Robust Selective Sampling from Single and Multiple Teachers

Ofer Dekel
Microsoft Research
oferd@microsoft.com

Claudio Gentile
DICOM, Università dell'Insubria
claudio.gentile@uninsubria.it

Karthik Sridharan
TTI-Chicago
karthik@ttic.edu

Abstract

We present a new online learning algorithm in the selective sampling framework, where labels must be actively queried before they are revealed. We prove bounds on the regret of our algorithm and on the number of labels it queries when faced with an adaptive adversarial strategy of generating the instances. Our bounds both generalize and strictly improve over previous bounds in similar settings. Using a simple online-to-batch conversion technique, our selective sampling algorithm can be converted into a statistical (pool-based) active learning algorithm. We extend our algorithm and analysis to the multiple-teacher setting, where the algorithm can choose which subset of teachers to query for each label.

1 Introduction

A *selective sampling* algorithm (Cohn et al., 1990; Freund et al., 1997) is an online learning algorithm that actively decides which labels to query. More precisely, learning takes place in a sequence of rounds. On round t, the online learner receives an instance $\mathbf{x}_t \in \mathbb{R}^d$ and predicts a binary label $\hat{y}_t \in \{-1, +1\}$. Then, the learner decides whether or not to *query* the true label y_t associated with $\mathbf{x}_t$. If the label is queried, the learner incurs a unit cost and uses the label to improve his future predictions. If the label is not queried, the learner never knows whether his prediction was correct. Nevertheless, the accuracy of the learner is evaluated on both queried and unqueried instances. We say that a selective sampling algorithm is *robust* if it works even when the instance sequence $\mathbf{x}_1, \mathbf{x}_2, ...$ is generated by an *adaptive adversary*. Robustness thereby implies a high level of adaptation to the learning environment.

Inspired by known online ridge regression algorithms (e.g., (Hoerl & Kennard, 1970; Lai & Wei, 1982; Vovk, 2001; Azoury & Warmuth, 2001; Cesa-Bianchi et al., 2003; Cesa-Bianchi et al., 2005; Li et al., 2008; Strehl & Littman, 2008; Cavallanti et al., 2009; Cesa-Bianchi et al., 2009)), we begin by presenting a new robust selective sampling algorithm within the label-noise setting considered in (Cavallanti et al., 2009; Cesa-Bianchi et al., 2009; Strehl & Littman, 2008). We measure the predictive accuracy of our learner using the game-theoretic notion of *regret* (formally defined below) and prove formal bounds on this quantity. We also prove bounds on the number of queries issued by the learner. Our bounds are strictly better than the best available bounds in the robust selective sampling setting, and can be shown to be optimal with respect to certain parameters. A detailed comparison of our results with the results of the predominant previous papers on this topic (Cesa-Bianchi et al., 2006; Strehl & Littman, 2008; Cesa-Bianchi et al., 2009) is given in Section 2.5, after our results are presented.

Selective sampling can be viewed as an online-learning variant of active learning. The literature on active learning is vast, and we can hardly do it justice here. Recent papers on active learning include (Balcan et al., 2006; Balcan et al., 2007; Castro & Nowak, 2008; Dasgupta et al., 2008; Dasgupta et al., 2005; Hanneke, 2007; Hanneke, 2009). All of these papers consider the case where instances are drawn i.i.d. from a fixed distribution (either known or unknown). As a by-product of our adversarial analysis, we also obtain a tight regret bound in the case where the instances $\mathbf{x}_t$ are generated i.i.d. according to a fixed and unknown distribution. Moreover, using a simple online-to-batch conversion technique, our online learner becomes a randomized statistical pool-based active-learning algorithm, with a high-probability risk bound.

In the second part of this paper, we extend our algorithm and analysis to the case where the learner has access to multiple teachers, each one with a different area of expertise and a different level of overall competence. In other words, the learner is free to query any subset of teachers and each teacher is capable of providing accurate labels only within some subset of the instance space. The learner is not given any information on the expertise region of each teacher, and must infer this information directly from the labels. Roughly speaking, the goal of the learner is to perform as well as each teacher in his respective area of expertise. We first present an online learner that either queries all of the teachers or does not query any teacher. We then enhance this learner to query only those teachers it believes to be experts on $\mathbf{x}_t$.

The general aim of this line of research is to provide algorithms of practical utility for which we can also prove formal performance guarantees. The motivation behind selective sampling is the same as the motivation behind any active learning algorithm: human-generated labels are expensive and therefore we only want labels that improve our ability to make accurate predictions. Our work within the multiple teacher setting is motivated by an Internet search company that uses online learning techniques to determine the results of its search engine. More concretely, the instance $\mathbf{x}_t$ represents the pairing of a search-engine query with a candidate web page; the goal of the online learner is to determine whether or not this pair constitutes a good match. The company employs human teachers to provide the correct answer for any instance. Clearly, there is no way to manually label the millions of daily search engine queries, and some intelligent mechanism of choosing which instances to label is required. Each teacher provides labels of different quality in different regions of the instance space. To make accurate predictions, the learner must figure out which teachers to trust for each instance.

A learning framework sharing similar motivations to ours is the proactive learning setting (Donmez & Carbonell, 2008; Yang & Carbonell, 2009a), where the learner has access to teachers of different quality, with associated costs per label. Yang and Carbonell (2009b) presents a theoretical analysis of proactive learning, however, this analysis relies on the strong assumption that each teacher gives the correct label most of the time. We make no such assumption in our analysis. Moreover, our setting supports the realistic scenario where each teacher has a very narrow area of expertise and gives useless labels outside of this area.

2 The Single Teacher Case

In this section, we focus on the standard online selective sampling setting, where the learner has to learn an accurate predictor while determining whether or not to query the label of each instance it observes. In this setting, the learner has no control over where the label comes from.

2.1 Preliminaries and Notation

As mentioned above, on round t of the online learning process, the learner receives input $\mathbf{x}_t \in \mathbb{R}^d$, predicts $\hat{y}_t \in \{-1,+1\}$, and chooses whether or not to query the correct label $y_t \in \{-1,+1\}$. We set $Z_t = 1$ if a query is issued and $Z_t = 0$ otherwise. The only assumption we make on the process that generates $\mathbf{x}_t$ is that $\|\mathbf{x}_t\| \le 1$; for all we know instances may be generated by an *adaptive* adversary. Note that most of the previous work on this topic makes stronger assumptions on the process that generates $\mathbf{x}_t$, leading to a less general setting. As for the labels, we adopt the standard stochastic linear noise[1] model for this problem (Cesa-Bianchi et al., 2003; Cavallanti et al., 2009; Cesa-Bianchi et al., 2009; Strehl & Littman, 2008) and assume that each $y_t \in \{-1,+1\}$ is sampled according to the law $P(y_t = 1\,|\,\mathbf{x}_t\,) = (1 + \mathbf{u}^\top \mathbf{x}_t)/2$, where $\mathbf{u} \in \mathbb{R}^d$ is a fixed but unknown vector with $\|\mathbf{u}\| \le 1$. Note that under this setup, $\mathbb{E}\,[y_t\,|\mathbf{x}_t\,] = \mathbf{u}^\top \mathbf{x}_t$, and we denote the latter by Δ_t. The learner uses hyperplanes to predict the label on each round. That is, on round t the learner predicts $\hat{y}_t = \mathrm{sign}(\hat{\Delta}_t)$ where $\hat{\Delta}_t = \mathbf{w}_{t-1}{}^\top \mathbf{x}_t$. Let P_t denote the conditional probability $\mathbb{P}(\,\cdot\,|\,\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}, \mathbf{x}_t, y_1, \ldots, y_{t-1})$. We evaluate the accuracy of the learner's predictions using its cumulative *regret*, defined as

$$R_T \;=\; \textstyle\sum_{t=1}^T \Big(P_t(y_t\hat{\Delta}_t < 0) - P_t(y_t\Delta_t < 0) \Big) \quad .$$

Additionally, we are interested in the number of queries issued by the learner $N_T = \sum_{t=1}^T Z_t$. Our goal is to simultaneously bound the cumulative regret R_T and the number of queries N_T with high probability over the random draw of labels.

