Online Learning: Stochastic, Constrained, and Smoothed Adversaries

Authored by:

Ambuj Tewari
Alexander Rakhlin
Karthik Sridharan

Abstract

Learning theory has largely focused on two main learning scenarios: the classical statistical setting where instances are drawn i.i.d. from a fixed distribution, and the adversarial scenario whereby at every time step the worst instance is revealed to the player. It can be argued that in the real world neither of these assumptions is reasonable. We define the minimax value of a game where the adversary is restricted in his moves, capturing stochastic and non-stochastic assumptions on data. Building on the sequential symmetrization approach, we define a notion of distribution-dependent Rademacher complexity for the spectrum of problems ranging from i.i.d. to worst-case. The bounds let us immediately deduce variation-type bounds. We study a smoothed online learning scenario and show that exponentially small amount of noise can make function classes with infinite Littlestone dimension learnable.

1 Paper Body

In the papers [1, 10, 11], an array of tools has been developed to study the minimax value of diverse sequential problems under the worst-case assumption on Nature. In [10], many analogues of the classical notions from statistical learning theory have been developed, and these have been extended in [11] for performance measures well beyond the additive regret. The process of sequential symmetrization emerged as a key technique for dealing with complicated nested minimax expressions. In the worst-case model, the developed tools give a unified treatment to such sequential problems as regret minimization, calibration of forecasters, Blackwell?s approachability, ?-regret, and more. Learning theory has been so far focused predominantly on the i.i.d. and the worst-case learning scenarios. Much less is known about learnability in-between these two extremes. In the present paper, we make progress towards filling this gap by proposing a framework in which it is possible to variously restrict the behavior of Nature.

By restricting Nature to play i.i.d. sequences, the results boil down to the classical notions of statistical learning in the supervised learning scenario. By not placing any restrictions on Nature, we recover the worst-case results of [10]. Between these two endpoints of the spectrum, particular assumptions on the adversary yield interesting bounds on the minimax value of the associated problem. Once again, the sequential symmetrization technique arises as the main tool for dealing with the minimax value, but the proofs require more care than in the i.i.d. or completely adversarial settings. 1

Adapting the game-theoretic language, we will think of the learner and the adversary as the two players of a zero-sum repeated game. Adversary?s moves will be associated with ?data?, while the moves of the learner ? with a function or a parameter. This point of view is not new: game-theoretic minimax analysis has been at the heart of statistical decision theory for more than half a century (see [3]). In fact, there is a well-developed theory of minimax estimation when restrictions are put on either the choice of the adversary or the allowed estimators by the player. We are not aware of a similar theory for sequential problems with non-i.i.d. data. The main contribution of this paper is the development of tools for the analysis of online scenarios where the adversary?s moves are restricted in various ways. In additional to general theory, we consider several interesting scenarios which can be captured by our framework. All proofs are deferred to the appendix.

2

Value of the Game

Let F be a closed subset of a complete separable metric space, denoting the set of moves of the learner. Suppose the adversary chooses from the set X . Consider the Online Learning Model, defined as a T -round interaction between the learner and the adversary: On round t = 1, . . . , T , the learner chooses ft ? F, the adversary simultaneously picks xt ? X , and the learner suffers loss ft (xt). The goal of the learner is to minimize regret, defined as PT PT t=1 ft (xt) ? inf f ?F t=1 f (xt). It is a standard fact that simultaneity of the choices can be formalized by the first player choosing a mixed strategy; the second player then picks an action based on this mixed strategy, but not on its realization. We therefore consider randomized learners who predict a distribution qt ? Q on every round, where Q is the set of probability distributions on F, assumed to be weakly compact. The set of probability distributions on X (mixed strategies of the adversary) is denoted by P. We would like to capture the fact that sequences (x1 , . . . , xT) cannot be arbitrary. This is achieved by defining restrictions on the adversary, that is, subsets of ?allowed? distributions for each round. These restrictions limit the scope of available mixed strategies for the adversary. Definition 1. A restriction P1:T on the adversary is a sequence P1 , . . . , PT of mappings Pt : X t?1 7? 2P such that Pt (x1:t?1) is a convex subset of P for any x1:t?1 ? X t?1 . Note that the restrictions depend on the past moves of the adversary, but not on those of the player. We will write Pt instead of Pt (x1:t?1) when x1:t?1 is clearly defined. Using the notion of restrictions, we can give names to several types of adversaries that we will study in this paper. (1) A worst-case adversary is defined by vacuous restrictions Pt (x1:t?1) = P. That

