Stochastically Constrained Adversarial Lifelong Learning

by

Mica Grant-Hagen
B.Sc., University of Victoria, 2021

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science



We acknowledge and respect the Lək̓ʷəŋən (Songhees and Xʷsepsəm/Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day.

Stochastically Constrained Adversarial Lifelong Learning

by

Mica Grant-Hagen
B.Sc., University of Victoria, 2021

Supervisory Committee

Dr. N. Mehta, Supervisor
(Department of Computer Science)

Dr. J.P. Champati, Departmental Member
(Department of Computer Science)

## ABSTRACT

Lifelong learning is a setting of online learning where there is a sequence of tasks and each task is an online learning problem. The lifelong learning algorithm is split into two sections: the meta-level and the within-task level. In this thesis, we explore an online learning setting with a stochastically-constrained adversarial (SCA) assumption. The SCA assumption is that there is not a singular representation that works the best for each of the tasks in terms of minimizing expected losses, but that there is a representation that will become better than all other representations on average over all tasks after enough tasks have been complete. We show that in the single task setting, Decreasing Hedge with perturbed losses achieves best-of-both-worlds regret bounds under the SCA assumption. In the lifelong learning setting, we propose an algorithm with Decreasing Hedge in the meta-level and Squint in the within-task level. Under the SCA assumption in the meta-level and the Bernstein condition in the within-task level, this lifelong learning algorithm achieves best-of-both-worlds regret bounds.

# Contents

# Lifelong Learning Glossary

| Notation | Description |
|---|---|
| $n$ | Length of a task. |
| $K$ | The number of tasks. |
| $\mathcal{G}$ | The set of representations. |
| $g$ | A representation. |
| $g^*$ | The representation $g \in \mathcal{G}$ with the lowest average cumulative expected losses over all tasks. |
| $\mathcal{P}_k$ | The probability distribution over the set of representations $\mathcal{G}$ for task $k$. |
| $\mathcal{H}$ | The set of hypotheses. |
| $h$ | A hypothesis. |
| $h_k^{*,g}$ | The hypothesis $h \in \mathcal{H}$ with the smallest expected loss for the representation-task pair $(g, k)$. |
| $\hat{h}_k^g$ | The empirical risk minimizer (ERM) for the representation-task pair $(g, k)$. |
| $\mathcal{M}_{a,k}^g$ | The probability distribution over the set of hypotheses $\mathcal{H}$ for the representation-task pair $(g, k)$ at timestep $a$. |
| $\tau^*$ | The threshold of tasks to be completed for the stochastically-constrained adversarial (SCA) assumption to hold. |
| $\Delta_{\mathcal{G}}$ | The sub-optimality gap between $g^*$ and all other representations $g \in \mathcal{G}$. |
| $\ell_{a,k}^{g,h}$ | The loss that hypothesis $h$ suffered at timestep $a$ for the representation-task pair $(g, k)$. |
| $\tilde{L}_k^g$ | The summation of the average losses of the ERM over tasks $1, ..., k$ for the representation-task pair $(g, k)$. |
| $L_{a,k}^{g,h}$ | A representation. |

| **Notation** | **Description** |
|---|---|
| $L_{a,k}^{g,*}$ | The cumulative loss of $h_k^{*,g}$ for the representation-task pair $(g, k)$. |
| $R_k^g(h)$ | The true risk of hypothesis $h$ for the representation-task pair $(g, k)$. |
| $\hat{R}_k^g(h)$ | The empirical risk of hypothesis $h$ for the representation-task pair $(g, k)$. |

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Nishant Mehta for his support and mentorship in my research journey. Thank you for shaping me into the researcher that I am today. Thank you to all of my lab mates Ali, Andrea, Quan, and Steve for all of the laughs and good memories over the years. I would like to thank my family for their love and support.

I would like to give a special thanks to my husband, Bryce, for his constant encouragement, for always lifting my spirits, and for making sure that I always had lots of snacks.

I would like to give a special thanks to my mom for her never-ending support and love. Thank you for pushing me to follow my dreams and passions.

# Chapter 1

# Introduction

The ability to apply our accumulated knowledge to learn new things makes us intelligent beings. We would not consider robots who can only perform one action as intelligent. The robots must be able to adapt to new situations and utilize knowledge previously acquired. In traditional machine learning problems, we use multiple copies of a learning algorithm to deal with different tasks, but in lifelong learning, we use one learning algorithm on a sequence of tasks and leverage its gained knowledge as it progresses through the sequence.

This way of learning is more in line with how humans learn new things. If we were to learn how to skip for the first time, we would use our knowledge of walking and jumping to aid in easing the learning process of skipping. We can apply this thought process to image recognition. Imagine that a learning algorithm was to be trained on identifying dogs from a set of images. It would be logical to think that we could use the same features to identify cats in images or at least use these features to benefit the learning algorithm instead of starting from scratch for this new problem.

Online learning is a framework where the data arrives sequentially. At each timestep, the learning algorithm has to make a prediction based on the previously seen data, and then Nature will reveal a loss vector. The learning algorithm will suffer a loss based on their prediction and Nature's loss vector, and then the learning algorithm updates their prediction strategy to try to minimize the future losses accumulated. There are two main settings in online learning: the stochastic setting and the adversarial setting. The difference between these two settings is that in the stochastic setting, the losses are independent and identically distributed (i.i.d.) according to some distribution, whereas in the adversarial setting, no assumptions are made about how Nature chooses the losses. In the stochastic setting, we typically assume a sub-optimality gap so that one expert will always be better than the rest in terms of expected losses. The difference in cumulative losses generated by the learning algorithm and the best expert is called regret and is the performance measure of classical online learning that we will be focusing on in this work.

Lifelong learning is a setting introduced by Thrun [1995] to propose the idea of blending how humans learn and how machine learning algorithms learn. The learning algorithm is given a sequence of tasks. Each of these tasks is an online learning problem. The lifelong learning algorithm transfers knowledge from each task by means of a set of feature representations that are applied on the tasks. The lifelong learning algorithm is split into two levels: the within-task level and the meta-level. The within-task level handles each task individually and the meta-level takes information generated from the within-task level to make decisions on the representations for the future tasks. Similar to the single-task online learning problems of obtaining low cumulative loss relative to the best expert, the lifelong learning algorithm is trying to minimize its losses through a strategic selection of a representation to achieve low regret.

Our previous image recognition example would be analogous to the lifelong learning algorithm going through a sequence of images. The learning algorithm is directed to identify if the image is a dog. Once that sequence is done, the learning algorithm has to go through another sequence of images and is tasked with identifying the cats. After that sequence is complete, the learning algorithm gets another sequence of images to identify the rabbits. This sequence of sequences continues on. In the classical stochastic setting with the sub-optimally gap assumption, there would be a single representation that is the best for each task. Logically, it does not make sense that one representation would be optimal for identifying dog, cats, rabbits, and more in image recognition, so what if we relax this assumption of a single best representation for all tasks? Assume that there is not a single feature map that works the best for each of the tasks, but after learning on enough tasks, the learning algorithm finds one representation that on average works the best. This best-on-average property for the feature map is a more realistic idea than having a single mapping that simultaneously is the best for all of the different animals that we task the image recognition learning algorithm to identify.

In this thesis, we will introduce a setting with a stochastically-constrained adversarial (SCA) assumption. This assumption is based on the idea that there is not a singular representation that works the best for each of the tasks in terms of minimizing expected losses, but that there is a representation that will become better than all other representations on average over all tasks after enough tasks have been completed. We introduce an algorithm for this setting that is able to achieve best-of-both-worlds regret guarantees, which means that it can get near-optimal regret bounds in both the stochastic setting and the adversarial setting. Wu et al. [2019] produce bounds for the stochastic setting and the adversarial setting, but they use a different algorithm for each setting instead of one algorithm for both settings. Wu et al. [2021] establish a best-of-both-worlds algorithm, though their assumptions on the sub-optimality gap of the representations is much stronger than our SCA assumption. Our framework is similar to Alquier et al. [2017] in which the learning algorithm updates at the meta-level (i.e., the representation) at the end of the task. Alquier et al. [2017] proves a general expected regret bound for a general loss function, but they do not prove any stochastic guarantees.

In Chapter 3, we detail the problem setting for this thesis and give the basis of lifelong learning. We formally introduce the SCA assumption.

In Chapter 4, we prove a regret bound for the Decreasing Hedge algorithm in the single task setting under the SCA assumption. We show results for both bounded perturbed losses and bounded non-perturbed losses.

In Chapter 5, we introduce our best-of-both-worlds lifelong learning algorithm. We map the results of Section 4.1 to the meta-level of the algorithm and prove a complete lifelong learning pseudo-regret bound with the meta-level under the SCA assumption and the within-task level under a relaxation of the sub-optimality gap assumption (the Bernstein condition, see Section 3.4). We prove a worst-case regret bound for the fully-adversarial setting.

# Chapter 2

# Background

## 2.1 Decision Theoretic Online Learning

Decision theoretic online learning (DTOL) [Freund and Schapire, 1997], which is often referred to as the Hedge setting, is a setting where Learner makes predictions based on the losses that each expert $i \in [M]$ suffered. In each timestep $t = 1, 2, ..., T$, Learner will play a probability vector $\mathbf{v}_t$ over all of the experts $i \in [M]$.

**Algorithm 1:** Protocol of DTOL

```
for t=1,2,...,T do
    Learner plays a probability distribution v_t
    Nature plays a loss vector ℓ_t
    Learner suffers loss ⟨v_t, ℓ_t⟩
```

In the Hedge algorithm, the probability is based on the weights assigned to each experts and those weights are updated at the beginning of each timestep based on the cumulative losses $L_{i,t} = \sum_{a=1}^{t} \ell_{i,a}$ and a learning rate $\eta > 0$. For each expert, these weights $\mathbf{v}_{i,t}$ are changed based on the update rule

$$\mathbf{v}_{i,t} = \frac{e^{-\eta L_{i,t-1}}}{\sum_{j \in [M]} e^{-\eta L_{j,t-1}}}. \tag{2.1}$$

The performance measure of DTOL is regret, which is the the difference in cumulative losses generated by the Learner and the best expert. Regret $R_T$ is defined as

$$R_T = \sum_{t=1}^{T} \mathbb{E}_{i\sim \mathbf{v}_t}\left[\ell_{i,t}\right] - \min_{i\in[M]} \sum_{t=1}^{T} \ell_{i,t}.$$

If we set the learning rate to be $\eta = \sqrt{\frac{8\log(M)}{T}}$, then the Hedge algorithm has a regret bound of Cesa-Bianchi and Lugosi [2006]

$$R_T \leq \sqrt{\frac{T\log(M)}{2}}.$$

### 2.1.1 Decreasing Hedge

The next algorithm to discuss in the DTOL setting is Decreasing Hedge [Auer et al., 2000]. We will be using Decreasing Hedge in the single-task setting in Chapter 4 and as the meta-level algorithm in Chapter 5.

