle\, w, x \,\rangle$ for some $w \in \mathcal{W}$. Take the vector in $\mathcal{W}'$ that is closest to w and call it w'. Thus, $\|w - w'\|_* \le \varepsilon$. Let $g \in \mathcal{F}'$ be given by $g(x) = \langle\, x, w' \,\rangle$. Then,

$$\frac{1}{T} \sum_{t=1}^T |f(x_t) - g(x_t)| = \frac{1}{T} \sum_{t=1}^T |\langle\, w - w', x_t \,\rangle| \le \varepsilon X, \tag{8}$$

where in the last step we used Hölder's inequality and that by assumption $x_t \in \mathcal{X}$ and thus $\|x_t\| \le X$. This argument thus shows that to get an α-covering of $\mathcal{F}$, we need an ε-covering of $\mathcal{W}$ with $\varepsilon = \alpha/X$ and therefore $\mathcal{N}_T(\mathcal{F}, \alpha) \le \mathcal{N}(\mathcal{W}, \alpha/X)$. As it is well known, $\mathcal{N}(\mathcal{W}, \varepsilon) \le (2W/\varepsilon + 1)^d$ and thus $\ln \mathcal{N}_T(\mathcal{F}, \alpha) \le d \ln(1 + 2WX/\alpha)$. □

Theorem 2.3. *There exist an instance of* free-label online probing *such that the minimax regret of any algorithm is* $\Omega\left(\sqrt{\binom{d}{d/2} T}\right)$.

Proof. Let $\mathcal{X} = \{0,1\}^d$ and let $\mathcal{F}$ contain the XOR functions applied to all possible subsets of features: $\mathcal{F} = \{\mathrm{XOR}_S \mid S \subset \{1, \ldots, d\}\}$, where $\mathrm{XOR}_S(x) = \otimes_{i \in S} x_i$ ($\otimes$ denotes the Boolean XOR operator). Let the input vector $x_t \in \mathcal{X}$ at time t be such that its components are generated

independently from each other from the uniform distribution and let them be independent of inputs generated at different time indices. The prediction space is restricted to $\mathcal{Y} = \{0, 1\}$ and the loss is defined to be the zero-one loss: $\ell_t(y) = \mathbb{1}_{\{y \neq \mathrm{XOR}_{S^*}(x_t)\}}$, where $S^* \subset \{1, \dots, d\}$. The cost of the individual features is uniform. In particular, $c_i = 1/(2d)$.

Note that if an algorithm chooses *not* to observe some feature $i \in \{1, \dots, d\}$ at some time step t, there is now way the algorithm can find out the result of $\mathrm{XOR}_S(x_t)$ for any S containing i.[3] Hence, no algorithm can infer the losses for such functions. It is also clear that if the feature values for some set of features S are observed then the loss for any function $\mathrm{XOR}_{S'}$ with $S' \subset S$ can be inferred from the observed loss function. Thus, the directed graph over $\mathcal{F}$ that connects $f \in \mathcal{F}$ to $f' \in \mathcal{F}$ when given the loss for f, one also learns the loss of f', is isomorphic to the graph obtained from the lattice structure of subsets of $\{1, \dots, d\}$. This latter graph has an independent set of size $\binom{d}{d/2}$ (*i.e.,* the set of all functions that use exactly $d/2$ features) and thus we can apply the same method that Mannor and Shamir (2011) uses to prove their Theorem 4 to get the desired lower bound for this problem. □

A.2 Free Label Probing: Linear Prediction

A.2.1 Upper Bound on the Regret

Theorem 2.4. *Let $w_{\lim}, x_{\lim} > 0$, $c \in [0, \infty)^d$ be given, $\mathcal{W} \subset B_{\|\cdot\|_*}(0, w_{\lim})$ convex, $\mathcal{X} \subset B_{\|\cdot\|}(0, x_{\lim})$ and fix $T \geq 1$. Then, there exist a parameter setting for* LQDEXP3 *such that the following holds: Let R_T denote the regret of* LQDEXP3 *against the best linear predictor from $\mathrm{Lin}(\mathcal{W}, \mathcal{X})$ when* LQDEXP3 *is used in an online free-label probing problem defined with the sequence $((x_t, y_t))_{1 \leq t \leq T}$ ($\|x_t\| \leq x_{\lim}$, $|y_t| \leq y_{\lim}$, $1 \leq t \leq T$), quadratic losses $\ell_t(y) = (y - y_t)^2$, and feature-costs given by the vector c. Then,*

$$\mathbb{E}\left[R_T\right] \leq C \sqrt{Td\left(4y_{\lim}^2 + \|c\|_1\right)\left(w_{\lim}^2 x_{\lim}^2 + 2y_{\lim} w_{\lim} x_{\lim} + 4y_{\lim}^2 + \|c\|_1\right) \ln(E_{\mathcal{G}} T)}\,,$$

where $C > 0$ is a universal constant (i.e., the value of C does not depend on the problem parameters).

Before stating the proof, we state a lemma that will be needed in the proof of this theorem. The lemma gives a bound on the regret of an exponential weights algorithm as a function of some "statistics" of the losses fed to the algorithm. Since the result is essentially extracted from the paper by Auer et al. (2002), its proof is omitted.

Lemma A.2.1. *Fix the integers $N, T > 0$, the real numbers $0 < \gamma < 1$, $\eta > 0$ and let μ be a probability mass function over the set $\underline{N} = \{1, \dots, N\}$. Let $\ell_t : \underline{N} \to \mathbb{R}$ be a sequence of loss functions such that*

$$\eta \ell_t(i) \geq -1 \tag{9}$$

holds for all $1 \leq t \leq T$ and $i \in \underline{N}$. Define the sequence of functions $(u_t)_{1 \leq t \leq T}, (p_t)_{1 \leq t \leq T}$ ($u_t : \underline{N} \to \mathbb{R}^+$, $p_t : \underline{N} \to [0, 1]$) by $u_t \equiv 1$,

$$u_t(i) = \exp\left(\eta \sum_{s=1}^{t-1} \ell_s(i)\right), \quad p_t(i) = (1 - \gamma)\frac{u_t(i)}{\sum_{j \in \underline{N}} u_t(j)} + \gamma \mu(i), \quad (i \in \underline{N},\, 1 \leq t \leq T + 1)\,.$$

Let $\hat{L}_T = \sum_{t=1}^T \sum_{j \in \underline{N}} p_t(j) \ell_t(j)$ and $L_T(i) = \sum_{t=1}^T \ell_t(i)$. Then, for any $i \in \underline{N}$,

$$\hat{L}_T - L_T(i) \leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \sum_{j \in \underline{N}} p_t(j) \ell_t^2(j) + \gamma \sum_{t=1}^T \sum_{j \in \underline{N}} \mu(j) \left\{\ell_t(j) - \ell_t(i)\right\}\,.$$

Proof of Theorem 2.4. Fix the sequence of $((x_t, y_t))_{1 \leq t \leq T}$ as in the statement of the theorem and let $\ell_t(y) = (y - y_t)^2$. Remember that (with a slight abuse of notation), the loss of using weight $w \in \mathcal{W}$ in time step t is

$$\ell_t(w) = \ell_t(\langle w, x_t \rangle), \quad 1 \leq t \leq T\,.$$

[3] Here, and in what follows we will identify features with their indices.

