

# Low-Regret Active Learning

Cenk Baykal<sup>1</sup> Lucas Liebenwein<sup>1</sup> Dan Feldman<sup>2</sup> Daniela Rus<sup>1</sup>

## Abstract

We develop an online learning algorithm for identifying unlabeled data points that are most informative for training (i.e., active learning). By formulating the active learning problem as the prediction with sleeping experts problem, we provide a framework for identifying informative data with respect to any given definition of informativeness. At the core of our work is an efficient algorithm for sleeping experts that is tailored to achieve low regret on predictable (easy) instances while remaining resilient to adversarial ones. This stands in contrast to state-of-the-art active learning methods that are overwhelmingly based on greedy selection, and hence cannot ensure good performance across varying problem instances. We present empirical results demonstrating that our method (i) instantiated with an informativeness measure consistently outperforms its greedy counterpart and (ii) reliably outperforms uniform sampling on real-world data sets and models.

## 1. Introduction

Modern neural networks have been highly successful in a wide variety of applications ranging from Computer Vision (Feng et al., 2019) to Natural Language Processing (Brown et al., 2020). However, these successes have come on the back of training large models on massive labeled data sets, which may be costly or even infeasible to obtain in other applications. For instance, applying deep networks to the task of cancer detection requires medical images that can only be labeled with the expertise of healthcare professionals, and a single accurate annotation may come at the cost of a biopsy on a patient (Shen et al., 2019).

*Active learning* focuses on alleviating the high label-cost of learning by only querying the labels of points that are deemed to be the most informative. The notion of infor-

<sup>1</sup>Massachusetts Institute of Technology, Cambridge, USA  
<sup>2</sup>University of Haifa, Haifa, Israel. Correspondence to: Cenk Baykal <baykal@mit.edu>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML).

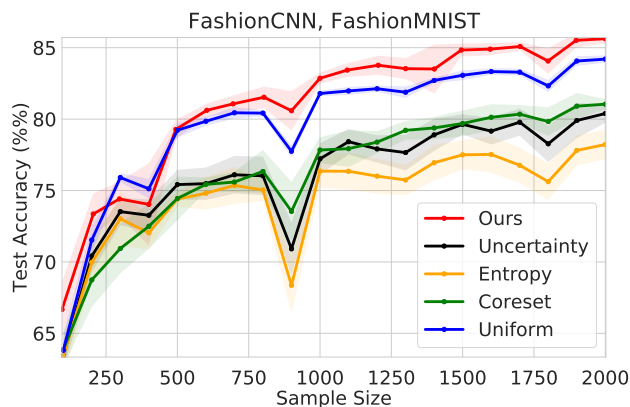


Figure 1. Evaluations on FashionMNIST (Xiao et al., 2017) with data augmentation and normalization where we train a new network from scratch after each data acquisition step. The initial set of points and the sequence of random network initializations (one per sample size) are fixed across all algorithms to ensure fairness. Existing approaches based on greedy selection are not robust and may perform significantly worse than uniform sampling.

mativeness is not concrete and may be defined in a task-specific way. Unsurprisingly, prior work in active learning has primarily focused on devising proxy metrics to appropriately quantify the informativeness of each data point in a tractable way. Examples include proxies based on model uncertainty (Gal et al., 2017), clustering (Sener & Savarese, 2017), and margin proximity (Ducoffe & Precioso, 2018) (see (Ren et al., 2020) for a complete survey).

An overwhelming majority of existing methods are based on greedy selection of the points that are ranked as most informative with respect to the proxy criterion. Despite the intuitiveness of this approach, it is known to be highly sensitive to outliers and to occasionally perform significantly worse than uniform sampling on certain tasks (Ebrahimi et al., 2020) – as Fig. 1 also depicts. In fact, this shortcoming manifests itself even on reportedly redundant data sets, such as MNIST, where existing approaches can lead to models with up to 15% (absolute terms) higher test error (Muthakana, 2019) than those obtained with uniform sampling. In sum, the general lack of robustness guarantees of prior approaches impedes their widespread applicability.

to high-impact deep learning tasks.

In this paper, we propose a low-regret active learning framework that can be applied to any task with any user-specified notion of informativeness. Our approach deviates from the standard greedy paradigm and instead formulates the active learning problem as that of learning with expert advice in an adversarial environment. We propose and analyze a novel regret minimization algorithm tailored to the active learning setting. In this regard, our work aims to advance the development of effective and robust active learning strategies that can be widely applied to modern deep learning tasks.

In this work, we:

1. Show how sequential data acquisition learning can be viewed as the prediction with sleeping experts problem,
2. Develop an efficient, predictive algorithm for low-regret active learning and establish bounds on its regret,
3. Demonstrate that our algorithm can be used off-the-shelf to improve existing greedy heuristics,
4. Evaluate and compare the effectiveness our approach on modern networks and data sets.

## 2. Problem Formulation

We consider the setting where we are given a set of  $n$  unlabeled data points  $\mathcal{P} \subset \mathcal{X}^n$  from the input space  $\mathcal{X} \subset \mathbb{R}^d$ . We assume that there is an oracle ORACLE that maps each point  $x \in \mathcal{P}$  to one of  $k$  categories. Given a network architecture and sampling budget  $b \in \mathbb{N}_+$ , our goal is to generate a subset of points  $\mathcal{S} \subset \mathcal{P}$  with  $|\mathcal{S}| = b$  such that training on  $\{(x, \text{ORACLE}(x))_{x \in \mathcal{S}}\}$  leads to the most accurate model  $\theta$  among all other choices for a subset  $\mathcal{S} \subset \mathcal{P}$  of size  $b$ .

The iterative variant of acquisition procedure is shown as Alg. 1, where ACQUIRE is an active learning algorithm that identifies (by using  $\theta_{t-1}$ )  $b_t$  unlabeled points to label at each iteration  $t \in [T]$  and TRAIN trains a model initialized with  $\theta_{t-1}$  using the labeled set of points. We emphasize that prior work has overwhelmingly used the *scratch* option (Line 6, Alg. 1), which entails discarding the model information  $\theta_{t-1}$  from the previous iteration and training a randomly initialized model from scratch on the set of labeled points acquired thus far,  $\mathcal{S}$ .

### 2.1. Background & Greedy Selection

Consider an *informativeness function*  $g : \mathcal{X} \times \Theta \rightarrow [0, 1]$  that quantifies the informativeness of each point  $x \in \mathcal{X}$  with respect to the model  $\theta \in \Theta$ , where  $\Theta$  is the set of all possible parameters for the given architecture. An example of the gain function is the maximum variation ratio (also

called the *uncertainty metric*) defined as

$$g(x, \theta) = 1 - \max_{i \in [k]} f_{\theta}(x)_i.$$

where  $f_{\theta}(x) \in \mathbb{R}^k$  is the softmax output of the model  $\theta$  given input  $x$ . As examples with this metric, the *gain*  $g(x, \theta)$  of point  $x$  is 1 if the network is absolutely certain about the label of  $x$  and  $1 - 1/k$  in the other extreme case where the network predicts a uniform distribution.

---

#### Algorithm 1 ACTIVELEARNING

---

**Input:** Set of points  $\mathcal{P} \subseteq \mathbb{R}^{d \times n}$ , ACQUIRE: an active learning algorithm for selecting labeled points

- 1:  $\mathcal{S} \leftarrow \emptyset$ ;
  - 2:  $\theta_0 \leftarrow$  Randomly initialized network model;
  - 3: **for**  $t \in [T] = \{1, \dots, T\}$  **do**
  - 4:  $\mathcal{C}_t \leftarrow$  ACQUIRE( $\mathcal{P} \setminus \mathcal{S}, b_t, \theta_{t-1}$ ) {Get new batch of  $b_t \in \mathbb{N}_+$  points to label using the active learning algorithm ACQUIRE}
  - 5:  $\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{C}_t$  {Add new points}
  - 6: (*scratch option*)  $\theta_{t-1} \leftarrow$  Randomly initialized network
  - 7:  $\theta_t \leftarrow$  TRAIN( $\theta_{t-1}, \{(x, \text{ORACLE}(x))_{x \in \mathcal{S}}\}$ ) {Obtain new labels (by querying ORACLE) and train network on the labeled samples thus far}
  - 8: **end for**
  - 9: **return**  $\theta_T$
- 

In the context of Alg. 1, prior work on active learning (Muthakana, 2019; Geifman & El-Yaniv, 2017; Gal et al., 2017; Sener & Savarese, 2017) has generally focused on greedy acquisition strategies (ACQUIRE in Alg. 1) that rank the remaining unlabeled points  $x \in \mathcal{P} \setminus \mathcal{S}$  by their informativeness  $g(x, \theta_{t-1})$  as a function of the model  $\theta_{t-1}$ , and subsequently pick the top  $b_t$  points to label.

**Why greedy can fail** Greedy approaches to data acquisition have shown promise in certain active learning applications and tasks (Gal et al., 2017; Sener & Savarese, 2017)), however, as noted in Sec. 1 – as well as our empirical results in Sec. 5 – these approaches are highly sensitive to outliers and at times perform significantly worse than naive uniform sampling. In fact, Fig. 1 depicts a scenario where various popular active learning approaches perform significantly worse than uniform sampling.

To understand why this could be happening, note that at iteration  $t \in [T]$  the greedy approach makes a judgment about the informativeness of each point using only the model  $\theta_{t-1}$  (i.e., a single snapshot). However, in the deep learning setting where stochastic elements such as random initialization, stochastic optimization, (randomized) data augmentation, and dropout are commonly present,  $\theta_{t-1}$  is itself a random variable with non-negligible variance. This means that, for

example, we could get unlucky with our training and obtain a deceptive model  $\theta_{t-1}$  that assigns high gains (informativeness) to points that may not truly be helpful towards obtaining a highly accurate model at the end of  $T$  active learning iterations.

This problem is perhaps more clearly seen by considering an adversary that generates the gains at each round in a way intended to fool us into greedily picking points that are ultimately not very uninformative in the grand scheme of things, i.e., with respect to the entire time horizon  $T$ . The reported effectiveness of using ensembles in active learning further substantiates our conjecture that different models can in fact significantly disagree about the informativeness of points, which leads to high-variance estimates, and a misguided acquisition of points.

## 2.2. Active Learning as Prediction with Expert Advice

Rather than attempting to model this randomness in the gains observed, we will assume that the gains can be generated by a *non-oblivious adversary* that has knowledge of our actions in the preceding rounds. This formulation leads us to the well-studied learning with experts problem as we outline in this subsection. We note that at the expense of formulating a seemingly more difficult problem involving an adversary, we obtain the benefit of generality and widespread applicability without imposing any assumptions on how the gains are defined or how they are generated.

**Setting** We let  $g_{t,i}$  denote  $g(x_i, \theta_{t-1})$  (see Alg. 1) in round  $t \in [T]$  where  $x_i$  is the  $i^{\text{th}}$  point in  $\mathcal{P}$ . For ease of presentation, assume for the time being that  $b_t = 1$ , i.e., we select a single new point at each iteration. Rather than picking this point deterministically, consider selecting sampling this point with respect to a probability distribution  $p \in \Delta$  where  $\Delta = \{p \in [0, 1]^n : \sum_{j=1}^n p_j = 1\}$  is the probability simplex. We note that there have been prior attempts to make greedy approaches robust by sampling in this very same way (rather than deterministic selection), however, it was observed that this led to *inferior* performance in practice (Gissin & Shalev-Shwartz, 2019).

To map this problem to the canonical learning with experts problem, we consider the problem of minimizing the sum of losses  $\ell_{t,i} = 1 - g_{t,i} \in [0, 1]$  over the  $T$  active learning iterations. Under this setting a natural first attempt at a formulation of expected regret is the following

$$\text{Regret}(p_1, \dots, p_T) = \sum_{t=1}^T \langle p_t, l_t \rangle - \min_{p \in \Delta} \sum_{t=1}^T \langle p, l_t \rangle,$$

where  $\langle p_t, l_t \rangle = \mathbb{E}_{i \sim p_t | t-1} [l_{t,i}]$  is the expected loss of our random choice  $i \sim p_t$  conditioned on our past history.

