Online Contextual Learning with Limited Feedback

Sayak Ray Chowdhury SAYAK@BU.EDU
Boston University, USA

Aditya Gangrade GANGRADE@BU.EDU
Boston University, USA

Ashok Cutkosky CUTKOSKY@BU.EDU
Boston University, USA

Venkatesh Saligrama SRV@BU.EDU
Boston University, USA

Abstract

We consider a smooth online *contextual* convex optimization problem, where the learner, given the context, has flexibility to decide whether to predict, get feedback and suffer a loss, or to abstain and suffer no loss but get no feedback in every round. The learner is compared against a competitor that gets feedback, and suffers a loss, at every round. This paves the way for designing an abstention rate for the learner that obtains *zero* cumulative regret. We design an epoch-based online gradient descent algorithm, where, at every epoch, abstention rates are computed adaptively in order to minimize abstentions while ensuring zero regret. We bound the excess abstention needed by our algorithm compared to a certain benchmark abstention schedule. The ratio of excess abstentions to the benchmark in every epoch m decays at a rate approximately $O\Big(\max_{c=1}^{K}\sqrt{\frac{1}{\tau_{c,m}}}\Big)$, where K is the number of contexts and $\tau_{c,m}$ is the total number of observations for context c in epoch m.

Keywords: Online learning, Limited feedback

1. Introduction

Consider a resource constrained (e.g., low-power, battery-limited) edge device, such as a sensor or a smartphone that receives a stream of learning tasks (e.g., classification, regression). Due to resource limitations, such a device cannot locally implement most modern day accurate but resource-intensive models (e.g., deepNNs) that are needed to make accurate decisions. Instead the device has access to such a complex model implemented on a cloud server, to which it can send queries in order to maintain accuracy. This incurs communication costs such as latency and battery drain. By leveraging this ability, we should hope that our device would perform *better* than the best possible on-device model (which could be interpreted as the best compression of the cloud model). Of course, we need to do this while minimizing the number of queries to the cloud, i.e., minimizing the total resource consumption (Xu et al., 2014; Nan and Saligrama, 2017). Furthermore, the complexity of the tasks may vary depending on some external factors. For example, regression tasks assigned to a sensor can be hard during some specific times of the day, and can be easy during other times. These scenarios can be modeled by incorporating contextual information in the learning tasks.

We are interested in devising on-device learning algorithms that address these situations. However, we also recognize an additional wrinkle: in many such on-device prediction tasks, obtaining any sort of feedback on the predictions is extremely difficult (e.g. a smartphone user may not indicate whether a dining recommendation was good or not). However, if we believe that complex cloud model is highly accurate, we can use a query to the cloud not only to obtain better predictions, but also as a way of obtaining *labels*. Thus, the decision to query or not is based not only on problem "difficulty", but also on a desire to gather data.

Selective online optimization is a paradigm of relevance to such settings (Cesa-Bianchi et al., 2005; Alon et al., 2015). At each round, the learning setup allows an agent to *optionally* observe feedback (typically, a convex and smooth loss function; depending on the generated context at that round) before picking a decision with some probability. This allows us to model a scenario in which feedback is expensive or scarce. In a different twist, we allow the agent to also obtain this expensive information *before* taking any action. Hence, the agent enjoys flexibility to pick the a postiori best decision corresponding to the observed feedback, and abstain from suffering a loss. Conceptually, this captures querying the cloud model and picking an accurate decision for the learning task of that round. If the agent decides to not query/abstain, then it doesn't receive any feedback, and it picks the decision suggested by a standard online learning algorithm (e.g. online gradient descent), which is run locally on the edge device.

Given a certain amount of prediction error budget, we would like to minimize the total number of abstentions made by the learning. Thus, there should be some ideal pareto-frontier of possible performances. However, we will focus on a particular intuitively-desirable budget: the total loss of the learning should be *strictly smaller* than the loss of the best-in-hindsight on-device model. That is, given this additional flexibility of abstaining, we would like to ensure a negative regret for the learner while minimizing the total number of abstentions (i.e. number of queries to the cloud).

Our problem is related to that of online selective classification (Cortes et al., 2018; Li et al., 2011; Gangrade et al., 2021). In this setting, data (feature, label) is generated sequentially by an adversary; and a learner uses the features to produce a decision which can either be a label, or an abstention. Typically, feedback is provided if the learner abstains; and the learner is said to incur a mistake if it didn't abstain and its prediction did not equal the true label. The emphasis is on ensuring very few mistakes, and bound the regret suffered compared to a classifier that makes no mistakes, while abstaining the fewest number of times. However, it is not straightforward to extend these results to the contextual learning setting since it presents an additional challenge of adapting abstention probabilities with contexts.