The feature that makes Decreasing Hedge different from Hedge is that it uses a time-varying learning rate $\eta_t = \sqrt{\frac{8\log(M)}{t}}$.

In online learning, there are different ways that Nature can assign the losses: stochastic and adversarial. In the stochastic setting, the losses are drawn i.i.d. In the adversarial setting, the losses are either assigned before the game begins, which means Nature is an oblivious adversary, or the losses are assigned in reaction to the Learner's actions, which means Nature is an adaptive adversary.
In the stochastic setting, we use pseudo-regret as a performance measure:

$$\mathcal{R}_T := \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_{i\sim \mathbf{v}_t}\left[\ell_{i,t}\right]\right] - \min_{i\in[M]} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{i,t}\right].$$

Mourtada and Gaïffas [2019] prove a pseudo-regret bound for Decreasing Hedge with a learning rate of $\eta_t = 2\sqrt{\frac{\log(M)}{t}}$ and a sub-optimality gap $\Delta$ between the best expert in hindsight $i^*$ and all other experts in terms of expected losses, such that $\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t}\right] \geq \Delta$. They show that under these conditions, the pseudo-regret of

Decreasing Hedge is upper bounded by

$$\mathcal{R}_T \leq \frac{4\log(M) + 25}{\Delta}.$$

### 2.1.2 Squint

Squint [Koolen and Van Erven, 2015] is a self-tuning online learning algorithm from the Hedge setting. We will be using Squint as the within-task algorithm in Chapter 5.

We define the regret after timestep $T$ compared to expert $i$ as

$$R_T^i = \sum_{t=1}^{T} \left(\langle \mathbf{v}_t, \ell_t \rangle - \ell_{i,t}\right).$$

Squint has a second-order regret bound, which means that we are able to reduce the $T$ in Hedge's regret to the variance, $V_T^i = \sum_{t=1}^{T} \left(\langle \mathbf{v}_t, \ell_t \rangle - \ell_{i,t}\right)^2$.

Squint has a prior distribution on the experts, $\pi(i)$, and a prior over the learning rate, $\gamma(\eta)$, with $\eta \in [0, \frac{1}{2}]$. With these two priors, Squint's update rule is

$$\mathbf{v}_{i,t+1} = \pi(i) \frac{\mathbb{E}_{\eta\sim\gamma}\left[\exp\left(\eta R_t^i - \eta^2 V_t^i\right)\eta\right]}{\sum_{j\in V} \pi(j)\, \mathbb{E}_{\eta\sim\gamma}\left[\exp\left(\eta R_t^j - \eta^2 V_t^j\right)\eta\right]}.$$

For this thesis, we will assume that the prior over the experts is a uniform distribution and that the prior over the learning rate is an improper prior in the form of $\gamma(\eta) = \frac{1}{\eta}$. With these two priors fixed, Squint achieves a second-order regret upper bound of

$$R_T^i \leq \sqrt{V_T^i K_T^i} + K_T^i,$$

where $K_T = \mathcal{O}\left(-\ln(\frac{1}{M}) + \ln\ln T\right)$.

## 2.2 Lifelong Learning

Lifelong learning, also known as online-within-online lifelong learning, is setting where Learner is given a sequence of tasks to learn on and each task is a sequential decision problem [Thrun, 1995]. The algorithm is split into two levels: the meta-level and the within-task level. The within-task level works with one task at a time and provides the meta-level with the necessary information from that task. The meta-level manages a distribution over a set of feature representations shared by every task.

The first paper that derived a regret bound for this online-within-online lifelong learning setting was Alquier et al. [2017] who considered the adversarial setting. Before this paper, the main theoretical achievements were for the learning-to-learn setting in Baxter [2000], where the tasks are presented simultaneously, or for the batch-within-online lifelong learning setting in Pentina and Lampert [2015] and Pentina and Urner [2016], where the tasks come in sequentially, but each task is a batch learning problem. In Alquier et al. [2017], and later on in Denevi et al. [2019], they separate the regrets of the two levels, meta and within-task, by only updating the meta-level at the end of each task and using a task-average regret. Alquier et al. [2017] propose an exponentially weighted aggregation for lifelong learning (EWA-LL) algorithm in the meta-level. EWA-LL updates at the end of each task based on the average losses accumulated by the within-task algorithm. Alquier et al. [2017] bounded the expected regret in the adversarial setting.

A different approach in which the meta-level is updated on every timestep of the tasks was introduced by Wu et al. [2019] to tackle the problem of requiring a large number of tasks to achieve small regret. Wu et al. [2019] prove a conditional expected regret bound[1] in the adversarial setting. They use the Hedge algorithm update rule (2.1) for both the within-task level and the meta-level. They modify their algorithm for the stochastic setting by using the adversarial setting algorithm as an "exploration" phase, then they commit to a single representation in the "exploitation" phase. They prove a pseudo-regret bound in the stochastic setting.

Wu et al. [2021] use the same framework for updating the meta-level as Wu et al. [2019], but propose a best-of-both-worlds algorithm. In this paper, they explore lifelong learning in the branching experts problem setting [Gofer et al., 2013]. In the branching experts problem for lifelong learning, the full set of representations and the full set of hypotheses are not initially given. As time progresses, new representations and hypotheses branch off of already revealed representations and hypotheses in a tree structure. Their algorithm uses a variant of the Prod algorithm [Cesa-Bianchi

[1]See equation (3.1) for the definition of conditional expected regret.

et al., 2006] in the meta-level and a variant of Decreasing Hedge in the within-task level. Wu et al. [2021] prove a pseudo-regret bound in the stochastic setting and a conditional expected regret bound in the adversarial setting. In the stochastic setting, Wu et al. [2021] use a strong assumption that there is a single representation that is the best in terms of minimizing expected loss for each task uniformly. In this thesis, we relax that assumption and assumes that there is a representation that is the best on average over all tasks.

Table 2.1 highlights the differences between the lifelong settings of the papers referenced above.

| | | Within-task | |
|---|---|---|---|
| | | Batch | Online |
| Meta | Batch | Baxter [2000] | — |
| | Online | Pentina and Lampert [2015]<br>Pentina and Urner [2016] | Alquier et al. [2017]<br>Wu et al. [2019]<br>Denevi et al. [2019]<br>Wu et al. [2021] |

Table 2.1: Batch vs Online in the meta and within-task levels.

# Chapter 3

# Problem Setup

## 3.1 Lifelong Learning

Lifelong learning is an online learning setting where the learner is given a sequence of tasks to learn on. We will be in the more specific setting of online-within-online lifelong learning: the tasks are given in a sequence and each task corresponds to making a sequence of decisions. We denote tasks as $k \in [K]$. At the meta-level, the algorithm deals with a set of representations $\mathcal{G}$ and at the within-task level, the algorithm deals with a set of hypotheses $\mathcal{H}$. For this thesis, we assume that every task shares a common set of hypotheses. We also assume that every task will be the same length, $n$. At each timestep within a task, $a = 1, 2, ..., n$, the learner will play a representation and hypothesis. The within-task level deals with each task individually like a traditional online learning problem and provides a loss for both the within-task and meta algorithms based on the learner's selection of representation $g \in \mathcal{G}$ and hypothesis $h \in \mathcal{H}$.

There are two main ways in online learning that the losses are played: the stochastic setting and the adversarial setting. In the stochastic setting, losses are independent and identically distributed (i.i.d.) according to some distribution. In the adversarial setting, the adversary could be either oblivious or adaptive. An oblivious adversary will choose and fix the loss vectors beforehand, whereas an adaptive adversary can select a loss vector to play at each timestep based on Learner's previous decisions. For this thesis, when we are working in the stochastic setting for lifelong learning, all the within-task losses will be bounded $\ell_{a,k}^{g,h} \in [0, 1]$ and i.i.d., where $\ell_{a,k}^{g,h}$ is the loss

**Algorithm 2:** Simple Protocol of Lifelong Learning

**for** $k = 1$ *to* $K$ **do**
  **for** $a = 1$ *to* $n$ **do**
    Learner plays a representation and hypothesis pair, $(g, h)$, where $g \in \mathcal{G}$ and $h \in \mathcal{H}$
    Nature plays loss vector $\ell_{a,k} \in [0, 1]^{|\mathcal{G}| \times |\mathcal{H}|}$
    Learner suffers $\ell_{a,k}^{g,h}$

**Algorithm 3:** Protocol of Lifelong Learning in Hedge Setting

Initialize probability vector over representations $\mathcal{P}_1$
**for** $k = 1$ *to* $K$ **do**
  Initialize probability vector over hypotheses $\mathcal{M}_{1,k}^{g}$ for all $g \in \mathcal{G}$
  **for** $a = 1$ *to* $n$ **do**
    Learner plays a joint distribution by drawing $g \sim \mathcal{P}_k$, then $h \sim \mathcal{M}_{a,k}^{g}$
    Nature plays loss vector $\ell_{a,k} \in [0, 1]^{|\mathcal{G}| \times |\mathcal{H}|}$
    Learner suffers $\mathbb{E}_{g \sim \mathcal{P}_k}\left[\mathbb{E}_{h \sim \mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right]$
    Update $\mathcal{M}_{a+1,k}^{g}$ for all $g \in \mathcal{G}$
  Update $\mathcal{P}_k$

that hypothesis $h$ suffered at timestep $a$ for the representation-task pair $(g, k)$. Across tasks, we only assume losses to be independent, so that the tasks could have different distributions.

We will be working in the Hedge setting, which is also referred to as decision-theoretic online learning [Freund and Schapire, 1997], for both the within-task and meta algorithms. Instead of playing a single representation and hypothesis at each timestep, our learner will be playing a distribution $\mathcal{P}_k$ over the representations and a distribution $\mathcal{M}_{a,k}^{g}$ over the hypotheses. The algorithm will update $\mathcal{M}_{a,k}^{g}$ at the end of each timestep and update $\mathcal{P}_k$ at the end of each task.

Our performance measure is regret. Regret measures the amount of loss that our learner suffers compared to using the representation and hypotheses with the smallest losses in hindsight. Our main goal is to minimize the amount of regret that our algorithm incurs.