is, any mixed strategy is available to the adversary, including any deterministic point distribution. (2) A constrained adversary is defined by Pt (x1:xt?1) being the set of all distributions supported on the set {x ? X : Ct (x1 , . . . , xt?1 , x) = 1} for some deterministic binary-valued constraint Ct . The deterministic constraint can, for instance, ensure that the length of the path determined by the moves x1 , . . . , xt stays below the allowed budget. (3) A smoothed adversary picks the worst-case sequence which gets corrupted by i.i.d. noise. Equivalently, we can view this as restrictions on the adversary who chooses the ?center? (or a parameter) of the noise distribution. Using techniques developed in this paper, we can also study the following adversaries (omitted due to lack of space): (4) A hybrid adversary in the supervised learning game picks the worst-case label yt , but is forced to draw the xt -variable from a fixed distribution [6]. (5) An i.i.d. adversary is defined by a time-invariant restriction Pt (x1:t?1) = {p} for every t and some p ? P. For the given restrictions P1:T , we define the value of the game as ?

VT (P1:T) =
inf
sup
E
q1 ?Q p1 ?P1 f1 ,x1
inf
sup
E ? ? ? inf
q2 ?Q p2 ?P2 f2 ,x2
sup
E
qT ?Q p ?P fT ,xT T T
”
T X t=1
ft (xt) ? inf
f ?F
T X t=1
#
f (xt)
(1)

where ft has distribution qt and xt has distribution pt . As in [10], the adversary is adaptive, that is, chooses pt based on the history of moves f1:t?1 and x1:t?1 . At this point, the only difference from 2

the setup of [10] is in the restrictions Pt on the adversary. Because these restrictions might not allow point distributions, suprema over pt ?s in (1) cannot be equivalently written as the suprema over xt ?s. A word about the notation. In [10], the value of the game is written as VT (F), signifying that the main object of study is F. In [11], it is written as VT (?, ?T) since the focus is on the complexity of the set of transformations ?T and the payoff mapping ?. In the present paper, the main focus is indeed on the restrictions on the adversary, justifying our choice VT (P1:T) for the notation. The first step is

to apply the minimax theorem. To this end, we verify the necessary conditions. Our assumption that F is a closed subset of a complete separable metric space implies that Q is tight and Prokhorov?s theorem states that compactness of Q under weak topology is equivalent to tightness [14]. Compactness under weak topology allows us to proceed as in [10]. Additionally, we require that the restriction sets are compact and convex. Theorem 1. Let F and X be the sets of moves for the two players, satisfying the necessary conditions for the minimax theorem to hold. Let P1:T be the restrictions, and assume that for any x1:t?1 , Pt (x1:t?1) satisfies the necessary conditions for the minimax theorem to hold. Then " T # T X X inf Ext ?pt [ft (xt)] ? inf f (xt) . (2) VT (P1:T) = sup Ex1 ?p1 . . . sup ExT ?pT p1 ?P1

pT ?PT

t=1

ft ?F

f ?F

t=1

The nested sequence of suprema and expected values in Theorem 1 can be re-written succinctly as VT (P1:T) = sup Ex1 ?p1 Ex2 ?p2 (?—x1) . . . ExT ?pT (?—x1:T ?1) p?P

= sup E p?P

"

T X t=1

inf Ext ?pt [ft (xt)] ? inf

ft ?F

f ?F

T X

"

T X t=1

#

f (xt)

t=1

inf Ext ?pt [ft (xt)] ? inf

ft ?F

f ?F

T X t=1

#

f (xt)