Now, observe that $\hat{y}_t = \sum_{i \in \mathrm{s}(W_t)} W_{t,i} x_{t,i} = \langle\, W_t, x_t \,\rangle$ holds thanks to the definition of $\mathrm{s}(W_t)$ and thus $\ell_t(\hat{y}_t) = \ell_t(W_t)$. Hence, the total loss of the algorithm can be written as

$$\hat{L}_T = \sum_{t=1}^{T} \left[\langle\, \mathrm{s}(W_t), c \,\rangle + \ell_t(W_t) \right].$$

Let

$$L_T(w) = T \langle\, \mathrm{s}(w), c \,\rangle + \sum_{t=1}^{T} \ell_t(w), \qquad w \in \mathbb{R}^d,$$

be the total loss of using the weight vector w. Then the regret of LQDEXP3 up to time T on the sequence $((x_t, y_t))_{1 \le t \le T}$ can be written as

$$R_T = \max_{w \in \mathcal{W}} R_T(w)\,,$$

where

$$R_T(w) \doteq \hat{L}_T - L_T(w), \qquad w \in \mathbb{R}^d.$$

Using the discretized weight vector set, $\mathcal{W}'$, the regret can be written as

$$\begin{aligned} R_T &= \max_{w \in \mathcal{W}} R_T(w) \\ &= \left\{ \hat{L}_T - \min_{w' \in \mathcal{W}'} L_T(w') \right\} + \left\{ \min_{w' \in \mathcal{W}'} L_T(w') - \min_{w \in \mathcal{W}} L_T(w) \right\} \\ &= \left\{ \hat{L}_T - \min_{w' \in \mathcal{W}'} L_T(w') \right\} + \max_{w \in \mathcal{W}} \min_{w' \in \mathcal{W}'} \left\{ L_T(w') - L_T(w) \right\}. \end{aligned} \tag{10}$$

Now, fix $w \in \mathcal{W}$. By construction, $\mathcal{W}'$ is such that for any $s \in \{0,1\}$, there exists some vector $w' \in \mathcal{W}'$ such that $\mathrm{s}(w') = s$. Then,

$$\min_{w' \in \mathcal{W}'} \left\{ L_T(w') - L_T(w) \right\} \le \min_{w' \in \mathcal{W}' : \mathrm{s}(w') = \mathrm{s}(w)} \left\{ L_T(w') - L_T(w) \right\} = \min_{w' \in \mathcal{W}' : \mathrm{s}(w') = \mathrm{s}(w)} \sum_{t=1}^{T} \ell_t(w') - \ell_t(w).$$

Let us first deal with the second term. A simple calculation shows that $\ell_t : [-y_{\lim}, y_{\lim}] \to \mathbb{R}$, $y \mapsto (y - y_t)^2$ is $4y_{\lim}$-Lipschitz. Hence, as long as $w' \in W'$ is such that $\mathrm{s}(w') = \mathrm{s}(w)$,

$$L_T(w') - L_T(w) = \sum_{t=1}^{T} \ell(\langle\, w', x_t \,\rangle, y_t) - \ell_t(\langle\, w, x_t \,\rangle, y_t) \le 4T y_{\lim} \left(\frac{1}{T} \sum_{t=1}^{T} |\langle\, w - w', x_t \,\rangle| \right).$$

For $s \in \mathcal{G}$, define $\mathcal{W}'(s) = \{ w \in \mathcal{W}' \mid \mathrm{s}(w) = s \}$ and $W(s) = \{ w \in \mathcal{W} \mid \mathrm{s}(w) = s \}$. For $\alpha > 0$, let $W_\alpha(s) \subset \mathcal{W}$ be the minimal cardinality subset of $\mathcal{W}(s)$ such that $\mathrm{Lin}(\mathcal{X}, \mathcal{W}_\alpha(s))$ is an α-cover of $\mathrm{Lin}(\mathcal{X}, \mathcal{W}(s))$ w.r.t. $d_{\mathcal{Y}}(y, y') = |y - y'|$. Choose

$$\mathcal{W}' = \cup_{s \in \mathcal{G}} \mathcal{W}_\alpha(s)\,.$$

Then, by construction,

$$\min_{w' \in \mathcal{W}'} L_T(w') - L_T(w) \le 4T y_{\lim} \alpha \tag{11}$$

and since this holds for any $w \in W$, we get that the same bound applies to $\max_{w \in \mathcal{W}} \min_{w' \in \mathcal{W}'} L_T(w') - L_T(w)$. Before we turn to bounding the first term of (10), let us bound the cardinality of $\mathcal{W}'$, which we will need later.

Notice that

$$|\mathcal{W}'| \le \sum_{s \in \mathcal{G}} |\mathcal{W}_\alpha(s)| \le |\mathcal{G}| \max_{s \in \mathcal{G}} |\mathcal{W}_\alpha(s)|.$$

Now, note also that thanks to the definition of $E_{\mathcal{G}}$, for any $s \in \mathcal{G}$, $w \in \mathcal{W}$, $\|w\|_* \le E_{\mathcal{G}} \cdot \|w \odot s\|_*$. Let $\mathcal{W}_\alpha$ denote a minimum cardinality α-cover of $\mathcal{W}$. Then, it is easy to see that for any $s \in \mathcal{G}$, $\mathrm{Lin}(\mathcal{X}, \mathcal{W}_{\alpha/E_{\mathcal{G}}})$ is an α-cover of $\mathrm{Lin}(\mathcal{X}, \mathcal{W}(s))$ w.r.t. $d_{\mathcal{Y}}(y, y') = |y - y'|$. Hence, by the minimum cardinality property of $\mathcal{W}_\alpha(s)$, we have $|\mathcal{W}_\alpha(s)| \le |\mathcal{W}_{\alpha/E_{\mathcal{G}}}|$ and, by Lemma 2.1, we get that $\ln^+ |\mathcal{W}_\alpha(s)| \le d \ln(1 + 2E_{\mathcal{G}} y_{\lim}/\alpha)$. Hence,

$$\ln |\mathcal{W}'| \le \ln(|\mathcal{G}|) + d \ln(1 + 2E_{\mathcal{G}} y_{\lim}/\alpha). \tag{12}$$

Let us now turn to bounding the expectation of the first term of (10). We have,

$$\mathbb{E}\left[\hat{L}_T - \min_{w\in\mathcal{W}'} L_T(w)\right] = \max_{w\in\mathcal{W}'} \mathbb{E}\left[\hat{L}_T - L_T(w)\right],$$

where we have exploited that $L_T(w)$ is deterministic. Therefore, it suffices to bound $\mathbb{E}\left[\hat{L}_T - L_T(w)\right]$ for any fixed $w \in \mathcal{W}'$. Thus, fix $w \in \mathcal{W}'$.