We construct an epoch-based online gradient descend (OGD) scheme with doubling epoch lengths. At the beginning of every epoch, we adaptively compute abstention probabilities for every context without having to rely on the future loss functions. With these abstention probabilities, we query the cloud model to ensure non-positive regret of OGD in the current epoch. With the emphasis on non-positive regret, we bound the learner's excess abstentions over an optimal abstention strategy that has the knowledge of cumulative competitor loss for all contexts. We show that at every epoch m, the excess abstentions decays at a rate approximately $O\Big(\max_{c=1}^{K}\sqrt{\frac{1}{\tau_{c,m}}}\Big)$, where K is the number of contexts and $\tau_{c,m}$ is the total number of observations for context c in epoch m.

2. Preliminaries

We consider a modification of the online convex optimization problem with *contexts* and *abstentions*. First, we describe a more standard setting without abstentions before describing our full setting. The learning game proceeds as follows. At reach round $s \geq 1$, we (i) observe a context $c \in \mathcal{C}$, where $|\mathcal{C}| = K$, (ii) play a point $w_{c,s} \in W \subset \mathbb{R}^d$ from some closed and non-empty set W of diameter D, (iii) observe convex loss function $f_{c,s} : \mathbb{R}^d \to \mathbb{R}$, and (iv) pay a loss of value $f_{c,s}(w_{c,s})$. The aim of this game is to minimize the cumulative regret with respect to any competitor $u = (u_1, \dots, u_K) \in W^K$, where $u_c \in W$ is the competitor associated with context c. The cumulative regret after T rounds is defined as

$$R(u) = \sum_{c=1}^{K} R_c(u_c), \quad R_c(u_c) = \sum_{s=1}^{\tau_{c,T}} (f_{c,s}(w_{c,s}) - f_{c,s}(u_c)),$$

where $\tau_{c,T}$ denotes the number of times context c arrives in T rounds.

In what follows, we will work with losses that are smooth and stochastic. Specifically, we assume that each loss function $f_{c,s}$ is H-smooth in open sets containing W and satisfies $f_{c,s}(w) \in [0,1]$ for all $w \in \mathbb{R}$. This implies $\|\nabla f_{c,s}(w)\|^2 \leq 2H f_{c,s}(w)$ for all w, c and s. Furthermore, for each context c, the loss functions $\{f_{c,s}\}_{s \geqslant 1}$ are drawn i.i.d. from some distribution $\mathcal{D}_c$ supported over $[0,1]^W$. We also assume that contexts generated independently of all other randomness with (known) probabilities $\pi = (\pi_1, \dots, \pi_K)$ so that $\sum_{c=1}^{K} \pi_c = 1$.

Adaptive learning rate for OGD. Define, for each context $c \in \mathcal{C}$, and competitor $u_c \in W$, its cumulative loss in T rounds $L^*_{c,T} = \sum_{s=1}^{\tau_{c,T}} f_{c,s}(u_c)$. Then, standard analysis of OGD (Orabona, 2019) for losses corresponding to context c with learning rate $\eta_{c,s} = \frac{D}{\sqrt{2\sum_{s'=1}^{s} \|\nabla f_{c,s'}(w_{c,s'})\|^2}}$, $s = 1, \dots, \tau_{c,T}$, yields the regret

$$R_c(u_c) \leqslant 2D\sqrt{H \sum_{s=1}^{\tau_{c,T}} f_{c,s}(w_{c,s})} = \sqrt{2} \min_{\eta_c > 0} \frac{D^2}{2\eta_c} + H\eta_c \sum_{s=1}^{\tau_{c,T}} f_{c,s}(w_{c,s}), \tag{1}$$

which can be further expressed in terms of $L^*_{c,T}$ as $R_c(u_c) \leqslant 4HD^2 + 4D\sqrt{HL^*_{c,T}}$.