(3)

where the supremum is over all joint distributions p over sequences, such that p satisfies the restrictions as described below. Given a joint distribution p on sequences (x1 , . . . , xT) ? X T , we denote the associated conditional distributions by pt (?—x1:t?1). We can think of the choice p as a sequence of oblivious strategies {pt : X t?1 7? P}Tt=1 , mapping the prefix x1:t?1 to a conditional distribution pt (?—x1:t?1) ? Pt (x1:t?1). We will indeed call p a ?joint distribution? or an ?oblivious strategy? interchangeably. We say that a joint distribution p satisfies restrictions if for any t and any x1:t?1 ? X t?1 ,

pt (?—x1:t?1) ? Pt (x1:t?1). The set of all joint distributions satisfying the restrictions is denoted by P. We note that Theorem 1 cannot be deduced immediately from the analogous result in [10], as it is not clear how the restrictions on the adversary per each round come into play after applying the minimax theorem. Nevertheless, it is comforting that the restrictions directly translate into the set P of oblivious strategies satisfying the restrictions. Before continuing with our goal of upper-bounding the value of the game, we state the following interesting facts. Proposition 2. There is an oblivious minimax optimal strategy for the adversary, and there is a corresponding minimax optimal strategy for the player that does not depend on its own moves. The latter statement of the proposition is folklore for worst-case learning, yet we have not seen a proof of it in the literature. The proposition holds for all online learning settings with legal restrictions P1:T , encompassing also the no-restrictions setting of worst-case online learning [10]. The result crucially relies on the fact that the objective is external regret.

3

Symmetrization and Random Averages

Theorem 1 is a useful representation of the value of the game. As the next step, we upper bound it with an expression which is easier to study. Such an expression is obtained by introducing Rademacher random variables. This process can be termed sequential symmetrization and has been exploited in [1, 10, 11]. The restrictions Pt , however, make sequential symmetrization considerably more involved than in the papers cited above. The main difficulty arises from the fact that the set Pt (x1:t?1) depends on the sequence x1:t?1 , and symmetrization (that is, replacement of xs with x?s) has to be done with care as it affects this dependence. Roughly speaking, in the process of symmetrization, a tangent sequence x?1 , x?2 , . . . is introduced such that xt and x?t are independent and 3

identically distributed given ?the past?. However, ?the past? is itself an interleaving choice of the original sequence and the tangent sequence. Define the ?selector function? ? : X ?X ?{?1} 7? X by ?(x, x? , ?) = x? if ? = 1 and ?(x, x? , ?) = x if ? = ?1. When xt and x?t are understood from the context, we will use the shorthand ?t (?) := ?(xt , x?t , ?). In other words, ?t selects between xt and x?t depending on the sign of ?. Throughout the paper, we deal with binary trees, which arise from symmetrization [10]. Given some set Z, an Z-valued tree of depth T is a sequence z = (z1 , . . . , zT) of T mappings zi : {?1}i?1 7? Z. The T -tuple ? = (?1 , . . . , ?T) ? {?1}T defines a path. For brevity, we write zt (?) instead of zt (?1:t?1). T ?1

Given a joint distribution p, consider the ?(X ? X) ? = (?1 , . . . , ?T) defined by

7? P(X ? X)?- valued probability tree

$

?t (?1:t?1) (x1 , x?1), . . . , (xT ?1 , x?T ?1) = (pt (?—?1 (?1), . . . , ?t?1 (?t?1)), pt (?—?1 (?1), . . . , ?t?1 (?t?1))).

In other words, the values of the mappings ?t (?) are products of conditional distributions, where conditioning is done with respect to a sequence

made from xs and x?s depending on the sign of ?s . We note that the difficulty in intermixing the x and x? sequences does not arise in i.i.d. or worstcase symmetrization. However, in-between these extremes the notational complexity seems to be unavoidable if we are to employ symmetrization and obtain a version of Rademacher complexity. As an example, consider the ?left-most? path ? = ?1 in a binary tree of depth T , where 1 = (1, . . . , 1) is a T -dimensional vector of ones. Then all the selectors ?(xt , x?t , ?t) choose the sequence x1 , . . . , xT . The probability tree ? on the ?left-most? path is, therefore, defined by the conditional distributions pt (?—x1:t?1); on the path ? = 1, the conditional distributions are pt (?—x?1:t?1). $

Slightly abusing the notation, we will write ?t (?) (x1 , x?1), . . . , (xt?1 , x?t?1) for the probability tree since ?t clearly depends only on the prefix up to time t ? 1. Throughout the paper, it will be understood that the tree ? is obtained from p as described above. Since all the conditional distributions of p satisfy the restrictions, so do the corresponding distributions of the probability tree ?. By saying that ? satisfies restrictions we then mean that p ? P.