By the construction of $\tilde{\ell}_t$, $\mathbb{E}\left[\tilde{\ell}_t(w)|p_t\right] = \ell_t(w)$ holds for any $w \in \mathbb{R}^d$. Hence, $\mathbb{E}\left[L_T(w)\right] = \mathbb{E}\left[\sum_{t=1}^T \tilde{\ell}_t(w)\right]$. Furthermore, $\mathbb{E}\left[\ell_t(W_t)|p_t\right] = \sum_{w\in\mathcal{W}'} p_t(w)\ell_t(w) = \sum_{w\in\mathcal{W}'} p_t(w)\mathbb{E}\left[\tilde{\ell}_t(w)|p_t\right] = \mathbb{E}\left[\sum_{w\in\mathcal{W}'} p_t(w)\tilde{\ell}_t(w)|p_t\right]$.

Introduce $\hat{\tilde{\ell}}_t(w) = \tilde{\ell}_t(w) + \langle\, \mathrm{s}(w), c\,\rangle$, Then, we see that it suffices to bound

$$\mathbb{E}\left[\hat{L}_T - L_T(w)\right] = \mathbb{E}\left[\sum_{t=1}^T \sum_{w'\in W'} p_t(w')\hat{\tilde{\ell}}_t(w') - \sum_{t=1}^T \hat{\tilde{\ell}}_t(w)\right].$$

Now, by Lemma A.2.1, under the assumption that $0 < \gamma \le 1$, $0 < \eta$ are such that for any $w' \in \mathcal{W}'$, $1 \le t \le T$ the inequality

$$\eta\hat{\tilde{\ell}}_t(w') \ge -1 \tag{13}$$

holds, we have

$$\begin{aligned}
&\sum_{t=1}^T \sum_{w'\in W'} p_t(w')\hat{\tilde{\ell}}_t(w') - \sum_{t=1}^T \hat{\tilde{\ell}}_t(w) \\
&\le \frac{\ln|\mathcal{W}'|}{\eta} + \eta\sum_{t=1}^T \sum_{w'\in\mathcal{W}'} p_t(w')\hat{\tilde{\ell}}_t^2(w') + \gamma\sum_{t=1}^T \sum_{w'\in\mathcal{W}'} \mu(w')(\hat{\tilde{\ell}}_t(w') - \hat{\tilde{\ell}}_t(w)).
\end{aligned}$$

Let us assume for a moment that η, γ can be chosen to satisfy the above quoted conditions – we shall return to the choice of these parameters soon. Taking expectations of both sides of the last inequality, we get

$$\mathbb{E}\left[\hat{L}_T - L_T(w)\right] \le \frac{\ln|\mathcal{W}'|}{\eta} + \eta\sum_{t=1}^T \sum_{w'\in\mathcal{W}'} \mathbb{E}\left[p_t(w')\hat{\tilde{\ell}}_t^2(w')\right] + \gamma\sum_{t=1}^T \sum_{w'\in\mathcal{W}'} \mu(w')(\ell_t(w') + \langle\, \mathrm{s}(w'), c\,\rangle),$$

where we have used that $\mathbb{E}\left[\hat{\tilde{\ell}}_t(w)\right] = \ell_t(w) + \langle\, \mathrm{s}(w), c\,\rangle \ge 0$. Thus, we see that it remains to bound $\mathbb{E}\left[p_t(w')\hat{\tilde{\ell}}_t^2(w')\right]$, which is done in the following lemma.

Lemma A.2.2. *Let $\mathcal{W}'$, $\tilde{\ell}_t$, p_t be as in* LQDExp3. *Then, there exist a constant $C > 0$ such that the following equation holds:*

$$\sum_{w\in\mathcal{W}'} p(w)\mathbb{E}\left[\hat{\tilde{\ell}}^2(w)\,|\,p\right] \le (4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + y_{\lim}^2 + \|c\|_1).$$

Proof. By the tower rule, we have

$$\mathbb{E}\left[\sum_{w\in\mathcal{W}'} p_t(w)\hat{\tilde{\ell}}_t^2(w)\right] = \mathbb{E}\left[\sum_{w\in\mathcal{W}'} p_t(w)\mathbb{E}\left[\hat{\tilde{\ell}}_t^2(w)\,|\,p_t\right]\right]$$

Therefore, it suffices to bound

$$\sum_{w\in\mathcal{W}'} p_t(w)\mathbb{E}\left[\hat{\tilde{\ell}}_t^2(w)\,|\,p_t\right].$$

For simplifying the presentation, since t is fixed, from now on we will remove the subindex t from the quantities involved and write $\hat{\tilde{\ell}}$ instead of $\hat{\tilde{\ell}}_t$, p instead of p_t, etc.

The plan of the proof is as follows: We construct a deterministic upper bound $h(w)$ on $|\hat{\tilde{\ell}}(w)|$ and an upper bound B on $\sum_{w\in\mathcal{W}'} p(w)h(w)$. Then, we provide an upper bound B' on $\mathbb{E}\left[\hat{\tilde{\ell}}(w)|p\right]$ so that

$$\sum_{w\in\mathcal{W}'} p(w)\mathbb{E}\left[\hat{\tilde{\ell}}^2(w)\,\middle|\,p\right] \le \sum_{w\in\mathcal{W}'} p(w)h(w)\mathbb{E}\left[\hat{\tilde{\ell}}(w)\,\middle|\,p\right] \le B'\sum_{w\in\mathcal{W}'} p(w)h(w) \le BB'.$$