Abstentions. In this paper we make a modification to the contextual online convex optimization game: after seeing the context c, our learner may *optionally* choose to see the loss $f_{c,s}$ *before* picking $w_{c,s}$ (we call this choice "abstaining", or "receiving feedback"). Otherwise, we get *no* feedback. Further, we make the assumption that for all s, there is some $w^*_{c,s} \in W$ such that $f_{c,s}(w^*_{c,s}) = 0$. We will be focusing our attention on randomized strategies which abstain with some c-dependent probability p_c, and always play $w^*_{c,s}$ to suffer zero loss when abstaining. We can show that a simple variant of adaptive gradient descent with such randomized feedback obtains:

$$\mathbb{E}[R(u)] \leqslant \sum_{c=1}^{K} \left(\frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} - p_c L^*_{c,T} \right), \tag{2}$$

where the expectation is over the randomness introduced due to abstentions. From this, it is clear that appropriate choices of p_c will lead to negative regret.

The key problem is determining what p_c should be: contexts with large $L^*_{c,T}$ are inherently "difficult" to predict on-device, and so we should increase p_c. Note that this "difficulty" arises from two different phenomena. If a context c is very common (has high π_c), then $L^*_{c,T}$ becomes larger. However, even for a relatively rare context, $L^*_{c,T}$ may still be large if it is impossible for the on-device model to accurately predict in this context.

Note that this problem may be interesting even for a *single* context. However, when multiple contexts are involved then it is a significantly richer study: we now have the ability to "hedge" different contexts against each other. For example, by abstaining slightly more than may be naively required on a "hard", but possibly slightly less common context, we obtain a more negative regret on that context, which can be used to cancel out a positive regret on an easier but more common context, saving a large number of abstentions on this easier context in the process. Thus, the goal is to determine how to balance the these abstention rates p_c given that we do not know $L^*_{c,T}$ apriori.

3. Algorithm and Result

We consider an epoch-based strategy, where the length of epoch m is given by τ_m with $\tau_{m+1} = 2\tau_m$, $m \geqslant 1$. Intuitively, in each epoch, we select conservative p_c using an estimate of L^*_c generated in the *previous* epochs. Then, we will obtain negative regret in this epoch while refining our estimates of L^*_c in order to reduce p_c further in subsequent epochs.

Let $\tau_{c,m}$ denote the number of occurrence of context c in epoch m and $L^*_{c,m} := \sum_{s=1}^{\tau_{c,m}} f_{c,s}(u_c)$ denotes the cumulative loss of a competitor u_c in epoch m. Given probabilities $p = (p_1, \dots, p_K)$, total number of abstentions in epoch m is given by $A_m(p) = \sum_{c=1}^K \pi_c p_c \tau_m$. Furthermore, from (2), the regret in epoch m, denoted by $R_m(u;p)$, satisfies $\mathbb{E}[R_m(u;p)] \leqslant \sum_{c=1}^K \left(\frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,m}}{p_c}} - p_c L^*_{c,m} \right)$. If the cumulative competitor loss $L^*_{c,m}$, $c = 1, \dots, K$, were known apriori, the abstention probabilities guaranteeing non-positive regret in epoch m could have been computed as follows:

$$p^*_m = \operatorname*{argmin}_{0 \prec p \preceq 1} A_m(p) \quad \text{s.t.} \quad \sum_{c=1}^K \left(\frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,m}}{p_c}} - p_c L^*_{c,m} \right) \leqslant 0. \tag{3}$$

Since the competitor losses are not known in advance, we estimate those using data generated in epoch $m-1$. To this end, let $I_{c,m-1,s}$ denotes the indicator of abstaining for context c at round $s \leqslant \tau_{c,m-1}$, which occurs with probability $\widehat{p}_{c,m-1}$. Define the estimate $\widehat{L}_{c,m} = 2\sum_{s=1}^{\tau_{c,m-1}} \frac{I_{c,m-1,s}}{\widehat{p}_{c,m-1}} f_{c,s}(w_{c,s})$. Also, define, for $\delta \in (0,1], \lambda_c \in (0,1)$, the deviations

$$\Delta^U_{c,m} = \alpha_{c,m} + 2\beta_{c,m} - 4\sqrt{2\tau_{c,m-1}\log(8K/\delta)}, \quad \Delta^L_{c,m} = \alpha_{c,m} + 2\lambda_c\beta_{c,m} + 2D^2H/(1-\lambda_c) \ ,$$