2.2 Algorithm

The single teacher algorithm is a margin-based selective sampling procedure. The algorithm "Selective Sampler" (Algorithm 1) depends on a confidence parameter $\delta \in (0,1]$. As in known online ridge-regression-like algorithms (e.g., (Hoerl & Kennard, 1970; Vovk, 2001; Azoury & Warmuth, 2001; Cesa-Bianchi et al., 2003; Cesa-Bianchi et al., 2005; Li et al., 2008; Strehl & Littman, 2008; Cavallanti et al., 2009; Cesa-Bianchi et al., 2009)), our algorithm maintains a weight vector $\mathbf{w}_t$ (initialized as $\mathbf{w}_0 = \mathbf{0}$) and a data correlation matrix A_t (initialized as $A_0 = I$). After receiving $\mathbf{x}_t$ and predicting $\hat{y}_t = \mathrm{sign}(\hat{\Delta}_t)$, the algorithm computes an adaptive data-dependent threshold θ_t, defined as

$$\theta_t^2 \;=\; \mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t \Big(1 + 4 \textstyle\sum_{i=1}^{t-1} Z_i r_i + 36 \log(t/\delta) \Big) \;\; ,$$

[1] The noise model we are adopting here not only can be made more general (i.e., highly nonlinear) by the use of kernel functions (see Section 2.2), but has also undergone a rather thorough experimental validation on real-world data (Cavallanti et al., 2009; Cesa-Bianchi et al., 2009).

where $r_i = \mathbf{x}_i^\top A_i^{-1} \mathbf{x}_i$. The definition of θ_t derives from our analysis, and can be interpreted as the algorithm's uncertainty in its own predictions. More precisely, the learner believes that $|\hat{\Delta}_t - \Delta_t| \le \theta_t$. A query is issued only if[2] $|\hat{\Delta}_t| \le \theta_t$, or in other words, when the algorithm is unsure about the sign of Δ_t.

It is important to stress how θ_t depends on the three terms $\sum_{i=1}^{t-1} Z_i r_i$, $\log(t/\delta)$, and $\mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t$. We can prove that $\sum_{i=1}^{t} Z_i r_i$ grows only logarithmically with the number of queries N_t, and obviously $\log(t/\delta)$ grows logarithmically with t. The behavior of the third term, $\mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t$, depends on the relationship between the current instance $\mathbf{x}_t$ and the previous instances. If $\mathbf{x}_t$ lies along the directions spanned by the previous instances then we can show that $\mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t$ tends to shrink as $1/N_t$. As a result, the threshold θ_t is on the order of $\log(t/\delta)/N_t$ and the algorithm keeps querying labels at a slow logarithmic rate. On the other hand, if the adversary chooses $\mathbf{x}_t$ to lie outside of the subspace spanned by the previous examples, then the term $\mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t$ causes θ_t to be large, and the algorithm is more likely to issue a query. Overall, to ensure a small uncertainty threshold θ_t over all input directions determined by the adversarial choice of $\mathbf{x}_t$, the algorithm must query on the order of $\log(t)$ labels for each such direction in the instance space.

If the label is not queried, ($Z_t = 0$) then the algorithm does not update its internal state. If the label is queried ($Z_t = 1$), then the algorithm computes the intermediate vector $\mathbf{w}'_{t-1}$ in such a way that $\hat{\Delta}'_t = {\mathbf{w}'_{t-1}}^\top \mathbf{x}_t$ is at most one in magnitude. Observe that $\hat{\Delta}_t$ and $\hat{\Delta}'_t$ have the same sign and only their magnitudes can differ. In particular, it holds that

$$\hat{\Delta}'_t = \begin{cases} \mathrm{sgn}(\hat{\Delta}_t) & \text{if } |\hat{\Delta}_t| > 1 \\ \hat{\Delta}_t & \text{otherwise} \end{cases} .$$

Next, the algorithm defines the new vector $\mathbf{w}_t$ so that $A_t \mathbf{w}_t$ undergoes an additive update, where A_t is a rank-one adjustment of A_{t-1}.

It is not hard to show that this algorithm has a quadratic running time per round, where quadratic means $O(d^2)$ if it is run in primal form, and $O(N_t^2)$ if it is run in dual form (i.e., in a reproducing kernel Hilbert space). In the dual case, since the algorithm updates only when $Z_t = 1$, the number of labels N_t corresponds to the number of support vectors used to define the current hypothesis.

2.3 Analysis

We now prove formal guarantees on the regret of the algorithm and the number of labels it queries. Some details are omitted due to space constraints, and the interested reader is referred to (Dekel et al., 2010) for a more complete analysis. Following (Cesa-Bianchi et al., 2009), the bounds we give depend on how many of the (adversarially chosen) inputs $\mathbf{x}_t$ are close to being complete noise. To capture this dependence, for any $\epsilon > 0$, define

$$T_\epsilon = \sum_{t=1}^{T} \mathbb{1}\{|\Delta_t| \le \epsilon\} . \tag{1}$$

Note that if $|\Delta_t| \le \epsilon$ then $P_t(y_t = 1) \in [1/2 + \epsilon, 1/2 - \epsilon]$. In short, T_ϵ is a "hardness" parameter which is essentially controlled by the adversary. This need not be the case when data is i.i.d. (see Section 2.4). The following theorem is the main result of this section, and is stated so as to emphasize both the data-dependent and the time-dependent aspects of our bounds.

Theorem 1 *Assume that Selective Sampler is run with confidence parameter $\delta \in (0, 1]$. Then with probability at least $1 - \delta$ it holds that for all $T > 0$ that*

$$\begin{aligned} R_T &\le \inf_{\epsilon > 0} \Big\{ \epsilon\, T_\epsilon + \frac{2 + 8 \log|A_T| + 144 \log(T/\delta)}{\epsilon} \Big\} = \inf_{\epsilon > 0} \Big\{ \epsilon\, T_\epsilon + O\Big(\frac{d \, \log T + \log(T/\delta)}{\epsilon} \Big) \Big\} \\ N_T &\le \inf_{\epsilon > 0} \Big\{ T_\epsilon + O\Big(\frac{\log|A_T| \log(T/\delta) + \log^2|A_T|}{\epsilon^2} \Big) \Big\} = \inf_{\epsilon > 0} \Big\{ T_\epsilon + O\Big(\frac{d^2 \log^2(T/\delta)}{\epsilon^2} \Big) \Big\} , \end{aligned}$$

where $|A_T|$ is the determinant of the matrix A_T.

As in (Cesa-Bianchi et al., 2009) it is easy to see that the algorithm can also be run in an infinite dimensional reproducing kernel Hilbert space. In this case, the dimension d in the bounds above is replaced by a quantity that depends on the spectrum of the data's Gram matrix.

The proof of Theorem 1 splits into a series of lemmas. For every $T > 0$ and $\epsilon > 0$, we define

$$U_{T,\epsilon} = \sum_{t=1}^{T} \bar{Z}_t \, \mathbb{1}\Big\{ \Delta_t \hat{\Delta}_t < 0 \Big\} \qquad \text{and} \qquad Q_{T,\epsilon} = \sum_{t=1}^{T} Z_t \, \mathbb{1}\Big\{ \Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2 \Big\} |\Delta_t| ,$$

[2] This is denoted by $Z_t = \mathbb{1}\left\{ |\hat{\Delta}_t| \le \theta_t \right\}$ in the algorithm's pseudocode. Here and throughout $\mathbb{1}\left\{ \cdot \right\}$ denotes the indicator function.

Algorithm 1: Selective Sampler

```
input confidence level $\delta \in (0,1]$
initialize $\mathbf{w}_0 = \mathbf{0},\ A_0 = I$
for $t = 1, 2, \ldots$
    receive $\mathbf{x}_t \in \mathbb{R}^d : ||\mathbf{x}_t|| \le 1$, and set $\hat{\Delta}_t = \mathbf{w}_{t-1}{}^\top \mathbf{x}_t$
    predict $\hat{y}_t = \mathrm{sgn}(\hat{\Delta}_t) \in \{-1, +1\}$
    $\theta_t^2 = \mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t \Big(1 + 4\sum_{i=1}^{t-1} Z_i r_i + 36\log(t/\delta)\Big)$
    $Z_t = \mathbb{1}\left\{\hat{\Delta}_t^2 \le \theta_t^2\right\} \in \{0, 1\}$
    if $Z_t = 1$
        query $y_t \in \{-1, +1\}$
        $\mathbf{w}'_{t-1} = \begin{cases} \mathbf{w}_{t-1} - \left(\frac{|\hat{\Delta}_t| - 1}{\mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t}\right) A_{t-1}^{-1}\mathbf{x}_t & \text{if } |\hat{\Delta}_t| > 1 \\ \mathbf{w}_{t-1} & \text{otherwise} \end{cases}$
        $A_t = A_{t-1} + \mathbf{x}_t \mathbf{x}_t^\top,\ \ r_t = \mathbf{x}_t^\top A_t^{-1} \mathbf{x}_t,\ \ \mathbf{w}_t = A_t^{-1}(A_{t-1}\mathbf{w}'_{t-1} + y_t \mathbf{x}_t)$
    else
        $A_t = A_{t-1},\ \mathbf{w}_t = \mathbf{w}_{t-1},\ r_t = 0$
```

where $\bar{Z}_t = 1 - Z_t$. In the above, $U_{T,\epsilon}$ deals with rounds where the algorithm does not make a query, while $Q_{T,\epsilon}$ deals with rounds where the algorithm does make a query. The proof exploits the potential-based method (e.g., (Cesa-Bianchi & Lugosi, 2006)) for online ridge-regression-like algorithms introduced in (Azoury & Warmuth, 2001). See also (Hazan et al., 2006; Dani et al., 2008) for a similar use in different contexts. The potential function we use is the (quadratic) Bregman divergence $d_t(\mathbf{u}, \mathbf{w}) = \frac{1}{2}\,(\mathbf{u} - \mathbf{w})^\top A_t (\mathbf{u} - \mathbf{w})$, where A_t is the matrix computed by Selective Sampler at time t. The proof structure is as follows. First, Lemma 2 below decomposes the regret R_T into 3 parts: $R_T \le \epsilon T_\epsilon + U_{T,\epsilon} + Q_{T,\epsilon}$. The bound on $U_{T,\epsilon}$ is given by Lemma 3. For the bound on $Q_{T,\epsilon}$ and the bound on the number of queries N_T, we use Lemmas 4 and 5, respectively. However, both of these lemmas require that $(\Delta_t - \hat{\Delta}_t)^2 \le \theta_t^2$ for all t. This assumption is taken care of by the subsequent Lemma 6. Since ϵ is a positive free parameter, we can take the infimum over $\epsilon > 0$ to get the required results.

