6.883: Online Methods in Machine Learning

Alexander Rakhlin

# LECTURE 23

## 1. SOME CONSEQUENCES OF ONLINE NO-REGRET METHODS

In this lecture, we explore some consequences of the developed techniques.

### 1.1 Convex optimization

Whenever we have an online linear optimization guarantee

$$\sum_{t=1}^{n} \langle \widehat{y}_t, z_t \rangle - \min_{v \in K} \sum_{t=1}^{n} \langle v, z_t \rangle \leq \mathbf{Reg}_n \tag{1}$$

that holds for all sequences $z_1, \ldots, z_n \in K'$, then for any convex function $f : K \to \mathbb{R}$ with gradients in $K'$

$$\{\nabla f(v) : v \in K\} \subseteq K',$$

it holds that

$$f(\bar{y}) - \min_{v \in K} f(v) \leq \frac{1}{n} \mathbf{Reg}_n \tag{2}$$

for

$$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} \widehat{y}_t.$$

In other words, by running an online linear optimization method and averaging the trajectory (called Polyak averaging), we obtain a point in $K$ with suboptimality bounded by normalized regret $\frac{1}{n}\mathbf{Reg}_n$. For instance, if $K = B_2$ and gradients of $f$ are also bounded by 1 in Euclidean norm, methods developed earlier give $\mathbf{Reg}_n = \sqrt{n}$, and thus the average of the trajectory $\bar{y}$ satisfies

$$f(\bar{y}) - \min_{v \in K} f(v) \leq \frac{1}{\sqrt{n}}.$$

Recall that this is a bound we already derived in the beginning of the course.

Let's prove the claim (2). By convexity, for any $v \in K$,

$$f(\bar{y}) - f(v) \leq \frac{1}{n} \sum_{t=1}^{n} f(\widehat{y}_t) - f(v) \leq \frac{1}{n} \sum_{t=1}^{n} \langle \widehat{y}_t - v, \nabla f(\widehat{y}_t) \rangle. \tag{3}$$

Then (2) follows from (1) by taking $z_t = \nabla f(\widehat{y}_t)$. This step is valid because $z_t$ can be chosen based on $\widehat{y}_t$ (as per the online linear optimization protocol).

We remark that (2) holds with $\mathbf{Reg}_n = \mathbf{Reg}_n(z_1, \ldots, z_n)$. Indeed, adaptive data-dependent bounds for optimization have been derived via this approach.

Finally, observe that a version of (3)

$$\frac{1}{n}\sum_{t=1}^{n} f_t(\widehat{y}_t) - \frac{1}{n}\sum_{t=1}^{n} f_t(v) \le \frac{1}{n}\sum_{t=1}^{n}\langle \widehat{y}_t - v, \nabla f_t(\widehat{y}_t)\rangle \tag{4}$$

holds with time-varying convex functions $f_t$. Hence, a regret guarantee for online *linear* optimization translates into a regret guarantee for this online *convex* optimization scenario.

### 1.2 Zero-sum games

Let $M$ be a $d_1 \times d_2$ real matrix with bounded entries. We think of $M_{i,j}$ (resp., $-M_{i,j}$) as the cost for Player I (resp., Player II) when the two players choose their moves as $i$ and $j$. As discussed a few lectures ago, the minimax value is

$$\min_{q\in\Delta_{d_1}} \max_{j\in[d_2]} q^\mathsf{T} M \boldsymbol{e}_j = \max_{p\in\Delta_{d_2}} \min_{i\in[d_1]} \boldsymbol{e}_i^\mathsf{T} M p = \min_{q\in\Delta_{d_1}} \max_{p\in\Delta_{d_2}} q^\mathsf{T} M p. \tag{5}$$

It turns out, both players can find a near-minimax-optimal strategy simply by employing an online linear optimization method with a non-trivial regret guarantee. More precisely, we assume that on round $t$, Player $I$ picks a probability distribution $q_t \in \Delta_{d_1}$ on rows, while Player II picks a probability distribution $p_t \in \Delta_{d_2}$ on columns of $M$. The first player then receives $Mp_t$ (or a random draw $Me_{J_t}$, $J_t \sim p_t$), while the second player receives $q_t^\mathsf{T} M$ (or the corresponding random draw).