The following is the regret based on the simple protocol (Algorithm 2), where $g_{a,k}$ and $h_{a,k}$ are the representation and hypothesis that the algorithm plays in round $a$ of task $k$:

$$\tilde{R}_K := \sum_{k=1}^{K}\sum_{a=1}^{n} \ell_{a,k}^{g_{a,k},h_{a,k}} - \min_{\substack{g\in\mathcal{G} \\ h_1,h_2,\ldots,h_K\in\mathcal{H}}} \sum_{k=1}^{K}\sum_{a=1}^{n} \ell_{a,k}^{g,h_k}.$$

The next notions of regret will be based on the Hedge version of the protocol (Algorithm 3). The first one is similar to the previous form of regret, but deals with the conditional expectation of the losses with respect to the probability distributions played. We define conditional expected regret as:

$$R_K := \sum_{k=1}^{K}\sum_{a=1}^{n} \mathbb{E}_{g\sim\mathcal{P}_{a,k}}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right] - \min_{\substack{g\in\mathcal{G} \\ h_1,h_2,\ldots,h_K\in\mathcal{H}}} \sum_{k=1}^{K}\sum_{a=1}^{n} \ell_{a,k}^{g,h_k}. \tag{3.1}$$

In the stochastic setting, we work with expected regret and pseudo-regret instead of the previous forms of regret due to the losses being i.i.d. and we take an expectation over the losses to handle the randomness. The following definition is the expected regret:

$$\mathbb{E}\left[R_K\right] := \mathbb{E}\left[\sum_{k=1}^{K}\sum_{a=1}^{n} \mathbb{E}_{g\sim\mathcal{P}_{a,k}}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right] - \min_{\substack{g\in\mathcal{G} \\ h_1,h_2,\ldots,h_K\in\mathcal{H}}} \sum_{k=1}^{K}\sum_{a=1}^{n} \ell_{a,k}^{g,h_k}\right].$$

The last notion of regret is pseudo-regret. In plain terms, this is the regret based on the expected losses, unlike the normal notion of regret, which uses the actual loss:

$$\mathcal{R}_K := \mathbb{E}\left[\sum_{k=1}^{K}\sum_{a=1}^{n} \mathbb{E}_{g\sim\mathcal{P}_{a,k}}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right]\right] - \min_{g\in\mathcal{G}} \sum_{k=1}^{K} \min_{h\in\mathcal{H}} \sum_{a=1}^{n} \mathbb{E}\left[\ell_{a,k}^{g,h}\right].$$

A key relationship between the regrets is that the expected regret is always an upper bound on the pseudo-regret, i.e., $\mathcal{R}_K \leq \mathbb{E}\left[R_K\right]$.

We will bound the pseudo-regret in the stochastic setting (Theorem 2) and conditional expected regret in the adversarial setting (Theorem 3).

## 3.2 Stochastically-Constrained Adversarial Assumption

In the stochastic setting, we will be assuming the stochastically-constrained adversarial (SCA) assumption. The SCA assumption is that there is a clear "best" expert or representation after a certain time threshold has passed. After this threshold is passed, there is a single expert or representation with the smallest average expected loss. We will define the SCA assumption for the single-task setting and the lifelong learning setting.

For the lifelong learning setting, we define $g^*$ to be the representation with the lowest average cumulative expected losses over all tasks when the hypothesis with the lowest expected losses is used for each task, $h_k^{*,g^*}$. We assume that there exists some number of tasks, $\tau^*$, such that after our learning algorithm completes that number of tasks, there exists a strictly positive sub-optimality gap $\Delta_{\mathcal{G}}$ between the cumulative losses of the representation $g^*$ and some representation $g \neq g^*$ when using $h_k^{*,g}$ for all representation-task pairings $(g, k)$.

The SCA assumption is a weaker version of an assumption from Wu et al. [2019] and Wu et al. [2021], where $g^*$ is the representation with the smallest expected loss for each individual task. The SCA assumption is inspired by prior work in single-task online learning. In particular, it is motivated by the ""adversarial with a gap" assumption of Mourtada and Gaïffas [2019], which defines the sub-optimality gap in terms of cumulative losses rather than cumulative expected losses.

Recall that we assume that the within-task losses are i.i.d. and bounded $\ell_{a,k}^{g,h} \in [0, 1]$ and that the meta-level losses are independent.

**Assumption 1** (SCA assumption)**.** *In the meta-level, assume that there is some representation $g^* \in \mathcal{G}$ and $\tau^* \geq 1$ and $\Delta_{\mathcal{G}} > 0$ such that, for all $g \neq g^*$ and $k \geq \tau^*$, the following holds:*

$$\mathbb{E}\left[L_k^{g,*}\right] - \mathbb{E}\left[L_k^{g^*,*}\right] \geq k\Delta_{\mathcal{G}},$$

*where $L_k^{g,*} = \sum_{b=1}^{k} \sum_{a=1}^{n} \frac{\ell_{a,b}^{g,h_a^{*,g}}}{n}$ and $h_k^{*,g} = \arg\min_{h \in \mathcal{H}} \mathbb{E}\left[\ell_{a,k}^{g,h}\right]$, for all $a \in [n]$.*

The strict $g^*$ sub-optimality gap assumption from Wu et al. [2019] implies the SCA assumption with $\tau^* = 1$. For a fixed representation $g \in \mathcal{G}$ and for task $k = 1, ..., K$ and for all $a = 1, ..., n$, the strict sub-optimality gap assumption is:

$$\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right] - \mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right] \geq \Delta_{\mathcal{G}}. \tag{3.2}$$

When we take a summation over all tasks and within-task timesteps, we recover the SCA assumption with $\tau^* = 1$, since

$$\begin{aligned}\sum_{b=1}^{k}\sum_{a=1}^{n}\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right] - \mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right] &\geq \sum_{b=1}^{k}\sum_{a=1}^{n}\Delta_{\mathcal{G}} \\ n\left(\mathbb{E}\left[L_k^{g,*}\right] - \mathbb{E}\left[L_k^{g^*,*}\right]\right) &\geq kn\Delta_{\mathcal{G}} \\ \mathbb{E}\left[L_k^{g,*}\right] - \mathbb{E}\left[L_k^{g^*,*}\right] &\geq k\Delta_{\mathcal{G}}.\end{aligned}$$

For the single-task setting, we define the cumulative loss of expert $i$ as $L_{i,t} = \sum_{a=1}^{t} \ell_{i,a}$, where $i \in M$ and $M$ is the set of experts. For the SCA assumption, we define $i^* \in M$ to be the expert with the lowest average expected cumulative loss over all timesteps $t = 1, ..., T$. We assume that there is a strictly positive sub-optimality gap $\Delta$ between the average expected cumulative loss of expert $i^*$ and all other experts $i \neq i^*$ after the learning algorithm has learned on at least $\tau^*$ number of timesteps.

We assume the losses are bounded $\ell_{i,t} \in [0, 1]$ and independent, for all $i \in M$ and $t = 1, ..., T$. We now formally present the SCA assumption for the single-task setting.

**Assumption 2** (SCA assumption for single-task)**.** *In the single-task setting, assume that there is some expert $i^* \in M$ and $\tau^* \geq 1$ and $\Delta > 0$ such that, for all $i \neq i^*$ and $t \geq \tau^*$, the following holds:*

$$\mathbb{E}\left[L_{i,t}\right] - \mathbb{E}\left[L_{i^*,t}\right] \geq \Delta t. \tag{3.3}$$

Our goal is to achieve best-of-both-worlds regret bounds, meaning that we want a single lifelong learning algorithm that achieves low regret in the adversarial setting and low pseudo-regret in the stochastic setting under the SCA assumption.

## 3.3 Within-task Gap

An assumption that is made in Wu et al. [2021] is that when using representative $g^*$, there will be a sub-optimality gap $\Delta_h$ between the best hypothesis and second best,

$$\Delta_h := \min_k \min_{\substack{h \neq h_k^{*,g^*} \\ h \in \mathcal{H}}} \left( \mathbb{E}\left[\ell_{a,k}^{g^*,h}\right] - \mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right] \right), \tag{3.4}$$

where $h_k^{*,g} := \arg\min_{h \in \mathcal{H}} \mathbb{E}\left[\ell_{a,k}^{g,h}\right]$ and $a$ is some arbitrary within-task timestep.

The above is a very strong assumption, especially with our relaxation of the $g^*$ assumption. Since our $g^*$ is not guaranteed to be best representation for every task individually, as it is under the strict $g^*$ sub-optimality gap assumption (3.2), it is a strong assumption to assume that each task has a within-task gap for every representation.

Instead of using this sub-optimality gap assumption in the within-task level, we will assume that every representation-task pairing satisfies the Bernstein condition, which is a relaxation of the sub-optimality gap assumption.

## 3.4 Bernstein condition

The Bernstein condition [Bartlett and Mendelson, 2006] is a generalization of the gap assumption, which relaxes the strictness of the $\Delta_h$ assumption.

Fix a representation $g \in \mathcal{G}$ and a task $k \in [K]$. For a fixed $B \geq 0$ and $\kappa \in [0, 1]$, the $(B, \kappa)$-Bernstein condition is satisfied if there exists an hypothesis $h_k^{*,g} \in \mathcal{H}$, such that for all other $h \in \mathcal{H}$, $h \neq h_k^{*,g}$ for all $a \leq n$:

$$\mathbb{E}\left[\left(\ell_{a,k}^{g,h} - \ell_{a,k}^{g,h_k^{*,g}}\right)^2\right] \leq B\mathbb{E}\left[\ell_{a,k}^{g,h} - \ell_{a,k}^{g,h_k^{*,g}}\right]^\kappa \tag{3.5}$$

We will focus on the case of the Bernstein condition where $\kappa = 1$.

Koolen et al. [2016] give an example of when the $(B, 1)$-Bernstein condition holds is under the within-task sub-optimality gap assumption (3.4) when $B = \frac{1}{\Delta_h}$. For a fixed representation $g \in \mathcal{G}$ and task $k = 1, ..., K$, we have:

$$\begin{aligned}\mathbb{E}\left[\left(\ell_{a,k}^{g,h} - \ell_{a,k}^{g,h_k^{*,g}}\right)^2\right] &\leq 1 \\ &= \frac{1}{\Delta_h}\Delta_h \\ &\leq B\mathbb{E}\left[\ell_{a,k}^{g,h} - \ell_{a,k}^{g,h_k^{*,g}}\right].\end{aligned}$$

### 3.4.1 Bernstein's Inequality

In this thesis, we use a one-sided Bernstein inequality, specifically, equation (2.10) from Boucheron et al. [2013].

**Lemma 1.** *Let $X_1, .., X_n$ be independent random variables such that, for some $b > 0$, $X_i \leq b$ almost surely for all $i \leq n$.*

*For any $\epsilon > 0$,*

$$\Pr\left(\sum_{i=1}^{n}\left(X_i - \mathbb{E}\left[X_i\right]\right) \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2\left(\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] + \frac{b\epsilon}{3}\right)}\right). \tag{3.6}$$

# Chapter 4

# Single Task

In this section, we show the pseudo-regret bound for Decreasing Hedge under the SCA assumption. This chapter will focus on the single-task setting. In Chapter 5, we will create a mapping between the single-task setting and the meta-level of the lifelong learning setting. This mapping will allow us to use the results proven in this chapter and extend them to the meta-level.

The SCA assumption is similar to that addressed by Mourtada and Gaïffas [2019], where they bound the regret of the Decreasing Hedge algorithm in their "adversarial with a gap" assumption. They define their sub-optimality gap between the best expert for their setting and all other experts in terms of the average loss suffered by the respective expert, given that the learner has completed more than $\tau^*$ timesteps. Their "adversarial with a gap" setting is as follows. Assume the losses are bounded $\ell_{i,t} \in [0, 1]$, and define the cumulative loss of expert $i$ as $L_{i,t} = \sum_{a=1}^{t} \ell_{i,a}$, where $i \in M$ and $M$ is the set of experts. At timestep $t$, assume that there exists $\Delta > 0$ and an expert $i^* \in M$ such that for all $i \neq i^*$ and all $t \geq \tau^*$, the following holds:

$$L_{i,t} - L_{i^*,t} \geq \Delta t.$$

This means that after some time threshold of $\tau^*$, there will be an expert $i^*$ that is optimal in the average losses by at least some gap, $\Delta$.

