Hybrid Stochastic-Adversarial On-line Learning

Alessandro Lazaric
SequeL Project,
INRIA Lille - Nord Europe, France
alessandro.lazaric@inria.fr

Rémi Munos
SequeL Project,
INRIA Lille - Nord Europe, France
remi.munos@inria.fr

Abstract

Most of the research in online learning focused either on the problem of adversarial classification (i.e., both inputs and labels are arbitrarily chosen by an adversary) or on the traditional supervised learning problem in which samples are independently generated from a fixed probability distribution. Nonetheless, in a number of domains the relationship between inputs and labels may be adversarial, whereas input instances are generated according to a constant distribution. This scenario can be formalized as an hybrid classification problem in which inputs are stochastic, while labels are adversarial. In this paper, we introduce this hybrid stochastic-adversarial classification problem, we propose an online learning algorithm for its solution, and we analyze its performance. In particular, we show that, given a hypothesis space $\mathcal{H}$ with finite VC dimension, it is possible to incrementally build a suitable finite set of hypotheses that can be used as input for an exponentially weighted forecaster achieving a cumulative regret of order $O(\sqrt{nVC(\mathcal{H})\log n})$ with overwhelming probability. Finally, we discuss extensions to multi-label classification, learning from experts and bandit settings with stochastic side information, and application to games.

1 Introduction

Motivation and relevance. The problem of classification has been intensively studied in supervised learning both in the stochastic and adversarial settings. In the former, inputs and labels are jointly drawn from a fixed probability distribution, while in the latter no assumption is made on the way the sequence of input-label pairs is generated. Although the adversarial setting allows to consider a wide range of problems by dropping any assumption about data, in many applications it is possible to consider an hybrid scenario in which inputs are drawn from a probability distribution, while labels are adversarialy chosen. Let us consider a problem in which a company tries to predict whether a user is likely to buy an item or not (e.g., a new model of mobile phone, a new service) on the basis of a set of features describing her profile (e.g., sex, age, salary, etc.). In the medium-term, user profiles can be well assumed as coming from a fixed probability distribution. In fact, features such as age and salary are almost constant and their distribution in a sample set does not change in time. On the other hand, user preferences may rapidly change in an unpredictable way (e.g., because of competitors who released a new product). This scenario can be formalized as a classification problem with stochastic inputs and adversarial labels. Alternatively, the problem can be casted as a two-player games in which the structure of the game (i.e., the payoffs) is determined by a stochastic event x (e.g., a card, a dice). Each player selects a strategy h defined over all the possible events and plays action $h(x)$. In general, the resulting payoff is a function of the actions and the stochastic event x. The Nash equilibrium in such a game is a pair of mixed strategies (i.e., a probability distribution over the set of pure strategies) such that their expected payoff (where expectation is taken on strategies randomization and the event distribution) cannot be improved by unilateral deviations from equilibrium strategies.

1: **for** $t = 1, 2, \dots$ **do**
2: A sample $x_t \overset{iid}{\sim} P$ is revealed to both the learner and the adversary
3: Simultaneously,
- Adversary chooses a loss function $\ell_t : \mathcal{Y} \to [0, 1]$
- Learner chooses a hypothesis $h_t \in \mathcal{H}$
4: Learner predicts $\widehat{y}_t = h_t(x_t) \in \mathcal{Y}$
5: Learner observes the feedback:
- $\ell_t(\widehat{y}_t)$ (in case of *bandit* information)
- *or* $\ell_t(\cdot)$ (in case of *full* information)
6: Learner incurs a loss $\ell_t(\widehat{y}_t)$
7: **end for**

Figure 1: The protocol of the general hybrid stochastic-adversarial setting.

Definition of the general problem. More formally, we consider the general prediction problem summarized in the *protocol* in Figure 1. At each round t an input x_t is drawn from a fixed distribution P (unknown from the learner) and revealed to both the learner and the adversary. Simultaneously, the adversary chooses a loss function ℓ_t and the learner chooses a hypothesis h_t in a set of available hypotheses $\mathcal{H}$ and predicts

$\widehat{y}_t = h_t(x_t)$. The feedback returned to the learner can be either the loss function ℓ_t (i.e., *full* information) or just the loss $\ell_t(\widehat{y}_t)$ of the chosen prediction (i.e., *bandit* information). The objective of the learner is to minimize her regret, that is to incur a cumulative loss that is almost as small as the one obtained by the best hypothesis in $\mathcal{H}$ on the same sequence of input-label pairs. More formally, for any $n > 0$, the regret of an algorithm $\mathcal{A}$ is

$$R_n(\mathcal{A}) = \sum_{t=1}^{n} \ell_t(h_t(x_t)) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell_t(h(x_t)), \qquad (1)$$

where h_t is the hypothesis chosen by $\mathcal{A}$ at time t.

Results so far. Many theoretical results are available for a number of online learning algorithms in the adversarial setting with full information. Given a finite set of N experts (i.e., hypotheses) as input, at each round the Exponentially Weighted Forecaster (EWF) [LW94, CBFH$^+$97, Vov98] randomizes on experts' predictions with a probability concentrated on experts which had a good performance so far (i.e., low cumulative loss). Despite its simplicity, the EWF achieves an upper-bound regret of $O(\sqrt{n \log N})$, where n is the time horizon of the problem. Although the mild dependency on the number of experts allows to use a large number of experts, the EWF cannot be directly extended to the case of infinite sets of experts. Many margin based algorithms with linear hypotheses have been proposed for adversarial classification [Ros58, WW99, CS03]. The simplest example of this class of algorithms is the perceptron [Ros58] in which a weight matrix W is updated whenever a prediction mistake is made. The number of classification mistakes of the perceptron is bounded [FSSSU06] by $L + D + \sqrt{LD}$ where L is the cumulative loss and D is the complexity of any weight matrix. In the linearly separable case (i.e., input-label pairs can be perfectly classified by a linear predictor, that is $L = 0$), the number of mistake is finite (for any time horizon n) and depends on the complexity D of the weight matrix corresponding to the optimal predictor. The agnostic online learning algorithm recently proposed in [SS08] successfully merges the effectiveness of the EWF with the general case of an infinite hypothesis set $\mathcal{H}$. Under the assumption that the Littlestone dimension [Lit88] of $\mathcal{H}$ is finite ($Ldim(\mathcal{H}) < \infty$), it is possible to define a suitable finite subset of the hypothesis space such that the EWF achieves a regret of the order of $O(Ldim(\mathcal{H}) + \sqrt{nLdim(\mathcal{H}) \log n})$.

The problem of classification with bandit information (also known as contextual bandit problem) is of major interest in applications in which the true label is not revealed and only the loss for the chosen label is returned to the learner (e.g., recommendation systems). This scenario is analyzed in [LZ07] in the fully stochastic setting. They introduce an epoch-based online learning algorithm whose regret can be bounded by merging supervised sample bounds with bandit bounds. In [KSST08] a modification of the perceptron is proposed (i.e., the *banditron*) to solve the online multi-label classification problem in the fully adversarial case. In particular, they analyze the performance of the banditron in terms of mistake bounds with particular attention to the linearly separable case.