Let $\bar{q} = \frac{1}{n}\sum_{t=1}^{n} q_t$ and $\bar{p} = \frac{1}{n}\sum_{t=1}^{n} p_t$ be the averages of the trajectories. Since each player employs a no-regret strategy,

$$\frac{1}{n}\sum_{t=1}^{n} q_t^\mathsf{T} M p_t - \min_{q\in\Delta_{d_1}} \frac{1}{n}\sum_{t=1}^{n} q^\mathsf{T} M p_t \le \frac{1}{n}\mathbf{Reg}_n^{(1)} \tag{6}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}(-q_t^\mathsf{T} M) p_t - \min_{p\in\Delta_{d_2}} \frac{1}{n}\sum_{t=1}^{n}(-q_t^\mathsf{T} M) p \le \frac{1}{n}\mathbf{Reg}_n^{(2)}. \tag{7}$$

We put a minuses in the second expression because the second player can pretend she is the minimizing player in the game with losses given according to $-M$. Adding the two expressions and using the definition of $\bar{q}$ and $\bar{p}$,

$$\max_{p\in\Delta_{d_2}} \bar{q}^\mathsf{T} M p - \min_{q\in\Delta_{d_1}} q^\mathsf{T} M \bar{p} \le \frac{1}{n}(\mathbf{Reg}_n^{(1)} + \mathbf{Reg}_n^{(2)}) \tag{8}$$

However, we know that

$$\min_{q\in\Delta_{d_1}} q^\mathsf{T} M \bar{p} \le \max_{p\in\Delta_{d_2}} \min_{q\in\Delta_{d_1}} q^\mathsf{T} M p = \min_{q\in\Delta_{d_1}} \max_{p\in\Delta_{d_2}} q^\mathsf{T} M p \le \max_{p\in\Delta_{d_2}} \bar{q}^\mathsf{T} M p. \tag{9}$$

In view of (8) and (9), the pair $(\bar{q}, \bar{p})$ attains a value within $\frac{1}{n}(\mathbf{Reg}_n^{(1)} + \mathbf{Reg}_n^{(2)})$ from minimax. In other words, the average of the trajectory of both players converges (in the sense of the minimax value) to a set of minimax optima (Nash equilibria). This technique extends to multi-player games, whereby each decision-maker runs a no-regret algorithm against the (complicated) set of moves of the other players.

### 1.2.1 What if Player II best-responds

As an alternative (which will be used in the next section) suppose Player I employs a regret-minimization strategy, while Player II best responds to the observed $q_t^\mathsf{T} M$:

$$p_t = \operatorname*{argmin}_{j\in[d_2]} q_t^\mathsf{T} M \boldsymbol{e}_j.$$

This behavior also leads to the minimax strategy since

$$\max_p \bar{q}^\mathsf{T} M p \le \frac{1}{n}\sum_{t=1}^n \max_p q_t^\mathsf{T} M p = \frac{1}{n}\sum_{t=1}^n q_t^\mathsf{T} M p_t \tag{10}$$

$$\le \min_{q\in\Delta_{d_1}} \frac{1}{n}\sum_{t=1}^n q^\mathsf{T} M p_t + \frac{1}{n}\mathbf{Reg}_n^{(1)} = \min_{q\in\Delta_{d_1}} q^\mathsf{T} M \bar{p} + \frac{1}{n}\mathbf{Reg}_n^{(1)} \tag{11}$$

In view of (9), the pair of strategies $(\bar{q}, \bar{p})$ is within $\frac{1}{n}\mathbf{Reg}_n^{(1)}$ from minimax optimal. In particular, note that $d_2$ does not affect the rate of convergence – this will be important in the next section.