Lemma 2 *For any $\epsilon > 0$ it holds that* $R_T \le \epsilon T_\epsilon \ + \ U_{T,\epsilon} \ + \ Q_{T,\epsilon}$.

Proof: We have

$$\begin{aligned}
P_t(\hat{\Delta}_t y_t < 0) - P_t(\Delta_t y_t < 0) &\le \mathbb{1}\Big\{\hat{\Delta}_t \Delta_t \le 0\Big\}\Big|2P_t(y_t = 1) - 1\Big| \ = \ \mathbb{1}\Big\{\hat{\Delta}_t \Delta_t \le 0\Big\}|\Delta_t| \\
&= \mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 \le \epsilon^2\Big\}|\Delta_t| + \mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2\Big\}|\Delta_t| \\
&\le \epsilon\,\mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 \le \epsilon^2\Big\} + \mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2\Big\}|\Delta_t| \qquad (2) \\
&\le \epsilon\,\mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 \le \epsilon^2\Big\} + \mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2, Z_t = 0\Big\}|\Delta_t| \\
&\qquad\qquad + \mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2, Z_t = 1\Big\}|\Delta_t| \\
&\le \epsilon\,\mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 \le \epsilon^2\Big\} + \bar{Z}_t\,\mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2\Big\} + Z_t\,\mathbb{1}\Big\{\Delta_t \hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2\Big\}|\Delta_t|.
\end{aligned}$$

Summing over $t = 1 \ldots T$ completes the proof. ■

Lemma 3 *For any $\epsilon > 0$ and $T > 0$, with probability at least $1 - \delta$ it holds that*

$$Q_{T,\epsilon} \ \le \ \frac{2 + 8\log|A_T| + 144\log(T/\delta)}{\epsilon} \ = \ O\left(\frac{d\,\log T + \log(T/\delta)}{\epsilon}\right) \ .$$

Proof Sketch: Using the fact that $\mathrm{sgn}(\hat{\Delta}_t) = \mathrm{sgn}(\hat{\Delta}'_t)$ and the inequality $\mathbb{1}\big\{\Delta_t^2 > \epsilon^2\big\} \le |\Delta_t|/\epsilon$, we upper bound $Q_{T,\epsilon} \le \frac{1}{\epsilon}\sum_{t=1}^T Z_t\,\mathbb{1}\Big\{\hat{\Delta}'_t \Delta_t < 0\Big\}\Delta_t^2$. Moreover, $\hat{\Delta}'_t \Delta_t < 0$ implies $\Delta_t^2 \le (\Delta_t - \hat{\Delta}'_t)^2$ and therefore $Q_{T,\epsilon} \le \frac{1}{\epsilon}\sum_{t=1}^T Z_t(\Delta_t - \hat{\Delta}'_t)^2$. Next, we define $M_t = Z_t(\Delta_t - y_t)(\Delta_t - \hat{\Delta}'_t)$ and note that

$$\sum_{t=1}^T M_t \ = \ \frac{1}{2}\sum_{t=1}^T Z_t(\Delta_t - \hat{\Delta}'_t)^2 - \frac{1}{2}\sum_{t=1}^T Z_t\ \left((y_t - \hat{\Delta}'_t)^2 - (y_t - \Delta_t)^2\right) \ .$$

We also note that $(M_t)_{t=1}^T$ is a martingale difference sequence with $|M_i| \le 4$. Using martingale tail bounds from (Kakade & Tewari, 2008), we obtain a high probability upper bound on $\sum_{t=1}^T M_t$, which implies that

$$\frac{1}{\epsilon}\sum_{t=1}^T Z_t(\Delta_t - \hat{\Delta}_t')^2 \;\le\; \frac{2}{\epsilon}\sum_{t=1}^T Z_t\;\Big((y_t - \hat{\Delta}_t')^2 - (y_t - \Delta_t)^2\Big) + \frac{144}{\epsilon}\;\log(T/\delta)\;\;.$$

Finally, we use techniques from (Azoury & Warmuth, 2001) to upper bound the above by

$$\frac{4}{\epsilon}\sum_{t=1}^T Z_t\left(d_{t-1}(\mathbf{u},\mathbf{w}_{t-1}') - d_t(\mathbf{u},\mathbf{w}_t') + 2\log|A_t| - 2\log|A_{t-1}|\right) + \frac{144}{\epsilon}\;\log(T/\delta)\;\;.$$

We are left with a telescoping sum that collapses to the desired upper bound. ■

Lemma 4 *Assume that $(\Delta_t - \hat{\Delta}_t)^2 \le \theta_t^2$ holds for all t. Then, for any $\epsilon > 0$, we have $U_{T,\epsilon} = 0$*

Proof: We rewrite our assumption $(\Delta_t - \hat{\Delta}_t)^2 \le \theta_t^2$ as $\Delta_t\hat{\Delta}_t \ge \frac{\hat{\Delta}_t^2 + \Delta^2 - \theta_t^2}{2} \ge \frac{\hat{\Delta}_t^2 - \theta_t^2}{2}$. However, if $\bar{Z}_t = 1$, then $\hat{\Delta}_t^2 > \theta_t^2$ and so $\Delta_t\hat{\Delta}_t \ge 0$. Hence, under the above assumption, we can guarantee that for any t, $\bar{Z}_t\,\mathbb{1}\Big\{\Delta_t\hat{\Delta}_t < 0\Big\} = 0$, thereby implying $U_{T,\epsilon} = \sum_{t=1}^T \bar{Z}_t\,\mathbb{1}\Big\{\Delta_t\hat{\Delta}_t < 0, \Delta_t^2 > \epsilon^2\Big\} = 0$. ■

Lemma 5 *Assume that $(\Delta_t - \hat{\Delta}_t)^2 \le \theta_t^2$ holds for all t. Then, for any $\epsilon > 0$, we have*

$$N_T \;\le\; T_\epsilon + O\left(\frac{\log|A_T|\log(T/\delta) + \log^2|A_T|}{\epsilon^2}\right) \;=\; T_\epsilon + O\left(\frac{d^2\log^2(T/\delta)}{\epsilon^2}\right)\;\;.$$

Proof Sketch: Define $\beta_t = \frac{\epsilon^2\,\mathbf{x}_t^\top A_{t-1}^{-1}\mathbf{x}_t}{8\,r_t}$ and rewrite

$$Z_t \;=\; Z_t\,\mathbb{1}\{\theta_t^2 < \beta_t\} \;+\; Z_t\,\mathbb{1}\{\theta_t^2 \ge \beta_t\} \;=\; \mathbb{1}\Big\{\hat{\Delta}_t^2 \le \theta_t^2,\; \theta_t^2 < \beta_t\Big\} \;+\; Z_t\,\mathbb{1}\{\theta_t^2 \ge \beta_t\}\;\;.$$

We begin by dealing with the first term on the right-hand side above. Our assumption implies that whenever $\hat{\Delta}_t^2 \le \theta_t^2$ it also holds that $\Delta_t^2 \le 4\theta_t^2$. Hence we can upper bound the first term by $\mathbb{1}\{\Delta_t^2 \le 4\theta_t^2,\; \theta_t^2 < \beta_t\}$. Using technical results from (Azoury & Warmuth, 2001), we have that $\beta_t \le \epsilon^2/4$, and we can further upper bound $\mathbb{1}\{\Delta_t^2 \le 4\theta_t^2,\; \theta_t^2 < \beta_t\} \le \mathbb{1}\{\Delta_t^2 \le \epsilon^2\}$. Summing over t gives

$$N_T \;=\; \sum_{t=1}^T Z_t \;\le\; T_\epsilon + \sum_{t=1}^T Z_t\,\mathbb{1}\{\theta_t^2 \ge \beta_t\}\;\;.$$