where $\alpha_{c,m} = 2\sqrt{2\pi_c\tau_{m-1}\log(8K/\delta)} + \frac{2}{3}\log(8K/\delta)$, $\beta_{c,m} = \frac{1-\widehat{p}_{c,m-1}}{\widehat{p}_{c,m-1}}\sqrt{2\tau_{c,m-1}\log(8K/\delta)}$. Using these estimates, we compute abstention probabilities $\widehat{p}_m = (\widehat{p}_{1,m}, \dots, \widehat{p}_{K,m})$ as

$$\widehat{p}_m = \operatorname*{argmin}_{0 \prec p \preceq 1} A_m(p) \text{ s.t.} \sum_{c=1}^K \left(\frac{4HD^2}{p_c} + 4D\sqrt{\frac{H(\widehat{L}_{c,m} + \Delta^U_{c,m})}{p_c}} - p_c(\lambda_c\widehat{L}_{c,m} - \Delta^L_{c,m}) \right) \leqslant 0. \tag{4}$$

Excess abstentions. Define, for each context c, the competitor $u_c = \operatorname{argmin}_{w \in W} \mathbb{E}[f_{c,s}(w)]$. First, we obtain the following concentration bound for the competitor loss $\tau_{c,m} L^*_{c,m}$.

Algorithm 1 Epoch-Based OGD with abstentions

Parameter: Confidence level $\delta \in (0, 1]$, constants $\lambda_1, \ldots, \lambda_K \in (0, 1)$
Initialize: $\widehat{p}_{c,1} = 1$ for all c, initial epoch length τ_1
for epoch $m = 1, 2, \ldots$ **do**
  Initialize $w_{c,1} \in W$ and $\tau_{c,m} = 0$ for all c
  for time $s = 1, 2, \ldots, \tau_m$ **do**
    Observe context c_s and set $\tau_{c_s,m} = \tau_{c_s,m} + 1$
    Sample $I_{c_s,m,\tau_{c_s,m}} \sim \mathcal{B}(\widehat{p}_{c_s,m})$
    if $I_{c_s,m,\tau_{c_s,m}} = 1$ **then**
      Output $w^*_{c_s,\tau_{c_s,m}}$ and pay loss 0
    else
      Output $w_{c_s,\tau_{c_s,m}}$ and pay loss $f_{c_s,\tau_{c_s,m}}(w_{c_s,\tau_{c_s,m}})$
    end if
    Set $w_{c_s,\tau_{c_s,m}+1} = w_{c_s,\tau_{c_s,m}} - \frac{D}{\sqrt{2\sum_{s'=1}^{\tau_{c_s,m}} \|\nabla f_{c_s,s'}(w_{c,s'})\|^2}} \nabla f_{c_s,\tau_{c_s,m}}(w_{c_s,\tau_{c_s,m}})$
  end for
  Set size of next epoch $\tau_{m+1} = 2\tau_m$
  Compute estimates $\widehat{L}_{c,m+1}$, $\Delta^U_{c,m+1}$, $\Delta^L_{c,m+1}$ for all c
  Compute abstention rates $\widehat{p}_{c,m+1}$ for all c using (4)
end for

Lemma 1. *For any $m \geqslant 1, \delta \in (0, 1], \lambda_1, \ldots \lambda_K \in (0, 1)$, with probability at least $1 - \delta$,*

$$\forall c = 1, \ldots, K, \quad \lambda_c \widehat{L}_{c,m} - \Delta^L_{c,m} \leqslant L^*_{c,m} \leqslant \widehat{L}_{c,m} + \Delta^U_{c,m} \ .$$

Armed with this result, we have the following bound on regret and excess abstentions.

Theorem 2. *Define, for each $c \in [K]$ and $m \geqslant 1$, $\varepsilon^U_{c,m} = \frac{\Delta^U_{c,m}}{L^*_{c,m}}$ and $\varepsilon^L_{c,m} = \frac{\Delta^L_{c,m}}{L^*_{c,m}}$. Then, for each epoch $m \geqslant 1$, with probability at least $1 - \delta$, we have*

$$\mathbb{E}[R_m(u; \widehat{p}_m)] \leqslant 0, \quad \frac{A_m(\widehat{p}_m) - A_m(p^*_m)}{A_m(p^*_m)} \leqslant \max_{c=1}^{K} \left(1 + \frac{1}{\lambda_c}\right) \left(\frac{(1 - \lambda_c) + (\lambda_c \varepsilon^U_{c,m} + \varepsilon^L_{c,m})}{\lambda_c - (\lambda_c \varepsilon^U_{c,m} + \varepsilon^L_{c,m})}\right) \ .$$