Before providing these bounds, let's review some basic relations. Remember that $W_\infty = \sup_{w\in\mathcal{W}}\|w\|_\infty$ and $X_1 = \sup_{x\in\mathcal{X}}\|x\|_1$. Further, note that for any $1 \le j, j' \le d$, we have

$$\mathbb{E}\left[\mathbb{1}_{\{j\in \mathrm{s}(W)\}}\,\middle|\,p\right] = \sum_{w\in\mathcal{W}':j\in\mathrm{s}(w)} p(w) = \sum_{w\in\mathcal{W}'} \mathbb{1}_{\{j\in\mathrm{s}(w)\}}p(w) = q(j), \tag{14}$$

$$\mathbb{E}\left[\mathbb{1}_{\{j,j'\in \mathrm{s}(W)\}}\,\middle|\,p\right] = \sum_{w\in\mathcal{W}':j,j'\in\mathrm{s}(w)} p(w) = \sum_{w\in\mathcal{W}'} \mathbb{1}_{\{j,j'\in\mathrm{s}(w)\}}p(w) = q(j,j'). \tag{15}$$

As to the upper bound $h(w)$ on $|\hat{\tilde{\ell}}(w)|$, we start with

$$|\hat{\tilde{\ell}}(w)| \le |w^\top \tilde{X} w| + 2|y|\,|w^\top \tilde{x}| + |y|^2 + \|c\|_1. \tag{16}$$

Now, $|y| \le y_{\lim}$ and

$$|w^\top \tilde{x}| \le W_\infty \sum_{j=1}^{d} \mathbb{1}_{\{j\in\mathrm{s}(w)\}}\frac{|x_j|}{q(j)} \doteq g(w,x),$$

$$|w^\top \tilde{X} w| \le W_\infty^2 \sum_{j,j'} \mathbb{1}_{\{j,j'\in\mathrm{s}(w)\}}\frac{|x_j x_{j'}|}{q(j,j')} \doteq G(w,x).$$

Hence,

$$|\hat{\tilde{\ell}}(w)| \le G(w,x) + 2y_{\lim}g(w,x) + y_{\lim}^2 + \|c\|_1 \doteq h(w)$$

which is indeed a deterministic upper bound on $|\hat{\tilde{\ell}}(w)|$. To bound $\sum_{w\in\mathcal{W}'} p(w)h(w)$, it remains to upper bound $\sum_{w\in\mathcal{W}'} p(w)g(w,x)$ and $\sum_{w\in\mathcal{W}'} p(w)G(w,x)$. To upper bound these, we move the sum over the weights w inside the other sums in the definitions of g and G to get:

$$\sum_{w\in\mathcal{W}'} p(w)g(w,x) = W_\infty \sum_{j=1}^{d}\frac{|x_j|}{q(j)}\sum_{w\in\mathcal{W}'} p(w)\mathbb{1}_{\{j\in\mathrm{s}(w)\}} = W_\infty X_1, \qquad \text{(by (14) and } \|x\|_1 \le X_1)$$

$$\begin{aligned}\sum_{w\in\mathcal{W}'} p(w)G(w,x) &= W_\infty^2 \sum_{j,j'}\frac{|x_j x_{j'}|}{q(j,j')}\sum_{w\in\mathcal{W}'} p(w)\mathbb{1}_{\{j,j'\in\mathrm{s}(w)\}} = W_\infty^2\sum_{j,j'}|x_j x_{j'}| \qquad \text{(by (15))}\\ &= W_\infty^2\|\,x\,\|_1^2 \le W_\infty^2 X_1^2.\end{aligned}$$

Hence,

$$\sum_{w\in\mathcal{W}'} p(w)h(w) \le W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + y_{\lim}^2 + \|c\|_1.$$

Let us now turn to bounding $\mathbb{E}\left[|\hat{\tilde{\ell}}(w)|\,\middle|\,p\right]$. From (16), it is clear that it suffices to upper bound $\mathbb{E}\left[|w^\top \tilde{X} w|\,|\,p\right]$ and $\mathbb{E}\left[|w^\top \tilde{x}|\,|\,p\right]$. From (14) and (15), the definitions of $\tilde{x}$ and $\tilde{X}$ and because by assumption $|||w|||_*|||x||| \le w_{\lim}x_{\lim} = y_{\lim}$, we obtain

$$\mathbb{E}\left[|w^\top \tilde{x}|\,\middle|\,p\right] = \sum_j |w_j x_j| \le y_{\lim} \quad \text{and}$$

$$\mathbb{E}\left[|w^\top \tilde{X} w|\,\middle|\,p\right] = \sum_{j,j'} |w_j w_{j'} x_j x_{j'}| = \Big(\sum_j |w_j x_j|\Big)^2 \le y_{\lim}^2.$$

Thus,

$$\mathbb{E}\left[|\hat{\tilde{\ell}}(w)| \,\middle|\, p\right] \le \mathbb{E}\left[|w^\top \tilde{X} w| + 2y_{\lim}|w^\top \tilde{x}| + y_{\lim}^2 + \|c\|_1 \,\middle|\, p\right] \le 4y_{\lim}^2 + \|c\|_1.$$

Putting together all the bounds, we get

$$\sum_{w\in\mathcal{W}'} p(w)\mathbb{E}\left[\hat{\tilde{\ell}}^2(w)\,|\,p\right] \le (4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + y_{\lim}^2 + \|c\|_1).$$

□

It remains to bound $\sum_{w'\in\mathcal{W}'} \mu(w')\left(\ell_t(w') + \langle\, \mathrm{s}(w'), c\,\rangle\right)$. Because of the bounds on weight vectors in $\mathcal{W}'$ and $((x_t, y_t))_{(1\le t\le T)}$, we know that $\ell_t(w') + \langle\, \mathrm{s}(w'), c\,\rangle \le 4y_{\lim}^2 + \|c\|_1$. Combining the inequalities obtained so far, we get

$$\begin{aligned}\mathbb{E}\left[\hat{L}_T - L_T(w)\right] \le \frac{\ln|\mathcal{W}'|}{\eta} + \eta(4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + y_{\lim}^2 + \|c\|_1) \\ + \gamma T(4y_{\lim}^2 + \|c\|_1).\end{aligned} \tag{17}$$

Thus, it remains to select η, γ such that the earlier imposed conditions, amongst them (13), hold and the above bound on the expected regret is minimized. To ensure $\eta\hat{\tilde{\ell}}_t(w) \ge -1$, we start with a lower bound on $\tilde{\ell}_t(w)$:

$$\begin{aligned}\tilde{\ell}_t(w) &= w^\top \tilde{X}_t\, w - 2\, w^\top \tilde{x}_t\, y_t + y_t^2 \\ &\ge w^\top \tilde{X}_t\, w - 2\, w^\top \tilde{x}_t\, y_t \\ &= \sum_{i,j=1}^{d} w_i w_j (\tilde{X}_t)_{i,j} - 2y_t \sum_{j=1}^{d} w_j \tilde{x}_{t,j} \\ &\ge -\frac{\sum_{i,j} \mathbb{1}_{\{i\in \mathrm{s}(w)\}}\mathbb{1}_{\{j\in \mathrm{s}(w)\}}|x_{t,i}x_{t,j}w_i w_j|}{\gamma} - 2y_{\lim}\frac{\sum_i \mathbb{1}_{\{i\in\mathrm{s}(w)\}}|x_{t,i}w_i|}{\gamma} \\ &\ge -\frac{\|w\|_*^2\|x_t\|^2}{\gamma} - 2y_{\lim}\frac{\|w\|_*\|x_t\|}{\gamma} \\ &\ge -\frac{3y_{\lim}^2}{\gamma}\ .\end{aligned}$$