Next, we use the definitions of Z_t, β_t and θ_t to get,

$$\begin{aligned}\sum_{t=1}^T Z_t\,\mathbb{1}\{\theta_t^2 \ge \beta_t\} &= \sum_{t=1}^T Z_t\,\mathbb{1}\Big\{8\,r_t\left(1 + 4\textstyle\sum_{i=1}^{t-1} Z_i r_i + 36\log(t/\delta)\right) \ge \epsilon^2\Big\}\\ &\le \frac{8}{\epsilon^2}\;\sum_{t=1}^T Z_t r_t\left(1 + 4\textstyle\sum_{i=1}^{t-1} Z_i r_i + 36\log(t/\delta)\right)\;\;.\end{aligned}$$

Once again relying on results from (Azoury & Warmuth, 2001), we have that $Z_i r_i \le \log|A_i| - \log|A_{i-1}|$ and the above can be upper bounded by

$$\frac{8}{\epsilon^2}\;(1 + 36\log(T/\delta))\log|A_T| + \frac{16}{\epsilon^2}\;\log^2|A_T|\;\;.$$

This concludes the proof. ■

Lemma 6 *If Selective Sampler is run with confidence parameter $\delta \in (0,1]$, then with probability at least $1-\delta$, the inequality $(\Delta_t - \hat{\Delta}_t)^2 \le \theta_t^2$ holds simultaneously for all t.*

Proof Sketch: First note that by Hölder's inequality,

$$(\Delta_t - \hat{\Delta}_t)^2 = ((\mathbf{w}_{t-1} - \mathbf{u})^\top\mathbf{x}_t)^2 \le 2\,\mathbf{x}_t^T A_{t-1}^{-1}\mathbf{x}_t\;d_{t-1}(\mathbf{w}_{t-1},\mathbf{u})\;. \tag{3}$$

The algorithm only performs an update when $Z_t = 1$. Since this update is that of online ridge regression, we can use techniques in (Azoury & Warmuth, 2001) to show that

$$\frac{1}{2}\sum_{i=1}^{t-1} Z_i \left((y_i - \hat{\Delta}'_i)^2 - (y_i - \Delta_i)^2\right) \;\le\; \frac{1}{2} - d_{t-1}(\mathbf{u}, \mathbf{w}_{t-1}) + 2\textstyle\sum_{i=1}^{t-1} Z_i r_i \;\;.$$

Plugging back into (3) gives

$$(\Delta_t - \hat{\Delta}_t)^2 \;\le\; \mathbf{x}_t^T A_{t-1}^{-1}\mathbf{x}_t\Big(1 + 4\textstyle\sum_{i=1}^{t-1} Z_i r_i - \sum_{i=1}^{t-1} Z_i\big((y_i - \hat{\Delta}'_i)^2 - (y_i - \Delta_i)^2\big)\Big) \;\;. \qquad (4)$$

As in the proof of Lemma 3, we construct the martingale difference sequence $M_i = Z_i(\Delta_i - y_i)(\Delta_i - \hat{\Delta}'_i)$ and use tail bounds from (Kakade & Tewari, 2008) to prove that for any given $t > 1$, with probability at least $1 - \delta/t^2$,

$$-\frac{1}{2}\sum_{i=1}^{t-1} Z_i \left((y_i - \hat{\Delta}'_i)^2 - (y_i - \Delta_i)^2\right) \;\le\; 36\log(t/\delta) \;\;.$$

Plugging the above into Eq. (4) and recalling the definition of θ_t, we have that $(\Delta_t - \hat{\Delta}_t)^2 \le \theta_t^2$. A union bound over all t concludes the proof. ■

Remark 1 *Computing the intermediate vector $\mathbf{w}'_{t-1}$ from $\mathbf{w}_{t-1}$, as defined in the Selective Sampler pseudocode, corresponds to projecting $\mathbf{w}_{t-1}$ onto the convex set $C_t = \{\mathbf{w} \in \mathbb{R}^d \,:\, |\mathbf{w}^\top \mathbf{x}_t| \le 1\}$ w.r.t. the Bregman divergence d_{t-1}, i.e., $\mathbf{w}'_{t-1} = \mathrm{argmin}_{\mathbf{u}\in C_t} d_{t-1}(\mathbf{u}, \mathbf{w}_{t-1})$. Notice that C_t includes the unit ball since $\mathbf{x}_t$ is normalized. This projection step is needed for technical purposes during the construction of our bounded martingale difference sequence (see previous lemmas). Unlike similar constructions (e.g. (Hazan et al., 2006; Dani et al., 2008)), we do not project onto the unit ball, which would involve a line search over matrices and would slow down the algorithm to a significant extent. Moreover, we can prove that the total number of times that Selective Sampler projects onto C_t is $O\left(d^2\log^2(T/\delta)\right)$.*

2.4 An Online-to-Batch Conversion

It is instructive to see what the bound in Theorem 1 looks like when we assume that the instances $\mathbf{x}_t$ are drawn i.i.d. according to an unknown distribution over the Euclidean unit sphere, and to compare this bound to standard statistical learning bounds. We model the distribution of the instances near the hyperplane $\{\mathbf{x} : \mathbf{u}^\top\mathbf{x} = 0\}$ using the well-known *Mammen-Tsybakov low noise condition*[3] (Tsybakov, 2004):

There exist $c > 0$ and $\alpha \ge 0$ such that $P\big(|\mathbf{u}^\top\mathbf{x}| < \epsilon\big) \le c\,\epsilon^\alpha$ for all $\epsilon > 0$.

We now describe a simple randomized algorithm which, with high probability over the sampling of the data, returns a linear predictor with a small expected risk (expectation is taken over the randomization of the algorithm). The algorithm is as follows:

1. Run Algorithm 1 with confidence level δ on the data $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_T, y_T)$, and obtain the sequence of predictors $\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{T-1}$
2. Pick $r \in \{0, 1, \ldots, T-1\}$ uniformly at random and return $\mathbf{w}_r$.

Due to the unavailability of all labels, standard conversion techniques that return a single deterministic hypothesis (e.g., (Cesa-Bianchi & Gentile, 2008)) do not readily apply here. The following theorem states a high probability bound on the risk and the label complexity of our algorithm. We omit the proof due to space constraints.

Theorem 7 *Let $\mathbf{w}_r$ be the linear hypothesis returned by the above algorithm. Then with probability at least $1 - \delta$ we have*

$$\mathbb{E}_r\left[P'_r(y\,\mathbf{w}_r^\top\mathbf{x} < 0)\right] \le P(y\,\mathbf{u}^\top\mathbf{x} < 0) + \mathcal{O}\left((d\log(T/\delta))^{\frac{\alpha+1}{\alpha+2}}\,T^{-\frac{\alpha+1}{\alpha+2}} + \log\left(\frac{\log T}{\delta}\right)/T\right) \;\;,$$

$$N_T = \mathcal{O}\left((d^2\log^2(T/\delta))^{\frac{\alpha}{\alpha+2}}\;T^{\frac{2}{\alpha+2}} + \log(1/\delta)\right) \;\;,$$

where $\mathbb{E}_r$ is the expectation over the randomization in the algorithm, and $P'_r(\cdot)$ denotes the conditional probability[4] $P(\cdot\,|\,\mathbf{x}_1, \ldots, \mathbf{x}_{r-1}, y_1, \ldots, y_{r-1})$.

[3] The constant c might actually depend on the input dimension d. For notational simplicity, Theorem 7 regards c as a constant, hence it is hidden in the big-oh notation.

[4] Notice the difference with the conditional probability $P_r(\cdot)$ defined in Section 2.1.

As α goes from 0 (no assumptions on the noise) to ∞ (hard separation assumption), the above bound on the average regret roughly interpolates between $1/\sqrt{T}$ and $1/T$. Correspondingly, the bound on the number of labels N_T goes from T to $\log^2 T$. In particular, observe that, viewed as a function of N_T (and disregarding log factors), the instantaneous regret is of the form $N_T^{-\frac{\alpha+1}{2}}$. These bounds are sharper than those in (Cavallanti et al., 2009) and, in fact, no further improvement is generally possible (see Castro and Nowak (2008)). The same rates are obtained by (Hanneke, 2009) under much more general conditions, for less efficient algorithms that are based on empirical risk minimization.

One might wonder whether an adaptively adversarial model of learning might somehow be overkill for obtaining i.i.d. results. As a matter of fact, the way our algorithm works makes an adaptively adversarial analysis a very natural one even for deriving the above i.i.d. results.