**Sleeping Experts** However, the previous formulation is ill-equipped for active learning, since (i) it does not make sense to compete with a fixed distribution when set of actions is decreasing over time and (ii) we should not be picking points that have already been labeled in prior active learning iterations. To impose this constraint, we generalize the prior formulation to one with *sleeping experts* (Saha et al., 2020; Luo & Schapire, 2015; Gaillard et al., 2014; Kleinberg et al., 2010). More concretely, let  $\mathcal{I}_{t,i} \in \{0, 1\}$  denote whether expert  $i \in [n]$  is sleeping in round  $t$ . The sleeping expert problem imposes the constraint that  $\mathcal{I}_{t,i} = 0 \Rightarrow p_{t,i} = 0$ . For the data acquisition setting, we define for each  $i \in [n]$

$$\mathcal{I}_{t,i} = \mathbb{1}\{x_i \text{ not picked in any of the preceding rounds}\},$$

and formulate the dynamic active learning regret as

$$\text{Regret}(\rho) = \sum_{t=1}^T \langle p_t, l_t \odot \mathcal{I}_t \rangle - \sum_{t=1}^T \min_{p \in \mathcal{A}_t} \langle p, l_t \odot \mathcal{I}_t \rangle \quad (1)$$

where  $\rho = (p_1, \dots, p_T)$  is the sequence of sampling probability distributions over  $T$  rounds and  $\mathcal{A}_t$  is the constrained probability simplex with respect to  $\mathcal{I}_t \in \{0, 1\}^n$

$$\mathcal{A}_t = \{p \in \Delta : \forall i \in [n] p_i = 0 \text{ if } \mathcal{I}_{t,i} = 0\}.$$

## 3. Method

In this section we motivate and present Alg. 2, an efficient online learning algorithm with instance-dependent guarantees that performs well on predictable sequences while remaining resilient to adversarial ones.

### 3.1. Background

Algorithms for the prediction with sleeping experts problem have been extensively studied in literature (Gaillard et al., 2014; Luo & Schapire, 2015; Saha et al., 2020; Kleinberg et al., 2010; Shayestehmanesh et al., 2019; Koolen & Van Erven, 2015). These algorithms enjoy strong guarantees in the adversarial setting; however, they suffer from (i) sub-optimal regret bounds in predictable settings<sup>1</sup> and/or (ii) exceedingly high computational complexity.

Our approach hinges on the observation that the active learning setting is not entirely adversarial in practice as discussed in Sec. 2. More reasonably, we can expect the informativeness of the points to resemble a predictable sequence plus random noise which models the random components (see

<sup>1</sup>The works of (Gaillard et al., 2014; Koolen & Van Erven, 2015) already do achieve second-order regret bounds for easy instances, and (Luo & Schapire, 2015) achieves first-order quantile bounds, however, these approaches cannot exploit predictions, to, e.g., have constant regret in predictable environments (Orabona, 2019) as other optimistic algorithms can.

Sec. 2) at each time step. This predictability in the corresponding losses motivates an algorithm that can leverage predictions about the loss for the next time step to achieve lower regret when the losses do not vary significantly over time. This is especially pertinent to the active learning problem as the informativeness of certain points may not change too drastically from one acquisition step to the next.

### 3.2. AdaProd<sup>+</sup>

To this end, we extend<sup>2</sup> the Optimistic Adapt-ML-Prod algorithm (Wei et al., 2017) to the active learning setting where the number of total experts are unknown in advance<sup>3</sup>. Optimistic online learning algorithms are capable of incorporating predictions  $\hat{\ell}_{t+1}$  for the loss in the next round  $\ell_{t+1}$  and guaranteeing regret as a function of the predictions' accuracy, i.e., as a function of  $\sum_{t=1}^T \|\ell_t - \hat{\ell}_t\|_\infty$ . Although we could have attempted to extend other optimistic approaches (Steinhardt & Liang, 2014; Orabona, 2019; Mohri & Yang, 2015; Rakhlin & Sridharan, 2013), the work of (Wei et al., 2017) ensures – to the best of our knowledge – the smallest regret in predictable environments when compared to related approaches.

Our algorithm ADAPROD<sup>+</sup> is shown as Alg. 2. Besides relaxing Optimistic Adapt-ML-Prod's requirement of known number of experts, we draw inspiration from the Bernstein Online Aggregation algorithm (Wintenberger, 2017) – an approach that is adaptive to the range of the suffered losses – in order to improve the empirical effectiveness of our algorithm. Our insight is that our predictions can be leveraged to improve practical performance by allowing larger learning rates to be used without sacrificing theoretical guarantees (Line 10 of Alg. 2). Empirical comparisons with Adapt-ML-Prod can be found in Sec. C of the supplementary.

### 3.3. Generating Predictions

Our approach can be used with general predictors  $\hat{\ell}_t$  for the true loss  $\ell_t$  at round  $t$ , however, for the purposes of obtaining bounds in terms of the temporal variation in the losses, we follow (Wei et al., 2017) and use the most recently observed loss as our prediction for the next round, i.e.,  $\hat{\ell}_t = \ell_{t-1}$ . However, a subtle point is that our algorithm requires a prediction  $\hat{r}_t \in \mathbb{R}^n$  for the instantaneous regret at round  $t$ , i.e.,  $r_t = \langle p_t, \ell_t \rangle - \ell_t$ . However, using  $\hat{r}_t = \langle p_{t-1}, \hat{\ell}_t \rangle - \hat{\ell}_t$  is not sufficient since the term  $(\hat{r}_t - r_t)^2$  cannot be upper bounded by an expression of the form  $\|\ell_t - \hat{\ell}_t\|_\infty^2 = \|\ell_t - \ell_{t-1}\|_\infty^2$  (Wei et al., 2017).

<sup>2</sup>Our extension is inspired by the AdaNormalHedge algorithm (Luo & Schapire, 2015).

<sup>3</sup>This is because we may not exactly pick  $b_t$  points due to the randomness of selection (see Sec. 3.4), and hence the number of experts seen up to time  $\tau$ ,  $\sum_{t=1}^{\tau} \sum_{i=1}^n \mathcal{I}_{t,i}$ , is a random variable.

To achieve this bound, we follow the procedure of (Wei et al., 2017) as follows. We define the mapping  $\hat{r}_t : \alpha \mapsto (\alpha - \ell_t) \in [-1, 1]^n$  and perform a binary search over the update rule in Lines 4-5 of Alg. 2 so that  $\alpha$  is such that  $\alpha = \langle p_t(\hat{r}_t(\alpha)), \hat{\ell}_t \rangle$ , where  $p_t(\hat{r}_t(\alpha))$  is the distribution obtained when  $\hat{r}_t(\alpha)$  is used as the optimistic prediction in Lines 4-5. Note that the existence of such an  $\alpha$  follows by the continuity of the update step with respect to  $\alpha$  and the intermediate value theorem.

---

#### Algorithm 2 ADAPROD<sup>+</sup>

---

- 1: For all  $i \in [n]$ , initialize  $R_{1,i} \leftarrow 0$ ;  $C_{1,i} \leftarrow 0$ ;  
 $\eta_{0,(1,i)} \leftarrow 1$ ;  $w_{0,(1,i)} = 1$ ;  $\hat{r}_{1,i} = 0$ ;
- 2: **for** each round  $t \in [T]$  **do**
- 3:  $\mathcal{A}_t \leftarrow \{i \in [n] : \mathcal{I}_{t,i} = 1\}$  {Set of awake experts}
- 4:  $p_{t,i} \leftarrow \sum_{s \in [t]} \eta_{t-1,(s,i)} w_{t-1,(s,i)} \exp(\eta_{t-1,(s,i)} \hat{r}_{t,i})$   
 for each  $i \in \mathcal{A}_t$
- 5:  $p_{t,i} \leftarrow p_{t,i} / \sum_{j \in \mathcal{A}_t} p_{t,j}$  for each  $i \in \mathcal{A}_t$  {Normalize}
- 6: Adversary reveals  $\ell_t$  and we suffer loss  $\tilde{\ell}_t = \langle \ell_t, p_t \rangle$
- 7: For all  $i \in \mathcal{A}_t$ ,  $r_{t,i} \leftarrow \tilde{\ell}_t - \ell_{t,i}$  and  $C_{t,i} \leftarrow 0$
- 8: For all  $i \in \mathcal{A}_t$  and  $s \in [t]$ , set  $C_{s,i} \leftarrow C_{s,i} + (\hat{r}_{t,i} - r_{t,i})^2$
- 9: Get prediction  $\hat{r}_{t+1} \in [-1, 1]^n$  for next round (see Sec. 3.3)
- 10: For all  $i \in \mathcal{A}_t$ , set  $w_{t-1,(t,i)} \leftarrow 1$ ,  $\eta_{t-1,(t,i)} \leftarrow 2/3$ , and for all  $s \in [t]$ , let

$$\eta_{t,(s,i)} \leftarrow \min \left\{ \eta_{t-1,(s,i)}, \frac{2}{3(1 + \hat{r}_{t+1,i})}, \sqrt{\frac{2 \log(n)}{1 + C_{s,i}}} \right\},$$

and perform the two-step update

$$w_{t,(s,i)} \leftarrow \exp(\eta_{t-1,(s,i)} r_{t,i} - \eta_{t-1,(s,i)}^2 (r_{t,i} - \hat{r}_{t,i})^2)$$

$$w_{t,(s,i)} \leftarrow (w_{t-1,(s,i)} w_{t,(s,i)})^{\eta_{t,(s,i)} / \eta_{t-1,(s,i)}}$$

11: **end for**

---

### 3.4. Back to Active Learning

To unify ADAPROD<sup>+</sup> with Alg. 1, observe that we can define the ACQUIRE function to be a procedure that at time step  $t$  first samples a point by sampling with respect to probabilities  $p_t$ , obtains the (user-specified) losses  $\ell_t$  with respect to the model  $\theta_{t-1}$ , and passes them to our algorithm 2 as if they were obtained from the adversary as in Line 6. This yields an updated probability distribution  $p_{t+1}$  and we repeat the process.

In practice, to generalize this to sampling a batch of  $b_t$  points instead, we can instead consider sampling each point  $i \in [n]$  with probability  $\tilde{p}_{t,i} = \min\{b_t p_{t,i}, 1\}$ . This implies that the expected number of points sampled is roughly  $b_t$  if we assume that the probability  $p_t$  is not heavily concentrated on a single point (i.e.,  $b_t p_{t,i} \leq 1$ )<sup>4</sup>. Note that this is a very

<sup>4</sup>Alternatively, we could perform a (binary) search over  $\beta \geq b_t$

mild assumption in active learning where the number of unlabeled points  $n$  is assumed to be much larger than the batch size, i.e.,  $n \gg b_t$ . Incorporating principled tools from combinatorial online learning (e.g., Component iProd algorithm of (Koolen & Van Erven, 2015)) in an efficient way is an avenue for future work.

### 3.5. Flexibility via Proprietary Loss

We end this section by exemplifying the advantage of our approach in its flexibility to be applied off-the-shelf to any definition of (bounded) loss  $\ell \in [0, 1]^n$  (i.e., 1 - informativeness). To this end, we can consider augmenting the popular *uncertainty* metric as defined in Sec. 2. To this end, consider the loss

$$\ell_{t,i} = \underbrace{\Delta(f_t(x_i), \mathcal{U}_k)}_{\text{Prediction certainty}} \cdot \underbrace{(1 - \Delta(f_t(x_i), f_{t-1}(x_i)))}_{\text{"Ensemble" agreement}}$$

where  $f_t(x_i) \in \mathbb{R}^k$  is the softmax output of the model at iteration  $t$  (see Sec. 2),  $\Delta(p, q)$  is the total variation distance between distributions  $p$  and  $q$   $\Delta(p, q) = \max_i |p_i - q_i|$ , and  $\mathcal{U}_k = (1/k, \dots, 1/k)$  is the uniform distribution.

Note that this loss conveys information about both the prediction uncertainty of point  $x_i$  and the ‘ensemble’ disagreement across consecutive models in the active learning process, and this additional flexibility may lead to significantly improved results over just the uncertainty metric (see Sec. 5). We also highlight that this flexibility means that our approach can always be instantiated with any state-of-the-art notion of informativeness, and consequently, can scale with future advances in appropriate notions of informativeness widely studied in literature.

## 4. Analysis

In this section, we present the theoretical guarantees of our algorithm in the learning with sleeping experts setting. Our main result is an instance-dependent bound on the dynamic regret – the strongest notion of regret – of our approach in the active learning setting. We focus on the key ideas in this section and refer the reader to the Sec. A of the appendix for the full proofs and a detailed discussion of the generalization to the batch setting.

The main idea of our analysis is to show that ADAPROD<sup>+</sup> (Alg. 2), which itself is an extension of Optimistic AdaptML-Prod (Wei et al., 2017), retains the adaptive regret guarantees of the time-varying variant of their algorithm without having to know the number of experts a priori of (Wei et al., 2017). Inspired by AdaNormalHedge (Luo & Schapire, 2015), we show that our algorithm can efficiently ensure adaptive regret. In particular, at time step  $t \in [T]$  the run-

so that  $\sum_i \min\{\beta p_{t,i}, 1\} = b_t$  assuming that such a  $\beta$  is finite.

ning time per update step (i.e., a full update of the distribution  $p_t$ ) of Alg. 2 is dominated by running a binary search over  $\alpha \in [0, 1]$  as described in 3; each binary step evaluation takes  $\mathcal{O}(tN_t)$  with  $N_t = \sum_{i=1}^n \mathcal{I}_{t,i} \leq n$  for all  $t \in [T]$  where the inequality comes from the fact that we select a new unlabeled data point at time step  $t$ . Hence,  $\alpha$  can be approximated within error  $1/T^2$  in  $2 \log(T)$  iterations, leads to a total runtime of  $\tilde{\mathcal{O}}(tN_t)$  per update step, where we use  $\tilde{\mathcal{O}}(\cdot)$  to suppress  $\log T$  and  $\log n$  factors.