Observe that $\varepsilon^U_{c,m} \approx O(1/\sqrt{\tau_{c,m}})$ and $\varepsilon^L_{c,m} \approx O(\lambda_c/\sqrt{\tau_{c,m}})$. Hence, we the excess factor in abstentions correspondinf to context c behaves approximately as $O\left(a_{c,m} + b_{c,m}/\sqrt{\tau_{c,m}}\right)$ for appropriate constants $a_{c,m}, b_{c,m}$ (depending on $\lambda_c, D, H, K, \delta, \pi_c, \widehat{p}_{c,m-1}$). Therefore, the total excess abstentions in epoch m decays at a rate roughly $O\left(\max_{c=1}^K 1/\sqrt{\tau_{c,m}}\right)$.

Remark 3. *Instead of epoch-based scheme, one can consider abstaining for first half of the horizon T, and use data generated to compute abstention rates for the next half. As a corollary of Theorem 2, this scheme attains $O\left(\max_{c=1}^K 1/\sqrt{\tau_{c,T}}\right)$ bound on excess abstentions.*

Proof. We work under the good event of Lemma 1. We have the expected regret

$$\begin{aligned}
\mathbb{E}[R_m(u; \widehat{p}_m)] &\leqslant \sum_{c=1}^{K} \left(\frac{4HD^2}{\widehat{p}_{c,m}} + 4D\sqrt{\frac{HL^*_{c,m}}{\widehat{p}_{c,m}}} - \widehat{p}_{c,m} L^*_{c,m}\right) \\
&\leqslant \sum_{c=1}^{K} \left(\frac{4HD^2}{\widehat{p}_{c,m}} + 4D\sqrt{\frac{H(\widehat{L}_{c,m} + \Delta^U_{c,m})}{\widehat{p}_{c,m}}} - \widehat{p}_{c,m}(\lambda_c \widehat{L}_{c,m} - \Delta^L_{c,m})\right) \leqslant 0 \ ,
\end{aligned}$$

where the last inequality follows from definition of $\widehat{p}_m$. Denote, for each context c, an abstention probability (assuming it is well-defined)

$$\tilde{p}_{c,m} := p^*_{c,m}\frac{1/\lambda_c + \varepsilon^U_{c,m} + \varepsilon^L_{c,m}/\lambda_c}{\lambda_c - \lambda_c\varepsilon^U_{c,m} - \varepsilon^L_{c,m}} = p^*_{c,m}\frac{1 + \lambda_c\varepsilon^U_{c,m} + \varepsilon^L_{c,m}}{\lambda_c(\lambda_c - \lambda_c\varepsilon^U_{c,m} - \varepsilon^L_{c,m})}.$$

It holds that $\tilde{p}_m = (\tilde{p}_{1,m}, \ldots, \tilde{p}_{K,m})$ is a feasible point of (4). To see this, note that

$$\begin{aligned}
&\sum_{c=1}^K \left(\frac{4HD^2}{\tilde{p}_{c,m}} + 4D\sqrt{\frac{H(\widehat{L}_{c,m} + \Delta^U_{c,m})}{\tilde{p}_{c,m}}} - \tilde{p}_{c,m}(\lambda_c\widehat{L}_{c,m} - \Delta^L_{c,m})\right)\\
&\leqslant \sum_{c=1}^K \left(\frac{4HD^2}{\tilde{p}_{c,m}} + 4D\sqrt{\frac{H(L^*_{c,m}/\lambda_c + \Delta^U_{c,m} + \Delta^L_{c,m}/\lambda_c)}{\tilde{p}_{c,m}}} - \tilde{p}_{c,m}(\lambda_c L^*_{c,m} - \lambda_c\Delta^U_{c,m} - \Delta^L_{c,m})\right)\\
&\leqslant \sum_{c=1}^K \left(\frac{4HD^2}{\tilde{p}_{c,m}} + 4D\sqrt{\frac{HL^*_{c,m}(1/\lambda_c + \varepsilon^U_{c,m} + \varepsilon^L_{c,m}/\lambda_c)}{\tilde{p}_{c,m}}} - \tilde{p}_{c,m}L^*_{c,m}(\lambda_c - \lambda_c\varepsilon^U_{c,m} - \varepsilon^L_{c,m})\right)\\
&= \sum_{c=1}^K \left(\frac{4HD^2}{p^*_{c,m}}\frac{\lambda_c - \lambda_c\varepsilon^U_{c,m} - \varepsilon^L_{c,m}}{1/\lambda_c + \varepsilon^U_{c,m} + \varepsilon^L_{c,m}/\lambda_c} + 4D\sqrt{\frac{HL^*_{c,m}(\lambda_c - \lambda_c\varepsilon^U_{c,m} - \varepsilon^L_{c,m})}{p^*_{c,m}}} - p^*_{c,m}L^*_{c,m}\Big(\frac{1}{\lambda_c} + \varepsilon^U_{c,m} + \frac{\varepsilon^L_{c,m}}{\lambda_c}\Big)\right)\\
&\leqslant \sum_{c=1}^K \left(\frac{4HD^2}{p^*_{c,m}} + 4D\sqrt{\frac{HL^*_{c,m}}{p^*_{c,m}}} - p^*_{c,m}L^*_{c,m}\right) \leqslant 0\ ,
\end{aligned}$$