Sampling of a pair of X -valued trees from ?, written as (x, x?) ? ?, is defined as the following recursive process: for any ? ? {?1}T , (x1 (?), x?1 (?)) ? ?1 (?) and (xt (?), x?t (?)) ? ?t (?)((x1 (?), x?1 (?)), . . . , (xt?1 (?), x?t?1 (?)))

for 2 ? t ? T

(4)

To gain a better understanding of the sampling process, consider the first few levels of the tree. The roots x1 , x?1 of the trees x, x? are sampled from p1 , the conditional distribution for t = 1 given by p. Next, say, ?1 = +1. Then the ?right? children of x1 and x?1 are sampled via x2 (+1), x?2 (+1) ? p2 (?—x?1) since ?1 (+1) selects x?1 . On the other hand, the ?left? children x2 (?1), x?2 (?1) are both distributed according to p2 (?—x1). Now, suppose ?1 = +1 and ?2 = ?1. Then, x3 (+1, ?1), x?3 (+1, ?1) are both sampled from p3 (?—x?1 , x2 (+1)). The proof of Theorem 3 reveals why such intricate conditional structure arises, and Proposition 5 below shows that this structure greatly simplifies for i.i.d. and worst-case situations. Nevertheless, the process described above allows us to define a unified notion of Rademacher complexity for the spectrum of assumptions between the two extremes. Definition 2. The distribution-dependent sequential Rademacher complexity of a function class F ? RX is defined as # " T X ? ?t f (xt (?)) RT (F, p) = E(x,x?)?? E? sup f ?F t=1

where ? = (?1 , . . . , ?T) is a sequence of i.i.d. Rademacher random variables and ? is the probability tree associated with p.

We now prove an upper bound on the value VT (P1:T) of the game in terms of this distributiondependent sequential Rademacher complexity. The result cannot be deduced directly from [10], and it greatly increases the scope of problems whose learnability can now be studied in a unified manner. Theorem 3. The minimax value is bounded as VT (P1:T) ? 2 sup RT (F, p). p?P

4

(5)

More generally, for any measurable function Mt such that Mt (p, f, x, x? , ?) = Mt (p, f, x? , x, ??), ”

VT (P1:T) ? 2 sup E(x,x?)?? E? sup p?P

T X

?

#

?t (f (xt (?)) ? Mt (p, f, x, x , ?))

f ?F t=1

The following corollary provides a natural ?centered? version of the distribution-dependent Rademacher complexity. That is, the complexity can be measured by relative shifts in the adversarial moves. Corollary 4. For the game with restrictions P1:T , # ” T

X ?t f (xt (?)) ? Et?1 f (xt (?)) VT (P1:T) ? 2 sup E(x,x?)?? E? sup p?P

f ?F t=1

where Et?1 denotes the conditional expectation of xt (?). Example 1. Suppose F is a unit ball in a Banach space and f (x) = hf, xi. Then

T

X

VT (P1:T) ? 2 sup E(x,x?)?? E? ?t xt (?) ? Et?1 xt (?)

p?P t=1

Suppose the adversary plays a simple random walk (e.g., pt (x—x1 , . . . , xt?1) = pt (x—xt?1) is uniform on a unit sphere). For simplicity, suppose this is the only strategy allowed by the set P. Then xt (?) ? Et?1 xt (?) are independent increments when conditioned

on the history. Further, the in

PT crements do not depend on ?t . Thus, VT (P1:T) ? 2E t=1 Yt where {Yt } is the corresponding random walk.