Our second contribution, inspired by the work of (Wintberger, 2017), hinges on the simple observation that a vast amount of prior work in prediction with expert advice relies on the celebrated Prod inequality  $\log(1+x) \geq x - x^2$  for  $x \geq -1/2$  (Cesa-Bianchi & Lugosi, 2006; Hoeven et al., 2018) to establish second-order regret bounds. The constraint that  $x \geq -1/2$  imposes an upper bound on the learning rates to be at most  $1/4$  (Wei et al., 2017), which may lead to overly-conservative plays in practice (see comparisons with (Wei et al., 2017) in Sec. B of the appendix). A closer inspection of the analysis for optimistic algorithms reveals that we can leverage predictions to get away with larger learning rates, which leads to significant gains in practice without compromising any of the theoretical guarantees.

We then combine the adaptive regret bounds inherited from (Wei et al., 2017) with Theorem 4 of (Luo & Schapire, 2015) to conclude the main regret guarantee of our algorithm adapted to the setting where we predict the next round’s loss to be the most recently observed loss, i.e.,  $\hat{\ell}_t = \ell_{t-1}$ .

**Theorem 1.** *The (expected) dynamic regret of AdaProd<sup>+</sup> (Alg. 2) as in (1) over  $T$  active learning iterations (with batch size  $b_t = 1$ ) is bounded by*

$$\text{Regret}(p_1, \dots, p_t) \leq \tilde{\mathcal{O}} \left( \sqrt{(1 + \mathcal{V}(\ell_{1:T})) \mathcal{V}_+(u_{1:T}^*)} \right),$$

where  $u_{1:T}^* = (u_1^*, \dots, u_T^*)$  is such that  $\forall t \in [T]$ ,  $u_t^* = \arg\min_{u \in \mathcal{A}_t} \langle u, \ell_t \rangle$ ,  $\mathcal{V}(\ell_{1:T}) = \sum_{t=1}^T \|\ell_t - \ell_{t-1}\|_\infty^2$ , is the temporal variation in losses, and finally,  $\mathcal{V}_+(u_{1:T}^*) = \sum_{t=1}^T \sum_{i=1}^n \max\{0, u_{t,i}^* - u_{t-1,i}^*\}$  is the total (positive) variation in the time-varying optimal distribution.

## 5. Results

In this section, we present evaluations of our algorithm and its variants, and compare its performance to popular active learning strategies on common vision tasks. We refer the interested reader to the supplementary material for additional experimental results on ImageNet and CIFAR10 as well as comparisons of ADAPROD<sup>+</sup> to existing online learning algorithms in the active learning setting.

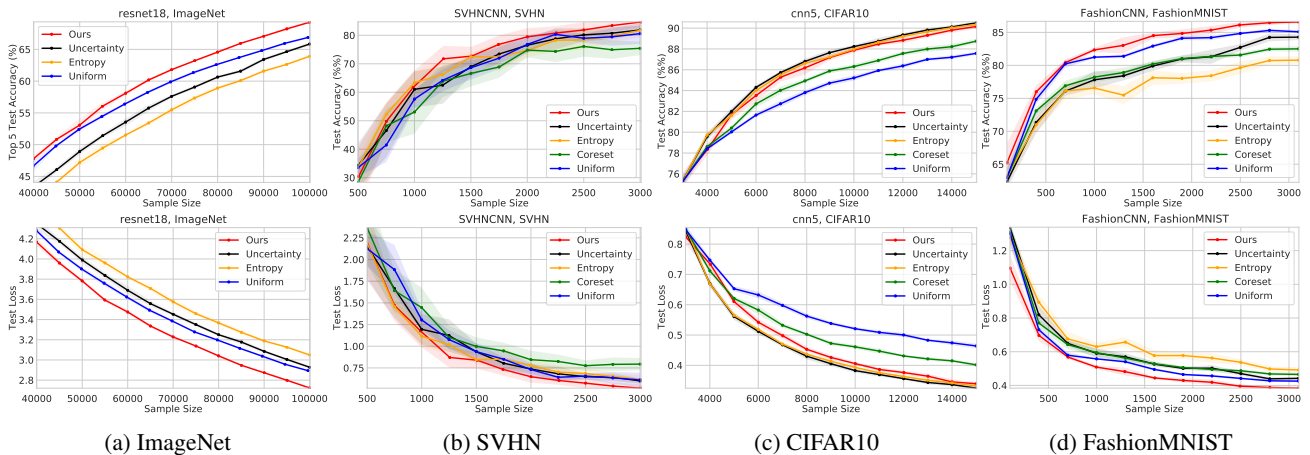


Figure 2. Evaluations on popular computer vision benchmarks trained on modern convolutional neural networks. Our algorithm is consistently ensures higher performance than that of uniform sampling and that of competing approaches, with the exception of CIFAR10. This is in contrast to the highly varying performance of prior (greedy) approaches, which leads to unreliable performance.

### 5.1. Setup

We compare our active learning algorithm Alg. 2 (labeled **Ours**) with the loss described in Sec. 3 (unless otherwise stated) to **Uncertainty**: a greedy selection approach based on the informativeness described in Sec. 2; **Entropy**: a greedy approach that defines informativeness by the entropy of the network’s softmax output; **Coreset**<sup>5</sup>: the clustering-based active learning algorithm of (Sener & Savarese, 2017; Geifman & El-Yaniv, 2017); and naive **Uniform** sampling. The algorithms were implemented in Python and we used the PyTorch (Paszke et al., 2017) library for our deep learning evaluations. The full codebase containing the implemented algorithms and code to run and replicate our results can be found in the supplementary material.

We consider the following popular vision data sets trained on modern convolutional networks for our evaluations:

1. **FashionMNIST**(Xiao et al., 2017): 60,000 grayscale images of size  $28 \times 28$
2. **CIFAR10** (Krizhevsky et al., 2009): 50,000 color images ( $32 \times 32$ ) each belonging to one of 10 classes
3. **SVHN** (Netzer et al., 2011): 73,257  $32 \times 32$  real-world images of digits taken from Google Street View
4. **ImageNet** (Deng et al., 2009): more than 1.2 million images spanning 1000 classes

We used standard convolutional networks for training FashionMNIST (Xiao et al., 2017) and SVHN (Chen, 2020), and the CNN5 architecture (Nakkiran et al., 2019) and residual networks (resnets) (He et al., 2016) for our evaluations

<sup>5</sup>Following standard methodology (Sinha et al., 2019), we implemented the greedy version of the algorithm (Gal et al., 2017) proposed in (Sener & Savarese, 2017).

on CIFAR10 and ImageNet. The networks were trained with optimized hyper-parameters from literature and with data augmentation and normalization. The full set of hyper-parameter specifications and additional details about the experimental setup are in the supplementary material. All results are averaged over 10 trials unless otherwise stated.

### 5.2. Evaluations on Vision Tasks

As our initial experiment, we evaluate and compare the performance of our approach on benchmark computer vision applications. Fig. 2 depicts the results of our experiments on the data sets evaluated with respect to test accuracy and test loss of the obtained network. For these experiments, we used the standard methodology (Ren et al., 2020; Gal et al., 2017) of retraining the network from scratch as the option in Alg. 1 (i.e., starting from a randomly initialized network) after each data acquisition step. We note that we were not able to evaluate the Coreset algorithm on ImageNet due to the extensive memory requirements of the pairwise distance computations that even the greedy variant requires, and in a practical sense, this underscores the scalability of our approach in practice.

Note that for all data sets, our algorithm (shown in red) consistently outperforms uniform sampling, and in fact, also leads to reliable and significant improvements over existing approaches for all data sets, except CIFAR10, where we fare only slightly worse than greedy strategies (but even here, we perform roughly 3 – 4% better than uniform in terms of absolute test accuracy). Perhaps most notably, our approach shines brightest on the most difficult scenario: ImageNet. On ImageNet, we consistently perform at least 3 – 6% better than competitors, especially uniform sampling, as measured by test accuracy, and orders of magnitude better in terms

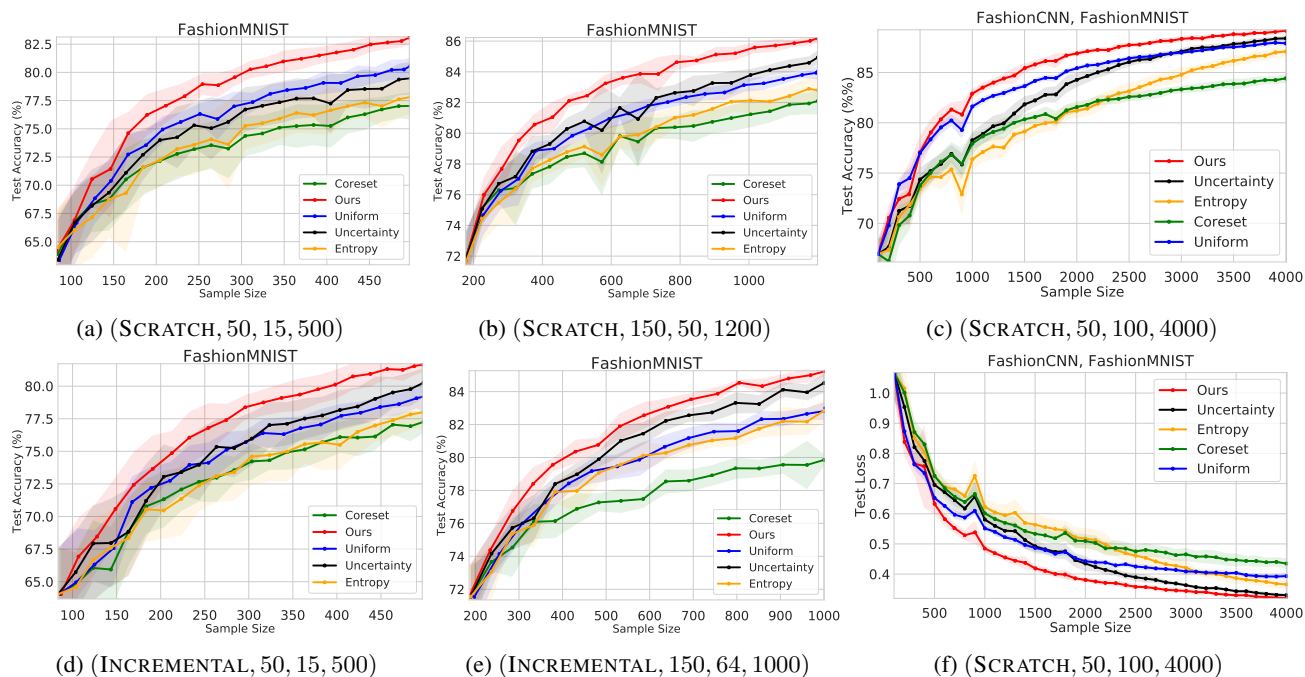


Figure 3. Our evaluations on the FashionMNIST data set with varying data acquisition configurations (OPTION,  $n_{\text{start}}$ ,  $b$ ,  $n_{\text{end}}$ ) and incremental or scratch options. In addition to depicting the test accuracy, we show the **test loss** for the longest time horizon setting in the bottom right figure. The performance of competing methods varies significantly across configurations even when the data set is fixed.

of test loss. Our findings support the scalability of our approach and show promise for its effectiveness on even larger data sets.

### 5.3. Robustness Evaluations

Next, we investigate the robustness of the considered approaches across varying data acquisition configurations evaluated on a fixed data set. To this end, we define a data acquisition configuration as the tuple (OPTION,  $n_{\text{start}}$ ,  $b$ ,  $n_{\text{end}}$ ) where OPTION is either SCRATCH or INCREMENTAL in the context of Alg. 1,  $n_{\text{start}}$  is the number of initial points at the first step of the active learning iteration,  $b$  is the fixed label budget per iteration, and  $n_{\text{end}}$  is the number of points at which the active learning process stops. Intuitively, we expect robust active learning algorithms to be resilient to changes in the data acquisition configuration and to outperform uniform sampling in a configuration-agnostic way.