We extend the results of Mourtada and Gaïffas [2019] by allowing for the losses to be stochastic and making the assumption be in terms of expected losses. We further expand on the results by bounding the pseudo-regret when the the losses seen by the algorithm are perturbed. We show the results of Decreasing Hedge under the SCA assumption without perturbed losses in Corollary 1.

We use two lemmas from Mourtada and Gaïffas [2019] in our proof of Theorem 1 (presented below). We will use a version of Hoeffding-Azuma's maximal inequality from Mourtada and Gaïffas [2019] [Proposition 5] and their Lemma 1.

**Lemma 2** (Hoeffding-Azuma's maximal inequality)**.** *Let $(Z_t)_{t\geq 1}$ be a sequence of random variables adapted to the filtration $(F_t)_{t\geq 1}$. Assume that $Z_t$ is a martingale difference sequence. We have a $F_{t-1}$-measurable random variable $A_t$, such that*

$$A_t - 1 \leq Z_t \leq A_t + 1.$$

*We define $S_n := \sum_{t=1}^{n} Z_t$. For every $n \geq 1$ and $q \geq 0$,*

$$\Pr\left(\sup_{m\geq n} \frac{S_m}{m} \geq q\right) \leq e^{\frac{-nq^2}{2}}.$$

**Lemma 3** (Lemma 1 of Mourtada and Gaïffas [2019])**.** *For every $\alpha > 0$,*

$$\sum_{t\geq 1} e^{-\alpha t} \leq \frac{1}{\alpha} \tag{4.1}$$

$$\sum_{t\geq 1} e^{-\alpha\sqrt{t}} \leq \frac{2}{\alpha^2}. \tag{4.2}$$

## 4.1 Single task with Noisy losses

This section will show the pseudo-regret bound for Decreasing Hedge under the SCA assumption (Assumption 2) with perturbed and bounded losses.

The perturbed losses that we are working with are the losses suffered by the learner plus some bounded noise, which is defined as $\tilde{\ell}_{i,t} = \ell_{i,t} + \nu_{i,t}$, where $\nu_{i,t}$ is noise added. We are dealing with perturbed losses because when we extend these results to the meta-level in the lifelong learning setting, the meta-level algorithm receives information that is analogous to perturbed losses.

**Algorithm 4:** Decreasing Hedge with perturbed losses

$t \leftarrow 1$
Set probability vector $\mathbf{v}_{i,t} = \frac{1}{|M|}$ for every expert $i \in M$
$L_{i,0} \leftarrow 0$ for every expert $i$
**while** $t \neq T$ **do**
  $\eta_t = 2\sqrt{\frac{\log(|M|)}{t}}$
  Update probability vector $\mathbf{v}_{i,t} = \frac{e^{-\eta_t \tilde{L}_{i,t-1}}}{\sum_{j \in M} e^{-\eta_t \tilde{L}_{i,t-1}}}$ for every expert $i$
  Learner plays probability distribution $\mathbf{v}_t$
  Nature plays perturbed losses $\tilde{\ell}_{i,t} = \ell_{i,t} + \nu_{i,t}$, for every expert $i$
  Learner suffers $\langle \mathbf{v}_t, \tilde{\ell}_t \rangle$
  Update cumulative perturbed losses $\tilde{L}_{i,t} = \tilde{L}_{i,t-1} + \tilde{\ell}_{i,t}$ for every expert $i$
  $t \leftarrow t + 1$

**Theorem 1.** *Assume that the SCA assumption (Assumption 2) holds. Assume that all losses are perturbed with added noise of* $\nu_{i,t} \in [-\epsilon_{ST}, \epsilon_{ST}]$*, such that* $\tilde{\ell}_{i,t} = \ell_{i,t} + \nu_{i,t}$*. Let* $0 \leq \epsilon_{ST} < \frac{\Delta}{4}$*.*
*Under these conditions, Decreasing Hedge (run on perturbed losses) with a learning rate of* $\eta_t = 2\sqrt{\frac{\log(|M|)}{t}}$ *has a pseudo-regret bound of*

$$\mathcal{R}_T \leq \mathcal{O}\left(\sqrt{\tau^* \log\left(|M|\right)} + \frac{\log\left(|M|\right)}{\Delta - 4\epsilon_{ST}} + \frac{\log\left(\Delta^{-1}\right)}{\Delta}\right).$$

*Proof.* First, we must define our perturbed cumulative losses for this setting, $\tilde{L}_{i,t} = \sum_{b=1}^{t} \tilde{\ell}_{i,b}$.

We define $t_0$ to be the smallest integer $t \geq 1$, such that:

$$|M| \exp\left(-\Delta\sqrt{\frac{t\log(|M|)}{2}}\right) \leq \Delta. \tag{4.3}$$

We rearrange to get $t_0 = \lceil \frac{2\log^2\left(\frac{|M|}{\Delta}\right)}{\Delta^2 \log(|M|)} \rceil \leq \frac{2\log^2\left(\frac{|M|}{\Delta}\right)}{\Delta^2 \log(|M|)} + 1$.

We want to show that the probability that, after some timestep, the cumulative perturbed losses of the best expert $i^*$ and all other experts are close together is very low.

We define $\tau = \sup\{t \geq 0, \exists i \neq i^*, \tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}\}$, which is the last timestep in which we do not have a large enough positive gap between the best expert $i^*$ and any other experts.

Since we are dealing with bounded noise, we can upper bound this probability by either taking the minimum or maximum value that the cumulative perturbed losses could be, which gives us an upper bound of

$$\begin{aligned}\Pr\left(\exists t \geq c, \tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{t\Delta}{2}\right) &\leq \Pr\left(\exists t \geq c, (L_{i,t} - t\epsilon_{ST}) - (L_{i^*,t} + t\epsilon_{ST}) \leq \frac{t\Delta}{2}\right) \\ &= \Pr\left(\exists t \geq c, L_{i,t} - L_{i^*,t} \leq \frac{t\Delta}{2} + 2t\epsilon_{ST}\right),\end{aligned}$$

where $c \geq \tau^*$ is some arbitrary timestep.

We assign $X_t = L_{i^*,t} - L_{i,t}$ and $Z_t = (X_t - X_{t-1}) - \mathbb{E}[X_t - X_{t-1}]$, for some arbitrary expert $i \in M$, and $X_0 = 0$. These terms will allow us to set up a bounded martingale difference sequence, so we are able to use Hoeffding-Azuma's maximal inequality (Lemma 2).

We work backwards to go from Hoeffding-Azuma's maximal inequality (Lemma 2) to the form of $\Pr\left(\exists t \geq c, L_{i,t} - L_{i*,t} \leq \frac{t\Delta}{2} + 2t\epsilon_{ST}\right)$, where $c \geq \tau^*$ is some arbitrary step. This will allow us to solve for the term $q$ and then get the probability bound

$$
\begin{aligned}
\Pr\left(\sup_{t \geq c} \frac{1}{t} \sum_{b=1}^{t} Z_b \geq q\right) &= \Pr\left(\sup_{t \geq c} \frac{1}{t} \left(\sum_{b=1}^{t} (X_b - X_{b-1}) - \mathbb{E}[X_b - X_{b-1}]\right) \geq q\right) \\
&= \Pr\left(\sup_{t \geq c} \frac{1}{t} \left(X_t - X_0 - \sum_{b=1}^{t} \mathbb{E}[X_b - X_{b-1}]\right) \geq q\right) \\
&= \Pr\left(\sup_{t \geq c} \frac{1}{t} (X_t - \mathbb{E}[X_t]) \geq q\right) \\
&\geq \Pr\left(\exists t \geq c, \frac{1}{t} (X_t - \mathbb{E}[X_t]) \geq q\right) \\
&= \Pr\left(\exists t \geq c, \frac{1}{t} X_t \geq q + \frac{\mathbb{E}[X_t]}{t}\right) \\
&= \Pr\left(\exists t \geq c, -X_t \leq \mathbb{E}[-X_t] - qt\right) \\
&\geq \Pr\left(\exists t \geq c, -X_t \leq t\Delta - qt\right) \\
&= \Pr\left(\exists t \geq c, L_{i,t} - L_{i^*,t} \leq t\Delta - qt\right).
\end{aligned}
$$

We set $t\Delta - qt = \frac{t\Delta}{2} + 2t\epsilon_{ST}$ and solve for $q$, which gives us $q = \frac{\Delta}{2} - 2\epsilon_{ST}$.

With this $q$ term, we will go back to the definition of Hoeffding-Azuma maximal inequality and use this term in the upper bound

$$
\Pr\left(\sup_{t \geq c} \frac{1}{t} \sum_{b=1}^{t} Z_b \geq \frac{\Delta}{2} - 2\epsilon_{ST}\right) \leq e^{\frac{-c\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}{2}}.
$$

We have an upper bound of $\Pr\left(\exists t \geq c, L_{i,t} - L_{i^*,t} \leq \frac{t\Delta}{2} + 2t\epsilon_{ST}\right)$, which means that we also have an upper bound of $\Pr\left(\exists t \geq c, \tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}\right)$. We take a union over all experts $i \neq i^*$ to get $\Pr\left(\bigcup_{i\neq i^*} \exists t \geq c, \tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}\right)$, which is equivalent to $\Pr(\tau \geq c)$. This is because if there exists an expert $i \neq i^*$ and a timestep $t \geq c$ that satisfy $\tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}$, then $\tau$ must also be greater than $c$ since $\tau$ is the largest $t$ that satisfies $\tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}$ for some $i \neq i^*$. Now, we can upper bound the probability that $\tau \geq c$ by

$$\begin{aligned}
\Pr(\tau \geq c) &= \Pr(\bigcup_{i\neq i^*} \exists t \geq c, \tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}) \\
&\leq \sum_{i\neq i^*} \Pr(\exists t \geq c, \tilde{L}_{i,t} - \tilde{L}_{i^*,t} \leq \frac{\Delta t}{2}) \\
&\leq \sum_{i\neq i^*} \Pr\left(\exists t \geq c, L_{i,t} - L_{i^*,t} \leq \frac{t\Delta}{2} + 2t\epsilon_{ST}\right) \\
&\leq |M| e^{\frac{-c\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}{2}}.
\end{aligned}$$

By using an exponential tail bound, we can take this probability bound and upper bound the expectation of $\tau$. We define $\gamma$ to be some arbitrary timestep, $\gamma \geq \tau^*$. The expectation of $\tau$ is upper bounded by

$$\begin{aligned}
\mathbb{E}[\tau] &= \int_0^\infty \Pr(\tau \geq c)\, dc \\
&= \int_0^\gamma \Pr(\tau \geq c)\, dc + \int_\gamma^\infty \Pr(\tau \geq c)\, dc \\
&\leq \gamma + \int_\gamma^\infty |M| e^{\frac{-c\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}{2}}\, dc \\
&= \gamma + |M| \left( \frac{-2}{\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2} e^{\frac{-c\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}{2}} \Big|_{c=\gamma}^{\infty} \right) \\
&= \gamma + |M| \frac{2}{\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2} e^{\frac{-\gamma\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}{2}}.
\end{aligned}$$

We tune $\gamma$ to minimize the bound given above. We will refer to the optimal non-constrained tuning as $\gamma^*$, where we have $\gamma^* = \frac{2\log(|M|)}{\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}$. Since we have the constraint of $\gamma \geq \tau^*$, we must take into consideration the scenario that $\gamma^* < \tau^*$. For this scenario, we can see that as we set $\gamma > \gamma^*$, the first term of the above bound becomes dominant as the second term grows smaller. Due to this, we can set an upper bound on the expectation of $\tau$ when using $\gamma = \tau^*$ to be $\mathbb{E}[\tau] \leq 2\tau^*$. We set $\gamma = \max\{\tau^*, \gamma^*\}$.