2.5 Related Work

Selective sampling is an online learning framework lying between passive learning (where the algorithm has no control over the learning sequence) and fully active learning (where the learning algorithm is allowed to select the instances $\mathbf{x}_t$). Recent papers on active learning include (Balcan et al., 2006; Bach, 2006; Balcan et al., 2007; Castro & Nowak, 2008; Dasgupta et al., 2008; Dasgupta et al., 2005; Hanneke, 2007; Hanneke, 2009). All of these papers consider the case when instances are drawn i.i.d. from a fixed distribution (either known or unknown). In particular, (Dasgupta et al., 2005) gives an efficient Perceptron-like algorithm for learning within accuracy ϵ the class of homogeneous d-dimensional half-spaces under the uniform distribution over the unit ball, with label complexity of the form $d \log \frac{1}{\epsilon}$. Still in the i.i.d. setting, more general results are given in (Balcan et al., 2007). A neat analysis of previously proposed general active learning schemes (Balcan et al., 2006; Dasgupta et al., 2008) is provided by the aforementioned paper (Hanneke, 2009). Due to their generality, many of the above results rely on schemes that are computationally prohibitive (exceptions being the results in (Dasgupta et al., 2005) and the realizable cases analyzed in (Balcan et al., 2007)). Finally, pool-based active learning scenarios are considered in (Bach, 2006, and the references therein), though the analysis is only asymptotic in nature and no quantification is given of the trade-off between risk and number of labels.

The results of Theorem 1 are more in line with the worst-case analyses in (Cesa-Bianchi et al., 2006; Strehl & Littman, 2008; Cesa-Bianchi et al., 2009). These papers present variants of Recursive Least Squares algorithms that operate on arbitrary instance sequences. The analysis in (Cesa-Bianchi et al., 2006) is completely worst case: the authors make no assumptions whatsoever on the mechanism generating instances or labels; however, they are unable to prove bounds on the label query rate. The setups in (Strehl & Littman, 2008; Cesa-Bianchi et al., 2009) are closest to ours in that they assume the same linear stochastic noise-model used in our analysis. The algorithm presented in (Strehl & Littman, 2008) approximates the Bayes margin to within a given accuracy ϵ, and queries $\tilde{O}(d^3/\epsilon^4)$ labels; this bound is significantly inferior to our bound, and it seems to hold only in the finite-dimensional case. A more precise comparison can be made to the (expectation) bounds presented in (Cesa-Bianchi et al., 2009), which are of the form $R_T \le \min_{0<\epsilon<1}\left(\epsilon\, T_\epsilon + \frac{T^{1-\kappa}}{\epsilon^2} + \frac{d}{\epsilon^2}\,\ln T\right)$, and $N_T = \mathcal{O}\left(d\,T^{\kappa}\ln T\right)$, where $\kappa \in [0,1]$ is a parameter of their algorithm. In contrast, our bound in Theorem 1 has a sharper dependence on ϵ, and a better trade-off between R_T and N_T. Moreover, unlike the analysis in (Cesa-Bianchi et al., 2009), our analysis covers the case where the instances are generated by an adaptive adversary.

3 The Multiple Teacher Case

The problem is still online binary classification, where at each time step $t = 1, 2, \ldots$ the learner receives an input $\mathbf{x}_t \in \mathbb{R}^d$, with $\|\mathbf{x}_t\| \le 1$, and outputs a binary prediction $\hat{y}_t$. However, there are now K available teachers, each with his own area of expertise. If $\mathbf{x}_t$ falls within the expertise region of teacher j, then that teacher can provide an accurate label. After making each binary prediction, the learner chooses if to issue a query to one or more of the K teachers. The learner is free to query any subset of teachers, but each teacher charges a unit cost per label. The expertise region of each teacher is unknown to the learner, and can only be inferred indirectly from the binary labels purchased from that teacher.

Formally, we assume that teacher j is associated with a weight vector $\mathbf{u}_j \in \mathbb{R}^d$, where $\|\mathbf{u}_j\| \le 1$. If teacher j is queried on round t, he stochastically generates the binary label $y_{j,t}$ according to the law $P_t(y_{j,t} = 1|\mathbf{x}_t) = (1 + \Delta_{j,t})/2$, where $\Delta_{j,t} = \mathbf{u}_j^\top \mathbf{x}_t$ and, as in Section 2, $\mathbf{x}_t$ can be chosen adversarially depending on previous $\mathbf{x}$'s and y_j's. We consider $|\Delta_{j,t}|$ to be the *confidence* of teacher j in his label for $\mathbf{x}_t$. When the learner issues a query, he receives nothing other than the binary label itself, and the confidence is only part of our theoretical model of the teacher. If $\mathbf{x}_t$ is almost orthogonal to $\mathbf{u}_j$ then teacher j has a very low confidence in his label, and we say that $\mathbf{x}_t$ lies outside the expertise region of teacher j.

It is no longer clear how we should evaluate the performance of the learner, since the K teachers will often give inconsistent labels on the given $\mathbf{x}_t$, and we do not have a well defined ground truth to compare against. Intuitively, we would like the learner to predict the label of $\mathbf{x}_t$ as accurately as the teachers who are experts on $\mathbf{x}_t$. To formalize this intuition, define the average margin of a generic subset of teachers[5] $C \subseteq [K]$ as $\Delta_{C,t} = \frac{1}{|C|}\sum_{i\in C}\Delta_{i,t}$. We define the set of experts for each instance using a user-specified parameter $\tau > 0$. Define

$$j_t^\star \;=\; \mathrm{argmax}_j |\Delta_{j,t}| \quad\text{and}\quad C_t \;=\; \{i : |\Delta_{i,t}| \geq |\Delta_{j_t^\star,t}| - \tau\}\ . \tag{5}$$

In words, $j_t^\star$ is the *most confident teacher* at time t, and C_t is the *set of confident teachers* at time t. This means that τ is a tolerance parameter that defines how confident a teacher must be, compared to the most confident teacher, to be considered a confident teacher. Although τ does not appear explicitly in the notation C_t, the reader should keep in mind that C_t and other sets defined later on in this section all depend on τ. Using the definitions above, $\Delta_{C_t,t}$ is the average margin of the confident teachers, and we abbreviate $\Delta_t = \Delta_{C_t,t}$.

Now, let y_t be the random variable that takes values in $\{-1, 1\}$, with $P_t(y_t = 1|\mathbf{x}_t) = (1 + \Delta_t)/2$. In words, y_t is the binary label generated according to the average margin of the confident teachers. We consider the sequence $y_1, \ldots, y_T$ to be our ad-hoc ground-truth, and the goal of our algorithm is to accurately predict this sequence. Note that an equivalent way of generating y_t is by picking a confident teacher j uniformly at random from C_t and setting $y_t = y_{j,t}$. Indeed there are other reasonable ways to define the ground-truth for this problem, however, we feel that our definition coincides with our intuitions on learning from teachers with different areas of expertise. If τ is set to be 1, the learner is compared against the average margin of all K teachers, while if $\tau = 0$, the learner is compared against the single most confident teacher.

We now describe and analyze two algorithms within the multiple teacher setting. We call these algorithms “first version” and “second version”. In the first version, the algorithm queries either all of the teachers or none of the teachers. The second version is more refined in that the algorithm may query a different subset of teachers on each round.

3.1 Algorithm, First Version

The learner attempts to model each weight vector $\mathbf{u}_j$ with a sequence of weight vectors $(\mathbf{w}_{j,t})_{t=1}^T$. As in the single teacher case, the learner maintains a variable threshold θ_t, which can be interpreted as the learner's confidence in its current set of weight vectors. The learner attempts to mimic the process of generating y_t by choosing its own set of confident teachers at each time step. Denoting $\hat\Delta_{j,t} = \mathbf{w}_{j,t}{}^\top\mathbf{x}_t$, the learner defines

$$\hat{j}_t \;=\; \mathrm{argmax}_j |\hat\Delta_{j,t}| \quad\text{and}\quad \hat{C}_t \;=\; \{i : |\hat\Delta_{i,t}| \geq |\hat\Delta_{\hat{j}_t,t}| - \tau - 2\theta_t\}\ ,$$

where $\hat{j}_t$ is the learner's estimate of the most confident teacher, and $\hat{C}_t$ is the learner's estimate of the set of confident teachers. Note that the definition of $\hat{C}_t$ is more inclusive than the definition of C_t in Eq. (5), in that it also includes teachers whose confidence falls below $|\hat\Delta_{\hat{j}_t,t}| - \tau$. This accounts for the uncertainty regarding the learner's set of weight vectors.