Fig. 3 shows the results<sup>6</sup> of our experiments on FashionMNIST. From the figures, we can see that our approach performs significantly better than the compared approaches in terms of both test accuracy and loss in all evaluated configurations. On the other hand, the performance of competing approaches fluctuates wildly. For instance, we can see that the uncertainty metric in Fig. 3 fares worse than naive uni-

form sampling in (a), but outperforms uniform sampling in settings (d) and (e); curiously, in (c), it is only better after an interesting cross-over point towards the end.

This inconsistency and sub-uniform performance is even more pronounced for the Entropy and Coreset algorithms that tend to perform significantly worse – up to -7% and -4% (see (a) and (e) in Fig. 3) absolute test accuracy when compared to that of our method and uniform sampling, respectively. We postulate that the poor performance of these competing approaches predominantly stems from their inherently greedy acquisition of data points in a setting with significant randomness as a result of stochastic training and data augmentation, among other elements. In contrast, our approach has provably low-regret with respect to the data acquisition objective, and this provable property of our algorithm translates to consistent performance across varying configurations and settings.

### 5.4. Boosting Prior Approaches

Despite the favorable results presented in the previous subsections, a lingering question still remains: to what extent is our choice of the loss function we formulated in Sec. 3.5 responsible for the effectiveness of our approach? More generally, can we expect our algorithm to perform well off-the-shelf – and even lead to improvements over greedy acquisition – with other choices for the loss?

<sup>6</sup>The full set of figures containing the results of additional configurations and can be found in the supplementary.

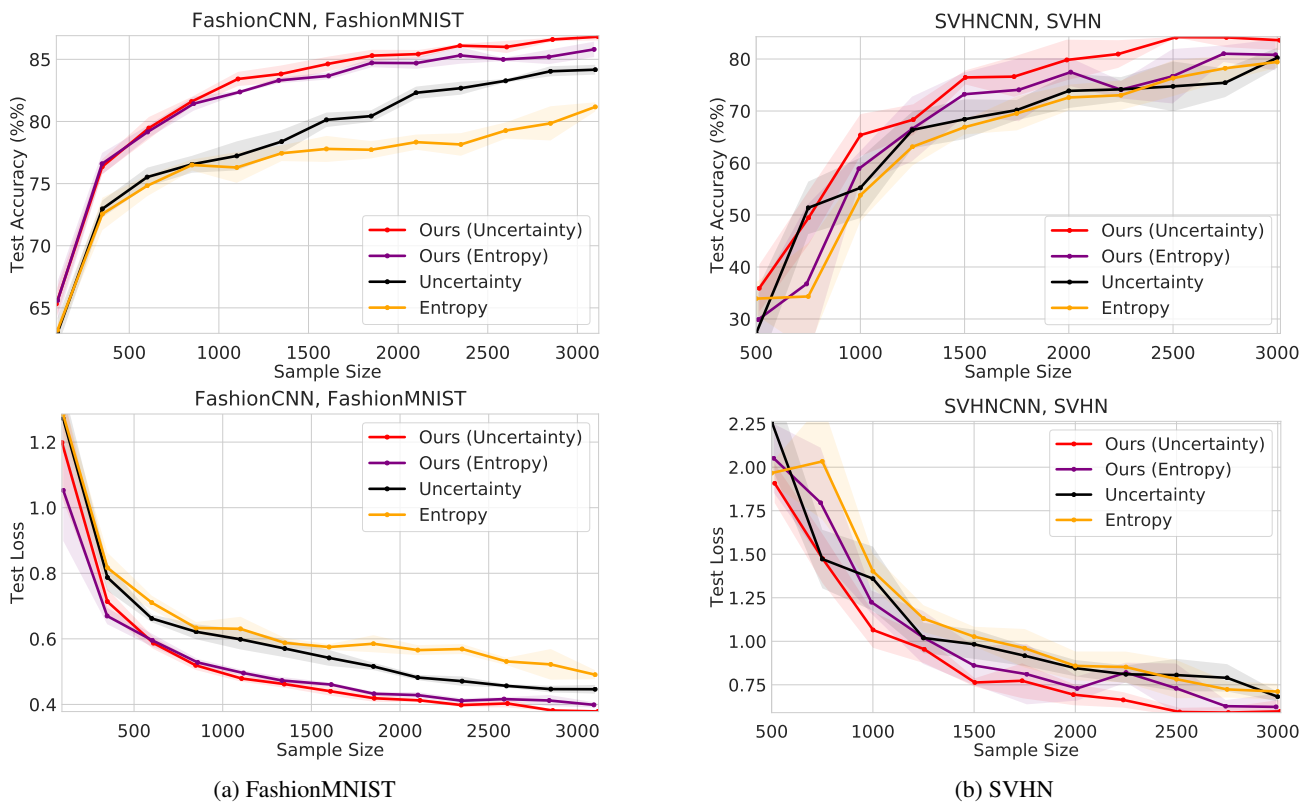


Figure 4. The performance of our algorithm when instantiated with informativeness metrics from prior work compared to that of existing greedy approaches. Despite no changes to the informativeness metric, using ADAPROD<sup>+</sup> off-the-shelf leads to improved performance.

To investigate, we implement two variants of our approach, *Ours (Uncertainty)* and *Ours (Entropy)*, that are instantiated with losses defined in terms of uncertainty and entropy, respectively, and compare them to the corresponding greedy variants on SVHN and FashionMNIST. We note that the uncertainty loss corresponds to  $\ell_{t,i} = \max_j f_t(x_i)_j \in [0, 1]$  and readily fits in our framework. For the entropy loss, the application is only slightly more nuanced in that we have to be careful that losses are bounded in the interval  $[0, 1]$ . This can be done by scaling the entropy by its maximum value  $\log(k)$  for a classification task with  $k$  classes, i.e.,  $\ell_{t,i} = 1 - \text{ENTROPY}(f_t(x_i)) / \log(k)$ .

The performance of the compared algorithms are shown in Fig. 4. Note that for both data sets and for both definition of losses, our approach is significantly better than its greedy counterpart (compare the red curve to black, and purple to orange). In other words, applying our approach off-the-shelf with existing informativeness measures from literature leads to strict improvements compared to existing work. As seen from Fig. 4(a), our approach has potential to yield up to a 6% increase in test accuracy for the same sample size. It is also worth noting that our approach also leads to consistently lower test loss in both of the evaluated scenarios.

## 6. Conclusion

In this paper, we introduced a low-regret active learning approach based on formulating the problem of data acquisition as that of prediction with experts. Building on our insights on the existing research gap in active learning, we introduced an efficient algorithm that is tailored to achieve low regret on predictable instances while remaining resilient to adversarial ones. Our empirical evaluations on large-scale real-world data sets and architectures substantiate the robustness of our approach and show that it leads to consistent and significant improvements over existing work.

Perhaps most importantly, our analysis and evaluations suggest that ADAPROD<sup>+</sup> can be applied off-the-shelf with existing informativeness measures to improve upon prior work. In fact, owing to the generality and applicability of our approach, we envision that it can be coupled with existing and future advances in uncertainty or informativeness quantification, e.g., (Amini et al., 2019), and applied to active learning. In this regard, we hope that the work presented here can contribute to the advancement of robust active learning, so that reliably effective data acquisition approaches can one day become an ordinary part of every practitioner’s toolkit, just like Adam and SGD have for stochastic optimization.



## References

- Amini, A., Schwarting, W., Soleimany, A., and Rus, D. Deep evidential regression. *arXiv preprint arXiv:1910.02600*, 2019. 8
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020. 1
- Cesa-Bianchi, N. and Lugosi, G. *Prediction, learning, and games*. Cambridge university press, 2006. 5, 12, 19
- Chen, A. Pytorch playground. <https://github.com/aaron-xichen/pytorch-playground>, 2020. 6, 16
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009. 6
- Ducoffe, M. and Precioso, F. Adversarial active learning for deep networks: a margin based approach. *arXiv preprint arXiv:1802.09841*, 2018. 1
- Ebrahimi, S., Gan, W., Salahi, K., and Darrell, T. Minimax active learning. *arXiv preprint arXiv:2012.10467*, 2020. 1
- Feng, X., Jiang, Y., Yang, X., Du, M., and Li, X. Computer vision algorithms and hardware implementations: A survey. *Integration*, 69:309–320, 2019. 1
- Gaillard, P., Stoltz, G., and Van Erven, T. A second-order bound with excess losses. In *Conference on Learning Theory*, pp. 176–196. PMLR, 2014. 3, 11, 12, 13, 14, 17, 19
- Gal, Y., Islam, R., and Ghahramani, Z. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pp. 1183–1192. PMLR, 2017. 1, 2, 6
- Geifman, Y. and El-Yaniv, R. Deep active learning over the long tail. *arXiv preprint arXiv:1711.00941*, 2017. 2, 6
- Gissin, D. and Shalev-Shwartz, S. Discriminative active learning. *arXiv preprint arXiv:1907.06347*, 2019. 3
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. 6, 16
- Hoeven, D., Erven, T., and Kotłowski, W. The many faces of exponential weights in online learning. In *Conference On Learning Theory*, pp. 2067–2092. PMLR, 2018. 5
- Kleinberg, R., Niculescu-Mizil, A., and Sharma, Y. Regret bounds for sleeping experts and bandits. *Machine learning*, 80(2):245–272, 2010. 3
- Koolen, W. M. and Van Erven, T. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pp. 1155–1175. PMLR, 2015. 3, 5, 11, 18, 19
- Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009. 6
- Liebenwein, L. Provable pruning. [https://github.com/lucaslie/provable\\_pruning](https://github.com/lucaslie/provable_pruning), 2021. 16
- Luo, H. Interval Regret. <https://haipeng-luo.net/courses/CSCI699/lecture9.pdf>, 2017. [Online; accessed November-2020]. 18
- Luo, H. and Schapire, R. E. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pp. 1286–1304, 2015. 3, 4, 5, 11, 14, 15, 18, 19
- Mohri, M. and Yang, S. Accelerating optimization via adaptive prediction. *arXiv preprint arXiv:1509.05760*, 2015. 4
- Muthakana, H. *Uncertainty and diversity in deep active image classification*. PhD thesis, Carnegie Mellon University Pittsburgh, PA, 2019. 1, 2
- Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., and Sutskever, I. Deep double descent: Where bigger models and more data hurt. *arXiv preprint arXiv:1912.02292*, 2019. 6, 16
- Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., and Ng, A. Y. Reading digits in natural images with unsupervised feature learning. 2011. 6
- Orabona, F. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019. 3, 4, 19
- Pankajj. Fashion MNIST with PyTorch. <https://www.kaggle.com/pankajj/fashion-mnist-with-pytorch-93-accuracy>, 2018. [Online; accessed November-2020]. 16
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. 2017. 6
- Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *Conference on Learning Theory*, pp. 993–1019. PMLR, 2013. 4
- Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., and Wang, X. A survey of deep active learning. *arXiv preprint arXiv:2009.00236*, 2020. 1, 6

- Saha, A., Gaillard, P., and Valko, M. Improved sleeping bandits with stochastic action sets and adversarial rewards. In *International Conference on Machine Learning*, pp. 8357–8366. PMLR, 2020. 3
- Sener, O. and Savarese, S. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017. 1, 2, 6
- Shayestehmanesh, H., Azami, S., and Mehta, N. A. Dying experts: Efficient algorithms with optimal regret bounds. *arXiv preprint arXiv:1910.13521*, 2019. 3
- Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., and Sieh, W. Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, 9(1):1–12, 2019. 1
- Sinha, S., Ebrahimi, S., and Darrell, T. Variational adversarial active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5972–5981, 2019. 6
- Steinhardt, J. and Liang, P. Adaptivity and optimism: An improved exponentiated gradient algorithm. In *International Conference on Machine Learning*, pp. 1593–1601. PMLR, 2014. 4
- Wei, C.-Y., Hong, Y.-T., and Lu, C.-J. Tracking the best expert in non-stationary stochastic environments. *arXiv preprint arXiv:1712.00578*, 2017. 4, 5, 11, 12, 14, 17
- Wintenberger, O. Optimal learning with bernstein online aggregation. *Machine Learning*, 106(1):119–141, 2017. 4, 5, 19
- Xiao, H., Rasul, K., and Vollgraf, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 1, 6

## A. Analysis

In this section, we present the full proofs and technical details of the claims made in Sec. 4. The outline of our analysis as follows. We first consider the *base* ADAPROD<sup>+</sup> algorithm (shown as Alg. 3), which is nearly the same algorithm as ADAPROD<sup>+</sup>, with the exception that it is meant to be a general purpose algorithm for a setting with  $K$  experts ( $K$  is not necessarily equal to the number of points  $n$ ). We show that this algorithm retains the original regret guarantees with respect to a stationary competitor of Adapt-ML-Prod.