What we have done. While all the previous approaches consider either the fully adversarial or fully stochastic setting, in this paper, we analyze the problem of prediction in case of stochastic inputs and adversarial loss functions. In Section 2, we consider a specific instance of the general problem, that is the problem of binary classification with full information. In Section 3 we devise an epoch-based algorithm that, given a hypothesis set $\mathcal{H}$ as input, incrementally builds a finite subset of $\mathcal{H}$ on the basis of the sequence of inputs experienced so far. At the beginning of each epoch, a new subset of $\mathcal{H}$ is generated and it is given as input to a EWF which is run until the end of the epoch. Because of the stochastic assumption about the generation of inputs, the complexity of the hypothesis space $\mathcal{H}$ can be measured according to the VC dimension instead of the Littlestone dimension like in the agnostic online learning algorithm. As a result, the algorithm performance can be directly obtained by merging the EWF performance in the adversarial setting and usual capacity measures for hypothesis spaces in stochastic problems (e.g., their VC dimension). The resulting algorithm is proved to incur a regret of order $O(\sqrt{nVC(\mathcal{H}) \log n})$ with overwhelming probability. A number of extensions are then considered in Section 4 for multi-label prediction, bandit information, and games with stochastic side information. Section 5 compares the proposed algorithm with existing online learning algorithms for the stochastic or adversarial setting. Finally, in Section 6 we draw conclusions.

1: **for** $t = 1, 2, \ldots$ **do**
2: A sample $x_t \overset{iid}{\sim} P$ is revealed to both the learner and the adversary
3: Simultaneously,
- Adversary chooses a label $y_t \in \mathcal{Y}$
- Learner chooses a hypothesis $h_t \in \mathcal{H}$
4: Learner predicts $\widehat{y}_t = h_t(x_t) \in \mathcal{Y}$
5: y_t is revealed
6: Learner incurs a loss $\ell(\widehat{y}_t, y_t) = \mathbb{I}\{\widehat{y}_t \neq y_t\}$
7: **end for**

Figure 2: The protocol of the hybrid stochastic-adversarial classification problem.

2 The Problem

Notation. In this section, we formally define the problem of binary classification and we introduce the notation used in the rest of the paper. Let $\mathcal{X}$ be the input space, P a probability distribution defined on $\mathcal{X}$, and $\mathcal{Y} = \{0, 1\}$ the set of labels. The learner is given as input a (possibly infinite) set $\mathcal{H}$ of hypotheses of the form $h : \mathcal{X} \to \mathcal{Y}$, mapping any possible input to a label. We define the distance between two hypotheses $h, h' \in \mathcal{H}$ as

$$\Delta(h, h') = \mathbb{E}_{x \sim P}\left[\mathbb{I}\{h(x) \neq h'(x)\}\right], \qquad (2)$$

(where $\mathbb{I}\{\xi\} = 1$ when event ξ is true, and 0 otherwise) that is, the probability that h and h' have different predictions given inputs drawn from P.

The protocol. The on-line classification problem we consider is summarized in Figure 2. The main difference with the general setting (Figure 1) is that at each round t the ad-

versary chooses a label y_t [1], and the learner incurs a loss $\ell(\widehat{y}_t, y_t)$ defined as $\mathbb{I}\{\widehat{y}_t \neq y_t\}$. In the following, we will use the short form $\ell_t(h)$ for $\ell(h(x_t), y_t)$ with $h \in \mathcal{H}$. Since at the end of each round the true label y_t is explicitly revealed (i.e., *full information* feedback), the learner can compute the loss for any hypothesis in $\mathcal{H}$. The objective of the learner is to minimize regret (1). As it can be noticed, the loss $\ell_t(h_t)$ is a random variable that depends on both the (randomized) algorithm and the distribution P. All the results obtained in the following will be stated in high-probability with respect to these two sources of stochasticity. In the next section, we introduce the Epoch-based Stochastic Adversarial (EStochAd) forecaster for the classification problem with stochastic inputs and adversarial labels.

3 Hybrid Stochastic-Adversarial Algorithms

3.1 Finite hypothesis space

Before entering in details about the algorithm, we briefly recall the EWF with side information with a finite number of experts. Let the hypothesis space $\mathcal{H}$ contain $N < \infty$ hypotheses (i.e., experts). At time t, for each hypothesis h_i ($i \in \{1, \ldots, N\}$), a weight is computed as

$$w_i^t = \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(h_i)\right) \tag{3}$$

where η is a strictly positive parameter. According to the previous definition, the smaller the cumulative loss the higher the weight for the hypothesis. At each step t, a loss function ℓ_t is adversarialy chosen and at the same time, the EWF draws a hypothesis h_t from a distribution $\mathbf{p}^t = (p_1^t, \ldots, p_N^t)$, where $p_i^t = \frac{w_i^t}{\sum_{j=1}^N w_j^t}$. As a result, it incurs a loss $\ell_t(h_t)$. At the end of each round, weights are updated according to (3). The following result provides an upper-bound on the regret for EWF .

Theorem 1 *[CBL06] Let $n, N \geq 1$, $0 \leq \beta \leq 1$, $\eta > 0$ and $w_i^1 = 1$, $i \in \{1, \ldots, N\}$. The exponentially weighted average forecaster satisfies*

$$\begin{aligned} R_n &= \sum_{t=1}^n \ell_t(h_t) - \min_{h \in \mathcal{H}} \sum_{t=1}^n \ell_t(h) \\ &\leq \frac{\log N}{\eta} + \frac{n\eta}{2} + \sqrt{\frac{n}{2} \log \frac{1}{\beta}}, \end{aligned}$$

with probability at least $1 - \beta$. Optimizing the parameter $\eta = \sqrt{2 \log N / n}$, the bound becomes

$$R_n \leq \sqrt{2n \log N} + \sqrt{\frac{n}{2} \log \frac{1}{\beta}}. \tag{4}$$

The implicit assumption in the previous theorem is that the time horizon n is known in advance. As usual, it is possible to obtain an anytime result for the previous algorithm by setting the learning parameter η to be a decreasing function of t (see e.g. [ACBFS03]). As it can be noticed, the EWF has a logarithmic dependency on the number of experts, thus allowing to consider large sets of experts. Nonetheless, the EWF cannot be directly applied when $\mathcal{H}$ contains an infinite number of hypotheses. In next sections we show that when inputs are drawn from a fixed distribution and the hypothesis space has a finite VC dimension, it is possible to incrementally define a finite subset of $\mathcal{H}$ that can be used as input for a EWF with a regret similar to (4).

[1] In the general case of a non-oblivious adversary, y_t may depend on past inputs $\{x_s\}_{s<t}$, predictions $\{\widehat{y}_s\}_{s<t}$, and current input x_t.

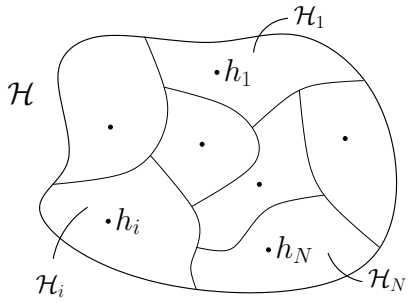

Figure 3: The hypothesis space $\mathcal{H}$ can be partitioned into classes containing hypotheses with the same sequence of prediction on inputs $\mathbf{x}_1^n$. The grid H_n is obtained by selecting one hypothesis for each class of the partition.