where the second last inequality is due to $\lambda_c < 1$ and the last inequality follows from (3). The bound on expected abstentions follows by noting that $A_m(\widehat{p}_m) \leqslant A_m(\tilde{p}_m)$. □

Proof sketch of Lemma 1. Since we consider doubling epochs, we first control the difference between competitor loss $L^*_{c,m}$ in epoch m and twice the competitor loss $L^*_{c,m}$ in epoch $m-1$. To this end, Bernstein's inequality yields with probability at least $1 - 2\delta$,

$$|L^*_{c,m} - 2L^*_{c,m-1}| = |\sum_{s=1}^{\tau_{c,m}} f_{c,s}(u_c) - 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(u_c)| \leqslant 2\sqrt{2\pi_c\tau_{m-1}\log(2/\delta)} + \frac{2}{3}\log(2/\delta)\ .$$

Since the competitor loss $L^*_{c,m-1}$ is unknown, we control its deviation from the loss of OGD. First, by Azuma-Hoeffding bound, we get $\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s}) - f_{c,s}(u_c) \geqslant -2\sqrt{2\tau_{c,m-1}\log(2/\delta)}$, with probability at least $1-\delta$. Next, by regret bound of OGD (see (1)), for any $\eta_c > 0$, we get $\sum_{s=1}^{\tau_{c,m-1}} (f_{c,s}(w_{c,s}) - f_{c,s}(u_c)) \leqslant \sqrt{2}\left(\frac{D^2}{2\eta_c} + H\eta_c\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s})\right)$. Choosing an $\eta_c > 0$ such that $\lambda_c := 1 - \sqrt{2}H\eta_c > 0$, we obtain

$$2\lambda_c\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s}) - \frac{2D^2H}{1-\lambda_c} \leqslant 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(u_c) \leqslant 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s}) - 4\sqrt{2\tau_{c,m-1}\log(2/\delta)}.$$

Since we get to see feedback in epoch $m-1$ only with probability $\widehat{p}_{c,m-1}$, we now need to control the loss of OGD with its importance-weighted version. To this end, by Azuma-Hoeffding inequality, we obtain $\left|\sum_{s=1}^{\tau_{c,m-1}}\left(1 - \frac{I_{c,m-1,s}}{\widehat{p}_{c,m-1}}\right) f_{c,s}(w_{c,s})\right| \leqslant \frac{1-\widehat{p}_{c,m-1}}{\widehat{p}_{c,m-1}}\sqrt{2\tau_{c,m-1}\log(2/\delta)}$, with probability at least $1-\delta$. Putting everything together, taking a union bound over all contexts and replacing δ with $\delta/(4K)$, the result follows.

Conclusion. We conclude by noting some future directions: (a) when losses are generated by an adversary, and (b) when context probabilities need to be learned adaptively.

References

Noga Alon, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Conference on Learning Theory*, pages 23–35. PMLR, 2015.

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

Corinna Cortes, Giulia DeSalvo, Claudio Gentile, Mehryar Mohri, and Scott Yang. Online learning with abstention. In *international conference on machine learning*, pages 1059–1067. PMLR, 2018.

Aditya Gangrade, Anil Kag, Ashok Cutkosky, and Venkatesh Saligrama. Online selective classification with limited feedback. *Advances in Neural Information Processing Systems*, 34, 2021.