We then consider the thought experiment where we use this standard version of our algorithm with the  $K = nT$  sleeping experts reduction shown in (Wei et al., 2017; Gaillard et al., 2014) to obtain guarantees for adaptive regret. This leads us to the insight (as in (Luo & Schapire, 2015; Koolen & Van Erven, 2015)) that we do not need to keep track of the full set of  $K$  experts, and can instead keep track of a much smaller (but growing) set of experts in an efficient way without compromising the theoretical guarantees.

---

### Algorithm 3 BASE ADAPROD<sup>+</sup>

---

- 1: For all  $i \in [K]$ ,  $C_{i,0} \leftarrow 0$ ;  $\eta_{0,i} \leftarrow 1$ ;  $w_{0,i} = 1$ ;  $\hat{r}_{1,i} = 0$ ;
- 2: **for** each round  $t \in [T]$  **do**
- 3:    $p_{t,i} \leftarrow \eta_{t-1,i} w_{t-1,i} \exp(\eta_{t-1,i} \hat{r}_{t,i})$  for each  $i \in [K]$
- 4:    $p_{t,i} \leftarrow p_{t,i} / \sum_{j \in [K]} p_{t,j}$  for each  $i \in [K]$  {Normalize}
- 5:   Adversary reveals  $\ell_t$  and we suffer loss  $\tilde{\ell}_t = \langle \ell_t, p_t \rangle$
- 6:   For all  $i \in [K]$ , set  $r_{t,i} \leftarrow \tilde{\ell}_t - \ell_{t,i}$
- 7:   For all  $i \in [K]$ , set  $C_{t,i} \leftarrow C_{t-1,i} + (\hat{r}_{t,i} - r_{t,i})^2$
- 8:   Get prediction  $\hat{r}_{t+1} \in [-1, 1]^K$  for next round (see Sec. 3.3)
- 9:   For all  $i \in [K]$ , update the learning rate

$$\eta_{t,i} \leftarrow \min \left\{ \eta_{t-1,i}, \frac{2}{3(1 + \hat{r}_{t+1,i})}, \sqrt{\frac{\log(K)}{1 + C_{t,i}}} \right\}$$

- 10:   For all  $i \in [K]$ , update the weights

$$w_{t,i} \leftarrow (w_{t-1,i} \exp(\eta_{t-1,i} r_{t,i} - \eta_{t-1,i}^2 (r_{t,i} - \hat{r}_{t,i})^2))^{\eta_{t,i} / \eta_{t-1,i}}$$

- 11: **end for**
- 

#### A.1. Recovering Optimistic Adapt-ML-Prod Guarantees for Alg. 3

We begin by observing that Alg. 3 builds on the standard Optimistic Adapt-ML-Prod algorithm (Wei et al., 2017) by using a different initialization of the variables (Line 1) and upper bound imposed on the learning rates (as in Alg. 2, and analogously, in Line 9 of Alg. 3). Hence, the proof is has the same structure as (Wei et al., 2017; Gaillard et al., 2014), and we prove all of the relevant claims (at times, in slightly different ways) below for clarity and completeness. We proceed with our key lemma about the properties of the learning rates.

**Lemma 2** (Properties of Learning Rates). *Assume that the losses are bounded  $\ell_t \in [0, 1]^K$  and that the learning rates  $\eta_{t,i}$  are set according to Line 9 of Alg. 3 for all  $t \in [T]$  and  $i \in [K]$ , i.e.,*

$$\eta_{t,i} \leftarrow \min \left\{ \eta_{t-1,i}, \frac{2}{3(1 + \hat{r}_{t+1,i})}, \sqrt{\frac{\log(K)}{1 + C_{t,i}}} \right\}.$$

Then, all of the following hold for all  $t + 1 \in [T]$  and  $i \in [K]$ :

1.  $\eta_{t,i} (r_{t+1,i} - \hat{r}_{t+1,i}) - \eta_{t,i}^2 (r_{t+1,i} - \hat{r}_{t+1,i})^2 \leq \log(1 + \eta_{t,i} (r_{t+1,i} - \hat{r}_{t+1,i}))$ ,
2.  $x \leq x^{\eta_{t+1,i} / \eta_{t,i}} + 1 - \frac{\eta_{t+1,i}}{\eta_{t,i}} \quad \forall x \geq 0$ ,

$$3. \frac{\eta_{t,i} - \eta_{t+1,i}}{\eta_{t,i}} \leq \log(\eta_{t,i}/\eta_{t+1,i}).$$

*Proof.* For the first claim, observe that the range of admissible values in the original Prod inequality (Cesa-Bianchi & Lugosi, 2006)

$$\forall x \geq -1/2 \quad x - x^2 \leq \log(1+x)$$

can be improved<sup>7</sup> to  $\forall x \geq -2/3$ . Now let  $x = \eta_{t,i}(r_{t+1,i} - \hat{r}_{t+1,i})$ , and observe that since  $\ell_t \in [0, 1]^K$ , we have  $r_{t+1,i} = \langle p_{t+1}, \ell_{t+1} \rangle - \ell_{t+1} \in [-1, 1]$ , and so

$$\begin{aligned} x &\geq \eta_{t,i}(-1 - \hat{r}_{t+1,i}) = -\eta_{t,i}(1 + \hat{r}_{t+1,i}) \\ &\geq -2/3, \end{aligned}$$

where in the last inequality we used the upper bound on  $\eta_{t,i} \leq 2/(3(1 + \hat{r}_{t+1,i}))$  which holds by definition of the learning rates.

For the second claim, recall Young's inequality<sup>8</sup> which states that for non-negative  $a, b$ , and  $p \geq 1$ ,

$$ab \leq a^p/p + b^{p/(p-1)}(1 - 1/p).$$

For our application, we set  $a = x$ ,  $b = 1$ , and  $p = \eta_{t,i}/\eta_{t+1,i}$ . Observe that  $p$  is indeed greater than 1 since the learning rates are non-increasing over time (i.e.,  $\eta_{t+1,i} \leq \eta_{t,i}$  for all  $t$  and  $i$ ) by definition. Applying Young's inequality, we obtain

$$x \leq x^{\eta_{t,i}/\eta_{t+1,i}}(\eta_{t+1,i}/\eta_{t,i}) + \frac{\eta_{t,i} - \eta_{t+1,i}}{\eta_{t,i}},$$

and the claim follows by the fact that the learning rates are non-increasing.

For the final claim, observe that the derivative of  $\log(x)$  is  $1/x$ , and so by the mean value theorem we know that there exists  $c \in [\eta_{t+1,i}, \eta_{t,i}]$  such that

$$\frac{\log(\eta_{t,i}) - \log(\eta_{t+1,i})}{\eta_{t,i} - \eta_{t+1,i}} = \frac{1}{c}.$$

Rearranging and using  $c \leq \max\{\eta_{t,i}, \eta_{t+1,i}\} = \eta_{t,i}$ , we obtain

$$\log(\eta_{t,i}/\eta_{t+1,i}) = \frac{\eta_{t,i} - \eta_{t+1,i}}{c} \geq \frac{\eta_{t,i} - \eta_{t+1,i}}{\eta_{t,i}}.$$

□

Having established our helper lemma, we now proceed to bound the regret with respect to a single expert as in (Wei et al., 2017; Gaillard et al., 2014). The main statement is given by the lemma below.

**Lemma 3** (BASE ADAPROD<sup>+</sup> Static Regret Bound). *The static regret of Alg. 3 with respect to any expert  $i \in [K]$ ,  $\sum_t r_{t,i}$ , is bounded by*

$$\mathcal{O}\left(\log K + \log \log T + (\sqrt{\log K} + \log \log T)\sqrt{(C_{T,i})}\right),$$

where  $C_{T,i} = \sum_{t \in [T]} (r_{t,i} - \hat{r}_{t,i})^2$ .

*Proof.* Consider  $W_t = \sum_{i \in [K]} w_{t,i}$  to be the sum of *potentials* at round  $t$ . We will first show an upper bound on the potentials and then show that this sum is an upper bound on the regret of any expert (plus some additional terms). Combining the upper and lower bounds will lead to the statement of the lemma. To this end, we first show that the sum of potentials does not increase too much from round  $t$  to  $t+1$ . To do so, we apply (2) from Lemma 2 with  $x = w_{t+1,i}$  to obtain for each  $w_{t+1,i}$

$$w_{t+1,i} \leq w_{t+1,i}^{\eta_{t,i}/\eta_{t+1,i}} + \frac{\eta_{t,i} - \eta_{t+1,i}}{\eta_{t,i}}.$$

<sup>7</sup>By inspection of the root of the function  $g(x) = \log(1+x) - x + x^2$  closest to  $x = -1/2$ , which we know exists since  $g(-1/2) > 0$  while  $g(-1) < 0$ .

<sup>8</sup>This follows by taking logarithms and using the concavity of the logarithm function.

Now consider the first term on the right hand side above and note that

$$\begin{aligned}
 w_{t+1,i}^{\eta_{t,i}/\eta_{t+1,i}} &= w_{t,i} \exp(\eta_{t,i} r_{t+1,i} - \eta_{t,i}^2 (r_{t+1,i} - \hat{r}_{t+1,i})^2) && \text{by definition; see Line 10} \\
 &= w_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i}) \exp(\eta_{t,i} (r_{t+1,i} - \hat{r}_{t+1,i}) - \eta_{t,i}^2 (r_{t+1,i} - \hat{r}_{t+1,i})^2) && \text{adding and subtracting } \eta_{t,i} \hat{r}_{t+1,i} \\
 &\leq w_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i}) (1 + \eta_{t,i} (r_{t+1,i} - \hat{r}_{t+1,i})) && \text{by (1) of Lemma 2} \\
 &= w_{t,i} \eta_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i}) r_{t+1,i} + w_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i}) \underbrace{(1 - \eta_{t,i} \hat{r}_{t+1,i})}_{\leq \exp(-\eta_{t,i} \hat{r}_{t+1,i})} && (1 + x \leq e^x \text{ for all real } x) \\
 &\leq \underbrace{w_{t,i} \eta_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i})}_{\propto p_{t+1,i}} r_{t+1,i} + w_{t,i}.
 \end{aligned}$$

As the brace above shows, the first part of the first expression on the right hand side is proportional to  $p_{t+1,i}$  by construction (see Line 3 in Alg. 3). Recalling that  $r_{t+1,i} = \langle p_{t+1}, \ell_{t+1} \rangle - \ell_{t+1,i}$ , we have by dividing and multiplying by the normalization constant,

$$\sum_{i \in [K]} w_{t,i} \eta_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i}) r_{t+1,i} = \left( \sum_{i \in [K]} w_{t,i} \eta_{t,i} \exp(\eta_{t,i} \hat{r}_{t+1,i}) \right) \sum_{i \in [K]} p_{t+1,i} r_{t+1,i} = 0,$$

since  $\sum_{i \in [K]} p_{t+1,i} r_{t+1,i} = 0$ . This shows that  $\sum_{i \in [K]} w_{t+1,i}^{\eta_{t,i}/\eta_{t+1,i}} \leq \sum_{i \in [K]} w_{t,i} = W_t$ .