Now that we have tuned $\gamma$, the expectation of $\tau$ is at most

$$\mathbb{E}[\tau] \leq \max\left\{2\tau^*, \frac{2}{\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}\left(\log\left(|M|\right) + 1\right)\right\}. \tag{4.4}$$

For this to hold, we must set a bound on $\epsilon_{ST}$, such that $\epsilon_{ST} \leq \frac{\Delta}{4}$.

We decompose the pseudo-regret into two sections: the "bad" and the "good". The "bad" section is up to timestep $\sigma + 1$, where $\sigma = \max\{\tau^*, \tau, t_0\}$. In the "bad" section, the algorithm suffers the worst case regret due to the cumulative losses of $i^*$ and all of the other experts being too close. This means that the algorithm has not passed timestep $\tau$, the algorithm is not passed timestep $\tau^*$ yet and the SCA assumption has not been satisfied, or the algorithm has not passed $t_0$.

We expand the regret into two sections, up to and after timestep $\sigma + 1$, which gives us the following decomposition of the pseudo-regret

$$\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^{T} \mathbb{E}\left[\hat{\ell}_t\right] - \min_{i \in [M]} \sum_{t=1}^{T} \mathbb{E}\left[\ell_{i,t}\right] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[\sum_{t=1}^{T} \hat{\ell}_t - \sum_{t=1}^{T} \ell_{i^*,t} | \sigma\right]\right] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[\sum_{t=1}^{\sigma+1} \hat{\ell}_t - \sum_{t=1}^{\sigma+1} \ell_{i^*,t} + \sum_{t=\sigma+2}^{T} \hat{\ell}_t - \sum_{t=\sigma+2}^{T} \ell_{i^*,t} | \sigma\right]\right] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[\sum_{t=1}^{\sigma+1} \hat{\ell}_t - \sum_{t=1}^{\sigma+1} \ell_{i^*,t} | \sigma\right]\right] + \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[\sum_{t=\sigma+2}^{T} \hat{\ell}_t - \sum_{t=\sigma+2}^{T} \ell_{i^*,t} | \sigma\right]\right].
\end{aligned}$$

Now, let us look at the "bad" case, which is if $t \leq \sigma + 1$. Since we are using Decreasing Hedge as our algorithm, we will use the worst case regret of Hedge,

$$
\begin{aligned}
\mathcal{R}_{\sigma+1} &\le \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[R_{\sigma+1}|\sigma\right]\right] \\
&\le \mathop{\mathbb{E}}_{\sigma}\left[\sqrt{(\sigma+1)\log\left(|M|\right)}\right].
\end{aligned}
$$

We use Jensen's inequality to bring the expectation inside of the function

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\sigma}\left[\sqrt{(\sigma+1)\log\left(|M|\right)}\right] &\le \sqrt{\left(\mathop{\mathbb{E}}_{\sigma}\left[\sigma\right]+1\right)\log\left(|M|\right)} \\
&= \sqrt{\left(\mathop{\mathbb{E}}_{\sigma}\left[\max\left\{\tau^*,\tau,t_0\right\}\right]+1\right)\log\left(|M|\right)} \\
&\le \sqrt{\tau^*\log\left(|M|\right)+\mathop{\mathbb{E}}_{\sigma}\left[\tau\right]\log\left(|M|\right)+t_0\log\left(|M|\right)+\log\left(|M|\right)} \\
&\le \sqrt{\tau^*\log\left(|M|\right)}+\sqrt{\mathop{\mathbb{E}}_{\sigma}\left[\tau\right]\log\left(|M|\right)}+\sqrt{t_0\log\left(|M|\right)}+\sqrt{\log\left(|M|\right)}.
\end{aligned}
$$

We use the upper bound on $\mathbb{E}_{\sigma}\left[\tau\right]$ (4.4) and get an upper bound for the "bad" section of the pseudo-regret:

$$
\begin{aligned}
&\leq \sqrt{\tau^* \log(|M|)} + \sqrt{\max\left\{2\tau^*, \frac{2}{\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}(\log(|M|)+1)\right\}\log(|M|)} \\
&\quad + \sqrt{\left(\frac{2\log^2\left(\frac{|M|}{\Delta}\right)}{\Delta^2 \log(|M|)} + 1\right)\log(|M|)} + \sqrt{\log(|M|)} \\
&\leq \sqrt{\tau^* \log(|M|)} + \sqrt{\left(2\tau^* + \frac{2}{\left(\frac{\Delta}{2} - 2\epsilon_{ST}\right)^2}(\log(|M|)+1)\right)\log(|M|)} \\
&\quad + \sqrt{\left(\frac{2\log^2\left(\frac{|M|}{\Delta}\right)}{\Delta^2 \log(|M|)} + 1\right)\log(|M|)} + \sqrt{\log(|M|)} \\
&\leq \sqrt{\tau^* \log(|M|)} + \sqrt{2\tau^* \log(|M|)} + \frac{\sqrt{2}\log(|M|)}{\frac{\Delta}{2} - 2\epsilon_{ST}} + \frac{\sqrt{2(\log(|M|))}}{\Delta - 4\epsilon_{ST}} + 2\sqrt{\log(|M|)} \\
&\quad + \frac{\sqrt{2}\log\left(\frac{|M|}{\Delta}\right)}{\Delta}.
\end{aligned}
$$

Next, we will look at the "good" case, which is when $t > \sigma + 1$. Initially, we fix some arbitrary timestep $t$ such that $T \geq t > \sigma + 1$. The instantaneous regret for this timestep $t$ is upper bounded by

$$\begin{aligned}
\hat{\ell}_t - \ell_{i^*,t} &= \hat{\ell}_t - \ell_{i^*,t} \\
&= \sum_{i \neq i^*} v_{i,t} \left( \ell_{i,t} - \ell_{i^*,t} \right) \\
&\leq \sum_{i \neq i^*} v_{i,t} \\
&= \sum_{i \neq i^*} \frac{e^{-\eta_t \left( \tilde{L}_{i,t-1} - \tilde{L}_{i^*,t-1} \right)}}{1 + \sum_{j \neq i^*} e^{-\eta_t \left( \tilde{L}_{j,t-1} - \tilde{L}_{i^*,t-1} \right)}} \\
&\leq \sum_{i \neq i^*} e^{-2\sqrt{\frac{\log(|M|)}{t}} \frac{\Delta(t-1)}{2}} && (4.5) \\
&\leq |M| e^{-\Delta \sqrt{(t-2)\log(|M|)}} \\
&\leq |M| e^{\frac{-\Delta \sqrt{(t_0)\log(|M|)}}{2}} e^{\frac{-\Delta \sqrt{(t-2)\log(|M|)}}{2}} && (4.6) \\
&\leq \Delta e^{\frac{-\Delta \sqrt{(t-2)\log(|M|)}}{2}} && (4.7) \\
&\leq \Delta e^{\frac{-\Delta \sqrt{(t-2)}}{2}}.
\end{aligned}$$

For the section of regret during the “good” case, we are able to use the fact that with $t > \tau$, we have $\tilde{L}_{i,t} - \tilde{L}_{i^*,t} > \frac{\Delta t}{2}$ which allows us to get (4.5). For (4.6), we use the fact that $t - 2 \geq t_0$, since $t > \sigma + 1$. We use the definition of $t_0$ (4.3) to get the next bound (4.7).

The next step is to sum these instantaneous regrets to get the full “good” case regret, which gives us

$$\begin{aligned}
\sum_{t=\sigma+2}^{T} \left( \hat{\ell}_t - \ell_{i^*,t} \right) &\leq \sum_{t=\sigma+2}^{T} \Delta e^{\frac{-\Delta\sqrt{(t-2)}}{2}} \\
&\leq \Delta \sum_{t \geq 1} e^{\frac{-\Delta\sqrt{(t)}}{2}} \\
&\leq \Delta \frac{2}{\left(\frac{\Delta}{2}\right)^2} && (4.8) \\
&= \frac{8}{\Delta},
\end{aligned}$$

where we use Lemma 3 to get (4.8).

We bring both sides of the regret decomposition together and finalize our regret for Decreasing Hedge in the single task setting under the SCA assumption to get

$$
\begin{aligned}
\mathcal{R}_T &= \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[\sum_{t=1}^{\sigma+1}\hat{\ell}_t - \sum_{t=1}^{\sigma+1}\ell_{i^*,t}|\sigma\right]\right] + \mathop{\mathbb{E}}_{\sigma}\left[\mathbb{E}\left[\sum_{t=\sigma+2}^{T}\hat{\ell}_t - \sum_{t=\sigma+2}^{T}\ell_{i^*,t}|\sigma\right]\right] \\
&\leq \sqrt{\tau^*\log\left(|M|\right)} + \sqrt{2\tau^*\log\left(|M|\right)} + \frac{\sqrt{2}\log\left(|M|\right)}{\frac{\Delta}{2}-2\epsilon_{ST}} + \frac{\sqrt{2\left(\log\left(|M|\right)\right)}}{\Delta-4\epsilon_{ST}} + 2\sqrt{\log\left(|M|\right)} \\
&\qquad + \frac{\sqrt{2}\log\left(\frac{|M|}{\Delta}\right)}{\Delta} + \frac{8}{\Delta}.
\end{aligned}
$$

□

We show the pseudo-regret bound for the case of non-perturbed losses, when $\epsilon_{ST} = 0$, under the SCA assumption in the following corollary.