We now show that the distribution-dependent sequential Rademacher complexity for i.i.d. data is precisely the classical Rademacher complexity, and further show that the distribution-dependent sequential Rademacher complexity is always upper bounded by the worst-case sequential Rademacher complexity defined in [10]. Proposition 5. First, consider the i.i.d. restrictions Pt = {p} for all t, where p is some fixed distribution on X , and let ? be the process associated with the joint distribution p = pT . Then ” # T X ? RT (F, p) = RT (F, p), where RT (F, p) = Ex1 ,...,xT ?p E? sup ?t f (xt) (6) f ?F t=1

is the classical Rademacher complexity. Second, for any joint distribution p, ” # T X ? RT (F, p) ? RT (F), where RT (F) = sup E? sup ?t f (xt (?)) x

(7)

f ?F t=1

is the sequential Rademacher complexity defined in [10].

In the case of hybrid learning, adversary chooses a sequence of pairs (xt , yt) where the instance xt ?s are i.i.d. but the labels yi ?s are fully adversarial. The distribution-dependent Rademacher complexity in such a hybrid case can be upper bounded by a vary natural quantity: a random average where expectation is taken over xt ?s and a supremum over Y-valued trees. So, the distribution dependent Rademacher complexity itself becomes a hybrid between

the classical Rademacher complexity and the worst case sequential Rademacher complexity. For more details, see Lemma 17 in the Appendix as another example of an analysis of the distribution-dependent sequential Rademacher complexity. Distribution-dependent sequential Rademacher complexity enjoys many of the nice properties satisfied by both classical and worst-case Rademacher complexities. As shown in [10], these properties are handy tools for proving upper bounds on the value in various examples. We have: (a) If F ? G, then R(F, p) ? R(G, p); (b) R(F, p) = R(conv(F), p); (c) R(cF, p) = —c—R(F, p) for all c ? R; (d) For any h, R(F + h, p) = R(F, p) where F + h = {f + h : f ? F}. In addition to the above properties, upper bounds on R(F, p) can be derived via sequential covering numbers defined in [10]. This notion of a cover captures the sequential complexity of a function class on a given X -valued tree x. One can then show an analogue of the Dudley integral bound, where the complexity is averaged with respect to the underlying process (x, x?) ? ?. 5

4

Application: Constrained Adversaries

In this section, we consider adversaries who are deterministically constrained in the sequences of actions they can play. It is often useful to consider scenarios where the adversary is worst case, yet has some budget or constraint to satisfy while picking the actions. Examples of such scenarios include, for instance, games where the adversary is constrained to make moves that are close in some fashion to the previous move, linear games with bounded variance, and so on. Below we formulate such games quite generally through arbitrary constraints that the adversary has to satisfy on each round. We easily derive several results to illustrate the versatility of the developed framework. For a T round game consider an adversary who is only allowed to play sequences x1 , . . . , xT such that at round t the constraint Ct (x1 , . . . , xt) = 1 is satisfied, where Ct : X t 7? {0, 1} represents the constraint on the sequence played so far. The constrained adversary can be viewed as a stochastic adversary with restrictions on the conditional distribution at time t given by the set of all Borel ? distributions on the set Xt (x1:t?1) = {x ? X : Ct (x1 , . . . , xt?1 , x) = 1}. Since this set includes all point distributions on each x ? Xt , the sequential complexity simplifies in a way similar to worst-case adversaries. We write VT (C1:T) for the value of the game with the given constraints. Now, assume that for any x1:t?1 , the set of all distributions on Xt (x1:t?1) is weakly compact in a way similar to compactness of P. That is, Pt (x1:t?1) satisfy the necessary conditions for the minimax theorem to hold. We have the following corollaries of Theorems 1 and 3. Corollary 6. Let F and X be the sets of moves for the two players, satisfying the necessary conditions for the minimax theorem to hold. Let {Ct : X t?1 7? {0, 1}}Tt=1 be the constraints. Then " T # T X X inf Ext ?pt [ft (xt)] ? inf f (xt) (8) VT (C1:T) = sup E p?P

t=1

ft ?F

f ?F

t=1

where p ranges over all distributions over sequences (x1 , . . . , xT) such

that ?t, Ct (x1:t?1) = 1.

Corollary 7. Let the set T be a set of pairs (x, x?) of X -valued trees with the property that for any ? ? {?1}T and any t ? [T], C(?1 (?1), . . . , ?t?1 (?t?1), xt (?)) = C(?1 (?1), . . . , ?t?1 (?t?1), x?t (?)) = 1 . The minimax value is bounded as VT (C1:T) ? 2

sup

RT (F, p).