Lihong Li, Michael L Littman, Thomas J Walsh, and Alexander L Strehl. Knows what it knows: a framework for self-aware learning. *Machine learning*, 82(3):399–443, 2011.

Feng Nan and Venkatesh Saligrama. Adaptive classification for prediction under a budget. *Advances in neural information processing systems*, 30, 2017.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Zhixiang Xu, Matt J Kusner, Kilian Q Weinberger, Minmin Chen, and Olivier Chapelle. Classifier cascades and trees for minimizing feature evaluation cost. *The Journal of Machine Learning Research*, 15(1):2113–2144, 2014.

Appendix

Proof of Equation (2)

Fix a context $c \in [K]$. Let $I_{c,s}$ be the indicator of whether we get feedback on round s so that $\mathbb{E}[I_{c,s}] = p_c$. Then, we run OGD using losses $\widehat{f}_{c,s} = \frac{I_{c,s}}{p_c} f_{c,s}$. Note that these losses are $\frac{H}{p_c}$ smooth. Further, on rounds where $I_{c,s} = 1$, we can ignore the output of the online gradient descent and instead play $w^*_{c,s}$. Then, the regret of OGD would be

$$\mathbb{E}\left[\sum_{s=1}^{\tau_{c,T}} \widehat{f}_{c,s}(w_{c,s}) - \widehat{f}_{c,s}(u_c)\right] \leq \frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}},$$

where the expectation is over the randomness due to abstentions. However, our regret in expectation is given by

$$\begin{aligned}
\mathbb{E}[R_c(u_c)] &\leq \frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} + \mathbb{E}\left[\sum_{s=1}^{\tau_{c,T}} I_{c,s}(f_{c,s}(w^*_{c,s}) - f_{c,s}(w_{c,s}))\right] \\
&\leq \frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} - \mathbb{E}\left[\sum_{s=1}^{\tau_{c,T}} p_c f_{c,s}(w_{c,s})\right] \\
&\leq \frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} - \mathbb{E}\left[\sum_{s=1}^{\tau_{c,T}} p_c(f_{c,s}(w_{c,s}) - f_{c,s}(u_c)) - p_c \sum_{s=1}^{\tau_{c,T}} f_{c,s}(u)\right] \\
&= \frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} - p_c\,\mathbb{E}[R_c(u_c)] - \mathbb{E}[p_c \sum_{s=1}^{\tau_{c,T}} f_{c,s}(u_c)].
\end{aligned}$$

Rearranging $-p_c\,\mathbb{E}[R_c(u_c)]$ and dividing by $1 + p_c$, we get

$$\begin{aligned}
\mathbb{E}[R_c(u_c)] &\leq \frac{1}{1+p_c}\left(\frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} - p_c\,\mathbb{E}\left[\sum_{s=1}^{\tau_{c,T}} f_{c,s}(u_c)\right]\right) \\
&\leq \frac{4HD^2}{p_c} + 4D\sqrt{\frac{HL^*_{c,T}}{p_c}} - p_c L^*_{c,T}.
\end{aligned}$$

The result now follows by summing over all $c = 1, \ldots, K$.

Proof of Lemma 1

Consider the random variable $Z = \sum_{s=1}^{\tau_{c,m}} f_{c,s}(u_c) - 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(u_c)$. Let $J_{c,m,s}$ denote the indicator of observing context c at round s of epoch m. With an abuse of notation, we write

$$Z = \underbrace{\sum_{s=1}^{\tau_m/2} J_{c,m,s} f_{c,s}(u_c) - \sum_{s=1}^{\tau_{m-1}} J_{c,m-1,s} f_{c,s}(u_c)}_{Z_1} + \underbrace{\sum_{s=\tau_m/2+1}^{\tau_m} J_{c,m,s} f_{c,s}(u_c) - \sum_{s=1}^{\tau_{m-1}} J_{c,m-1,s} f_{c,s}(u_c)}_{Z_2}.$$

Using independence of associated random variables and $\tau_m = 2\tau_{m-1}$, we note that Z_1 is sum of τ_{m-1} of random variables, independent of each other, with each having mean zero and absolute value upper bounded by 1. Furthermore, each having variance upper bounded by $2\pi_c$. Then, by Bernstein's inequality, with probability at least $1 - \delta$, $|Z_1| \leqslant$