**Corollary 1.** *Assume that the losses are bounded $\ell_{i,t} \in [0,1]$ for all $i \in M$ and $t \geq 1$. Assume that there exists $\Delta > 0$ and an expert $i^* \in M$, such that for all $i \neq i^*$ and all $t \geq \tau^*$, the stochastically constrained assumption (SCA) holds:*

$$\mathbb{E}\left[L_{i,t}\right] - \mathbb{E}\left[L_{i^*,t}\right] \geq \Delta t. \tag{4.9}$$

*Under these conditions, Decreasing Hedge with a learning rate $\eta_t = 2\sqrt{\frac{\log(|M|)}{t}}$ has a pseudo-regret bound of*

$$\mathcal{R}_T \leq \mathcal{O}\left(\sqrt{\tau^*\log\left(|M|\right)} + \frac{\log\left(|M|\right)}{\Delta} + \frac{\log\left(\Delta^{-1}\right)}{\Delta}\right).$$

# Chapter 5

# Lifelong Learning

In this chapter, we will introduce a lifelong learning algorithm and prove that it has best-of-both-world regret bounds. The single task setting results of Theorem 1 will be used as the foundation for the analysis. We will introduce a mapping from the single task setting to the meta-level to handle the deviation of the losses to ensure that what the meta-level receives from the within-task level is informative.

## 5.1 Algorithm

In lifelong learning, we work with two levels of the algorithm and make them cohesive: the within-task level and meta-level. The within-task algorithm is set up in a black box style in relation to the meta algorithm, such that within-task algorithm gets the task's data, learns on it, then outputs a single loss output for the meta algorithm to use. With this setup, we could use any online learning algorithm for the within-task algorithm. For this thesis, we will use Squint [Koolen and Van Erven, 2015] as the within-task algorithm since our goal is to minimize the regret and we get to experience fast rates while using Squint due to the Bernstein Condition assumption at this level. Fast rates means that the algorithm's average regret converges at a rate in the order of $\mathcal{O}\left(\frac{1}{n}\right)$.

Algorithm 5 is sectioned into two levels. At the meta-level, we are using Decreasing Hedge and at the within-task level, we are using Squint. At the start of each task $k \in K$, the algorithm updates the distribution $\mathcal{P}_k$ over the set of representations $\mathcal{G}$

**Algorithm 5:** Lifelong Learning with Delayed Meta-Update

Initialize $\mathcal{P}_k^1$ for all $g \in \mathcal{G}$
Set $\tilde{L}_0^g = 0$ for all $g \in \mathcal{G}$
**for** $k = 1$ *to* $K$ **do**
  $L_{0,k}^{g,h} = 0$ for all $h \in \mathcal{H}$ and $g \in \mathcal{G}$
  Update $\mathcal{P}_k^g$ for all $g \in \mathcal{G}$
  Initialize $\mathcal{M}_{k,1}^{g,h}$ for all $g \in \mathcal{G}$ and $h \in \mathcal{H}$
  **for** $a = 1$ *to* $n$ **do**
    **for** *all* $g$ **do**
      Update $\mathcal{M}_{a,k}^{g,h}$ for all $h \in \mathcal{H}$
    Suffer $\mathbb{E}_{g \sim \mathcal{P}_k^g}\left[\mathbb{E}_{h \sim \mathcal{M}_{a,k}^{g,h}}\left[\ell_{a,k}^{g,h}\right]\right]$
    Update $L_{a,k}^{g,h} = L_{a-1,k}^{g,h} + \ell_{a,k}^{g,h}$ for all $g \in \mathcal{G}$ and $h \in \mathcal{H}$
  $\tilde{L}_k^g = \tilde{L}_{k-1}^g + \frac{\min_{h \in \mathcal{H}} \sum_{a=1}^{n} \ell_{a,k}^{g,h}}{n}$, for all $g \in \mathcal{G}$

based on the meta-level algorithm. For the current task and for every representation $g \in \mathcal{G}$, at each timestep $a = 1, ..., n$, the algorithm updates the distribution $\mathcal{M}_{a,k}^g$ over the set of hypotheses $\mathcal{H}$ based on the within-task algorithm and suffers a loss. At the end of the task, the empirical risk of the ERM for every representation is given to the meta-level as the losses of task $k$, then a new task starts.

## 5.2 Regret Analysis

What we pass from the within-task level to the meta-level should be informative. We will bound the probability that the empirical risk of the ERM is $(\epsilon + \epsilon')$-within of the true risk of the true risk minimizer, where $\epsilon$ is an error term that is related to the distance between the true risk of the ERM and the true risk of the true risk minimizer and $\epsilon'$ is the error term related to the distance of the empirical risk of the ERM and the true risk of the ERM.

For the following lemma (Lemma 4), we will drop the $g$ superscript and $k$ subscript from the notation to show a more general result.

**Lemma 4.** *Let $B > 0$ and $\epsilon \geq 0$. For all hypotheses $h \in \mathcal{H}$, assume that the losses* *are i.i.d. and bounded,* $\ell_a^h \in [0,1]$*. We have the true risk of a hypothesis as*

$$R(h) = \mathbb{E}\left[\frac{1}{n}\sum_{a=1}^{n} \ell_a^h\right]$$

*and the empirical risk as*

$$\hat{R}(h) = \frac{1}{n}\sum_{a=1}^{n} \ell_a^h.$$

*We define two sets: the "good" set,* $\mathcal{H}_{\leq\epsilon} := \{h \in \mathcal{H} \mid R(h) - R(h^*) \leq \epsilon\}$*, and the "bad" set,* $\mathcal{H}_{>\epsilon} := \{h \in \mathcal{H} \mid R(h) - R(h^*) > \epsilon\}$*, where* $h^* = \arg\min_{h\in\mathcal{H}} R(h)$*.*

*The probability that the true risk of the ERM is not within* $\epsilon$ *of the true risk of the true risk minimizer,* $h^*$*, is*

$$\Pr\left(\arg\min_{h\in\mathcal{H}} \hat{R}(h) \in \mathcal{H}_{>\epsilon}\right) \leq |\mathcal{H}_{>\epsilon}| 2\exp\left(\frac{-n\epsilon}{2B+\frac{2}{3}}\right).$$

*Proof.* We will bound the probability that the ERM has a true risk, for some $\epsilon$, that is greater than the true risk minimizer's risk.

We bound the probability that the ERM is in the set of hypotheses whose true risk is $\epsilon$ more of the true risk of the true risk minimizer, which is the bad case

$$\Pr\left(\arg\min_{h\in\mathcal{H}} \hat{R}(h) \in \mathcal{H}_{>\epsilon}\right).$$

To upper bound this probability, we use that the probability that the smallest empirical risk of the "bad" set is less than or equal to the smallest empirical risk of the "good" set satisfies

$$\Pr\left(\arg\min_{h\in\mathcal{H}} \hat{R}(h) \in \mathcal{H}_{>\epsilon}\right) \leq \Pr\left(\min_{h\in\mathcal{H}_{>\epsilon}} \hat{R}(h) \leq \min_{h\in\mathcal{H}_{\leq\epsilon}} \hat{R}(h)\right).$$

From there, we can introduce the ERM of the "bad" set, $\hat{h}_\epsilon = \arg\min_{h\in\mathcal{H}_{>\epsilon}} \hat{R}(h)$,

$$\begin{aligned}
\Pr\left(\min_{h\in\mathcal{H}_{>\epsilon}} \hat{R}(h) \leq \min_{h\in\mathcal{H}_{\leq\epsilon}} \hat{R}(h)\right) &= \Pr\left(\hat{R}(\hat{h}_\epsilon) \leq \min_{h\in\mathcal{H}_{\leq\epsilon}} \hat{R}(h)\right) \\
&\leq \Pr\left(\hat{R}(\hat{h}_\epsilon) \leq \hat{R}(h^*)\right). \\
&= \Pr\left(\hat{R}(\hat{h}_\epsilon) - \hat{R}(h^*) \leq 0\right).
\end{aligned}$$

We can re-express the last line of the math display above to be able to use Lemma 1, Bernstein's inequality. For ease of reading, let us denote $\bar{X}_h = \hat{R}(h) - \hat{R}(h^*)$ for the following:

$$\begin{aligned}
\Pr\left(\bar{X}_{\hat{h}_\epsilon} \leq 0\right) &= \Pr\left(\bar{X}_{\hat{h}_\epsilon} - \mathbb{E}\left[\bar{X}_{\hat{h}_\epsilon}\right] \leq -\mathbb{E}\left[\bar{X}_{\hat{h}_\epsilon}\right]\right) \\
&\leq \Pr\left(\exists h \in \mathcal{H}_{>\epsilon} : \bar{X}_h - \mathbb{E}\left[\bar{X}_h\right] \leq -\mathbb{E}\left[\bar{X}_h\right]\right) \\
&\leq \sum_{h\in\mathcal{H}_{>\epsilon}} \Pr\left(\bar{X}_h - \mathbb{E}\left[\bar{X}_h\right] \leq -\mathbb{E}\left[\bar{X}_h\right]\right) \\
&= \sum_{h\in\mathcal{H}_{>\epsilon}} \Pr\left(\mathbb{E}\left[\bar{X}_h\right] - \bar{X}_h \geq \mathbb{E}\left[\bar{X}_h\right]\right).
\end{aligned}$$

We set $Y_h = -\bar{X}_h$ to get

$$\begin{aligned}
\sum_{h\in\mathcal{H}_{>\epsilon}} \Pr\left(\mathbb{E}\left[\bar{X}_h\right] - \bar{X}_h \geq \mathbb{E}\left[\bar{X}_h\right]\right) &= \sum_{h\in\mathcal{H}_{>\epsilon}} \Pr\left(\mathbb{E}\left[-Y_h\right] - (-Y_h) \geq \mathbb{E}\left[\bar{X}_h\right]\right) \\
&= \sum_{h\in\mathcal{H}_{>\epsilon}} \Pr\left(Y_h - \mathbb{E}\left[Y_h\right] \geq \mathbb{E}\left[\bar{X}_h\right]\right).
\end{aligned}$$

Let us rewrite $\bar{X}_h$ in terms of losses, such that $\bar{X}_h = \frac{1}{n}\sum_{a=1}^{n}\left(\ell_a^h - \ell_a^{h^*}\right)$ and define $X_{a,h} = \ell_a^h - \ell_a^{h^*}$, so that $\bar{X}_h = \frac{1}{n}\sum_{a=1}^{n} X_{a,h}$.