Putting it all together and applying (3) from Lemma 2 to bound  $\frac{\eta_{t,i}^{-\eta_{t+1,i}}}{\eta_{t,i}}$ , we obtain for the sum of potentials:

$$W_{t+1} \leq W_t + \sum_{i \in [K]} \log(\eta_{t,i}/\eta_{t+1,i}).$$

Given that  $W_0 = K$ , we have that for  $W_T$

$$\begin{aligned}
 W_T &\leq K + \sum_{t \in [T]} \sum_{i \in [K]} \log(\eta_{t,i}/\eta_{t+1,i}) \\
 &= K + \sum_{i \in [K]} \sum_{t \in [T]} \log(\eta_{t,i}/\eta_{t+1,i}) \\
 &= K + \sum_{i \in [K]} \log \left( \prod_{t=1}^{T-1} \eta_{t,i}/\eta_{t+1,i} \right) \\
 &= K + \sum_{i \in [K]} \log(\eta_{0,i}/\eta_{T,i}) \\
 &\leq K \left( 1 + \log \left( \max_{i \in [K]} \sqrt{1 + C_{T,i}} \right) \right) \\
 &\leq K (1 + \log(1 + 4T)/2).
 \end{aligned}$$

Now, we establish a lower bound for  $W_t$  in terms of the regret with respect to any expert  $i \in [K]$ . Taking the logarithm and using the fact that the potentials are always non-negative, we can show via a straightforward induction (as in (Gaillard et al., 2014)) that

$$\log(W_T) \geq \log(w_{T,i}) \geq \eta_{T,i} \sum_{t \in [T]} (r_{t,i} - \eta_{t-1,i} (r_{t,i} - \hat{r}_{t,i})^2).$$

Rearranging, and using the upper bound on  $W_T$  from above, we obtain

$$\sum_{t \in [T]} r_{t,i} \leq \eta_{T,i}^{-1} \log \left( K (1 + \log(\max_{i \in [K]} \sqrt{1 + C_{T,i}})) \right) + \sum_{t \in [T]} \eta_{t-1,i} (r_{t,i} - \hat{r}_{t,i})^2.$$

For the first term, consider the definition of  $\eta_{T,i}$  and note that  $\eta_{T,i} \geq \min\{1/3, \eta_{T-1,i}, \sqrt{\log(K)/(1 + C_{T,i})}\}$  since  $\hat{r}_{T+1,i} \leq 1$ . Now to lower bound  $\eta_{T,i}$ , consider the claim that  $\eta_{t,i} \geq \min\{1/3, \sqrt{\log(K)/(1 + C_{T,i})}\}$ . Note that this

claim holds trivially for the base cases where  $t = 0$  and  $t = 1$  since the learning rates are initialized to 1 and our optimistic predictions can be at most 1. By induction, we see that if this claim holds at time step  $t$ , we have for time step  $t + 1$

$$\begin{aligned} \eta_{t+1,i} &\geq \min\{1/3, \eta_{t,i}, \sqrt{\log(K)/(1 + C_{t+1,i})}\} \geq \min\{1/3, \eta_{t,i}, \sqrt{\log(K)/(1 + C_{T,i})}\} \\ &= \min\{\eta_{t,i}, \min\{1/3, \sqrt{\log(K)/(1 + C_{T,i})}\}\} \\ &\geq \min\left\{\min\{1/3, \sqrt{\log(K)/(1 + C_{T,i})}\}, \min\{1/3, \sqrt{\log(K)/(1 + C_{T,i})}\}\right\} \\ &= \min\{1/3, \sqrt{\log(K)/(1 + C_{T,i})}\}. \end{aligned}$$

Hence, we obtain  $\eta_{T,i} \geq \min\{1/3, C_{T,i}\}$ , and this implies that (by the same reasoning as in (Gaillard et al., 2014)) that

$$\eta_{T,i}^{-1} \log \left( K(1 + \log(\max_{i \in [K]} \sqrt{1 + C_{T,i}})) \right) \leq \mathcal{O} \left( (\sqrt{\log K} + \log \log T) \sqrt{1 + C_{T,i}} + \log K \right).$$

Now to bound the last term, we use the definition of  $\eta_{t-1,i}$  and Lemma 14 of (Gaillard et al., 2014) to obtain

$$\sum_{t \in [T]} \eta_{t-1,i} (r_{t,i} - \hat{r}_{t,i})^2 \leq 8 \sqrt{\log K (1 + C_{T,i})},$$

where  $C_{T,i} = \sum_{t \in [T]} (r_{t,i} - \hat{r}_{t,i})^2$  as before and this completes the proof.  $\square$

## A.2. Adaptive Regret

We now turn to establishing adaptive regret bounds via the sleeping experts reduction as in (Wei et al., 2017; Luo & Schapire, 2015) using the reduction of (Gaillard et al., 2014). The overarching goal is to establish an adaptive bound for the regret of every time interval  $[t_1, t_2]$ ,  $t_1, t_2 \in [T]$ , which is a generalization of the static regret which corresponds to the regret over the interval  $[1, T]$ . To do so, in the setting of  $n$  experts as in the main document, the main idea is to run the base algorithm (Alg. 3) on  $K = nT$  sleeping experts instead<sup>9</sup>. These experts will be indexed by  $(t, i)$  with  $t \in [T]$  and  $i \in [n]$ . Moreover, at time step  $t$ , only experts  $(s, i)$  such that  $s \leq t$ ,  $i \in [n]$  will be considered awake, and the remaining will be sleeping. This will generate a probability distribution  $\bar{p}_{t,(s,i)}$  over all  $K = nT$  experts  $(s, i)$ . Using this distribution, at round  $t$  we play

$$p_{t,i} = \sum_{s \in [t]} \bar{p}_{t,(s,i)} / Z_t,$$

where  $Z_t = \sum_{j \in [K]} \sum_{s' \in [t]} \bar{p}_{t,(s',j)}$ .

The main idea is to construct losses to give to the base algorithm so that that at any point  $t \in [T]$ , each expert  $(s, i)$  suffers the interval regret from  $s$  to  $t$  (which is defined to be 0 if  $s > t$ ), i.e.,  $\sum_{\tau=1}^t r_{\tau,(s,i)} = \sum_{\tau=s}^t r_{\tau,i}$ . To do so, we follow (Wei et al., 2017) and apply the base algorithm with the modified loss vectors  $\bar{\ell}_{t,(s,i)}$  for expert  $(s, i)$  as the original loss if the expert is awake, i.e.,  $\bar{\ell}_{t,(s,i)} = \ell_{t,i}$  if  $s \leq t$ , and  $\bar{\ell}_{t,(s,i)} = \langle p_{t,i}, \ell_t \rangle$ . The prediction vector is defined similarly:  $\bar{r}_{t,(s,i)} = \hat{r}_{t,i}$  if  $s \leq t$ , and 0 otherwise.

Note that this construction implies that the regret of the base algorithm with respect to the modified losses and predictions, i.e.,  $\bar{r}_{\tau,(s,i)} = \langle \bar{p}_{\tau,(s,i)}, \bar{\ell}_{\tau,(s,i)} \rangle$  is equivalent to  $r_{\tau,i}$  for rounds  $\tau > s$  where the expert is awake, and 0 otherwise. Thus,

$$\sum_{\tau \in [t]} \bar{r}_{\tau,(s,i)} = \sum_{\tau=s}^t r_{\tau,i},$$

which means that the regret of expert  $(s, i)$  with respect to the base algorithm is precisely regret of the interval  $[s, t]$ . Applying Lemma 3 to this reduction above (with  $K = nT$ ) immediately recovers the adaptive regret guarantee of Optimistic Adapt-ML-Prod.

<sup>9</sup>Note that this notion of sleeping experts is the same as the one we used for dealing with constructing a distribution over only the unlabeled data points remaining.

**Lemma 4** (Adaptive Regret of BASE ADAPROD<sup>+</sup>). *For any  $t_1 \leq t_2$  and  $i \in [n]$ , invoking Alg. 3 with the sleeping experts reduction described above ensures that*

$$\sum_{t=t_1}^{t_2} r_{t,i} \leq \hat{O} \left( \log(K) + \sqrt{C_{t_2,(t_1,i)} \log(K)} \right),$$

where  $C_{t_2,(t_1,i)} = \sum_{t=t_1}^{t_2} (r_{t,i} - \hat{r}_{t,i})^2$  and  $\hat{O}$  suppresses  $\log T$  factors.

### A.3. ADAPROD<sup>+</sup> and Dynamic Regret

To put it all together, we relax to requirement of having to update and keep track of  $K = NT$  experts and having to know  $T$ . To do so, observe that  $\log(K) \leq \log(nT) \leq 2 \log(n)$  since  $T \leq n / \min_{t \in [T]} b_t \leq n$ , where  $b_t \geq 1$  is the number of new points to label at active learning iteration  $t$ . This removes the requirement of having to know  $T$  or the future batch sizes beforehand, meaning that we can set the numerator of  $\eta_{t,(s,i)}$  to be  $\sqrt{2 \log(n)}$  instead of  $\sqrt{\log(K)}$  (as in 2 in Sec. 3). Next, observe that in the sleeping experts reduction above, we have

$$p_{t,i} = \sum_{s \in [t]} \bar{p}_{t,(s,i)} / Z_t,$$

where  $Z_t = \sum_{j \in [K]} \sum_{s' \in [t]} \bar{p}_{t,(s',j)}$ . But for  $s \leq t$ , by definition of  $\bar{p}_{t,(s,j)}$  and the fact that expert  $(s, i)$  is awake, we have  $\bar{p}_{t,(s,j)} \propto \eta_{t-1,(s,j)} w_{t-1,(s,j)} \exp(\eta_{t-1,(s,j)} \hat{r}_{t,j})$ , and so the normalization constant cancels from the numerator (from  $\bar{p}_{t,(s,j)}$ ) and the denominator (from the  $\bar{p}_{t,(s',j)}$  in  $Z_t = \sum_{j \in [K]} \sum_{s' \in [t]} \bar{p}_{t,(s',j)}$ ), leaving us with

$$p_{t,i} = \sum_{s \in [t]} \frac{\eta_{t,(s',j)} w_{t-1,(s',j)} \exp(\eta_{t,(s',j)} \hat{r}_{t,j})}{\gamma_t},$$

where  $\gamma_t = \sum_{j \in [K]} \sum_{s' \in [t]} \eta_{t,(s',j)} w_{t-1,(s',j)} \exp(\eta_{t,(s',j)} \hat{r}_{t,j})$ . Note that this corresponds precisely to the probability distribution played by ADAPROD<sup>+</sup>. Further, since ADAPROD<sup>+</sup> does not explicitly keep track of the experts that are asleep, and only updates the potentials  $W_{t,(s,i)}$  of those experts that are awake (for  $s \leq t$ ), ADAPROD<sup>+</sup> mimics the updates of the reduction described above involving passing of the modified losses to the base algorithm. Thus, we can conclude that ADAPROD<sup>+</sup> – overlooking the exception of a slight change in the constant of  $\eta$  (which only leads to a factor of 2 increase in regret) – leads to the same updates and generated probability distributions as the base algorithm for adaptive regret. This discussion immediately leads to the following lemma for the adaptive regret of our algorithm, very similar to the one established above except for  $\log n$  replacing  $\log T$  terms.

**Lemma 5** (Adaptive Regret of ADAPROD<sup>+</sup>). *For any  $t_1 \leq t_2$  and  $i \in [n]$ , Alg. 2 ensures that*

$$\sum_{t=t_1}^{t_2} r_{t,i} \leq \mathcal{O} \left( \log n + \log \log n + (\sqrt{\log n} + \log \log n) \sqrt{C_{t_2,(t_1,i)}} \right),$$

where  $C_{t_2,(t_1,i)} = \sum_{t=t_1}^{t_2} (r_{t,i} - \hat{r}_{t,i})^2$ .

Finally, applying the reduction from adaptive to dynamic regret (Theorem 4 of (Luo & Schapire, 2015)) and bounding

$$(r_{t,i} - \hat{r}_{t,i})^2 \leq 2(\langle p_{t,i}, \ell_{t,i} \rangle - \langle p_{t,i}, \ell_{t-1,i} \rangle)^2 + 2(\ell_{t,i} - \ell_{t-1,i})^2 \leq 4 \|\ell_t - \ell_{t-1}\|_\infty^2$$

by applying Hölder's inequality twice, we obtain the theorem established in the main manuscript.

## B. Experimental Setup & Additional Evaluations

In this section we (i) describe the experimental setup and detail hyper-parameters used for our experiments and (ii) provide additional evaluations and comparisons to supplement the results presented in the manuscript. Our code is included in the supplementary folder<sup>10</sup>.

## Low-Regret Active Learning

	FashionCNN	SVHNCNN	Resnet18	CNN5 (width=128)
loss	cross-entropy	cross-entropy	cross-entropy	cross-entropy
optimizer	Adam	Adam	SGD	Adam
epochs	60	60	80	60
epochs incremental	15	15	N/A	15
batch size	128	128	256	128
learning rate (lr)	0.001	0.001	0.1	0.001
lr decay	0.1@(50)	0.1@(50)	0.1@(30, 60)	0.1@(50)
lr decay incremental	0.1@(10)	0.1@(10)	N/A	0.1@(10)
momentum	N/A	N/A	0.9	N/A
Nesterov	N/A	N/A	No	N/A
weight decay	0	0	1.0e-4	0

Table 1. We report the hyperparameters used during training the convolutional architectures listed above corresponding to our evaluations on FashionMNIST, SVHN, CIFAR10, and ImageNet. except for the ones indicated in the lower part of the table. The notation  $\gamma@(n_1, n_2, \dots)$  denotes the learning rate schedule where the learning rate is multiplied by the factor  $\gamma$  at epochs  $n_1, n_2, \dots$  (this corresponds to MultiStepLR in PyTorch).

### B.1. Setup

Table 1 depicts the hyperparameters used for training the network architectures used in our experiments. Given an active learning configuration (OPTION,  $n_{\text{start}}$ ,  $b$ ,  $n_{\text{end}}$ ), these parameters describe the training process for each choice of OPTION as follows:

**INCREMENTAL** : we start the active learning process by acquiring and labeling  $n_{\text{start}}$  points chosen uniformly at random from the  $n$  unlabeled data points, and we train with the corresponding number of epochs and learning rate schedule listed in Table 1 under rows *epochs* and *lr decay*, respectively, to obtain  $\theta_1$ . We then proceed as in Alg. 1 to iteratively acquire  $b$  new labeled points based on the ACQUIRE function and *incrementally train* a model starting from the model from the previous iteration,  $\theta_{t-1}$ . This training is done with respect to the number of corresponding epochs and learning rate schedule shown in Table 1 under *epochs incremental* and *lr decay incremental*, respectively.