Thus, as long as $3\eta y_{\lim}^2 \le \gamma$, it follows that (13) holds. To minimize (17), we choose

$$\gamma = 3\eta y_{\lim}^2 \tag{18}$$

to get

$$\mathbb{E}\left[\hat{L}_T - L_T(w)\right] \le \frac{\ln|\mathcal{W}'|}{\eta} + \eta(4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + 4y_{\lim}^2 + \|c\|_1)\,.$$

Using $\eta = \sqrt{\frac{\ln|\mathcal{W}'|}{(4y_{\lim}^2+\|c\|_1)(W_\infty^2X_1^2+2y_{\lim}W_\infty X_1+4y_{\lim}^2+\|c\|_1)}}$, we get

$$\mathbb{E}\left[\hat{L}_T\right] - L_T(w) \le 2\sqrt{T\,(4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + 4y_{\lim}^2 + \|c\|_1)\ln|\mathcal{W}'|}\,.$$

Noting that here $w \in \mathcal{W}'$ was arbitrary, together with the regret decomposition (10), the bound (11) on the regret arising from discretization the bound (12) on $\ln|\mathcal{W}'|$ and that $\ln|\mathcal{G}| \le d\ln 2$ give

$$\begin{aligned}\mathbb{E}\left[R_T\right] &\le 2\sqrt{T\,(4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + 4y_{\lim}^2 + \|c\|_1)\ln|\mathcal{W}'|} + 4\,T\,y_{\lim}\alpha \\ &\le 2\sqrt{Td\,(4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + 4y_{\lim}^2 + \|c\|_1)\ln(2 + 4E_{\mathcal{G}}y_{\lim}/\alpha)} + 4\,T\,y_{\lim}\alpha\,.\end{aligned}$$

Choosing $\alpha = y_{\lim}T^{-1/2}$, we get the bound

$$\mathbb{E}\left[R_T\right] \le C\sqrt{Td\,(4y_{\lim}^2 + \|c\|_1)(W_\infty^2 X_1^2 + 2y_{\lim}W_\infty X_1 + 4y_{\lim}^2 + \|c\|_1)\ln(E_{\mathcal{G}}T)}\,. \tag{19}$$

for some constant $C > 0$. □

A.2.2 Lower Bound

Theorem 2.5. *Let $d > 0$, and consider the online free label probing problem with linear predictors, where $\mathcal{W} = \{w \in \mathbb{R}^d \,|\, \|w\|_1 \leq w_{\lim}\}$ and $\mathcal{X} = \{x \in \mathbb{R}^d \,|\, \|x\|_\infty \leq 1\}$. Assume, for all $t \geq 1$, that the loss functions are of the form $\ell_t(w) = (w^\top x_t - y_t)^2 + \langle \mathrm{s}(w), c \rangle$, where $|y_t| \leq 1$ and $c = 1/2 \times \mathbf{1} \in \mathbb{R}^d$. Then, for any prediction algorithm and for any $T \geq \frac{4d}{8\ln(4/3)}$, there exists a sequence $((x_t, y_t))_{1 \leq t \leq T} \in (\mathcal{X} \times [-1, 1])^T$ such that the regret of the algorithm can be bounded from below as*

$$\mathbb{E}[R_T] \geq \frac{\sqrt{2} - 1}{\sqrt{32 \ln(4/3)}} \sqrt{Td}\,.$$

Proof. The idea of the proof is similar to Mannor and Shamir (2011, Theorem 4). We will solve the problem of Multi-Armed Bandits with d arms using an algorithm that can solve free-label probing with examples having d features. We will use the lower bound proved in Cesa-Bianchi and Lugosi (2006, Theorem 6.11) for Multi-Armed Bandit game. They showed a method of choosing the losses and proved that there exist a universal constant C_{MAB} such that no algorithm can achieve a better regret than $C_{MAB}\sqrt{Td}$ in T rounds using d arms. In their method adversary chooses one of the arms beforehand and assign a random Bernoulli loss with parameter $1/2 + \varepsilon$ to that arm and a random Bernoulli loss with parameter $1/2$ to all other arms at each round. Then they proved that by choosing $\varepsilon = \sqrt{(1/(8\ln(4/3))d/T}$, no algorithm can achieve better expected regret bound than $C_{MAB}\sqrt{Td}$ in T rounds. Note that they use the fact that losses are in range $[0, 1]$. Without loss of generality we can add $1/2$ to all the losses and assume that the losses are now in range $[1/2, 1 + 1/2]$ and their result still hold.

Now we explain how we can solve that problem using an algorithm that solves free label probing game. More formally we will use the following lemma.

Lemma A.2.3. *Give any learner $\mathcal{A}$ for an online free-label probing game there exist a learner $\mathcal{A}'$ for Multi-Armed Bandit problem with the adversaries proposed at Cesa-Bianchi and Lugosi (2006, Theorem 6.11) and an adversary for online free-label probing game such that*

$$\mathbb{E}\left[R_{\mathcal{A}'}(T, MAB)\right] - 2d\sqrt{(1/(8\ln(4/3))} \leq \mathbb{E}\left[R_{\mathcal{A}}(T, OFLP)\right]\,,$$

holds where $R_{\mathcal{A}'}(T, MAB)$ is the regret of the learner $\mathcal{A}'$ in the Multi-Armed Bandit problem with the defined adversary and $R_{\mathcal{A}}(T, OFLP)$ is the regret of the learner $\mathcal{A}$ in the online free-label probing game.

Proof. We define the adversary in the online free label probing game. The adversary chooses $y_t = 1$ for all the rounds. Note the the challenge is finding a weight vector to predict the label and not only predicting the label. Consider the weight vector e_i which is a zero weight vector with a single one in its ith element for all $1 \leq i \leq d$. The adversary then chooses one of the components v, in advance and sets $x_{t,i}$ to be a Bernoulli random variable with parameter one for every $i \neq v$ and sets $x_{t,v}$ to be a Bernoulli random variable with parameter $1/2 + \varepsilon$. Note that this component v is the same arm as the adversary in multi-arm bandit chooses. Now we know that for each e_i the loss will be the cost of observing ith feature which is $1/2$ and a prediction error which is a Bernoulli random variable based on the assignments to the features. So you can easily see a correspondence between e_i and ith arm in multi-armed bandit problem with the adversary defined in Cesa-Bianchi and Lugosi (2006, Theorem 6.11).