3.2 Infinite hypothesis space

Sequence of inputs known in advance. First, we show that for any finite VC dimension hypothesis space $\mathcal{H}$ and any sequence of inputs, it is possible to define in *hindsight* a finite subset $H \subset \mathcal{H}$ that contains hypotheses with exactly the same performance as those in the full set $\mathcal{H}$. Let $VC(\mathcal{H}) = d < \infty$ and $\mathbf{x}_1^n = (x_1, \ldots, x_n)$ be a sequence of inputs drawn from P. On the basis of $\mathbf{x}_1^n$, we define a partition $\mathcal{P}_n = \{\mathcal{H}_i\}_{i \leq N}$ of $\mathcal{H}$, such that each class $\mathcal{H}_i$ contains hypotheses with the same sequence of predictions up to time n (i.e., $\forall h, h' \in \mathcal{H}_i, h(x_s) = h'(x_s), s \leq n$). From each class we pick an arbitrary hypothesis $h_i \in \mathcal{H}_i$ and we define the grid $H_n = \{h_i\}_{i \leq N}$. Since $\mathcal{H}$ has a finite VC dimension, for any $n > 0$ the cardinality of H_n is bounded by $N = |H_n| \leq \left(\frac{en}{d}\right)^d < \infty$ [BBL04]. The grid H built from partition $\mathcal{P}$ of $\mathcal{H}$ can also be incrementally refined as inputs are revealed. For instance, after observing x_1, $\mathcal{H}$ is partitioned in two classes containing hypotheses which predict 0 in x_1 and those which predict 1 respectively. The set H_1 is obtained by choosing arbitrarily any two hypotheses from the two classes. As new inputs are observed each class is further split (see Figure 3) and after n inputs the hypothesis space is partitioned into at most $O(n^d)$ classes. Finally, H_n is obtained by taking one hypothesis from each class. As a result, for any hypothesis in $\mathcal{H}$ there exists a corresponding hypothesis in H_n which has exactly the same sequence of predictions on $\mathbf{x}_1^n$ and, thus, the very same performance.

Lemma 2 *Let H_n be the grid defined above, then*

$$\inf_{h \in \mathcal{H}} \sum_{t=1}^n \ell_t(h) = \min_{h' \in H_n} \sum_{t=1}^n \ell_t(h'), \tag{5}$$

that is, the performance of the best hypothesis in $\mathcal{H}$ on $\mathbf{x}_1^n$ is exactly the same obtained by the best hypothesis in H_n.

According to the previous lemma, if the sequence of inputs is available before the learning to take place, then the regret defined in (1) (that compares the cumulative loss of the learner to the performance of the best hypothesis in the full set $\mathcal{H}$) can be minimized by a EWF run on H_n, thus obtaining exactly the same performance as in Theorem 1.

Lemma 3 *Let a sequence of inputs $x_1, \dots, x_n \overset{iid}{\sim} P$ be available before learning and let H_n be the grid defined above, then*

$$R_n \leq \sqrt{2nd \log \frac{en}{d}} + \sqrt{\frac{n}{2} \log \frac{1}{\beta}}$$

with probability $1 - \beta$.

Proof: The lemma immediately follows from Lemma 2, Theorem 1, and $N \leq \left(\frac{en}{d}\right)^d$. ■

Sequence of auxiliary inputs. Unfortunately, the sequence of inputs $\mathbf{x}_1^n$ is rarely available beforehand, thus preventing from building H_n before the actual learning process starts. Nonetheless, in the following we show that in case of stochastic inputs, the learner can take advantage of any sequence of inputs drawn from the same distribution P to build a set H that can be used as input for a EWF . We will further show in Section 3.3 that we do not even need to know a sequence of auxiliary inputs beforehand and the mere assumption that inputs are drawn from a fixed (and unknown) distribution is sufficient to learn efficiently.

But first, let us assume an auxiliary sequence of n' inputs $(\mathbf{x}')_1^{n'} = (x'_1, \dots, x'_{n'})$ is available to the learner before the classification problem actually begins and let $H_{n'}$ be the grid of $\mathcal{H}$ built on inputs $(\mathbf{x}')_1^{n'}$. The regret of EWF with experts in $H_{n'}$ can be decomposed as

$$\begin{aligned} R_n &= \sum_{t=1}^n \ell_t(h_t) - \inf_{h \in \mathcal{H}} \sum_{t=1}^n \ell_t(h) \\ &= \left(\sum_{t=1}^n \ell_t(h_t) - \min_{h' \in H_{n'}} \sum_{t=1}^n \ell_t(h') \right) \\ &+ \left(\min_{h' \in H_{n'}} \sum_{t=1}^n \ell_t(h') - \inf_{h \in \mathcal{H}} \sum_{t=1}^n \ell_t(h) \right) \\ &= R_{EWF} + R_H, \end{aligned} \tag{6}$$

where R_{EWF} is the regret due to EWF and R_H comes from the use of $H_{n'}$ instead of the full hypothesis space $\mathcal{H}$. While the first term can be bounded as in Theorem 1, the second term in general is strictly positive. In fact, since $H_{n'}$ is different from the set H_n that would be created according to inputs $\mathbf{x}_1^n$, equality (5) does not hold for $H_{n'}$. In particular, in the fully adversarial case, the sequence of inputs could be chosen so that hypotheses in $H_{n'}$ have an arbitrarily bad performance when used to learn on $\mathbf{x}_1^n$ (e.g., if the learner is shown the same input for n' steps, $H_{n'}$ would contain only two hypotheses!). The situation is different in our hybrid stochastic-adversarial setting. In fact, since all the inputs are sampled from the same distribution P, $H_{n'}$ is likely to contain hypotheses that are good to predict on any other sequence of inputs drawn from P. Therefore, under the assumption that n' inputs can be sampled from P beforehand, we prove that the regret (6) is bounded by $O(\sqrt{nd \log n'})$ with high probability.

Let

$$\Delta_n(h, h') = \frac{1}{n} \sum_{t=1}^n \mathbb{I}\{h(x_t) \neq h'(x_t)\} \tag{7}$$

be the empirical distance between two hypotheses $h, h' \in \mathcal{H}$ on a sequence of inputs $\mathbf{x}_1^n$ (and define similarly $\Delta_{n'}(h, h')$ as the empirical distance of h and h' on inputs $(\mathbf{x}')_1^{n'}$). The following result states the uniform concentration property of Δ_n around its expectation Δ.

Lemma 4 *For any sequence of inputs $x_1, \dots, x_n \overset{iid}{\sim} P$*

$$\sup_{(h,h') \in \mathcal{H}^2} |\Delta_n(h, h') - \Delta(h, h')| \leq \varepsilon_n = 2\sqrt{2\frac{2d \log \frac{en}{d} + \log \frac{4}{\beta}}{n}},$$

with probability $1 - \beta$.

Proof: $\Delta_n(h, h')$ and $\Delta(h, h')$ are the empirical average and expectation of the random variable $\mathbb{I}\{h(x) \neq h'(x)\}$ with $x \sim P$, which is bounded in $[0, 1]$. The pair (h, h') belongs to the set $\mathcal{H}^2$ whose VC dimension is $VC(\mathcal{H}^2) \leq 2VC(\mathcal{H}) = 2d$. Therefore, we deduce the stated uniform concentration property (see e.g., [BBL04]) ■

Using the previous lemma, it is possible to bound the difference in performance between the best hypothesis in $H_{n'}$ and the best in $\mathcal{H}$, and bound the regret in (6).

Theorem 5 *For any $0 < n \leq n'$, let $H_{n'}$ be a set of hypoheses built according to an auxiliary sequence of inputs $x'_1, \dots, x'_{n'} \overset{iid}{\sim} P$. An EWF with experts in $H_{n'}$ run on n new samples drawn from distribution P incurs a regret*

$$R_n \leq c_1 \sqrt{nd \log \frac{en'}{d}} + c_2 \sqrt{\frac{n}{2} \log \frac{12}{\beta}} \tag{8}$$

with probability $1-\beta$, where $c_1 = \left(8 + \sqrt{2}\right)$, and $c_2 = 9\sqrt{2}$.