**SCRATCH** : the only difference relative to the INCREMENTAL setting is that rather than training the model starting from  $\theta_{t-1}$ , we train a model from a randomly initialized network at each active learning iteration with respect to the training parameters under *epochs* and *lr decay* in Table 1.

**Architectures** We used the following convolutional networks on the specified data sets.

1. *FashionCNN* (Pankajj, 2018) (for FashionMNIST): a network with 2 convolutional layers with batch normalization and max pooling, 3 fully connected layers, and one dropout layer with  $p = 0.25$  in (Pankajj, 2018). This architecture achieves over 93% accuracy when trained with the whole data set.
2. *SVHNCNN* (Chen, 2020) (for SVHN): a small scale convolutional model very similar to FashionCNN except there is no dropout layer.
3. *Resnet18* (He et al., 2016) (for ImageNet): an 18 layer residual network with batch normalization.
4. *CNN5* (Nakkiran et al., 2019) (for CIFAR10): a 5-layer convolutional neural network with 4 convolutional layers with batch normalization. We used the width=128 setting in the context of (Nakkiran et al., 2019).

**Settings for experiments in Sec. 5** Prior to presenting additional results and evaluations in the next subsections, we specify the experiment configurations used for the experiments shown in the main document (Sec. 5). For the corresponding experiments in Fig. 2, we evaluated on the configuration (SCRATCH, 10k, 20k, 110k) for ImageNet, (SCRATCH, 10k, 20k, 110k),

<sup>10</sup>Our codebase builds on the publicly available codebase of (Liebenwein, 2021).



(SCRATCH, 250, 250, 3000) for SVHN, (SCRATCH, 3k, 1k, 15k) for CIFAR10, and (SCRATCH, 100, 300, 3000) for FashionMNIST. For the evaluations in Fig. 4, we used (SCRATCH, 100, 250, 3000) and (SCRATCH, 400, 250, 3000) for FashionMNIST and SVHN, respectively. The models were trained with standard data normalization with respect to the mean and standard deviation of the entire training set. For ImageNet, we used random cropping to  $224 \times 224$  and random horizontal flips for data augmentation; for the remaining data sets, we used random cropping to  $32 \times 32$  ( $28 \times 28$  for FashionMNIST) with 4 pixels of padding and random horizontal flips. All presented results were averaged over 10 trials with the exception of those for ImageNet, where we averaged over 3 trials. We used the proprietary loss metric defined in Sec. 3.5 for all of the experiments presented in this supplementary, as well as for the vision tasks and FashionMNIST comparisons in the main manuscript.

### B.2. Results on Data-Starved Settings

Figure 5 depicts the results of our additional evaluations on ImageNet and FashionMNIST in the *data-starved* setting where we begin with a very small (relatively) set of data points and can only query the labels of a very small set of points at each time step. For both data sets, our approach outperforms competing ones in the various metrics considered – yielding up to 5% improvements in test accuracy.

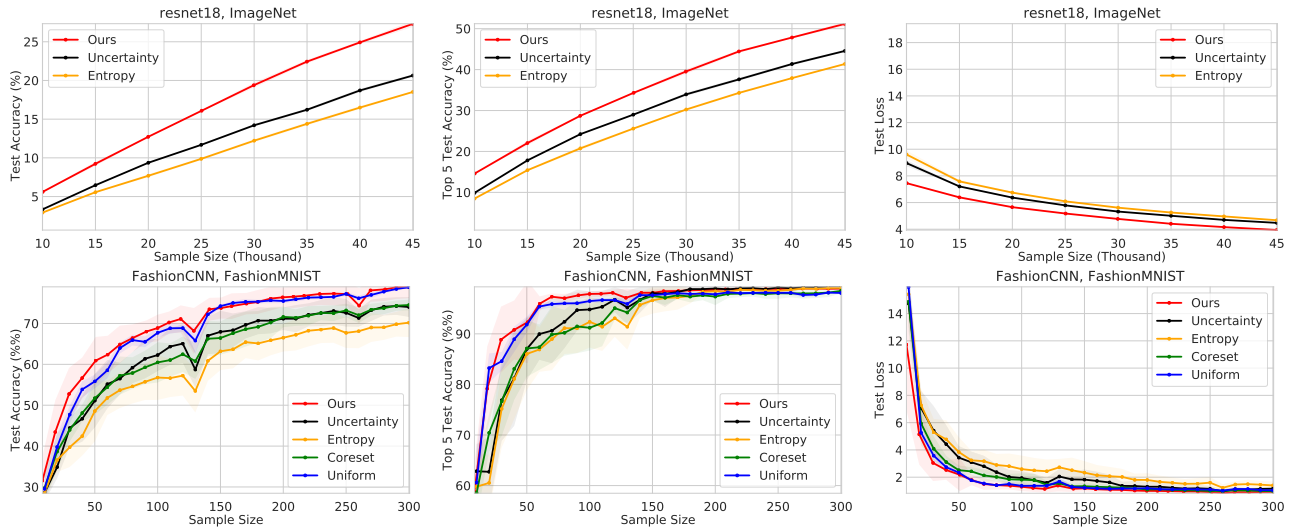


Figure 5. Results for the data-starved configuration (SCRATCH, 5k, 5k, 45k) on ImageNet (first row) and (SCRATCH, 10, 10, 300) on FashionMNIST (second row). Depicted from left to right are the results with respect to test accuracy, top-5 test accuracy, and test loss. Shaded region corresponds to values within one standard deviation of the mean.

### C. Comparison of Algorithms for Prediction with Expert Advice

In this section, we consider the performance of ADAPROD<sup>+</sup> relative to state-of-the-art algorithms for learning with prediction advice. Since our approach is intended to compete with time-varying competitors (see Sec. A), we compare it to existing methods that ensure low regret even with respect to this stronger notion of regret. In particular, we compare our approach to the following algorithms:

1. **Optimistic AMLProd** (Wei et al., 2017): we implement the (stronger) variant of Optimistic Adapt-ML-Prod that ensures dynamic regret (outlined at the end of Sec. 3.3 in (Wei et al., 2017)) that we extend in our work. This algorithm uses the sleeping experts reduction of (Gaillard et al., 2014) and consequently, requires initially creating  $\tilde{n} = nT$  sleeping experts and updating them with similar updates as in our algorithm (except the cost of the  $t^{\text{th}}$  update is  $\tilde{O}(nT)$  rather than  $\tilde{O}(N_t t)$  as in ours). Besides the computational costs, we emphasize that the only true functional difference between our algorithm and Optimistic AMLProd lies in the thresholding of the learning rates (Line 10 in Alg. 2). In our approach, we impose the upper bound  $\min\{\eta_{t-1,i}, 2/(3(1 + \hat{r}_{t+1,i}))\}$  for  $\eta_{t,i}$  for any  $t \in [T]$ , whereas (Wei et al., 2017) imposes the (smaller) bound of  $1/4$ .

2. **AdaNormalHedge(.TV)** (Luo & Schapire, 2015): we implement the time-varying version of AdaNormalHedge, AdaNormalHedge.TV as described in Sec. 5.1 of (Luo & Schapire, 2015). The only slight modification we make in our setting where we already have a sleeping experts problem is to incorporate the indicator  $\mathcal{I}_{t,i}$  in our predictions (as suggested by (Luo & Schapire, 2015) in their sleeping experts variant). In other words, we predict<sup>11</sup>  $p_{t,i} \propto \mathcal{I}_{t,i} \sum_{\tau=1}^t \frac{1}{\tau^2} w(R_{[\tau,t-1],i}, C_{[\tau,t-1]})$  rather than the original  $p_{t,i} \propto \sum_{\tau=1}^t \frac{1}{\tau^2} w(R_{[\tau,t-1],i}, C_{[\tau,t-1]})$ , where  $R_{[t_1,t_2],i} = \sum_{t=t_1}^{t_2} r_{t,i}$  and  $C_{[t_1,t_2],i} = \sum_{t=t_1}^{t_2} |r_{t,i}|$  (note that the definition of  $C$  is different than ours).
3. **Squint(.TV)** (Koolen & Van Erven, 2015): Squint is a parameter-free algorithm like AdaNormalHedge in that it can also be extended to priors over an initially unknown number of experts. Hence, we use the same idea as in AdaNormalHedge.TV (also see (Luo, 2017)) and apply the extension of the Squint algorithm for adaptive regret.

**Experiments** We conduct experiments on FashionMNIST and SVHN with SCRATCH and INCREMENTAL options and with and without data augmentation. The overarching goal of our evaluations is to assess the performance of ADAPROD<sup>+</sup> relative to competing algorithms for learning with expert advice in both easy (predictable) and hard (high temporal variation) environments. To this end, we consider the INCREMENTAL option with no (random) data augmentation to be the most benign/predictable environment, since the network changes only slightly from one active learning iteration to the next, and there is less randomness in the training and evaluation of losses. On the other hand, the SCRATCH option with data augmentation is an instance where losses may have high temporal variation, since the model may change significantly – due to training a new network from scratch and stochastic data augmentation – from one iteration to the next.

To get a sense of the general performance of the competing algorithms, we first evaluate the approaches on CIFAR10 with data augmentation and normalization. Figs. 6 and 7 depict the results of our evaluations with respect to the loss, top-1 test accuracy, and top-5 test accuracy. Despite the presence of stochastic data augmentation and the SCRATCH option, we found that on the losses did not vary too significantly on CIFAR10, which explains the favorable performance of our approach relative to non-optimistic algorithms. Moreover, these results in conjunction with the results presented in the main document (see CIFAR10 results in Fig. 2) shed light into the fact that the greedy approach indeed performs the best when the losses do not vary at all.

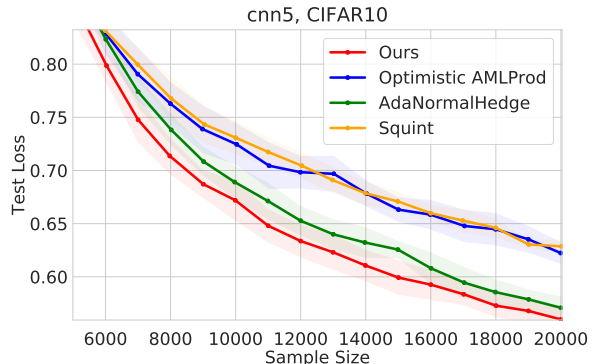


Figure 6. Evaluation of test loss in the (INCREMENTAL, 4k, 1k, 20k) configuration on the CIFAR10 dataset with data augmentation.

**Varying the Environment Variability** Figure 9 shows the results of our evaluations on the SVHN data set with two different configurations. In the easy (predictable) instance (left column), we can see that our approach leads to significant gains over competing approaches, especially non-optimistic ones. On the other hand, when there is high temporal variation in the losses, as with the SCRATCH option and data augmentation, ADAPROD<sup>+</sup> remains resilient and fares no worse (and in fact performs slightly better) than the baselines.

As further validation for the reliability of the proposed approach, we evaluate the algorithms on FashionMNIST with both INCREMENTAL and SCRATCH settings (see Fig. 8) with data augmentation. Note that despite the variability of the environment, ADAPROD<sup>+</sup> still has an edge on competing approaches. In sum, the presented results in both benign and highly varying environments support the claim that ADAPROD<sup>+</sup> can achieve the best of both worlds: it is sufficiently aggressive in picking points deemed to be highly informative on easy instances – which leads to practical gains (as seen in Figs. 6, 7, and 9) – and at the same time, is cautious enough so that it virtually never fares worse than state-of-the-art, non-optimistic algorithms with strong regret guarantees.

<sup>11</sup>We also implemented and evaluated the method with uniform prior over time intervals, i.e.,  $p_{t,i} \propto \mathcal{I}_{t,i} \sum_{\tau=1}^t w(R_{[\tau,t-1],i}, C_{[\tau,t-1]})$  (without the prior  $\frac{1}{\tau^2}$ ), but found that it performed worse than with the prior in practice. The same statement holds for the Squint algorithm.