(x,x?)?T

More generally, for any measurable function Mt such that Mt (f, x, x? , ?) = Mt (f, x? , x, ??), # " T X ?t (f (xt (?)) ? Mt (f, x, x? , ?)) . VT (C1:T) ? 2 sup E? sup (x,x?)?T

f ?F t=1

Armed with these results, we can recover and extend some known results on online learning against budgeted adversaries. The first result says that if the adversary is not allowed to move by more than ?t away from its previous average of decisions, the player has a strategy to exploit this fact and obtain lower regret. For the ?2 -norm, such ?total variation? bounds have been achieved in [4] up to a log T factor. Our analysis seamlessly incorporates variance measured in arbitrary norms, not just ?2 . We emphasize that such certificates of learnability are not possible with the analysis of [10]. Proposition 8 (Variance Bound). Consider the online linear optimization setting with F = {f : ?(f) ? R2 } for a ?-strongly function ? : F 7? R+ on F, and X = {x : kxk? ? 1}. Let f (x) = hf, xi for any f ? F and x ? X . Consider the sequence of constraints {Ct }Tt=1 given by Pt?1 1 Ct (x1 , . . . , xt?1 , x) = 1 if kx ? t?1 ? =1 x? k? ? ?t and 0 otherwise. Then ? q PT VT (C1:T) ? 2 2R ??1 t=1 ?t2 In particular, we obtain the following ?2 variance bound. Consider the case when ? : F 7? R+ is given by ?(f) = 12 kf k2 , F = {f : kf k2 ? 1} and X = {x : kxk 2 ? 1}. Consider the constrained Pt?1

1 game where the move xt played by adversary at time t satisfies xt ? t?1 ? =1 x? ? ?t . In 2 ? qPT 2 this case we can conclude that VT (C1:T) ? 2 2 t=1 ?t . We can also derive a variance bound 6

Pd over the simplex. Let ?(f) = i=1 fi log(dfi) is defined over the d-simplex F, and X = {x : kxk? ? 1}. Consider the constrained game where the move xt played by adversary at time t

Pt?1

1 satisfies maxj?[d] xt [j] ? t?1 ? =1 x? [j] ? ?t . For any f ? F, ?(f) ? log(d) and so we q ? PT conclude that VT (C1:T) ? 2 2 log(d) t=1 ?t2 .

The next Proposition gives a bound whenever the adversary is constrained to choose his decision from a small ball around the previous decision. Proposition 9 (Slowly-Changing Decisions). Consider the online linear optimization setting where adversary?s move at any time is close to the move during the previous time step. Let F = {f : ?(f) ? R2 } where ? : F 7? R+ is a ?-strongly function on F and X = {x : kxk? ? B}. Let f (x) = hf, xi for any f ? F and x ? X . Consider the sequence of constraints {Ct }Tt=1 given by Ct (x1 , . . . , xt?1 , x) = 1 if kx ? xt?1 k? ? ? and 0 otherwise. Then, p VT (C1:T) ? 2R? 2T /? .

In particular, consider the case of a Euclidean-norm restriction on the moves. Let ? : F 7? R+ is given by ?(f) = 12 kf k2 , F = {f : kf k2 ? 1} and X = {x : kxk2 ? 1}. Consider the constrained game where the move xt played by adversary at time t satisfies kxt ? xt?1 k2 ? ? . In this case ? we can conclude that VT (C1:T) ? 2? 2T . For the case of decision-making on the simplex, we Pd obtain the following result. Let ?(f) = i=1 fi log(dfi) is defined over the d-simplex F, and X = {x : kxk? ? 1}. Consider the constrained game where the move xt played by adversary at time t satisfies kxt ? xt?1 k? ? ?. pIn this case note that for any f ? F, ?(f) ? log(d) and so we can conclude that VT (C1:T) ? 2? 2T log(d) .