Proof: In (6) the regret is decomposed in two terms. By bounding the first term as in Theorem 1, we obtain

$$\begin{aligned} R_n &\leq \sqrt{2nd \log \frac{en'}{d}} + \sqrt{\frac{n}{2} \log \frac{1}{\beta}} \\ &+ \left(\min_{h' \in H_{n'}} \sum_{t=1}^n \ell_t(h') - \inf_{h \in \mathcal{H}} \sum_{t=1}^n \ell_t(h) \right), \end{aligned}$$

where the number of hypotheses in $H_{n'}$ is bounded by $|H_{n'}| \leq (en'/d)^d$. Since the sequence of inputs $(x'_1, \dots, x'_n)$ is drawn from the same distribution as that revealed during the learning process, the second term can be bounded as follows

$$\begin{aligned} R_H &= \left(\min_{h' \in H_{n'}} \sum_{t=1}^n \ell_t(h') - \inf_{h \in \mathcal{H}} \sum_{t=1}^n \ell_t(h) \right) \\ &= \sup_{h \in \mathcal{H}} \min_{h' \in H_{n'}} \sum_{t=1}^n \left(\ell_t(h') - \ell_t(h) \right) \\ &\leq \sup_{h \in \mathcal{H}} \min_{h' \in H_{n'}} n\Delta_n(h, h') \end{aligned} \tag{9}$$

Algorithm 1 The Epoch-based Stochastic Adversarial (EStochAd) forecaster

Input: hypothesis set $\mathcal{H}$
Initialize: $H_0 = \emptyset$ with any $h \in \mathcal{H}$
for $k = 0, 1, 2, \dots$ **do**
  Set $t_k = 2^k$, $t_{k+1} = 2^{k+1}$, $N_k = |H_k|$, and $\eta_k = \sqrt{2 \log N_k / n_k}$
  Initialize $w_i^{t_k} = 1, \;\; i \in \{1, \dots, N\}$
  for $t = t_k$ to $t_{k+1} - 1$ **do**
    Observe x_t
    Sample $h_i \sim \mathbf{p}^t$, with $p_i = w_i^t / (\sum_{j=1}^{N_k} w_j^t)$
    Predict $\widehat{y}_t = h_i(x_t)$
    Observe the true label y_t
    Update weights $w_i^{t+1} = w_i^t \exp\left(-\eta_k \ell_t(h_i)\right)$
  end for
  Build H_{k+1} according to inputs $\{x_1, \dots, x_{t_{k+1}-1}\}$
end for

$$
\begin{aligned}
&\leq \sup_{h \in \mathcal{H}} \min_{h' \in H_{n'}} n\Delta(h, h') + n\varepsilon_n && (10)\\
&\leq \sup_{h \in \mathcal{H}} \min_{h' \in H_{n'}} n\Delta_{n'}(h, h') + n\varepsilon_{n'} + n\varepsilon_n && (11)\\
&\leq 0 + n\varepsilon_{n'} + n\varepsilon_n \\
&\leq 4n\sqrt{2\frac{2d \log \frac{en'}{d} + \log \frac{4}{\beta'}}{n}} && (12)\\
&\leq 4\sqrt{4nd \log \frac{en'}{d}} + 4\sqrt{2n \log \frac{4}{\beta'}},
\end{aligned}
$$

with probability $1 - 2\beta'$. From the definition of the empirical distance in (7), we directly get (9). In fact, two hypotheses have different loss whenever their prediction is different. In both (10)-(11) we applied Lemma 4. The minimum distance $\Delta_{n'}(h, h')$ in (11) is zero for any hypothesis $h \in \mathcal{H}$. In fact, since $H_{n'}$ is built according to the same inputs $(x'_1, \dots, x'_n)$ on which $\Delta_{n'}(h, h')$ is measured, it is always possible to find a hypothesis $h' \in H_{n'}$ with exactly the same sequence of predictions as any $h \in \mathcal{H}$. Finally, (12) follows from the assumption $n' \geq n$ and from the definition of ε_n and $\varepsilon_{n'}$ in Lemma 4.

By joining the bound for R_{EWF} and R_H, and by setting $\beta = 3\beta'$ we obtain the statement of the theorem. ■

3.3 The Epoch-based Stochastic Adversarial (EStochAd) Forecaster

In the previous section we assumed a sequence of inputs $(x'_1, \dots, x'_n)$ could be sampled from P before starting the learning process. However, this assumption is often unrealistic when the distribution P is unknown and inputs are revealed only during the learning process. In this section we devise an epoch-based algorithm in which the hypothesis set is incrementally built in epochs according to the inputs experienced so far.

Let us divide the learning horizon into K epochs, such that epoch k is $n_k = t_{k+1} - t_k$ steps long, from time $t = t_k$ to $t_{k+1} - 1$. At the beginning of epoch k, a grid H_k is build on the basis of the sequence of inputs $\mathbf{x}_1^{t_k}$ and a EWF forecaster is run on H_k until the end of epoch k. The resulting algorithm is summarized in Algorithm 1. As it can be noticed, EStochAd is an anytime algorithm since the time horizon n does not need to be known in advance. In fact, the learning parameter η is set optimally at the beginning of each epoch, independently from the value of n.

According to Theorem 5, whenever $t_k \geq n_k$ the regret of an EWF with experts in H_k and parameter $\eta_k = \sqrt{2 \log N_k / n_k}$ in epoch k is

$$
\begin{aligned}
R_k &= \sum_{t=t_k}^{t_{k+1}-1} \ell_t(h_t) - \inf_{h \in \mathcal{H}} \sum_{t=t_k}^{t_{k+1}-1} \ell_t(h) \\
&\leq c_1 \sqrt{n_k d \log \frac{et_k}{d}} + c_2 \sqrt{\frac{n_k}{2} \log \frac{12}{\beta}} \qquad (13)
\end{aligned}
$$

with probability $1 - \beta$.

The next theorem shows that if the length of each epoch is set properly, then the regret of the EStochAd algorithm is bounded by $O\left(\sqrt{nd \log n}\right)$ with high probability.

Theorem 6 *For any $n > 0$, let $\mathcal{H}$ be a hypothesis space with finite VC dimension $d = VC(\mathcal{H}) < \infty$. The EStochAd algorithm described above satisfies*

$$
R_n \leq c_3 \sqrt{nd \log \frac{en}{d}} + c_4 \sqrt{n \log \frac{12(\lfloor \log_2 n \rfloor + 1)}{\alpha}} \qquad (14)
$$

with probability $1 - \alpha$, where $c_3 = 18 + 10\sqrt{2}$, and $c_4 = 18(\sqrt{2} + 1)$.