Let $R_{\mathcal{A}}(T, OFLP)$ denote the regret of the learner $\mathcal{A}$ in this online free-label probing. We know that if we make the set of competitors smaller, the regret can not be increased. Note that we does not change the set of actions that algorithm $\mathcal{A}$ can take. Let $R^*_{\mathcal{A}}(T, OFLP)$ denote the regret of the learner $\mathcal{A}$ in this online free-label probing when it competes only against e_i weight vectors for all $1 \leq i \leq d$. Since we make the set of competitors smaller we have

$$R^*_{\mathcal{A}}(T, OFLP) \leq R_{\mathcal{A}}(T, OFLP)\,. \tag{20}$$

Now consider the learner $\mathcal{A}$ that solves this online free-label probing game. We will construct another algorithm $\mathcal{A}'$ such that solves the multi-armed bandits problem. Let I_t denote the chosen arm by $\mathcal{A}'$ and $\ell_{t,i}$ denote the loss of arm i at round $t \geq 1$. Here are the different situations.

When $\mathcal{A}$ chooses $w_t = \mathbf{0} \in \mathbb{R}^d$ at round t, $\mathcal{A}'$ chooses one of the arms randomly in multi-armed bandit problem. By this choice, $\mathcal{A}$ does not observe any feature and predict zero for the label. Here is the expected regret at these types of rounds for $\mathcal{A}$.

$$\mathbb{E}\left[\ell_t(\mathbf{0}) - \ell_t(e_v)\right] = 1 - (1/2 + \mathbb{E}\left[(e_v^\top x_t - y_t)^2\right] = 1 - (1/2 + 1/2 - \varepsilon) = \varepsilon\,.$$

On the other hand, the expected regret of $\mathcal{A}'$ in the game of multi-armed bandits at each round is bounded by ε. By this we know that in the rounds that $\mathcal{A}$ chooses $w_t = \mathbf{0} \in \mathbb{R}^d$ we get

$$\mathbb{E}\left[\ell_{t,I_t} - \ell_{t,v}\right] = \mathbb{E}\left[\ell_t(e_{I_t}) - \ell_t(e_v)\right] \leq \varepsilon = \mathbb{E}\left[\ell_t(w_t) - \ell_t(e_t)\right] \qquad t \geq 1\,, \tag{21}$$

which means the regret of $\mathcal{A}'$ is not going to be increased more that regret of $\mathcal{A}$ in such rounds.

When $\mathcal{A}$ chooses a weight vector $w_t \neq \mathbf{0}$, $\mathcal{A}'$ chooses all arms i in the bandit game whose corresponding ith component of w_t is not zero in the free-label probing game in the consecutive rounds and after finding all required component values of x, it gives it to $\mathcal{A}$ as the feedback for calculating the loss. Note that the chosen weight vector by $\mathcal{A}$ requires either one feature or more than one feature. As a result $\mathcal{A}'$ plays the bandit games for T' rounds while $\mathcal{A}$ plays the online free-label probing game for T rounds. If it w_t needs only one feature due to the way the choice of $x_{t,i}$, the minimizer of expected loss is exactly e_i. Because if the ith component of w_t was α instead of one we get

$$\mathbb{E}\left[(w_t^\top x_t - y_t)^2\right] = \mathbb{E}\left[(\alpha x_{t,i} - 1)^2\right] = \mathbb{P}\left[x_{t,i} = 0\right] \times 1 + \mathbb{P}\left[x_{t,i} = 1\right] \times (1 - \alpha)^2\,.$$

which achieves its minimum for $\alpha = 1$. So we get Eq.(21) for these types of rounds as well. Now if w_t has more than one non-zero components as we said $\mathcal{A}'$ plays more rounds. At these extra rounds the expected regret of $\mathcal{A}'$ will be increased by at most ε. However $\mathcal{A}$ is also paying for those extra features that it needed. Since the cost of each feature is $1/2$ as well assuming that $\varepsilon \leq 1/2$, we can conclude that the regret of $\mathcal{A}$ for all these extra rounds is still less than or equal the regret of $\mathcal{A}$ on the rounds that it chooses w_t. Let T' denote the random number of rounds that $\mathcal{A}'$ is playing the bandits game. We know that this number is bounded by dT since at each round $\mathcal{A}$ can choose at most all the features. Putting the above results together with Eq.(21), we get

$$\mathbb{E}\left[R_{\mathcal{A}'}(T', MAB)\right] \leq \mathbb{E}\left[R^*_{\mathcal{A}}(T, OFLP)\right]\,.$$

Because the expected regret is increasing in the number of rounds we can use $\mathbb{E}\left[R_{\mathcal{A}'}(T, MAB)\right] \leq \mathbb{E}\left[R_{\mathcal{A}'}(T', MAB)\right]$ and also Eq.(20) to get

$$\mathbb{E}\left[R_{\mathcal{A}'}(T, MAB)\right] \leq \mathbb{E}\left[R_{\mathcal{A}}(T, OFLP)\right]\,.$$

Using the value of ε that Cesa-Bianchi and Lugosi (2006, Theorem 6.11) uses we get the lemma statement. Also $T >= \frac{4d}{8\ln(4/3)}$ in the lemma statement guarantees that $\varepsilon \leq 1/2$ which was needed in the middle of the proof. □

Using this lemma and also knowing that

$$\mathbb{E}\left[R_{\mathcal{A}'}(T, MAB)\right] \geq \sqrt{dT}\frac{\sqrt{2} - 1}{\sqrt{32\ln(4/3)}}$$

based on the result of Cesa-Bianchi and Lugosi (2006, Theorem 6.11), we can derive

$$\frac{\sqrt{2} - 1}{\sqrt{32\ln(4/3)}}\sqrt{dT} \leq \mathbb{E}\left[R_{\mathcal{A}}(T, OFLP)\right]\,.$$

□

A.3 Non-Free-Label Probing

A.3.1 Revealing Action Algorithm for Non-Free-Label Probing

Algorithm 3 Revealing action algorithm for non-free-label online probing

Parameters: Real numbers $0 \leq \eta, \gamma \leq 1$, Set of experts $\mathcal{F}$.
Initialization: $u_1(f) = 1$ ($f \in \mathcal{F}$).
for $t = 1$ **to** T **do**
  Draw $F_t \in \mathcal{F}$ from the probability mass function

$$p_t(f) = \frac{u_t(f)}{\sum_{f \in \mathcal{F}} u_t(f)}, \qquad f \in \mathcal{F}.$$