5

Application: Smoothed Adversaries

The development of smoothed analysis over the past decade is arguably one of the landmarks in the study of complexity of algorithms. In contrast to the overly optimistic average complexity and the overly pessimistic worst-case complexity, smoothed complexity can be seen as a more realistic measure of algorithm?s performance. In their groundbreaking work, Spielman and Teng [13] showed that the smoothed running time complexity of the simplex method is polynomial. This result explains good performance of the method in practice despite its exponential-time worst-case complexity. In this section, we consider the effect of smoothing on learnability. It is well-known that there is a gap between the i.i.d. and the worst-case scenarios. In fact, we do not need to go far for an example: a simple class of threshold functions on a unit interval is learnable in the i.i.d. supervised learning scenario, yet difficult in the online worst-case model [8, 2, 9]. This fact is reflected in the corresponding combinatorial dimensions: the Vapnik-Chervonenkis dimension is one, whereas the Littlestone dimension is infinite. The proof of the latter fact, however, reveals that the infinite number of mistakes on the part of the player is due to the infinite resolution of the carefully chosen adversarial sequence. We can argue that this infinite precision is an unreasonable assumption on the power of a real-world opponent. The idea of limiting the power of the malicious adversary through perturbing the sequence can be traced back to Posner and Kulkarni [9]. The authors considered on-line learning of functions of bounded variation, but in the so-called realizable setting (that is, when labels are given by some function in the given class). We define the smoothed online learning model as the following T -round interaction between the learner and the adversary. On round t, the learner chooses ft ? F; the adversary simultaneously ?t = ?(xt , st); chooses xt ? X , which is then perturbed by some noise st ? ?, yielding a value x and the player suffers ft (? xt). Regret is defined with respect to the perturbed sequence. Here ? : X ? S 7? X is some measurable mapping; for instance, additive disturbances can be written as x ? = ?(x, s) = x + s. If ? keeps xt unchanged, that is ?(xt , st) = xt , the setting is precisely the standard online learning model. In the full information version, we assume that the choice x ?t is revealed to the player at the end of round t. We now recognize that the setting is nothing but a particular way to restrict the adversary. That is, the choice xt ? X defines a parameter of a mixed strategy from which a actual

move ?(xt , st) is drawn; for instance, for additive zero-mean Gaussian noise, xt defines the center of the distribution from which xt + st is drawn. In other words, noise does not allow the adversary to play any desired mixed strategy. 7

The value of the smoothed online learning game (as defined in (1)) can be equivalently written as " T # T X X VT = inf sup E inf sup E ? ? ? inf sup E ft (?(xt , st)) ? inf f (?(xt , st)) q1

x1 f1 ?q1 q2 s1 ??

x2 f2 ?q2 s2 ??

qT

xT fT ?qT sT ??

t=1

f ?F

t=1

where the infima are over qt ? Q and the suprema are over xt ? X . Using sequential symmetrization, we deduce the following upper bound on the value of the smoothed online learning game. Theorem 10. The value of the smoothed online learning game is bounded above as " # T X VT ? 2 sup E E?1 . . . sup E E?T sup ?t f (?(xt , st)) x1 ?X s1 ??

xT ?X sT ??

f ?F t=1

We now demonstrate how Theorem 10 can be used to show learnability for smoothed learning of threshold functions. First, consider the supervised game with threshold functions on a unit interval (that is, non-homogenous hyperplanes). The moves of the adversary are pairs x = (z, y) with z ? [0, 1] and y ? {0, 1}, and the binary-valued function class F is defined by (9) F = {f? (z, y) = —y ? 1 {z ¡ ?}— : ? ? [0, 1]} , that is, every function is associated with a threshold ? ? [0, 1]. The class F has infinite Littlestone?s dimension and is not learnable in the worst-case online framework. Consider a smoothed scenario, with the z-variable of the adversarial move (z, y) perturbed by an additive uniform noise ? = Unif[??/2, ?/2] for some ? ? 0. That is, the actual move revealed to the player at time t is (zt + st , yt), with st ? ?. Any non-trivial upper bound on regret has to depend on particular noise assumptions, as ? = 0 corresponds to the case with infinite Littlestone dimension. For the uniform disturbance, the intuition tells us that noise implies a margin, and we should expect a 1/? complexity parameter appearing in the bounds. The next lemma quantifies the intuition that additive noise limits precision of the adversary. Lemma 11. Let ?1 , . . . , ?N be obtained by discretizing the interval [0, 1] into N = T a bins [?i , ?i+1) of length T ?a , for some a ? 3. Then, for any sequence z1 , . . . , zT ? [0, 1], with probability at least 1 ? ?T 1a?2 , no two elements of the sequence z1 + s1 , . . . , zT + sT belong to the same interval [?i , ?i+1), where s1 , . . . , sT are i.i.d. Unif[??/2, ?/2]. We now observe that, conditioned on the event in Lemma 11, the upper bound on the value in Theorem 10 is a supremum of N martingale difference sequences! We then arrive at: Proposition 12. For the problem of smoothed online learning of thresholds in 1-D, the value is p VT ? 2 + 2T (4 log T + log(1/?))