Proof: The theorem directly follows from Theorem 5 and from the definition of epochs. Given $t_k = n_k = 2^k$ the regret for each epoch can be rewritten as

$$
R_k \leq c_1 \sqrt{2^k d \log \frac{e2^k}{d}} + c_2 \sqrt{\frac{2^k}{2} \log \frac{12}{\beta}}
$$

Let $K = \lfloor \log_2 n \rfloor + 1$ be the index of the epoch containing the step n and $t_K = \min(2^K, n + 1)$. The total regret over all the K epochs can be bounded as follows

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \ell_t(h) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell_t(h) \\
&= \sum_{k=0}^{K-1} \sum_{t=t_k}^{t_{k+1}-1} \ell_t(h) - \inf_{h \in \mathcal{H}} \sum_{k=0}^{K-1} \sum_{t=t_k}^{t_{k+1}-1} \ell_t(h) \\
&\leq \sum_{k=0}^{K-1} \left(\sum_{t=t_k}^{t_{k+1}-1} \ell_t(h) - \inf_{h \in \mathcal{H}} \sum_{t=t_k}^{t_{k+1}-1} \ell_t(h) \right) \qquad (15)\\
&\leq \sum_{k=0}^{K-1} R_k = \sum_{k=0}^{\lfloor \log_2 n \rfloor} R_k \\
&\leq \left(c_1 \sqrt{d \log \frac{en}{d}} + c_2 \sqrt{\frac{1}{2} \log \frac{12}{\beta}} \right) \sum_{k=0}^{\lfloor \log_2 n \rfloor} \sqrt{2^k} \qquad (16)\\
&\leq \left(c_1 \sqrt{d \log \frac{en}{d}} + c_2 \sqrt{\frac{1}{2} \log \frac{12}{\beta}} \right) \frac{\sqrt{2n} - 1}{\sqrt{2} - 1} \\
&\leq c_3 \sqrt{nd \log \frac{en}{d}} + c_4 \sqrt{n \log \frac{12}{\beta}}.
\end{aligned}
$$

with probability $1 - \beta \log_2 n$. In (15) the regret is upper-bounded by considering the best hypothesis in each epoch rather than on the whole horizon of n steps. The inner term in the summation in (16) is the regret for epoch k and is bounded as in (13). Finally, by using a union bound and setting $\alpha = \beta(\lfloor \log_2 n \rfloor + 1)$ the result is obtained from the definition of the length of each epoch and some algebra. ■

It is worth noting that from a computational point of view the set of hypotheses H_k does not need to be regenerated from scratch at the beginning of each epoch k but it can be built incrementally as new inputs comes in. As a consequence, for each hypothesis h_i already available at the previous epoch, its weight w_i is initialized according to the cumulative loss up to time t. Similarly, new hypotheses can inherit the weight of hypotheses belonging to the same class before the refinement. Although no improvement in the bound can be proved, using the past performance to initialize the weight for new hypotheses is likely to have a positive impact in the performance.

4 Extensions

In this section, we discuss possible extensions of the proposed algorithm to different settings.

4.1 Multi-Label Classification

Although we analyzed the performance of EStochAd in the case of binary classification, the extension to the case of multi-label classification is straightforward. In order to measure the complexity of $\mathcal{H}$ we refer to the extension to multi-label classification of the VC dimension proposed by Natarajan in [Nat89] [2]. Let m be the total number of labels and $d = Ndim(\mathcal{H})$ be the Natarajan dimension of the hypothesis space. The number of hypotheses in H_n is now bounded by $|H_n| \leq (\frac{enm^2}{2d})^d$. In Lemma 4, instead of the VC dimension $Ndim(\mathcal{H})$ may be employed.

Furthermore, the equality in step (9) of the proof of Theorem 5 becomes a inequality since two hypotheses may have the same loss even when their prediction is different (in case of wrong prediction). The rest of the proofs remain unchanged and the next bound on the regret for EStochAd follows.

Theorem 7 *For any $n > 0$, let $m > 0$ be number of labels and $\mathcal{H}$ a hypothesis with finite Natarajan dimension $d = Ndim(\mathcal{H}) < infty$. The EStochAd algorithm satisfies*

$$R_n \leq c_3 \sqrt{nd \log \frac{enm^2}{2d}} + c_4 \sqrt{n \log \frac{3 \log_2 n}{\alpha}}, \qquad (17)$$

with probability $1 - \alpha$, with universal constants c_3, c_4.

4.2 Bandit Information

In the protocol in Figure 2 at the end of each episode the true label chosen by the adversary is explicitly revealed to the learner, thus defining a *full* information classification problem. However, in many applications (e.g., web advertisement systems) only the loss corresponding to the chosen hypothesis (*bandit* feedback) is available to the learner.

[2] For more details about complexity measures for m-values functions, refer to [BDCBHL95].

Algorithm 2 The Bandit-EStochAd forecaster

Input: hypothesis set $\mathcal{H}$
Initialize: EStochAd ($\mathcal{H}$)
for $t = 1, 2, \ldots$ **do**
  Observe x_t
  Sample h_i according to EStochAd (Algorithm 1)
  Sample $\widehat{y}_t \sim \mathbf{q}^t$, where

$$q_j^t = (1-\gamma)\mathbb{I}\{j = h_i(x_t)\} + \frac{\gamma}{m}, \;\; j \in \{1, \ldots, m\}$$

  Receive loss $\ell(\widehat{y}_t)$
  Define $\widetilde{\ell}_t(h_i) = \frac{\ell(\widehat{y}_t)}{q_{\widehat{y}_t}^t} \mathbb{I}\{h_i(x_t) = \widehat{y}_t\}$
  Update EStochAd weights with loss $\widetilde{\ell}_t(h_i)$
end for

The EStochAd algorithm can be extended to solve the hybrid stochastic-adversarial classification problem with bandit information simply by substituting the EWF with a bandit algorithm such as Exp4 [ACBFS03]. Let us consider the more general case illustrated in Figure 1 in which instead of selecting a label, at each round t the adversary chooses a bounded loss function $\ell : \mathcal{Y} \to [0, 1]$. At the end of each round, the learner incurs a loss $\ell_t(h(x_t))$ which is the only information revealed to the learner. In Theorem 5 the first part of the regret of EStochAd can be immediately derived from the bandit algorithm working on the set H_n. For instance, for Exp4 with N experts and m labels it is possible to prove the high-probability regret bound

$$R_n(\text{Exp4}) \leq 4 \sqrt{nm \log \frac{nN}{\beta}} + 8 \log \frac{nN}{\beta},$$

with probability $1 - \beta$. As discussed in the previous section, in case of m labels the number of experts at time n is bounded by $N = |H_n| \leq (\frac{enm^2}{2d})^d$. Besides, the second term in (6) is not affected by the different feedback in full and bandit settings and remains unchanged. The only difference is that the equality in step (9) of Theorem 5 becomes an inequality. Indeed, when two hypotheses have the same prediction their loss is the same. On the other hand, if the predictions are different, the difference between the losses cannot be greater than 1. Thus, $\ell_t(h) - \ell_t(h') \leq \mathbb{I}\{h(x_t) \neq h'(x_t)\}$. As a result, the leading term in the cumulative regret is due to Exp4 and we can prove the following regret bound for Bandit-EStochAd (Algorithm 2).

Theorem 8 *For any $n > 0$, let $m > 0$ be the number of arms (i.e., labels) and $\mathcal{H}$ a hypothesis set with finite Natarajan dimension $d = Ndim(\mathcal{H}) < infty$. The Bandit-EStochAd algorithm satisfies*

$$R_n \leq O\left(\sqrt{nmd \log \frac{nm^2}{\alpha}} + d \log \frac{nm^2}{\alpha} \right), \qquad (18)$$

with probability $1 - \alpha$.