Now we can apply Bernstein's inequality (3.6) on

$$\begin{aligned}\sum_{h\in\mathcal{H}_{>\epsilon}} \Pr\left(Y_h - \mathbb{E}\left[Y_h\right] \geq \mathbb{E}\left[\bar{X}_h\right]\right) &\leq \sum_{h\in\mathcal{H}_{>\epsilon}} \exp\left(\frac{-n\mathbb{E}\left[\bar{X}_h\right]^2}{2\left(\frac{1}{n}\sum_{a=1}^{n}\mathbb{E}\left[X_{a,h}^2\right]\right)+\frac{2\mathbb{E}\left[\bar{X}_h\right]}{3}}\right)\\ &\leq \sum_{h\in\mathcal{H}_{>\epsilon}} \exp\left(\frac{-n\mathbb{E}\left[\bar{X}_h\right]^2}{2B\mathbb{E}\left[\bar{X}_h\right]+\frac{2\mathbb{E}\left[\bar{X}_h\right]}{3}}\right) \qquad (5.1)\\ &= \sum_{h\in\mathcal{H}_{>\epsilon}} \exp\left(\frac{-n\mathbb{E}\left[\bar{X}_h\right]^2}{\mathbb{E}\left[\bar{X}_h\right]\left(2B+\frac{2}{3}\right)}\right)\\ &= \sum_{h\in\mathcal{H}_{>\epsilon}} \exp\left(\frac{-n\mathbb{E}\left[\bar{X}_h\right]}{2B+\frac{2}{3}}\right)\\ &\leq \left|\mathcal{H}_{>\epsilon}\right| \exp\left(\frac{-n\epsilon}{2B+\frac{2}{3}}\right).\end{aligned}$$

We get (5.1) because we are assuming the $(B, 1)$-Bernstein condition is satisfied. Using (3.5), we can upper bound this second moment by

$$\begin{aligned}\frac{1}{n}\sum_{a=1}^{n}\mathbb{E}\left[X_{a,h}^2\right] &= \frac{1}{n}\sum_{a=1}^{n}\left(\mathbb{E}\left[\left(\ell_a^h - \ell_a^{h^*}\right)^2\right]\right)\\ &\leq \frac{1}{n}\sum_{a=1}^{n}\left(B\mathbb{E}\left[\ell_a^h - \ell_a^{h^*}\right]\right)\\ &= B\mathbb{E}\left[\bar{X}_h\right].\end{aligned}$$

□

Let us write the risk and true risk for the meta-level setting now, such that the true risk for hypothesis $h_{a,k}^g \in \mathcal{H}$, $R_k^g(h) = \mathbb{E}\left[\frac{1}{n}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right]$ and the empirical risk $\hat{R}_k^g(h) = \frac{1}{n}\sum_{a=1}^{n}\ell_{a,k}^{g,h}$.

**Lemma 5.** *Let $n \geq 1$. For a fixed representation $g \in \mathcal{G}$, a fixed $k \in [K]$, the probability that the risk of the ERM $\hat{h}_k^g \in \mathcal{H}$ is not close to its true risk is*

$$\Pr\left(|R_k^g(\hat{h}_k^g) - \hat{R}_k^g(\hat{h}_k^g)| \geq \epsilon'\right) \leq |\mathcal{H}| 2\exp\left(-2n\left(\epsilon'\right)^2\right),$$

*for all $\epsilon' \geq 0$.*

*Proof.* For a fixed task $k$, representation $g$, and hypothesis $h \in \mathcal{H}$, we use Hoeffding's inequality with $0 \leq \ell_{a,k}^{g,h} \leq 1$ to get a bound of

$$\Pr\left(|R_k^g(h) - \hat{R}_k^g(h)| \geq \epsilon'\right) \leq 2\exp\left(-2n\left(\epsilon'\right)^2\right).$$

Now, we use a union bound over all hypotheses, which gives us

$$\Pr\left(\exists h \in \mathcal{H} : |R_k^g(h) - \hat{R}_k^g(h)| \geq \epsilon'\right) \leq |\mathcal{H}| 2\exp\left(-2n\left(\epsilon'\right)^2\right).$$

□

Using Lemma 4 and Lemma 5 to map the results of Theorem 1, we are able to get an upper bound on the pseudo-regret of Algorithm 5.

**Theorem 2.** *In the meta-level, assume that there is some representation $g^* \in \mathcal{G}$, $\tau^* \geq 1$, and $\Delta_{\mathcal{G}} > 0$ such that for all $g \neq g^*$ and $k \geq \tau^*$, the following holds:*

$$\mathbb{E}\left[L_k^{g,*}\right] - \mathbb{E}\left[L_k^{g^*,*}\right] \geq k\Delta_{\mathcal{G}},$$

*where $L_k^{g,*} = \sum_{b=1}^{k}\sum_{a=1}^{n} \frac{\ell_{a,b}^{g,h_a^{*,g}}}{n}$ and $h_k^{*,g} = \arg\min_{h\in\mathcal{H}} \mathbb{E}\left[\ell_{a,k}^{g,h}\right]$ for all $a \in [n]$.*

*In the within-task level, for all representations $g \in \mathcal{G}$, tasks $k \in [K]$, and timesteps $a \in [n]$ for $n \geq 1$, assume that the losses are i.i.d. and bounded $\ell_{a,k}^{g,h} \in [0,1]$ and that each task satisfies the $(B,1)$-Bernstein condition. For any $0 \leq \epsilon + \epsilon' \leq \frac{\Delta_{\mathcal{G}}}{4}$ and $\epsilon, \epsilon' \geq 0$, under the above conditions and with probability $1 - K|\mathcal{G}||\mathcal{H}|\left(2\exp\left(-2n\left(\epsilon'\right)^2\right) + \exp\left(\frac{-n\epsilon}{2B+\frac{2}{3}}\right)\right)$, Algorithm 5 has a pseudo-regret bound of*

$$\mathcal{R}_K \leq \mathcal{O}\left(\sqrt{\tau^* \log\left(|\mathcal{G}|\right)} + \frac{\log\left(|\mathcal{G}|\right)}{\Delta_{\mathcal{G}} - 4(\epsilon + \epsilon')} + \frac{\log\left(\Delta_{\mathcal{G}}^{-1}\right)}{\Delta_{\mathcal{G}}} + K|\mathcal{G}|\left(B+1\right)\left(\ln\left(|\mathcal{H}|\right) + \ln\ln n\right)\right).$$

*Proof.* We decompose the lifelong regret into two pieces: the meta-level and the within-task level. Our regret breakdown involves one intermediate step, which is to introduce the $\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right]\right]$ term like in the following:

$$\begin{aligned}\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^g}\left[\mathbb{E}\left[\ell_{a,k}^{g,h}\right]\right]\right]-\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right] &= \sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^g}\left[\mathbb{E}\left[\ell_{a,k}^{g,h}\right]\right]\right]\\ &-\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right]+\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right]\right]-\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right]\right]\\ &= \underbrace{\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right]-\sum_{k=1}^{K}\sum_{a=1}^{n}\ell_{a,k}^{g^*,h_k^{*,g^*}}}_{\text{Meta regret}}+\underbrace{\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^g}\left[\ell_{a,k}^{g,h}\right]-\ell_{a,k}^{g,h_k^{*,g}}\right]}_{\text{Within-task regret}}.\end{aligned}$$

The within-task regret focuses on the difference caused by the hypotheses for each representation-task pairing,

$$\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^g}\left[\mathbb{E}\left[\ell_{a,k}^{g,h}-\ell_{a,k}^{g,h_k^{*,g}}\right]\right]\right]. \tag{5.2}$$

The meta regret compares the representations based on the losses from the best hypothesis in expectation for each representation-task pair,

$$\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right]\right]-\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right]. \tag{5.3}$$

We begin by establishing a bound on the meta regret (5.3). We use Lemma 4 and Lemma 5, which allows us to extend the guarantees of Theorem 1 from the single-task setting to the meta-level.

The extension to the meta-level setting is made possible by identifying a mapping that aligns elements of the meta-level setting with their corresponding elements in the single-task setting. With this mapping, the representations $g \in \mathcal{G}$ play the role of the single task experts $i \in M$ and tasks $k \in [K]$ correspond to the time steps $t \in [T]$. The sub-optimality gap in the single task is denoted by $\Delta$ and in the meta-level, we denote it as $\Delta_{\mathcal{G}}$. For the losses used in the update rules, we define the meta-level losses as $\tilde{\ell}_k^g = \frac{\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}}{n}$ for each task-representation pair and the single task losses as $\tilde{\ell}_{i,t} = \ell_{i,a} + \nu_{i,a}$ for each expert at that timestep.

In the single-task setting, we have non-perturbed losses $\ell_{i,t}$ and the equivalent in the meta-level would be $R_k^g(h_k^{*,g})$. Since we are using the ERM from the tasks, we are not guaranteed to have the smallest loss in expectation for that task and representation, so we will account for the amount of deviation away from the true risk minimizer by using Lemma 4 to give a high-probability bound on the distance of the risk of the ERM to the risk of the true risk minimizer

$$|R_k^g(\hat{h}_k^g) - R_k^g(h_k^{*,g})| \leq \epsilon$$

and then using Lemma 5 to give a high-probability bound on the distance of the empirical risk to the risk for both the ERM and the true risk minimizer

$$|R_k^g(\hat{h}_k^g) - \hat{R}_k^g(\hat{h}_k^g)| \leq \epsilon'.$$

This gives us a high-probability bound of the form

$$|\hat{R}_k^g(\hat{h}_k^g) - R_k^g(h_k^{*,g})| \leq \epsilon + \epsilon'.$$

This $\epsilon + \epsilon'$ is the corresponding noise $\epsilon_{ST}$ in the single-task setting.

We take a union bound of Lemma 4 and Lemma 5 for all task and representation pairs. For Lemma 4, this gives us

$$\begin{aligned}
\Pr\left(\exists k \in [K], \exists g \in \mathcal{G} : \operatorname*{arg\,min}_{h_{a,k}^g \in \mathcal{H}} \hat{R}(h_{a,k}^g) \in \mathcal{H}_{>\epsilon}\right) &\leq \sum_{k=1}^{K} \sum_{g \in |\mathcal{G}|} \Pr\left(\operatorname*{arg\,min}_{h \in \mathcal{H}} \hat{R}(h_{a,k}^g) \in \mathcal{H}_{>\epsilon}\right) \\
&\leq K|\mathcal{G}||\mathcal{H}_{>\epsilon}| \exp\left(\frac{-n\epsilon}{2B + \frac{2}{3}}\right) \\
&\leq K|\mathcal{G}||\mathcal{H}| \exp\left(\frac{-n\epsilon}{2B + \frac{2}{3}}\right).
\end{aligned}$$

For Lemma 5, this union bound gives us

$$\begin{aligned}
\Pr\left(\exists k \in [K], \exists g \in \mathcal{G} : |R(h_{a,k}^g) - \hat{R}(h_{a,k}^g)| \geq \epsilon'\right) &\leq \sum_{k=1}^{K} \sum_{g \in |\mathcal{G}|} \Pr\left(|R(h_{a,k}^g) - \hat{R}(h_{a,k}^g)| \geq \epsilon'\right) \\
&\leq K|\mathcal{G}||\mathcal{H}| 2 \exp\left(-2n\left(\epsilon'\right)^2\right).
\end{aligned}$$

Now that we have our ERM for every representation and task pairing in the within-$\epsilon$ set and, moreover, at the end of the task, all hypotheses will have their risk close to their true risk. We can view the risk of the ERMs as the perturbed losses of Theorem 1.