What we found is somewhat surprising: for a problem which is not learnable in the online worstcase scenario, an exponentially small noise added to the moves of the adversary yields a learnable problem. This shows, at least in the given example, that the worst-case analysis and Littlestone?s dimension are brittle notions which might be too restrictive in the real world, where some noise is unavoidable. It is comforting that small additive noise makes the problem learnable! The proof for smoothed learning of half-spaces in higher dimension follows the same route as the one-dimensional exposition. For simplicity, assume the hyperplanes are homogenous and Z = Sd?1 ? Rd , Y = {?1, 1}, X = Z ?Y. Define F = {f? (z, y) = 1 {y hz, ?i ¿ 0} : ? ? Sd?1 }, and assume that the noise is distributed uniformly on a square patch with side-length ? on the surface of the sphere Sd?1 . We can also consider other distributions, possibly with support on a d-dimensional ball instead. Proposition 13. For the problem of smoothed online learning of half-spaces, s 3 ! d?1

1 1 3 VT = O dT log log T + vd?2 ? + ? d?1 ? where vd?2 is constant depending only on the dimension d.

We conclude that half spaces are online learnable in the smoothed model, since the upper bound of Proposition 13 guarantees existence of an algorithm which achieves this regret. In fact, for the two examples considered in this section, the Exponential Weights Algorithm on the discretization given by Lemma 11 is a (computationally infeasible) algorithm achieving the bound. 8

2 References

[1] J. Abernethy, A. Agarwal, P. Bartlett, and A. Rakhlin. A stochastic view of optimal regret through minimax duality. In COLT, 2009. [2] S. Ben-David, D. Pal, and S. Shalev-Shwartz. Agnostic online learning. In Proceedings of the 22th Annual Conference on Learning Theory, 2009. [3] J.O. Berger. Statistical decision theory and Bayesian analysis. Springer, 1985. [4] E. Hazan and S. Kale. Better algorithms for benign bandits. In SODA, 2009. [5] S.M. Kakade, K. Sridharan, and A. Tewari. On the complexity of linear prediction: Risk bounds, margin bounds, and regularization. NIPS, 22, 2008. [6] A. Lazaric and R. Munos. Hybrid Stochastic-Adversarial On-line Learning. In COLT, 2009. [7] M. Ledoux and M. Talagrand. Probability in Banach Spaces. Springer-Verlag, New York, 1991. [8] N. Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. Machine Learning, 2(4):285?318, 04 1988. [9] S. Posner and S. Kulkarni. On-line learning of functions of bounded variation under various sampling schemes. In Proceedings of the sixth annual conference on Computational learning theory, pages 439?445. ACM, 1993. [10] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Random averages, combinatorial parameters, and learnability. In NIPS, 2010. Full version available at arXiv:1006.1138. [11] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Beyond regret. In COLT, 2011. Full version available at arXiv:1011.3168. [12] S. Shalev-Shwartz, O. Shamir, N. Srebro, and K. Sridharan. Learnability, stability and uniform convergence. JMLR, 11:2635?2670,

Oct 2010. [13] D. A. Spielman and S. H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. Journal of the ACM, 51(3):385?463, 2004. [14] A. W. Van Der Vaart and J. A. Wellner. Weak Convergence and Empirical Processes : With Applications to Statistics. Springer Series, March 1996.

9