$\sqrt{2\pi_c\tau_{m-1}\log(2/\delta)} + \frac{1}{3}\log(2/\delta)$. A similar result holds for Z_2. Combining both, we get

$$|\sum_{s=1}^{\tau_{c,m}} f_{c,s}(u_c) - 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(u_c)| \leqslant |Z_1| + |Z_2| \leqslant 2\sqrt{2\pi_c\tau_{m-1}\log(2/\delta)} + \frac{2}{3}\log(2/\delta),$$

with probability at least $1-2\delta$. Now, by regret upper bound of OGD (1), for any $\eta_c > 0$,

$$2\sum_{s=1}^{\tau_{c,m-1}} (f_{c,s}(w_{c,s}) - f_{c,s}(u_c)) \leqslant \sqrt{2}\left(\frac{D^2}{\eta_c} + 2H\eta_c \sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s})\right) .$$

Furthermore, for the competitor $u_c = \operatorname{argmin}_{w\in W}\mathbb{E}[f_{c,s}(w)]$, we obtain

$$\sum_{s=1}^{\tau_{c,m-1}} (f_{c,s}(w_{c,s}) - f_{c,s}(u_c)) \geqslant \underbrace{\sum_{s=1}^{\tau_{c,m-1}} (f_{c,s}(w_{c,s}) - \mathbb{E}[f_{c,s}(w_{c,s})])}_{Z_3} + \underbrace{\sum_{s=1}^{\tau_{c,m-1}} (\mathbb{E}[f_{c,s}(u_c)] - f_{c,s}(u_c))}_{Z_4} .$$

Now, using Azuma-Hoeffding's inequality for martingale difference sequences, we obtain $Z_3 \geqslant -2\sqrt{2\tau_{c,m-1}\log(1/\delta)}$ with probability at least $1-\delta$. A similar bound holds for Z_4. Combining both using a union bound, we have with probability at least $1-\delta$,

$$\sum_{s=1}^{\tau_{c,m-1}} (f_{c,s}(w_{c,s}) - f_{c,s}(u_c)) \geqslant -2\sqrt{2\tau_{c,m-1}\log(2/\delta)}.$$

Now, choose an $\eta_c > 0$ such that $\lambda_c := 1 - \sqrt{2}H\eta_c > 0$. Then, we have the following upper and lower bounds on $\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(u)$:

$$2\lambda_c \sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s}) - \frac{2D^2H}{1-\lambda_c} \leqslant 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(u_c) \leqslant 2\sum_{s=1}^{\tau_{c,m-1}} f_{c,s}(w_{c,s}) - 4\sqrt{2\tau_{c,m-1}\log(2/\delta)}.$$

Now, for $s = 1, 2, \ldots, \tau_{m-1}$, define the random variable $X_s = \left(1 - \frac{I_{c,m-1,s}}{\widehat{p}_{c,m-1}}\right) f_{c,s}(w_{c,s})$. Note that $\{X_s\}_s$ is a martingale difference sequence. Also, $|X_s|^2 \leqslant \left(\frac{1-\widehat{p}_{c,m-1}}{\widehat{p}_{c,m-1}}\right)^2$. Hence, by Azuma-Hoeffding inequality, with probability at least $1-\delta$,

$$\left|\sum_{s=1}^{\tau_{c,m-1}} \left(1 - \frac{I_{c,m-1,s}}{\widehat{p}_{c,m-1}}\right) f_{c,s}(w_{c,s})\right| \leqslant \frac{1-\widehat{p}_{c,m-1}}{\widehat{p}_{c,m-1}}\sqrt{2\tau_{c,m-1}\log(2/\delta)}.$$

Putting everything together, with probability at least $1-4\delta$, we obtain

$$L^*_{c,m} \leqslant 2\sum_{s=1}^{\tau_{c,m-1}} \frac{I_{c,m-1,s}}{\widehat{p}_{c,m-1}} f_{c,s}(w_s) + \alpha_{c,m} + 2\beta_{c,m} - 4\sqrt{2\tau_{c,m-1}\log(2/\delta)},$$

$$L^*_{c,m} \geqslant 2\lambda_c \sum_{s=1}^{\tau_{c,m-1}} \frac{I_{c,m-1,s}}{\widehat{p}_{c,m-1}} f_{c,s}(w_s) - \alpha_{c,m} - 2\lambda_c\beta_{c,m} - 2D^2H/(1-\lambda_c).$$

The result follows by taking a union bound over all contexts and replacing δ with $\delta/(4K)$.