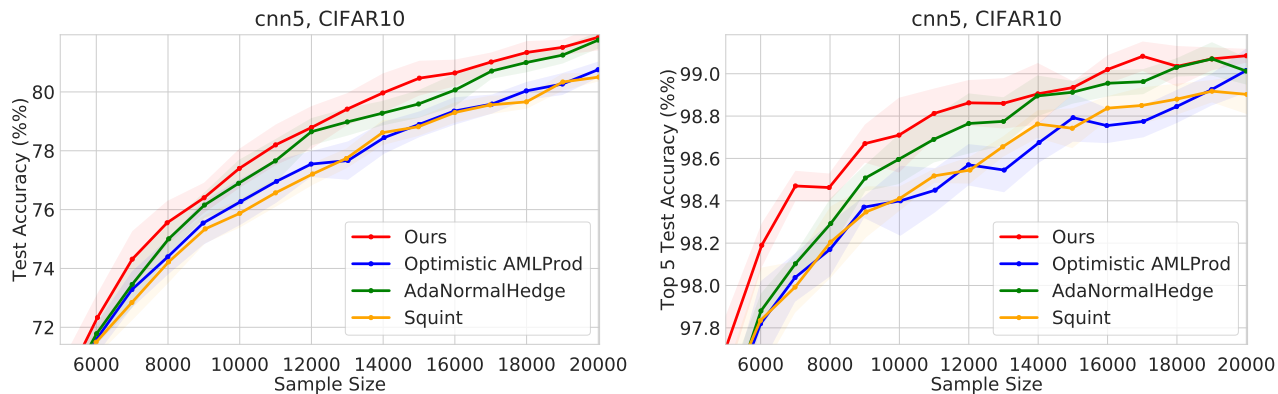


Figure 7. Evaluation of the (SCRATCH, 4k, 1k, 20k) configuration on the CIFAR10 dataset with data augmentation. Depicted from left to right are the results with respect to test accuracy, top-5 test accuracy, and test loss, averaged over 10 trials. Shaded region corresponds to values within one standard deviation of the mean.

### D. Discussion & Future Work

We discuss avenues for future work and potential improvements to our algorithm to conclude the supplementary material. Note that in this paper we introduced a prediction with expert advice algorithm that was particularly suited to active learning, where, as our empirical evaluations show, optimistic algorithms fare better on easy instances while remaining resilient to adversarial ones. Nevertheless, this does not mean that ADAPROD<sup>+</sup> is optimal for the active learning setting. In fact, we hope that this work can lead to the development of better active learning algorithms that improve on ADAPROD<sup>+</sup> and the techniques presented here.

For instance, one immediate improvement would be to extend the algorithm to be truly combinatorial, rather than solving an expert problem to obtain  $p_{t,i}$  and then modifying the obtained  $p_{t,i}$  in a way so that the expected number of points obtained is our budget  $b$ , i.e.,  $\tilde{p}_{t,i} = \min\{\beta_t p_{t,i}, 1\}$ , where  $\beta_t$  is such that  $\sum_{i \in [n]} \min\{\beta_t p_{t,i}, 1\} = b$ . Unfortunately, our attempts to bring in more principled tools from combinatorial learning, such as component iProd (Koolen & Van Erven, 2015), were to no avail because these approaches (from our experience and to the best of our knowledge) require solving for (Bregman) projections to the  $b$ -simplex, among other convex optimization routines that have to be performed *per update step*. Using off-the-shelf libraries such as CVX, as suggested by (Koolen & Van Erven, 2015), is computationally prohibitive and very memory intensive<sup>12</sup> when applied to, e.g., ImageNet. It is an open question whether the quasi-projection we perform significantly affects the order of our regret bounds (under the assumptions of active learning where  $n \gg b$ ), and/or whether more efficient (and implementable) approaches can be utilized to solve the true combinatorial problem more directly.

Other improvements could include relaxing the requirement that losses are in the  $[0, 1]$  and to have an algorithm whose regret scales with the range of the losses could lead to significant practical benefits, or even better, a scale-free algorithm (Orabona, 2019). We conjecture that this may be possible by extending Bernstein Online Aggregation (BOA) (Wintenberger, 2017) so that it is optimistic. In a similar vein, as evident from the results in Sec. B, AdaNormalHedge performs particularly well (relative to the other baseline algorithms) in the active learning scenario – despite having worse theoretical guarantees (second order vs. first order excess loss) than Squint. It would be interesting to investigate whether AdaNormalHedge can be extended to be optimistic, and whether that would yield improved results in practical settings<sup>13</sup>.

<sup>12</sup>We experienced out of memory errors with CVX on even moderately sized data sets.

<sup>13</sup>The authors of this manuscript originally tried extending the AdaNormalHedge algorithm to be optimistic to no avail. In fact, it is not even known (Luo & Schapire, 2015) whether AdaNormalHedge can even be extended (or proven) to enjoy the stronger second order excess loss bound, as Prod and Adapt-ML-Prod algorithms do (Gaillard et al., 2014; Cesa-Bianchi & Lugosi, 2006).

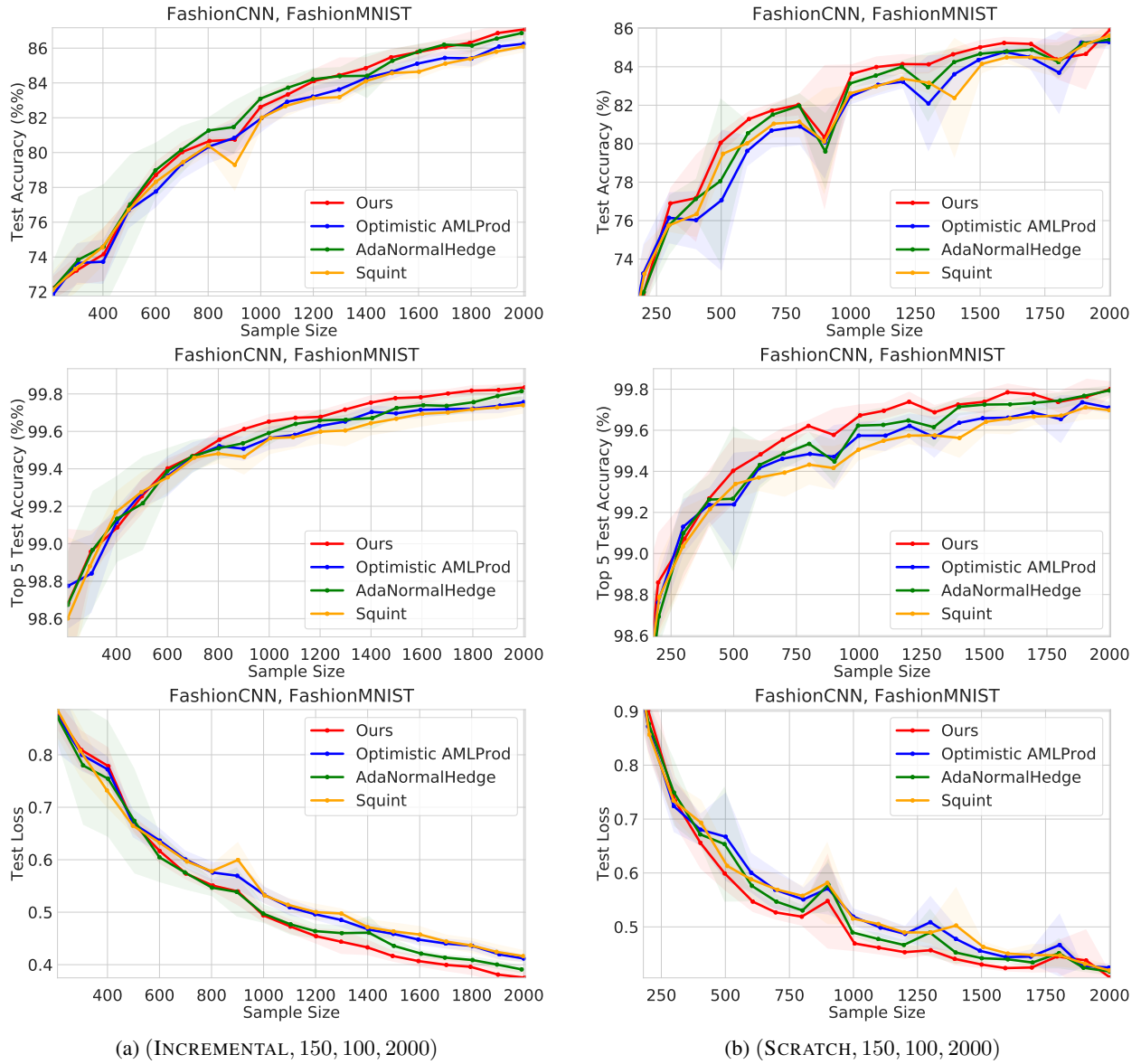


Figure 8. The results of experiments on FashionMNIST trained **with** data augmentation for the INCREMENTAL (left) and SCRATCH (right) options. The rows report the top-1 test accuracy, top-5 test accuracy, and test-loss, respectively, for each option. Results were averaged over 5 trials.

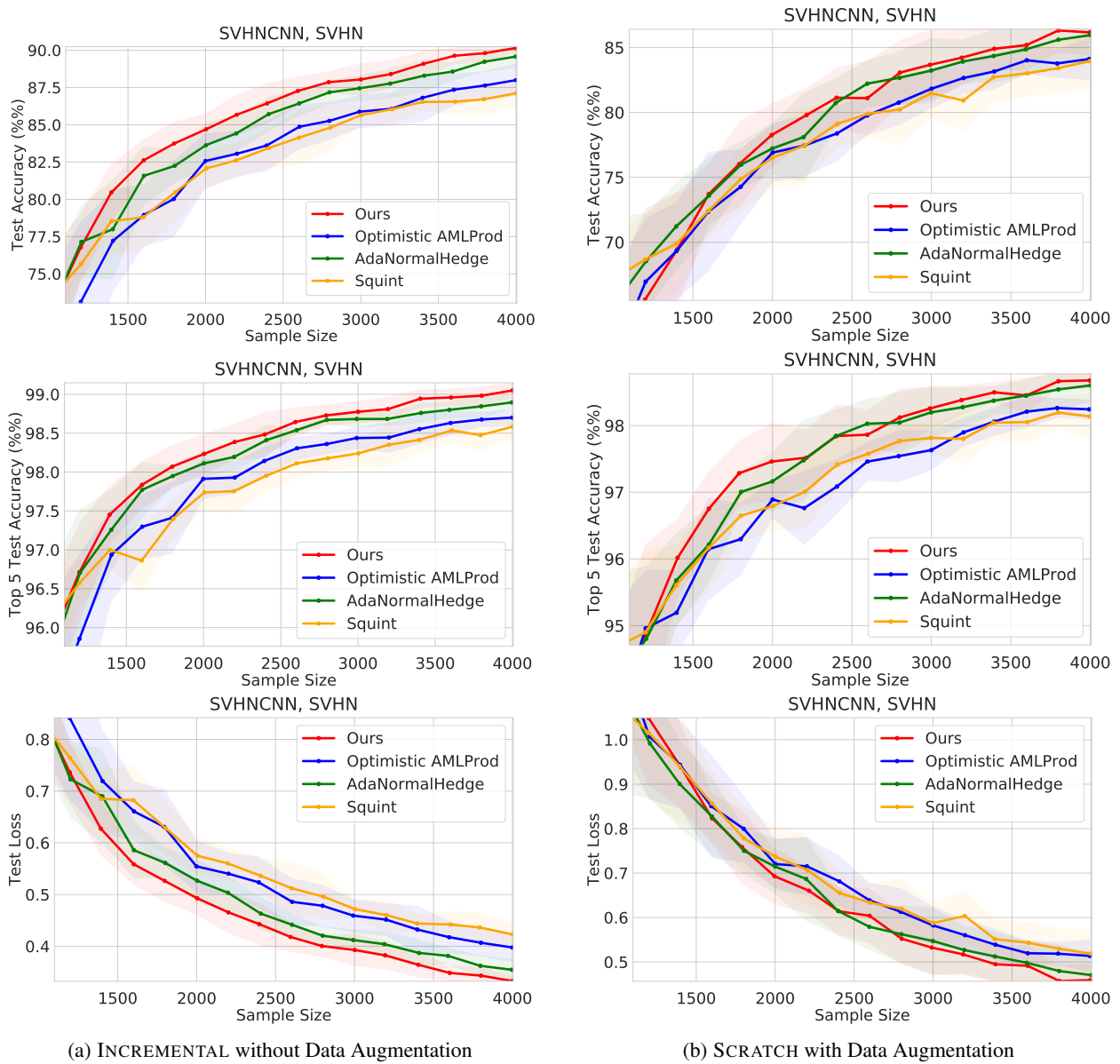


Figure 9. Evaluations of the (INCREMENTAL, 1000, 100, 4000) and (SCRATCH, 1000, 100, 4000) configurations on the SVHN dataset when trained with (left column) and without (right column) data augmentation. ADAPROD\* performs significantly better than competing approaches when the environment is predictable (left), and yet remains robust to environments with high variation (right), where it remains a top performer and is roughly tied for first with, or is slightly better than, AdaNormalHedge in all of the metrics shown above. Shaded region corresponds to values within one standard deviation of the mean.