1: **for** $t = 1, 2, \ldots$ **do**
2: Simultaneously,
- A stochastic input x_t is sampled from P
- Player A selects strategy $h_{A,t}$
- Player B selects strategy $h_{B,t}$
3: Player A (resp., B) plays action $\widehat{y}_{A,t} = h_{A,t}(x_t)$ (resp., $\widehat{y}_{B,t} = h_{B,t}(x_t)$)
4: Return feedback
- $\ell_A(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$ and $\ell_B(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$ (*bandit* information)
- *or* $\ell_A(\cdot, \widehat{y}_{B,t}, x_t)$ and $\ell_B(\widehat{y}_{A,t}, \cdot, x_t)$ (*full* information)
5: Player A (resp., B) incurs a loss $\ell_A(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$ (resp., $\ell_B(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$)
6: **end for**

Figure 4: The two-player strategic repeated game with stochastic side information.

4.3 Application in Games

In this section, we consider an extension of the stochastic-adversarial prediction problem to a two-player strategic repeated game with stochastic side information. Like in the general problem illustrated in Figure 1, the game could be either *full* or *bandit* information, depending on whether at the end of each round the learners receive the loss function $\ell_A(\cdot, \widehat{y}_{B,t}, x_t)$ (resp. $\ell_B(\widehat{y}_{A,t}, \cdot, x_t)$) or only the loss they incurred. Our main contribution here is to show that in the case of a zero-sum game, if both players play according to the (Bandit-)EStochAd algorithm, then the empirical frequencies of the strategies converge to the set of Nash equilibria.

For sake of simplicity we consider the same set of strategies for both players. Let A and B be two players and $\mathcal{H}$ be the set of strategies h mapping an input $x \in \mathcal{X}$ to an action in $\mathcal{Y} = \{1, \ldots, m\}$. The repeated game between player A and B is sketched in Figure 4. At each round t, an input x_t is drawn from P and, simultaneously, the players select strategies $h_{A,t} \in \mathcal{H}$ and $h_{B,t} \in \mathcal{H}$. As a result, they incur losses $\ell_A(h_{A,t}(x_t), h_{B,t}(x_t), x_t)$ and $\ell_B(h_{A,t}(x_t), h_{B,t}(x_t), x_t)$ respectively ($\ell_{A,t}(h_{A,t})$ and $\ell_{B,t}(h_{B,t})$ for short in the following). We define the expected loss for player A with respect to the input distribution P as

$$\overline{\ell}_A(h_A, h_B) = \mathbb{E}_{x \sim P}\left[\ell_A(h_A(x), h_B(x), x)\right].$$

Let $\mathcal{D}(\mathcal{H})$ be the set of distributions on the set of pure strategies $\mathcal{H}$. Given mixed strategies σ_A and σ_B in $\mathcal{D}(\mathcal{H})$ we define its corresponding expected loss (similarly for player B):

$$\overline{\ell}_A(\sigma_A, \sigma_B) = \mathbb{E}_{h_A \sim \sigma_A, h_B \sim \sigma_B}\left[\overline{\ell}_A(h_A, h_B)\right].$$

We say that a pair of strategies (σ_A^*, σ_B^*) is a Nash equilibrium if

$$\begin{aligned} \overline{\ell}_A(\sigma_A^*, \sigma_B^*) &\leq \overline{\ell}_A(\sigma_A, \sigma_B^*), \ \forall \sigma_A \in \mathcal{D}(\mathcal{H}) \\ \overline{\ell}_B(\sigma_A^*, \sigma_B^*) &\leq \overline{\ell}_B(\sigma_A^*, \sigma_B), \ \forall \sigma_B \in \mathcal{D}(\mathcal{H}). \end{aligned}$$

Now we consider the problem of approximating a Nash equilibrium in the zero-sum case (i.e., $\overline{\ell}_A(\cdot, \cdot) = -\overline{\ell}_B(\cdot, \cdot)$). In order to define the value of the game and apply the minimax theorem we need $\mathcal{D}(\mathcal{H})$ to be compact [CBL06]. In the following, we assume $\mathcal{H}$ to be a compact metric space, a sufficient condition for $\mathcal{D}(\mathcal{H})$ to be compact (see e.g., [SL07]). Under this assumption, the minimax theorem [CBL06] holds

$$\begin{aligned} V &= \sup_{\sigma_B \in \mathcal{D}(\mathcal{H})} \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B) \\ &= \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \sup_{\sigma_B \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B), \end{aligned} \tag{19}$$

where V is the value of the game. The following theorem proves that if both players run either EStochAd or Bandit-EStochAd (in full information and bandit information respectively), then their performance converges to the value of the game and the empirical frequencies of their strategies converge to the set of Nash equilibria.

Theorem 9 *Let losses ℓ_A, ℓ_B be bounded in $[0, 1]$, $\mathcal{H}$ be a compact metric set. If both players run (Bandit-)EStochAd in a zero-sum game with stochastic side information as defined above, then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \ell_A(h_{A,t}(x_t), h_{B,t}(x_t), x_t) = V \tag{20}$$

almost surely.

Proof: The proof is similar to the convergence proof for Hannan consistent strategies in zero-sum games [CBL06]. We first prove the following

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \ell_{A,t}(h_{A,t}) \leq V. \tag{21}$$

We note that the regret for both players can be bounded exactly as in (18). In fact, losses ℓ_A and ℓ_B are a special case of the adversarial loss function considered in Section 4.2. As a result, we have

$$\limsup_{n \to \infty} \left[\frac{1}{n} \sum_{t=1}^{n} \ell_{A,t}(h_{A,t}) - \inf_{h_A \in \mathcal{H}} \frac{1}{n} \sum_{t=1}^{n} \ell_{A,t}(h_A)\right] \leq 0, \tag{22}$$

with probability $1 - \alpha$, where $\ell_{A,t}(h_A) = \ell_A(h_A, h_{B,t}, x_t)$. Let $Z_t(h_A) = \ell_{A,t}(h_A) - \overline{\ell}_A(h_A, h_{B,t})$. By definition of the expected loss and by noticing that the hypothesis $h_{B,t}$ selected by the algorithm at time t does not depend on the input x_t, we have that Z_t is a martingale

$$\mathbb{E}_{x_t \sim P}\left[Z_t(h_A) | \mathcal{F}_{t-1}\right] = 0,$$

where $\mathcal{F}_{t-1}$ is the σ-algebra generated by all random variables up to time $t-1$ (i.e., past inputs and hypotheses for both players A and B). Thus, $Z_1, \cdots, Z_n$ is a martingale difference sequence and we may apply Hoeffding Azuma's inequality (see e.g., [DGL97]) and obtain

$$\frac{1}{n} \sum_{t=1}^{n} Z_t(h_A) \leq \sqrt{\frac{2}{n} \log \beta^{-1}},$$

with probability $1 - \beta$ for a fixed hypothesis h_A. Since the VC dimension of $\mathcal{H}$ is finite, a functional concentration inequality for martingales implies that $1/n \sum_{t=1}^{n} Z_t(h_A)$ concentrates around 0 uniformly for $h_A \in \mathcal{H}$. As a result, we have

$$\limsup_{n \to \infty} \left[\inf_{h_A \in \mathcal{H}} \frac{1}{n} \sum_{t=1}^{n} \ell_{A,t}(h_A) - \inf_{h_A \in \mathcal{H}} \frac{1}{n} \sum_{t=1}^{n} \overline{\ell}_A(h_A, h_{B,t})\right] \leq 0. \tag{23}$$