  Draw a Bernoulli random variable Z_t such that $\mathbb{P}[Z_t = 1] = \gamma$.
  if $Z_t = 0$ **then**
    $S_t = (\mathrm{s}(F_t), 0)$ (*i.e.,* $s_{t,d+1} = 0$).
    Obtain the features values, $(x_{t,i})_{i \in \mathrm{s}(F_t)}$.
    Predict $\hat{y}_t = F_t(x_t)$.
  else
    $S_t = \mathbf{1} \in \mathbb{R}^{d+1}$ (*i.e.,* all $d+1$ components are one).
    Observe all the features of x_t.
    Predict $\hat{y}_t = F_t(x_t)$.
    Receive the true label y_t.
  end if
  for each $f \in \mathcal{F}$ **do**
    $\tilde{\ell}_t(f) = \mathbb{1}_{\{Z_t = 1\}} \frac{\langle \mathrm{s}(f), c_{1:d} \rangle + \ell_t(\hat{y}_t)}{\gamma}$.
    $u_{t+1}(f) = u_t(f) \exp(-\eta \tilde{\ell}_t(f))$.
  end for
end for

A.3.2 Upper Bound

Lemma 3.1. *Given any non-free-label online probing with finitely many experts, Algorithm 3 with appropriately set parameters achieves*

$$\mathbb{E}[R_T] \leq C \max\left(T^{2/3}(\ell_{\max}^2 \|c\|_1 \ln|\mathcal{F}|)^{1/3}, \ell_{\max}\sqrt{T \ln|\mathcal{F}|}\right)$$

for some constant $C > 0$.

Proof. The regret of the algorithm is decomposed into two additive terms: (i) The extra loss suffered in exploration rounds. The cumulative expectation of this extra loss can be upper bounded by $T\gamma\|c\|_1$. (ii)The regret of the algorithm compared to each expert, excluding rounds that request the label and extra features. To upper bound this term, we follow the classical "exponential weights" proof (see *e.g.,* Cesa-Bianchi et al. (2006)).

First we make the trivial observation that for every time step t and $f \in \mathcal{F}$, $\mathbb{E}[\tilde{\ell}_t(f)] = \langle \mathrm{s}(f), c_{1:d} \rangle + \ell_t(f(s \odot x_t))$. That is, $\tilde{\ell}_t(f)$ is an unbiased estimate of the true loss of function f. Let $U_t = \sum_{f \in \mathcal{F}} u_t(f)$. Now we continue with lower and upper bounding the term U_T:

$$U_T \geq \sum_{f \in \mathcal{F}} u_T(f) \geq u_T(f^*) = \exp\left(-\eta \sum_{t=1}^{T} \tilde{\ell}_t(f^*)\right),$$

where f^* is and arbitrary expert in $\mathcal{F}$. For the upper bound we write

$$\begin{aligned}
\frac{U_t}{U_{t-1}} &= \sum_{f\in\mathcal{F}} \frac{u_{t-1}(f)\exp(-\eta\tilde{\ell}_t(f))}{U_{t-1}} \\
&= \sum_{f\in\mathcal{F}} p_t(f)(1-\eta\tilde{\ell}_t(f)+\eta^2\tilde{\ell}_t^2(f)) \qquad (22)\\
&= 1-\eta\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t(f)+\eta^2\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t^2(f) \\
&\le \exp\left(-\eta\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t(f)+\eta^2\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t^2(f)\right), \qquad (23)
\end{aligned}$$

where in (22) we used that $u_{t-1}(f)/U_{t-1} = p_t(f)$ and the inequality $e^x \le 1+x+x^2$ if $x \le 1$, and in (23) we used that $e^x \ge 1+x$. Multiplying the above inequality for $t = 1,\dots,T$ and also U_1 we get

$$U_T \le |\mathcal{F}|\exp\left(-\eta\sum_{t=1}^{T}\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t(f)+\eta^2\sum_{t=1}^{T}\sum_{(s,f(.))\in\mathcal{F}} p_t(f)\tilde{\ell}_t^2(f)\right).$$

We now merge the lower and upper bounds and take logarithm of both sides:

$$-\eta\sum_{t=1}^{T}\tilde{\ell}_t(f^*)-\ln|\mathcal{F}| \le -\eta\sum_{t=1}^{T}\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t(f)+\eta^2\sum_{t=1}^{T}\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t^2(f).$$

Rearranging gives

$$\sum_{t=1}^{T}\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t(f)-\sum_{t=1}^{T}\tilde{\ell}_t(f^*) \le \eta\sum_{t=1}^{T}\sum_{f\in\mathcal{F}} p_t(f)\tilde{\ell}_t^2(f)+\frac{\ln|\mathcal{F}|}{\eta}.$$

After taking expectation of both sides, the first term on the left hand side is the expected cumulative loss of the algorithm excluding the extra loss suffered in exploration rounds, while the second term is the expected cumulative loss of the any arbitrary expert f. The first term on the right hand side can be upper bounded as

$$\begin{aligned}
\eta\sum_{t=1}^{T}\sum_{f\in\mathcal{F}}\mathbb{E}[p_t(f)\tilde{\ell}_t^2(f)] &\le \eta\sum_{t=1}^{T}\sum_{f\in\mathcal{F}}\mathbb{E}[p_t(f)\tilde{\ell}_t(f)]\frac{\ell_{\max}}{\gamma} \\
&\le \frac{\eta\ell_{\max}^2 T}{\gamma},
\end{aligned}$$

where $\ell_{\max}$ is the maximum loss an action can suffer, ignoring the label cost c_{d+1}.

Adding up the two terms of the expected regret, we get

$$\mathbb{E}[R_T] \le T\gamma\|c\|_1+\frac{\eta\ell_{\max}^2 T}{\gamma}+\frac{\ln|\mathcal{F}|}{\eta}.$$

For setting the parameters optimally, we consider two cases.

(1) If $\|c\|_1 \ge \sqrt{\frac{\ln|\mathcal{F}|}{2T}}\ell_{\max}$, then we set

$$\eta = (\ln|\mathcal{F}|)^{2/3}T^{-2/3}(4\ell_{\max}^2\|c\|_1)^{-1/3} \qquad \gamma = \sqrt{\frac{\eta\ell_{\max}^2}{\|c\|_1}}$$

to get

$$\mathbb{E}[R_T] \le C_1 T^{2/3}(\ell_{\max}^2\|c\|_1\ln|\mathcal{F}|)^{1/3}$$

for some constant $C_1 > 0$. The condition on $\|c\|_1$ is needed for γ to be a probability. On the other hand,

(2) if $\|c\|_1 < \sqrt{\frac{\ln|\mathcal{F}|}{2T}}\ell_{\max}$, then we set

$$\eta = \sqrt{\frac{\ln|\mathcal{F}|}{T\ell_{\max}^2}} \qquad \gamma = 1$$

to get

$$\mathbb{E}[R_T] \leq C_2 \ell_{\max} \sqrt{T \ln|\mathcal{F}|}$$

for some constant C_2.