As above, we define the notation $\hat\Delta_{C,t} = \frac{1}{|C|}\sum_{i\in C}\hat\Delta_{i,t}$, and abbreviate $\hat\Delta_t = \hat\Delta_{\hat{C}_t,t}$. The learner predicts the binary label $\hat{y}_t = \mathrm{sgn}(\hat\Delta_t)$. Let P_t denote the conditional probability $P_t(\cdot) = \mathbb{P}(\,\cdot\,|\,\mathbf{x}_1, y_{1,1}\ldots, y_{K,1}, \mathbf{x}_2, y_{1,2}\ldots, y_{K,2}, \ldots \mathbf{x}_{t-1}, y_{1,t-1}, \ldots y_{K,t-1}, \mathbf{x}_t)$, and let the regret of the learner be

$$R_T \;=\; \sum_{t=1}^{T}\Big(P_t(y_t\hat\Delta_t < 0) - P_t(y_t\Delta_t < 0)\Big)\ . \tag{6}$$

Next, we proceed to describe our criterion for querying teachers. We present a simple criterion that either sets $Z_t = 1$ and queries all of the teachers or sets $Z_t = 0$ and queries none of them. Hence, the learner either incurs a cost of K or a cost of 0 on each round. We partition the set of confident teachers $\hat{C}_t$ into two sets,

$$\begin{aligned}\hat{H}_t \;&=\; \{i \,:\, |\hat\Delta_{i,t}| \;\geq\; |\hat\Delta_{\hat{j}_t,t}| - \tau + 2\theta_t\}\\ \hat{B}_t \;&=\; \{i \,:\, |\hat\Delta_{\hat{j}_t,t}| - \tau - 2\theta_t \;\leq\; |\hat\Delta_{i,t}| \;<\; |\hat\Delta_{\hat{j}_t,t}| - \tau + 2\theta_t\}\ .\end{aligned}$$

$\hat{H}_t$ is the set of teachers with especially high confidence, while $\hat{B}_t$ is the set of teachers with borderline confidence. Intuitively, the learner is unsure whether the teachers in $\hat{B}_t$ should or should not be included in $\hat{C}_t$. The learner issues a query (to all K teachers) if there exists a set $S \subseteq \hat{B}_t$ such that either $\hat\Delta_t\hat\Delta_{\hat{H}_t\cup S,\,t} < 0$ or $|\hat\Delta_{\hat{H}_t\cup S,\,t}| \leq \theta_t$. In other words, the learner searches for a subset of $\hat{B}_t$ such that replacing $\hat{B}_t$ with that subset would either flip the sign of $\hat\Delta_t$ or make it too small. If a query is issued, each weight vector $\mathbf{w}_{j,t}$ is updated as in the single teacher case. Pseudocode of this learner is given in Algorithm 2.

[5] Here and throughout, $[K] = \{1, 2, \ldots, K\}$.

Algorithm 2: Multiple Teacher Selective Sampler – first version

```
input  confidence level $\delta \in (0,1]$, tolerance parameter $\tau \geq 0$
initialize $A_0 = I,\ \forall j \in [K]\ \mathbf{w}_{j,0} = \mathbf{0}$
for $t = 1, 2, \ldots$
    receive $\mathbf{x}_t \in \mathbb{R}^d\ :\ ||\mathbf{x}_t|| \leq 1$
    $\theta_t^2 = \mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t \big(1 + 4\sum_{i=1}^{t-1} Z_i r_i + 36\log(Kt/\delta)\big)$
    $\forall j \in [K]\ \hat{\Delta}_{j,t} = \mathbf{w}_{j,t-1}{}^\top \mathbf{x}_t$   and   $\hat{j}_t = \text{argmax}_j |\hat{\Delta}_{j,t}|$
    predict $\hat{y}_t = \text{sgn}(\hat{\Delta}_t) \in \{-1,+1\}$
    $Z_t = \begin{cases} 1 & \text{if } \exists S \subseteq \hat{B}_t\ :\ \hat{\Delta}_t \hat{\Delta}_{S \cup \hat{H}_t,\,t} < 0 \ \text{ or } \ |\hat{\Delta}_{S \cup \hat{H}_t,\,t}| \leq \theta_t \\ 0 & \text{otherwise} \end{cases}$
    if $Z_t = 1$
        query $y_{1,t}, \ldots, y_{K,t}$
        $A_t = A_{t-1} + \mathbf{x}_t \mathbf{x}_t^\top,\ \ r_t = \mathbf{x}_t^\top A_t^{-1} \mathbf{x}_t$
        for $j = 1, \ldots, K$
            $\mathbf{w}'_{j,t-1} = \begin{cases} \mathbf{w}_{j,t-1} - \left(\frac{|\hat{\Delta}_{j,t}|-1}{\mathbf{x}_t^\top A_{t-1}^{-1} \mathbf{x}_t}\right) A_{t-1}^{-1} \mathbf{x}_t & \text{if } |\hat{\Delta}_{j,t}| > 1, \\ \mathbf{w}_{j,t-1} & \text{otherwise} \end{cases}$
            $\mathbf{w}_{j,t} = A_t^{-1}(A_{t-1} \mathbf{w}'_{j,t-1} + y_{j,t} \mathbf{x}_t)$
    else
        $A_t = A_{t-1},\ r_t = 0$  and  $\forall j \in [K]\ \mathbf{w}_{j,t} = \mathbf{w}_{j,t-1}$
```

3.2 Analysis, First Version

Our learning algorithm relies on labels it receives from a set of teachers, and therefore our bounds should naturally depend on the ability of those teachers to provide accurate labels for the concrete sequence $\mathbf{x}_1, \ldots, \mathbf{x}_T$. For example, if an input $\mathbf{x}_t$ lies outside the expertise regions of all teachers, we cannot hope to learn anything from the labels provided by the teachers for this input. Similarly, there is nothing we can do on rounds where the set of confident teachers is split between two equally confident but conflicting opinions. We count these difficult rounds by defining, for any $\epsilon > 0$,

$$T_\epsilon \; = \; \sum_{t=1}^{T} \mathbb{1}\{|\Delta_t| \leq \epsilon\} \; . \tag{7}$$

The above is just a multiple teacher counterpart to (1). However it is interesting to note that even in a case where most teachers have low confidence in their prediction on any given round, T_ϵ can still be small provided that the experts in the field have a confident opinion.

A more subtle difficulty presents itself when the collective opinion expressed by the set of confident teachers changes qualitatively with a small perturbation of the input $\mathbf{x}_t$ or one of the weight vectors $\mathbf{u}_j$. To state this formally, define for any $\epsilon > 0$

$$\begin{aligned} H_{\epsilon,t} \; &= \; \{i \ : \ |\Delta_{i,t}| \; \geq \; |\Delta_{j_t^\star,t}| - \tau + \epsilon\} \\ B_{\epsilon,t} \; &= \; \{i \ : \ |\Delta_{j_t^\star,t}| - \tau - \epsilon \; \leq \; |\Delta_{i,t}| \; < \; |\Delta_{j_t^\star,t}| - \tau + \epsilon\} \; . \end{aligned}$$

The set $H_{\epsilon,t}$ is the subset of teachers in C_t with especially high confidence, ϵ higher than the minimal confidence required for inclusion in C_t. In contrast, the set $B_{\epsilon,t}$ is the set of teachers with borderline confidence: either teachers in C_t that would be excluded if their margin were smaller by ϵ, or teachers that are not in C_t that would be included if their margin were larger by ϵ. We say that the average margin of the confident teachers is *unstable* with respect to τ and ϵ if $|\Delta_t| > \epsilon$ but we can find a subset $S \subseteq B_{\epsilon,t}$ such that either $\Delta_t \Delta_{S \cup H_{\epsilon,t},\,t} < 0$ or $|\Delta_{S \cup H_{\epsilon,t},\,t}| < \epsilon$. In other words, we are dealing with the situation where Δ_t is sufficiently confident, but a small ϵ-perturbation to the margins of the individual teachers can cause its sign to flip, or its confidence to fall below ϵ. We count the unstable rounds by defining, for any[6] $\epsilon > 0$,

$$T'_\epsilon \;=\; \sum_{t=1}^{T} \mathbb{1}\big\{|\Delta_t| > \epsilon\big\}\,\mathbb{1}\big\{\exists S \subseteq B_{\epsilon,t} \;:\; \Delta_t \Delta_{S\cup H_{\epsilon,t},\,t} < 0 \;\vee\; |\Delta_{S\cup H_{\epsilon,t},\,t}| \le \epsilon\big\}\,. \tag{8}$$

Intuitively T'_ϵ counts the number of rounds on which an ϵ-perturbation of the $\Delta_{t,j}$ of the teachers either changes the sign of the average margin or results in an average margin close to zero. Like T_ϵ, this quantity measures an inherent hardness of the multiple teacher problem.

The following theorem is the main theoretical result of this section. It provides an upper bound on the regret of the learner, as defined in Eq. (6), and on the total cost of queries, $N_T = K\sum_{t=1}^{T} Z_t$. Again, we stress both the data and the time-dependent aspects of the bound.