Now, since the mapping $\sigma_A \mapsto \overline{\ell}_A(\sigma_A, h_{B,t})$ is linear, this function admits a pure strategy as minimum, and we have

$$\inf_{h_A \in \mathcal{H}} \frac{1}{n} \sum_{t=1}^{n} \overline{\ell}_A(h_A, h_{B,t}) = \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \frac{1}{n} \sum_{t=1}^{n} \overline{\ell}_A(\sigma_A, h_{B,t}) = \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B^n)$$

where $\sigma_B^n(h) \in \mathcal{D}(\mathcal{H})$ is defined for any $h \in \mathcal{H}$ as $\sigma_B^n(h) = 1/n \sum_{t=1}^{n} \mathbb{I}\{h_{B,t} = h\}$. Finally, we have

$$\inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B^n) \leq \sup_{\sigma_B \in \mathcal{D}(\mathcal{H})} \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B), \quad (24)$$

Putting together (22), (23), and (24) we obtain (21). The same result can be obtained for ℓ_B. From the assumption $\overline{\ell}_A(\cdot,\cdot) = -\overline{\ell}_B(\cdot,\cdot)$, minimax theorem (19), and since this result holds for any α, then we have (20) with probability 1. ■

From the previous theorem and the compactness property of $\mathcal{D}(\mathcal{H})$ it also follows that the empirical frequencies of the mixed strategies σ_A^n and σ_B^n converge to the set of Nash strategies. Finally, it is interesting to notice that the convergence rate to the set of Nash equilibria is of the order $O(\sqrt{d/n \log(nm^2)})$ in the full information case, and $O(\sqrt{(md)/n \log(nm^2)})$ in the bandit information case.

5 Related Works

To the best of our knowledge this is the first work considering the hybrid stochastic-adversarial online learning problem. A similar setting is analyzed in [Rya06] for batch supervised learning where the sequence of labels is adversarial and inputs are *conditionally* independent and identically distributed (i.e., inputs are drawn from distributions conditioned to labels). In particular, they show that in such a scenario many learning bounds (derived in the pure stochastic setting) remain unchanged. The main difference with the setting illustrated in this paper is that we considered the problem of online learning instead of batch learning and inputs are i.i.d. and not conditioned to labels.

The possibility to convert batch algorithms for the fully stochastic into learning algorithm for the transductive online learning scenario is studied in [KK05]. In transductive online learning the samples are adversarialy generated and all the inputs are known to the learner beforehand. In this scenario, they prove that a batch algorithm can be efficiently translated into an online algorithm with a mistake bound of the order $n^{3/4}\sqrt{d \log n}$ with d the VC dimension of the hypothesis set. The transductive setting is very similar to the preliminary scenario we described in Section 3.2 in which we assume the sequence of inputs to be known in advance to the learner. In the rest of the paper we showed that in order to move from a transductive setting to a fully online problem and preserve similar results we need to assume the inputs to be drawn independently from some fixed distribution even if this distribution is not known.

A direct comparison with other algorithms for either fully adversarial or fully stochastic settings is difficult because of the different assumptions. Nonetheless, in the following we discuss similarities and differences between EStochAd and other existing algorithms for online prediction. In Table 5, we summarize the main approaches to the classification problem in both stochastic and adversarial settings. Unfortunately not all the bounds are immediately comparable. Some of the regret bounds are in expectation (with respect to either the distribution P or the randomized algorithm $\mathcal{A}$), while others are high-probability bounds. Perceptron performance is stated in terms of mistake bound.

It is interesting to notice that EStochAd incurs exactly the same regret rate as an empirical risk minimization algorithm run online in the fully stochastic case. [3] This means that under the assumption that inputs are i.i.d. from a fixed distribution P, the adversarial output does not cause any worsening in the performance with respect to a stochastic output. This result can be explained by the definition of the VC dimension itself. In fact, while VC definition requires samples to be generated from a distribution, no assumption is made on the way outputs are generated and any possible sequences of labels is considered. Therefore, it is not surprising that VC can be used as a complexity measure for both the case of stochastic and adversarial classification. However, the situation is significantly different in the case of a fully adversarial setting where also inputs can be arbitrarily chosen by an adversary.

Both EStochAd and the Agnostic Online Learning (AOL) algorithm proposed in [SS08] consider the problem of binary classification with adversarial outputs, an infinite number of hypotheses (experts), and they both build on the exponentially weighted forecaster [CBL06]. On the other hand, the main difference is that while with adversarial inputs it is necessary to consider the Littlestone dimension of $\mathcal{H}$ [Lit88], the stochastic assumption on the inputs allows EStochAd to refer to the VC dimension which is a more natural measure of complexity of the hypothesis space. Moreover, the dependency of the two algorithms on the hypothesis space complexity is different (see Table 5). While AOL has a linear dependency on $Ldim(\mathcal{H})$, in EStochAd the regret grows as $\sqrt{VC(\mathcal{H})}$. Furthermore, as proved in [Lit88], for any hypothesis space $\mathcal{H}$, $VC(\mathcal{H}) \leq Ldim(\mathcal{H})$. In the following we discuss an example showing how in some cases the difference between VC and Littlestone dimension may be arbitrarily large. Let consider a binary classification problem with $X = [0, 1]$ and a hypothesis space $\mathcal{H}$ containing functions of the form

$$h_\vartheta(x) = \begin{cases} 1 & \text{if } x \geq \vartheta \\ 0 & \text{otherwise,} \end{cases}$$

with $\vartheta \in [0, 1]$.

In the fully adversarial case the regret of AOL is linear in the time horizon (i.e., in the worst case it can make a mistake at each time step). In fact, it can be shown that the Littlestone dimension of $\mathcal{H}$ is infinite. According to [Lit88], the Littlestone dimension is the largest number of mistakes any learning algorithm could incur for any possible sequence of predictions in the realizable case when the adversary is allowed to choose the true label after observing the learner's prediction. Thus, the adversary selects inputs and labels so

[3] The online bound is obtained by summing on n steps the usual offline VC bounds [BBL04].

Algorithm	Setting	Hyp. space	Bound	Performance
Empirical Risk Minimization [BBL04]	S/S	$VC(\mathcal{H}) < \infty$	HP-Regret	$\mathbb{E}_{(x,y)\sim P}[R_n] \leq \sqrt{nVC(\mathcal{H})\log n} + \sqrt{n\log\beta^{-1}}$
Exp. Weighted Forecaster [CBL06]	A/A	$\lvert\mathcal{H}\rvert = N < \infty$	HP-Regret	$R_n \leq \sqrt{n\log N} + \sqrt{n\log\beta^{-1}}$
Perceptron [Ros58]	A/A	Linear	Mistake	$M_n \leq L + D + \sqrt{LD}$
Agnostic Online Learning [SS08]	A/A	$Ldim(\mathcal{H}) < \infty$	Exp-Regret	$\mathbb{E}_{\mathcal{A}}[R_n] \leq Ldim(\mathcal{H}) + \sqrt{nLdim(\mathcal{H})\log n}$
Transductive Online Learning [KK05]	Transd.	$VC(\mathcal{H}) < \infty$	Mistake	$M_n \leq L + n^{3/4}\sqrt{VC(\mathcal{H})\log n}$
EStochAd [*This paper*]	S/A	$VC(\mathcal{H}) < \infty$	HP-Regret	$R_n \leq \sqrt{nVC(\mathcal{H})\log n} + \sqrt{n\log\beta^{-1}}$

Table 1: Performance of algorithms for different classification scenarios. All the bounds are reported up to constant factors. In the setting column, the two letters specify how inputs and labels are generated, where *A* stands for *adversarial* and *S* for *stochastic*. In the bound column *HP* stands for high-probability bound and *Exp* stands for bound in expectation. In the parceptron bound M_n is the number of mistakes after n steps, L and D are the cumulative loss and the complexity of any weight matrix.