Combining the two bounds gives the result of the lemma. □

A.3.3 Lower Bound

In this section we prove the lower bound on the regret of the non-free-label probing game, stated in Section 3. The proof follows a standard lower bounding technique using a randomized construction for the loss functions. As such, we omit the proofs of two lemmas used in the derivation; the interested reader is referred to (Bartók, 2012) for these proofs.

Theorem 3.2. *There exists a constant C such that, for any non-free-label probing with linear predictors, quadratic loss, and $c_j > (1/d)\sum_{i=1}^{d} c_i - 1/2d$ for every $j = 1, \ldots, d$, the expected regret of any algorithm can be lower bounded by*

$$\mathbb{E}[R_T] \geq C(c_{d+1}d)^{1/3}T^{2/3}\,.$$

Proof. We construct a set of opponent strategies and show that the expected regret of any algorithm is high against at least one of them. The features $x_{t,i}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, d$ are generated by the iid random variables $X_{t,i}$ whose distribution is Bernoulli with parameter 0.5. Let $Z_t \in \{1, \ldots, d\}$ be random variables whose distribution will be specified later. The labels y_t are generated by the random variable defined as $Y_t = X_{t,Z_t}$.

To construct the distribution of Z_t we introduce the following notation. For every $i = 1, \ldots, d$, let

$$a_i = \frac{1}{d} + 2c_i - \frac{2}{d}\sum_{j=1}^{d} c_j\,.$$

The assumptions on c ensures that $a_i > 0$ for every $i = 1, \ldots, d$. For opponent strategy k, let the distribution of Z_t defined as

$$\mathbb{P}_k\left[Z_t = i\right] = \begin{cases} a_i - \varepsilon, & i \neq k; \\ a_i + (d-1)\varepsilon, & \text{i=k}\,, \end{cases}$$

with some $\varepsilon > 0$ to be defined later.

Lemma A.3.1. (Bartók 2012, Lemma 25) *Let e_k denote the k^{th} basis vector of dimension d. Against opponent strategy k, the instantaneous expected regret for any action such that $(s, s_\ell) \neq (e_k, 0)$ is at least $\frac{d\varepsilon}{2}$.*

For $i = 1, \ldots, d$, let N_i denote the number of times the player's action is (e_i, w, s_{d+1}). Similarly, let N_L denote the number of times the player requests the label. Now it is easy to see that the expected regret under opponent strategy k can be lower bounded by

$$\mathbb{E}_k[R_T] \geq (T - \mathbb{E}_k[N_k])\,\frac{d\varepsilon}{2} + c_{d+1}\mathbb{E}_k[N_L]\,.$$

The rest of the proof is devoted to show that for any algorithm, the average of the above value, $1/d\sum_{i=1}^{d}\mathbb{E}_i[R_T]$ can be lower bounded. We only show this for deterministic algorithms. The statement follows for randomizing algorithms with the help of a simple argument, see *e.g.,* Cesa-Bianchi and Lugosi (2006, Theorem 6.11).

A deterministic algorithm is defined as a sequence of functions $A_t(\cdot)$, where the argument of A_t is a sequence of observations up to time step $t-1$ and the value is the action taken at time step t. We

denote the observation at time step t by $h_t \in \{0,1,*\}^{d+1}$, where $h_{t,i} = x_{t,i}$ if $s_{t,i} = 1$ and $h_{t,i} = *$ if $s_{t,i} = 0$ for all $1 \leq i \leq d$. Similarly, $h_{t,d+1} = y_t$ if $s_{t,d+1} = 1$ and $h_{t,d+1} = *$ if $s_{t,d+1} = 0$. That is, $*$ is the symbol for not observing a feature or the label. The next lemma, which is the key lemma of the proof, shows that the expected value of N_i does not change too much if we change the opponent strategy.

Lemma A.3.2. (Bartók 2012, Lemma 26) *There exists a constant C_1 such that for any $i, j \in \{1, \ldots, d\}$,*

$$\mathbb{E}_i[N_i] - \mathbb{E}_j[N_i] \leq C_1 T\varepsilon\sqrt{d\mathbb{E}_j[N_L]}\,.$$

Now we are equipped to lower bound the expected regret. Let

$$j = \operatorname{argmin}_{k\in\{1,\ldots,d\}} \mathbb{E}_k[N_L].$$

By Lemma A.3.2,

$$\begin{aligned}\mathbb{E}_i[R_T] &\geq (T - \mathbb{E}_i[N_i])\,\frac{d\varepsilon}{2} + c_{d+1}\mathbb{E}_i[N_L] \\ &\geq \left(T - \mathbb{E}_j[N_i] - C_1 T\varepsilon\sqrt{d\mathbb{E}_j[N_L]}\right)\frac{d\varepsilon}{2} + c_{d+1}\mathbb{E}_j[N_L]\end{aligned}$$

Denoting $\sqrt{\mathbb{E}_j[N_L]}$ by ν we have

$$\begin{aligned}\frac{1}{d}\sum_{i=1}^{d}\mathbb{E}_i[R_T] &\geq \left(T - \frac{1}{d}\sum_{i=1}^{d}\mathbb{E}_j[N_i] - C_1 T\varepsilon\sqrt{d}\nu\right)\frac{d\varepsilon}{2} + c_{d+1}\nu^2 \\ &\geq \left(T - \frac{T}{d} - C_1 T\varepsilon\sqrt{d}\nu\right)\frac{d\varepsilon}{2} + c_{d+1}\nu^2\end{aligned}$$

What is left is to optimize this bound in terms of ν and ε. Since ν is the property of the algorithm, we have to minimize the expression in ν, with ε as a parameter. After simple algebra we get

$$\nu_{opt} = \frac{C_1 T\varepsilon^2 d^{3/2}}{4c_{d+1}}\,.$$

Substituting it back results in

$$\frac{1}{d}\sum_{i=1}^{d}\mathbb{E}_i[R_T] \geq (d-1)\frac{T\varepsilon}{2} - \frac{C_1^2 T^2\varepsilon^4 d^3}{16c_{d+1}}$$

Now we set

$$\varepsilon = \left(\frac{2}{C_1^2}\right)^{1/3}(c_{d+1})^{1/3}\,d^{-2/3}T^{-1/3}$$

to get

$$\mathbb{E}[R_T] \geq C_3\,(c_{d+1})^{1/3}\,d^{1/3}T^{2/3}$$

whenever $d \geq 2$. □