For all tasks $k = 1, ..., K$ and with a probability of $1-K|\mathcal{G}||\mathcal{H}|\left(2\exp\left(-2n\left(\epsilon'\right)^2\right)+\exp\left(\frac{-n\epsilon}{2B+\frac{2}{3}}\right)\right)$, we get a pseudo-regret bound for the meta-regret of

$$\begin{aligned}\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}\left[\ell_{a,k}^{g,h_k^{*,g}}\right]\right]-\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}\left[\ell_{a,k}^{g^*,h_k^{*,g^*}}\right] &= n\Bigg(\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}\left[\frac{\sum_{a=1}^{n}\ell_{a,k}^{g,h_k^{*,g}}}{n}\right]\right]\\ &\quad-\sum_{k=1}^{K}\mathbb{E}\left[\frac{\sum_{a=1}^{n}\ell_{a,k}^{g^*,h_k^{*,g^*}}}{n}\right]\Bigg)\\ &\leq \mathcal{O}\left(\sqrt{\tau^*\log\left(|\mathcal{G}|\right)}+\frac{\log\left(|\mathcal{G}|\right)}{\Delta_{\mathcal{G}}-4(\epsilon+\epsilon')}+\frac{\log\left(\Delta_{\mathcal{G}}^{-1}\right)}{\Delta_{\mathcal{G}}}\right).\end{aligned}$$

Now we bound the within-task regret (5.2).

The pseudo-regret bound that we acquire from Squint is from Theorem 3 of Koolen et al. [2016]. That theorem states that in any stochastic setting that satisfies the $(B, \kappa)$-Bernstein condition, Squint achieves fast rates in expected regret in the order of

$$\mathcal{R}\leq\mathbb{E}\left[R\right]\leq\left(1+4B\right)\left(\frac{F_n}{4}\right)^{\frac{1}{2-\kappa}}n^{\frac{1-\kappa}{2-\kappa}}+\left(5-\kappa\right)F_n,$$

where $F_n = \mathcal{O}\left(-\ln(\pi^{h_k^{*,g}})+\ln\ln n\right)$ and $\pi = (\pi^1, \pi^2, ...,)$ is the prior probability mass function.

If we take $\kappa = 1$ and assume that the prior over the experts is the uniform distribution, then the Squint pseudo regret is of order $\mathcal{O}\left(B\left(\ln(|\mathcal{H}|)+\ln\ln n\right)+\ln(|\mathcal{H}|)+\ln\ln n\right)$.

Now that we have bound both the meta-level regret (5.3) and the within-task regret (5.2), we can add them together and get the complete regret for Algorithm 5. With probability of $1-K|\mathcal{G}||\mathcal{H}|\left(2\exp\left(-2n\left(\epsilon'\right)^2\right)+\exp\left(\frac{-n\epsilon}{2B+\frac{2}{3}}\right)\right)$, the complete lifelong learning regret for the SCA setting and each task satisfying the $(B, 1)$-Bernstein condition is

$$
\begin{aligned}
R_K \leq & \mathcal{O}\Big(\sqrt{\tau^* \log\left(|\mathcal{G}|\right)} + \frac{\log\left(|\mathcal{G}|\right)}{\Delta_{\mathcal{G}} - 4(\epsilon+\epsilon')} + \frac{\log\left(\Delta_{\mathcal{G}}^{-1}\right)}{\Delta_{\mathcal{G}}} \\
& + K|\mathcal{G}|\left(B+1\right)\left(\ln\left(|\mathcal{H}|\right) + \ln\ln n\right)\Big).
\end{aligned}
$$

□

**Theorem 3.** *Assume the losses in the within-task level are bounded $\ell_{a,k}^{g,h} \in [0, 1]$. The conditional expected regret of Algorithm 5 in the fully adversarial setting is*

$$
R_K \leq \mathcal{O}\left(n\sqrt{K\log\left(|\mathcal{G}|\right)} + K\left(\sqrt{V_n\left(\ln(|\mathcal{H}|) + \ln\ln n\right)} + \ln(|\mathcal{H}|) + \ln\ln n\right)\right).
$$

*Proof.* We will decompose the conditional expected regret into two pieces by introducing a intermediate term $\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right]$, which gives us

$$
\begin{aligned}
\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right] - \min_{g\in\mathcal{G}}\sum_{k=1}^{K}\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h} &= \sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right] \\
- \min_{g\in\mathcal{G}}\sum_{k=1}^{K}\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h} &+ \sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right] - \sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right] \\
&= \underbrace{\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right] - \min_{g\in\mathcal{G}}\sum_{k=1}^{K}\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}}_{\text{Meta regret}} \\
&+ \underbrace{\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^{g}}\left[\ell_{a,k}^{g,h}\right]\right] - \sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right]}_{\text{Within-task regret}}. \qquad (5.4)
\end{aligned}
$$

Firstly, we bound the meta-regret portion of (5.4). Since we use the empirical risk of the ERM of each representation-task pairing as our losses in the meta level, we will rewrite the meta-regret in terms of empirical risk

$$
\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right] - \min_{g\in\mathcal{G}}\sum_{k=1}^{K}\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h} = n\left(\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\hat{R}_k^g\left(\hat{h}_k^g\right)\right] - \min_{g\in\mathcal{G}}\sum_{k=1}^{K}\hat{R}_k^g\left(\hat{h}_k^g\right)\right).
$$

We can bound the expression above with the adversarial guarantee of Decreasing Hedge [Mourtada and Gaïffas, 2019]

$$n\left(\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\hat{R}_k^g(\hat{h}_k^g)\right]-\min_{g\in\mathcal{G}}\sum_{k=1}^{K}\hat{R}_k^g(\hat{h}_k^g)\right)\leq n\sqrt{K\log\left(|\mathcal{G}|\right)}.$$

We bound the second part of (5.4) to get the within-task regret. This is Squint's regret in the adversarial setting (Koolen et al. [2016] [Theorem 3])

$$\sum_{k=1}^{K}\sum_{a=1}^{n}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\mathbb{E}_{h\sim\mathcal{M}_{a,k}^g}\left[\ell_{a,k}^{g,h}\right]\right]-\sum_{k=1}^{K}\mathbb{E}_{g\sim\mathcal{P}_k}\left[\min_{h\in\mathcal{H}}\sum_{a=1}^{n}\ell_{a,k}^{g,h}\right]\leq K\left(\sqrt{V_nF_n}+F_n\right),$$

where $F_n=\mathcal{O}\left(-\ln(\frac{1}{|\mathcal{H}|})+\ln\ln n\right)$.
Now that we have both pieces of our regret decomposition bounded, we combine them to get our full lifelong learning regret in the adversarial setting,

$$R_T\leq\mathcal{O}\left(n\sqrt{K\log\left(|\mathcal{G}|\right)}+K\left(\sqrt{V_n\left(\ln(|\mathcal{H}|)+\ln\ln n\right)}+\ln(|\mathcal{H}|)+\ln\ln n\right)\right).$$

□

# Chapter 6

# Conclusion and Future Work

In this thesis, we presented the SCA assumption in both the single task setting and the lifelong learning setting. Under the SCA assumption in the single task setting, we proved a pseudo-regret bound for Decreasing Hedge with perturbed losses that accumulates worst-case regret only until the algorithm passes timestep $\tau^*$. After the algorithm passes timestep $\tau^*$, Decreasing Hedge only picks up a constant (in $T$) amount of additional regret.

In the lifelong setting, we introduced a lifelong learning algorithm (Algorithm 5) with best-of-both-world regret guarantees. Since Algorithm 5 uses the empirical risk of the ERM of each representation-task pairing $(g, k)$ to update the meta-level at the end of each task, we created a mapping between the single task setting with perturbed losses and this lifelong learning setting. This mapping bounded the probability that for every representation-task pairing $(g, k)$, the empirical risk of the ERM is $(\epsilon + \epsilon')$-within the true risk of the true risk minimizer, since the true risk of the true risk minimizer is the meta-level equivalent of the non-perturbed losses in the single task setting.

Algorithm 5 uses a delayed-update procedure, where the meta-level is updated only at the end of each task. Wu et al. [2019] and Wu et al. [2021] instead use algorithms that update the meta-level at every timestep in order to reduce the number of tasks needed to achieve small regret. Incorporating a more frequent update procedure would be a promising direction for future work.

Another direction for future work is to reduce the computational complexity arising from running $|\mathcal{G}|$ number of copies of the within-task algorithm for each task. Dictionary learning could be used to reduce the dependency on the number of representations. Alquier et al. [2017] used dictionary learning at the meta-level to deal with the case of infinite representations. Another approach could be to use bandit-feedback. Wu et al. [2019] proved regret bounds for the stochastic bandit setting and the adversarial bandit setting, although with different algorithms for each setting. The cost of using bandit-feedback would be an increase to the regret.

# Bibliography

Pierre Alquier, Massimiliano Pontil, et al. Regret bounds for lifelong learning. In *Artificial Intelligence and Statistics*, pages 261–269. PMLR, 2017.

Peter Auer, Nicolò Cesa-Bianchi, and Claudio Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64:48–75, 01 2000. doi: 10.1006/jcss.2001.1795.

Peter L. Bartlett and Shahar Mendelson. Empirical minimization. *Probability Theory and Related Fields*, 135(3):311–334, 2006. URL https://doi.org/10.1007/s00440-005-0462-3.

J. Baxter. A model of inductive bias learning. *Journal of Artificial Intelligence Research*, 12:149–198, March 2000. ISSN 1076-9757. doi: 10.1613/jair.731. URL http://dx.doi.org/10.1613/jair.731.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolo Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice, 2006. URL https://arxiv.org/abs/math/0602629.

Giulia Denevi, Dimitris Stamos, Carlo Ciliberto, and Massimiliano Pontil. Online-within-online meta-learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.,

2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/e0e2b58d64fb37a2527329a5ce093d80-Paper.pdf.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997.

Eyal Gofer, Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Regret minimization for branching experts. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 618–638, Princeton, NJ, USA, 12–14 Jun 2013. PMLR. URL https://proceedings.mlr.press/v30/Gofer13.html.

Wouter M Koolen and Tim Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175. PMLR, 2015.

Wouter M Koolen, Peter Grünwald, and Tim Van Erven. Combining adversarial guarantees and stochastic fast rates in online learning. *Advances in Neural Information Processing Systems*, 29, 2016.

Jaouad Mourtada and Stéphane Gaïffas. On the optimality of the hedge algorithm in the stochastic regime. *Journal of Machine Learning Research*, 20:1–28, 2019.

Anastasia Pentina and Christoph H Lampert. Lifelong learning with non-i.i.d. tasks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/9232fe81225bcaef853ae32870a2b0fe-Paper.pdf.

Anastasia Pentina and Ruth Urner. Lifelong learning with weighted majority votes. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/f39ae9ff3a81f499230c4126e01f421b-Paper.pdf.

Sebastian Thrun. Is learning the n-th thing any easier than learning the first? In D. Touretzky, M.C. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8. MIT Press,

1995. URL https://proceedings.neurips.cc/paper_files/paper/1995/file/bdb106a0560c4e46ccc488ef010af787-Paper.pdf.

Yi-Shan Wu, Po-An Wang, and Chi-Jen Lu. Lifelong optimization with low regret. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 448–456. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/wu19a.html.

Yi-Shan Wu, Yi-Te Hong, and Chi-Jen Lu. Lifelong learning with branching experts. In *Asian Conference on Machine Learning*, pages 1161–1175. PMLR, 2021.