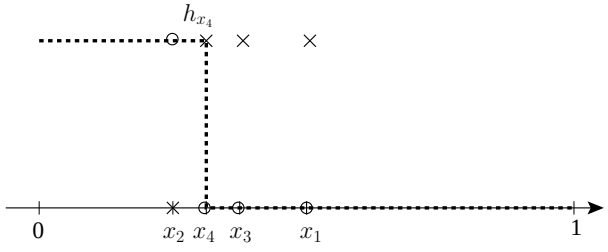

Figure 5: Example of a sequence of inputs and labels such that the adversary can force any learning algorithm to incur a mistake at each round. Circles represent the labels predicted by the learner and crosses the labels revealed by the adversary. h_{x_4} (in dotted-line) is an example of a hypothesis which perfectly classifies all the samples shown so far.

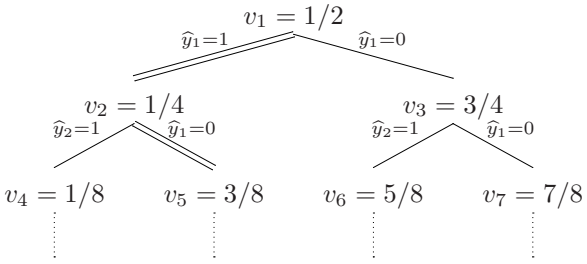

Figure 6: The mistake-tree is defined for any possible sequence of predictions. Double lines correspond to the example depicted in Figure 5.

as to force the learner to make as many mistakes as possible given the condition that there exists a hypothesis h^* in $\mathcal{H}$ such that $h^*(x_t) = y_t,\ \forall t \leq n$. In order to determine the Littlestone dimension of $\mathcal{H}$ we sketch how to build a shattered mistake-tree of depth n, for any $n > 0$ (see Figure 6). Nodes of the mistake-tree represent the inputs revealed by the adversary depending on the sequence of learner's predictions. Let $v_1 = \frac{1}{2}$ be the root of the mistake-tree, that is the first input x_1 revealed to the algorithm. Next, we label nodes v_2 and v_3 as the middle points of intervals $[0, v_1]$ and $[v_1, 1]$ respectively. The second input shown to the learner depends on the prediction at time $t = 1$. If the prediction is $\widehat{y}_1 = 1$ [4], then the adversary selects a label $y_1 = 0$ and the next input point is set to $x_2 = v_2$. If the algorithm predicts $\widehat{y}_2 = 0$ in x_2, it is still possible to force the algorithm to incur a mistake by setting $y_2 = 1$ without violating the realizability condition. In fact, any hypothesis with $x_2 \leq \vartheta < x_1$ perfectly classifies both y_1 and y_2. The next input x_3 is the middle point of interval $[x_2, x_1]$ and the algorithm is forced to make another mistake. The same process can be repeated at each round by choosing the next input to be the middle point of either the left or the right interval depending on the previous prediction and by revealing a label which is exactly the opposite of the one predicted by the learner. At each step the adversary can force the learner to make a mistake while guaranteeing that it is always possible to find a hypothesis in $\mathcal{H}$ that would make no mistakes (see Figure 5 for the sequence of inputs x_1, x_2, x_3, x_4). As a result, $Ldim(\mathcal{H}) = \infty$ and the AOL has a linear regret. On the other hand, when inputs cannot be arbitrarily chosen by an adversary but are sampled from a fixed distribution EStochAd can achieve a sub-linear regret. In fact, $\mathcal{H}$ could shatter at most one points, the VC dimension of $\mathcal{H}$ is 1, thus leading a regret for EStochAd of order $O(\sqrt{n\log n})$.

Therefore, even in very simple problems the possibility for the adversary to select the inputs may lead to an arbitrarily bad performance, while drawing inputs from a distribution allows the learner to achieve a sub-linear regret even if outputs are adversarial.

[4] The case $\widehat{y}_1 = 0$ is symmetric.

6 Conclusions

In this paper we introduced the hybrid stochastic-adversarial online prediction problem in which inputs are independently and identically generated and labels are arbitrarily chosen by an adversary. We devised an epoch-based algorithm for the specific problem of binary classification with full information and analyzed its regret. In particular, we noticed that while the stochastic assumption on inputs allows to use the well-known VC dimension as a measure of complexity for the hypothesis space, adversarial labels do not cause any worsening in the performance with respect to fully stochas-

tic algorithms. We believe that this analysis, together with its relationship with the results for the fully adversarial case, sheds light on the similarities and differences between batch stochastic learning and adversarial online learning along the line of [KK05]. Finally, we discussed extensions to multi-label classification, learning from experts and bandits settings with stochastic side information, and approximation of Nash equilibria in games.

References

[ACBFS03] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The non-stochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2003.

[BBL04] O. Bousquet, S. Boucheron, and G. Lugosi. Introduction to statistical learning theory. *Advanced Lectures on Machine Learning Lecture Notes in Artificial Intelligence*, 3176:169–207, 2004.

[BDCBHL95] S. Ben-David, N. Cesa-Bianchi, D. Haussler, and P. M. Long. Characterizations of learnability for classes of $\{0...n\}$-valued functions. *Journal of Computer and System Sciences*, 50:74–86, 1995.

[CBFH+97] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. Shapire, and M. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[CBL06] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[CS03] Koby Crammer and Yoram Singer. Ultraconservative online algorithms for multiclass problems. *J. Mach. Learn. Res.*, 3:951–991, 2003.

[DGL97] Luc Devroye, Laszlo Györfi, and Gabor Lugosi. *A Probabilistic Theory of Pattern Recognition (Stochastic Modelling and Applied Probability)*. Springer, February 1997.

[FSSSU06] Michael Fink, Shai Shalev-Shwartz, Yoram Singer, and Shimon Ullman. Online multiclass learning by interclass hypothesis sharing. In *Proceedings of the 23rd international conference on Machine learning*, pages 313–320, New York, NY, USA, 2006. ACM.

[KK05] Sham M. Kakade and Adam Kalai. From batch to transductive online learning. In *NIPS*, 2005.

[KSST08] Sham M. Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. Efficient bandit algorithms for online multiclass prediction. In *Proceedings of the 25th international conference on Machine learning*, pages 440–447, New York, NY, USA, 2008. ACM.

[Lit88] Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Mach. Learn.*, 2(4):285–318, 1988.

[LW94] N. Littlestone and M. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[LZ07] J. Langford and T. Zhang. The epoch greedy algorithm for contextual multi-armed bandits. In *Advances in Neural Information Processing Systems*, 2007.

[Nat89] B. K . Natarajan. On larning sets and functions. *Mach. Learn.*, 4:67–97, 1989.

[Ros58] F. Rosenblatt. The perceptron : A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[Rya06] Daniil Ryabko. Pattern recognition for conditionally independent data. *J. Mach. Learn. Res.*, 7:645–664, 2006.

[SL07] G. Stoltz and G. Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59(1):187 – 208, 2007.

[SS08] Shai Shalev-Shwartz. Agnostic online learnability. Technical Report TTIC-TR-2008-2, Toyota Technological Institute, 2008.

[Vov98] V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56:153–173, 1998.

[WW99] J. Weston and C. Watkins. Support vector machines for multi-class patter recognition. In *Proceedings of the Seventh European Symposium on Artificial Neural Networks*, 1999.