Theorem 8 *Assume Algorithm 2 is run with a confidence parameter $\delta > 0$. Then with probability at least $1-\delta$ it holds for all $T > 0$ that*

$$\begin{aligned} R_T \;&\le\; \inf_{\epsilon>0}\left\{\epsilon T_\epsilon + T'_\epsilon + \mathcal{O}\left(\frac{\log|A_T|\,\log(KT/\delta) + \log^2|A_T|}{\epsilon^2}\right)\right\}\\ &=\; \inf_{\epsilon>0}\left\{\epsilon T_\epsilon + T'_\epsilon + \mathcal{O}\left(\frac{d^2\log^2(KT/\delta)}{\epsilon^2}\right)\right\}\;,\\ N_T \;&\le\; K\inf_{\epsilon>0}\left\{T_\epsilon + T'_\epsilon + \mathcal{O}\left(\frac{\log|A_T|\,\log(KT/\delta) + \log^2|A_T|}{\epsilon^2}\right)\right\}\\ &=\; K\inf_{\epsilon>0}\left\{T_\epsilon + T'_\epsilon + \mathcal{O}\left(\frac{d^2\log^2(KT/\delta)}{\epsilon^2}\right)\right\}\;. \end{aligned}$$

As in the proof of Theorem 1, we begin by decomposing the regret. For any $\epsilon > 0$, Lemma 9 states that $R_T \;\le\; \epsilon T_\epsilon + T'_\epsilon + U_{T,\epsilon} + Q_{T,\epsilon}$, where T_ϵ is defined in Eq. (7), T'_ϵ is defined in Eq. (8), and

$$U_{T,\epsilon} = \textstyle\sum_{t=1}^{T} \bar{Z}_t\,\mathbb{1}\big\{\Delta_t\hat{\Delta}_t < 0\big\}, \quad Q_{T,\epsilon} = \sum_{t=1}^{T} Z_t\,\mathbb{1}\big\{\forall S \subseteq B_{\epsilon,t} \;:\; \Delta_t\Delta_{S\cup H_{\epsilon,t}} \ge 0,\; |\Delta_{S\cup H_{\epsilon,t}}| > \epsilon\big\}\,.$$

T_ϵ and T'_ϵ deal with time steps on which the ground truth itself is unreliable, $U_{T,\epsilon}$ sums over rounds where the learner does not make a query, and $Q_{T,\epsilon}$ sums over rounds where a query is issued. Similarly, for any $\epsilon > 0$, Lemma 10 upper bounds the number of time steps on which a query is issued by $T_\epsilon + T'_\epsilon + Q_{T,\epsilon}$. Lemma 11 upper bounds $Q_{T,\epsilon}$ and Lemma 12 upper bounds $U_{T,\epsilon}$. Both lemmas rely on the assumption that $(\Delta_{j,t} - \hat{\Delta}_{j,t})^2 \le \theta_t^2$ for all $t \in [T]$ and $j \in [K]$. A straightforward stratification of Lemma 6 in Section 2 over the K teachers verifies that this condition holds with high probability. The proofs of the mentioned lemmas are omitted.

Lemma 9 *For any $\epsilon > 0$ it holds that $R_T \;\le\; \epsilon T_\epsilon + T'_\epsilon + U_{T,\epsilon} + Q_{T,\epsilon}$.*

Lemma 10 *For any $\epsilon > 0$, it holds that $\sum_{t=1}^{T} Z_t \le T_\epsilon + T'_\epsilon + Q_{T,\epsilon}$.*

Lemma 11 *If $(\Delta_{j,t} - \hat{\Delta}_{j,t})^2 \le \theta_t^2$ holds for all $j \in [K]$ and $t \in [T]$, then*

$$Q_{T,\epsilon} = \mathcal{O}\left(\frac{\log|A_T|\,\log(KT/\delta) + \log^2|A_T|}{\epsilon^2}\right) = \mathcal{O}\left(\frac{d^2\log^2(KT/\delta)}{\epsilon^2}\right)\;.$$

Lemma 12 *If $(\Delta_{j,t} - \hat{\Delta}_{j,t})^2 \le \theta_t^2$ for all $j \in [K]$ and $t \in [T]$, then $U_{T,\epsilon} = 0$ for all $\epsilon > 0$.*

3.3 Algorithm, Second Version

The second version differs from the first one in that now each teacher j has its own threshold $\theta_{j,t}$, and also its own matrix $A_{j,t}$. As a consequence, the set of confident teachers $\hat{C}_t$ and the partition of $\hat{C}_t$ into highly confident ($\hat{H}_t$) and borderline ($\hat{B}_t$) teachers have to be redefined as follows:

$$\begin{aligned} \hat{C}_t &= \{j : |\hat{\Delta}_{j,t}| \ge |\hat{\Delta}_{\hat{j}_t,t}| - \tau - \theta_{j,t} - \theta_{\hat{j}_t,t}\}, \qquad \text{where } \hat{j}_t = \text{argmax}_j|\hat{\Delta}_{j,t}|,\\ \hat{H}_t \;&=\; \{i \;:\; |\hat{\Delta}_{i,t}| \;\ge\; |\hat{\Delta}_{\hat{j}_t,t}| - \tau + \theta_{j,t} + \max\nolimits_{j\in\hat{C}_t}\theta_{j,t}\},\\ \hat{B}_t \;&=\; \big\{i : |\hat{\Delta}_{\hat{j}_t,t}| - \tau - \theta_{j,t} - \theta_{\hat{j}_t,t} \;\le\; |\hat{\Delta}_{i,t}| \;<\; |\hat{\Delta}_{\hat{j}_t,t}| - \tau + \theta_{j,t} + \max\nolimits_{j\in\hat{C}_t}\theta_{j,t}\big\}\,. \end{aligned}$$

The pseudocode is given in Algorithm 3. Notice that the query condition defining Z_t now depends on an *average threshold* $\theta_{S\cup\hat{H}_t,\,t} = \frac{1}{|S\cup\hat{H}_t|}\sum_{j\in S\cup\hat{H}_t}\theta_{j,t}$.

[6] Notice that, up to degenerate cases, both T_ϵ and T'_ϵ tend to vanish as $\epsilon \to 0$. Hence, as in the single teacher case, the free parameter ϵ trades-off hardness terms against regret terms.

Algorithm 3: Multiple Teacher Selective Sampler – second version

```
input confidence level $\delta \in (0,1]$, tolerance parameter $\tau \geq 0$
initialize $A_{j,0} = I, \;\; \mathbf{w}_{j,0} = \mathbf{0}, \;\; \forall j \in [K]$
for $t = 1, 2, \ldots$
    receive $\mathbf{x}_t \in \mathbb{R}^d \; : \; ||\mathbf{x}_t|| \leq 1$
    $\forall j \in [K], \;\; \theta^2_{j,t} = \mathbf{x}_t^\top A^{-1}_{j,t-1}\mathbf{x}_t\big(1 + 4\sum_{i=1}^{t-1} Z_i r_{j,i} + 36\log(Kt/\delta)\big)$
    $\forall j \in [K], \;\; \hat{\Delta}_{j,t} = \mathbf{w}_{j,t-1}{}^\top \mathbf{x}_t$ and $\hat{j}_t = \text{argmax}_j |\hat{\Delta}_{j,t}|$
    predict $\hat{y}_t = \text{sgn}(\hat{\Delta}_t) \in \{-1, +1\}$
    $Z_t = \begin{cases} 1 & \text{if } \exists S \subseteq \hat{B}_t \; : \; \hat{\Delta}_t \hat{\Delta}_{S \cup \hat{H}_t,\, t} < 0 \;\; \text{or} \;\; |\hat{\Delta}_{S \cup \hat{H}_t,\, t}| \leq \theta_{S \cup \hat{H}_t,\, t} \\ 0 & \text{otherwise} \end{cases}$
    if $Z_t = 1$ and $j \in \hat{C}_t$
        query $y_{j,t}$
        $A_{j,t} = A_{j,t-1} + \mathbf{x}_t \mathbf{x}_t^\top, \;\; r_{j,t} = \mathbf{x}_t^\top A^{-1}_{j,t}\mathbf{x}_t$
        $\mathbf{w}'_{j,t-1} = \begin{cases} \mathbf{w}_{j,t-1} - \left(\frac{|\hat{\Delta}_{j,t}| - 1}{\mathbf{x}_t^\top A^{-1}_{j,t-1}\mathbf{x}_t}\right) A^{-1}_{j,t-1}\mathbf{x}_t & \text{if } |\hat{\Delta}_{j,t}| > 1, \\ \mathbf{w}_{j,t-1} & \text{otherwise} \end{cases}$
        $\mathbf{w}_{j,t} = A^{-1}_{j,t}(A_{j,t-1}\mathbf{w}'_{j,t-1} + y_{j,t}\mathbf{x}_t)$
    else
        $A_{j,t} = A_{j,t-1}, \;\; r_{j,t} = 0$ and $\mathbf{w}_{j,t} = \mathbf{w}_{j,t-1}$
```

3.4 Analysis, Second Version

The following theorem bounds the cumulative regret and the total number of queries with high probability. The proof is similar to the proof of Theorem 8. We keep the definitions of the sets $H_{\epsilon,t}$ and $B_{\epsilon,t}$ as given in Section 3.2, but in the bound on N_T in Theorem 13, we replace T'_ϵ with the more refined quantity T''_ϵ, where

$$T''_\epsilon \;=\; \sum_{t=1}^{T} \frac{|H_{\epsilon,t} \cup B_{\epsilon,t}|}{K} \, \mathbb{1}\big\{|\Delta_t| > \epsilon\big\} \, \mathbb{1}\big\{\exists S \subseteq B_{\epsilon,t} \;:\; \Delta_t \Delta_{S \cup H_{\epsilon,t},\, t} < 0 \;\vee\; |\Delta_{S \cup H_{\epsilon,t},\, t}| \leq \epsilon\big\} \;.$$

Note that T''_ϵ is similar to T'_ϵ except that while T'_ϵ only counts the number of times that perturbations to the $\Delta_{j,t}$'s lead to conflict or low confidence predictions, T''_ϵ counts the fraction of confident teachers involved in the conflict. If for most $\mathbf{x}_t$ only a few of the K teachers are experts (highly confident), then one would expect T''_ϵ to be much smaller than T'_ϵ and thus we expect the number of queries to be small.

Theorem 13 *Assume Algorithm 3 is run with a confidence parameter $\delta > 0$. Then with probability at least $1 - \delta$ it holds for all $T > 0$ that*

$$\begin{aligned} R_T &\leq \inf_{\epsilon > 0} \left\{ \epsilon T_\epsilon + T'_\epsilon + \mathcal{O}\left(\frac{K \, \log|A_T| \, \log(KT/\delta) + K \, \log^2|A_T|}{\epsilon^2} \right) \right\} \\ &= \inf_{\epsilon > 0} \left\{ \epsilon T_\epsilon + T'_\epsilon + \mathcal{O}\left(\frac{K \, d^2 \, \log^2(KT/\delta)}{\epsilon^2} \right) \right\} \quad , \\ N_T &\leq K \, \inf_{\epsilon > 0} \left\{ T_\epsilon + T''_\epsilon + \mathcal{O}\left(\frac{K \, \log|A_T| \, \log(KT/\delta) + K \, \log^2|A_T|}{\epsilon^2} \right) \right\} \\ &= K \, \inf_{\epsilon > 0} \left\{ T_\epsilon + T''_\epsilon + \mathcal{O}\left(\frac{K \, d^2 \, \log^2(KT/\delta)}{\epsilon^2} \right) \right\} \quad . \end{aligned}$$

Note that the above theorem holds at the cost of losing a factor K elsewhere in the regret terms, thereby making Theorem 8 and Theorem 13 incomparable.

4 Conclusions and Ongoing Research

We introduced a new Ridge-Regression-like algorithm operating in a robust selecting sampling environment, where the adversary can adapt on the fly to the algorithm's choices. We gave sharp bounds on the cumulative regret and the number of queries made by this algorithm, solving questions left open in previous investigations. We then lifted this machinery to solving the more involved problem where multiple unreliable teachers are available. We gave two algorithms and corresponding analyses.

We are currently running experiments on real-world data (the experimental setting is somewhat similar to the one described in (Donmez & Carbonell, 2008)) to see the performance of the multiple teacher algorithms compared to the simple baseline where K independent instances of the single teacher algorithm (Algorithm 1) are run in parallel. An implementation issue of the multiple teacher algorithms we have presented is the exponential explosion that seemingly arises when computing Z_t, due to the need to check all possible subsets $S \subseteq \hat{B}_t$. As a matter of fact, this check can be computed efficiently by sorting the teachers according to their estimated confidence $|\hat{\Delta}_{j,t}|$. Though preliminary, our experiments suggest that the multiple teacher algorithm largely outperforms the baseline, both in terms of accuracy and total number of requested labels.

On the theoretical side, a few points we are presently investigating are the following: i) The bound on N_T in Theorem 1 is tight w.r.t. ϵ (see the lower bound in (Cesa-Bianchi et al., 2009)), but need not be tight w.r.t. d. This might be due to the way we constructed our martingale argument to prove Lemma 6. ii) As a more general issue, we are trying to generalize our results to further label noise models, such as logistic models. iii) The bounds for the multiple teacher algorithms in Theorems 8 and 13 are likely to be suboptimal, and we are currently trying to better exploit the interaction structure among teachers. iv) Proactive learning, as presented in (Donmez & Carbonell, 2008; Yang & Carbonell, 2009b; Yang & Carbonell, 2009a), also allows for different costs for different teachers, the idea being that more expensive teachers may be more reliable. We are trying to see whether we can incorporate costs into our multiple teacher analysis.

Acknowledgments We thank the COLT 2010 reviewers for their helpful comments. This research was done while the second and the third authors were visiting Microsoft Research at Redmond. The second author acknowledges the PASCAL2 Network of Excellence under EC grant 216886 for supporting travel expenses to the conference.

References

Azoury, K. S., & Warmuth, M. K. (2001). Relative loss bounds for online density estimation with the exponential family of distributions. *Machine Learning*, *43*, 211–246.

Bach, F. (2006). *Active learning for misspecified generalized linear models* (Technical Report N15/06/MM). Ecole des mines de Paris.

Balcan, M., Beygelzimer, M., & Langford, J. (2006). Agnostic active learning. *Proc. of the 23th International Conference on Machine Learning* (pp. 65–72).

Balcan, M., Broder, A., & T. Zhang, T. (2007). Margin-based active learning. *Proc. of the 20th Annual Conference on Learning Theory* (pp. 35–50).

Castro, R., & Nowak, R. D. (2008). Minimax bounds for active learning. *IEEE Trans. IT*, *54*, 2339–2353.

Cavallanti, G., Cesa-Bianchi, N., & Gentile, C. (2009). Linear classification and selective sampling under low noise conditions. *Advances in Neural Information Processing Systems 21*.

Cesa-Bianchi, N., Conconi, A., & Gentile, C. (2003). Learning probabilistic linear-threshold classifiers via selective sampling. *Proc. of the 16th Annual Conference on Learning Theory* (pp. 373–387).

Cesa-Bianchi, N., Conconi, A., & Gentile, C. (2005). A second-order Perceptron algorithm. *SIAM Journal on Computing*, *43*, 640–668.

Cesa-Bianchi, N., & Gentile, C. (2008). Improved risk tail bounds for on-line algorithms. *IEEE Trans. IT*, *54*, 386–390.

Cesa-Bianchi, N., Gentile, C., & Orabona, F. (2009). Robust bounds for classification via selective sampling. *Proc. of the 26th International Conference on Machine Learning*.

Cesa-Bianchi, N., Gentile, C., & Zaniboni, L. (2006). Worst-case analysis of selective sampling for linear classification. *JMLR*, *7*, 1025–1230.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge University Press.

Cohn, R., Atlas, L., & Ladner, R. (1990). Training connectionist networks with queries and selective sampling. *Advances in Neural Information Processing Systems 2*.

Dani, V., Hayes, T. P., & Kakade, S. M. (2008). Stochastic linear optimization under bandit feedback. *Proc. of the 12th Annual Conference on Learning Theory*.

Dasgupta, S., Hsu, D., & Monteleoni, C. (2008). A general agnostic active learning algorithm. *Advances in Neural Information Processing Systems 21*.

Dasgupta, S., Kalai, A. T., & Monteleoni, C. (2005). Analysis of perceptron-based active learning. *Proc. of the 18th Annual Conference on Learning Theory*.

Dekel, O., Gentile, C., & Sridharan, K. (2010). *Robust selective sampling from single and multiple teachers* (Technical Report). Microsoft Research, Università dell'Insubria, TTI.

Donmez, P., & Carbonell, J. G. (2008). Proactive learning: Cost-sensitive active learning with multiple imperfect oracles. *CIKM*.

Freund, Y., Seung, S., Shamir, E., & Tishby, N. (1997). Selective sampling using the query by committee algorithm. *Machine Learning*, *28*, 133–168.

Hanneke, S. (2007). A bound on the label complexity of agnostic active learning. *Proc. of the 24th International Conference on Machine Learning* (pp. 353–360).

Hanneke, S. (2009). Adaptive rates of convergence in active learning. *Proc. of the 22th Annual Conference on Learning Theory*.

Hazan, E., Kalai, A., Kale, S., & Agarwal, A. (2006). Logarithmic regret algorithms for online convex optimization. *Proc. of the 19th Annual Conference on Learning Theory*.

Hoerl, A., & Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55–67.

Kakade, S., & Tewari, A. (2008). On the generalization ability of online strongly convex programming algorithms. *Advances in Neural Information Processing Systems*.

Lai, T. L., & Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 154–166.

Li, L., Littman, M., & Walsh, T. (2008). Knows what it knows: a framework for self-aware learning. *Proc. of the 25th International Conference on Machine Learning* (pp. 568–575).

Strehl, A., & Littman, M. (2008). Online linear regression and its application to model-based reinforcement learning. *Advances in Neural Information Processing Systems 20*.

Tsybakov, A. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. of Stat.*, *32*, 135–166.

Vovk, V. (2001). Competitive on-line statistics. *International Statistical Review*, *69*, 213–248.

Yang, L., & Carbonell, J. (2009a). *Adaptive proactive learning with cost-reliability tradeoff* (Technical Report CMU-ML-09-114). Carnegie Mellon University.

Yang, L., & Carbonell, J. (2009b). *Cost complexity of proactive learning via a reduction to realizable active learning* (Technical Report CMU-ML-09-113). Carnegie Mellon University.