## 1.3 Application: Boosting

We briefly describe a variant of Boosting, a procedure that aggregates weak classifiers into a strong one. This is a batch procedure, and the online aspect will appear as an iterative improvement of the aggregate.

More precisely, let $(x_1, y_1), \ldots, (x_m, y_m) \in \mathcal{X} \times \{\pm 1\}$ be a dataset, and $\mathcal{F}$ a class of $[-1, 1]$-valued functions on $\mathcal{X}$. Each function $f \in \mathcal{F}$ induces a binary prediction $\operatorname{sign}(f(x_t))$ at $x_t$. For the given pair $(x_t, y_t)$, a mistake occurs if $y_t f(x_t) < 0$.

Suppose the class $\mathcal{F}$ has finite cardinality $k$ (though this can be lifted at the end since the final bound does not depend on this value) and enumerate $\mathcal{F} = \{f_1, \ldots, f_k\}$. Define a matrix $M \in [-1, 1]^{m\times k}$ with

$$M_{i,j} = y_i f_j(x_i).$$

Consider a minimax value $q^\mathsf{T} M p$, where $q$ is a distribution on examples and $p$ is a distribution on functions. We have

$$\min_{q\in\Delta_m} \max_{j\in[k]} q^\mathsf{T} M \boldsymbol{e}_j = \max_{p\in\Delta_k} \min_{i\in[m]} \boldsymbol{e}_i^\mathsf{T} M p \quad \triangleq \gamma. \tag{12}$$

Suppose $\gamma > 0$. The two sides of (12) have interesting and distinct interpretations.

The left-hand side says: no matter what distribution $q$ on the dataset we choose, there is some function $f_j$ such that

$$\mathbb{E}_{i\sim q} y_i f_j(x_i) \ge \gamma.$$

If you've encountered the literature on boosting (search for it in the internet if you haven't), the above requirement is known as the weak learnability condition [FS95, SF12]. To be able to boost "weak learners", one must be able to find, for any discrete distribution on the $m$ datapoints, a hypothesis that beats random guess.

The right-hand side of (12) says: there exists a distribution $p$ over the $k$ functions, such that on any data point $(x_i, y_i)$, one has

$$y_i \bar{f}(x_i) \ge \gamma, \tag{13}$$

for $\bar{f} = \sum_{i=1}^k p_i f_i$. This can be recognized as a *separability* assumption with margin $\gamma$ (see Lecture 2).

Thus, separability with a margin is equivalent to the weak learnability condition. It is no surprise that weak learnability means one can create a mixture of functions that attains zero mistakes.

Equivalence aside, the goal of boosting is to furnish a distribution $p$ that nearly achieves (13). According to the previous subsection and, in particular, Eq. (11), one may achieve this by running an exponential weighting (or some other regret minimization procedure) over the $m$ choices for Player I, and best response for Player II. Best response exactly corresponds to the usual weak learnability assumption whereby a choice $f_j$ (a pure strategy) is returned to Player I for the given $q_t$, a key step in any boosting procedure.

It remains to check the number of iterations $n$ required. If margin is $\gamma$, we can aim for achieving a margin of $\gamma/2$. Then equating the Exponential Weights regret bound

$$\frac{1}{n}\mathbf{Reg}_n^{(1)} = \frac{1}{n}C\sqrt{n\log m}$$

to target accuracy $\gamma/2$, we find

$$n = O\left(\frac{1}{\gamma^2}\log m\right). \tag{14}$$

The typical analysis of AdaBoost claims exponentially decaying error, which is expressed as

$$\frac{1}{m}\sum_{i=1}^m \mathbf{1}\left\{y_i\bar{f}(x_i) \le 0\right\} \le \exp\{-c\gamma^2 n\} \tag{15}$$

after $n$ iterations. The setting of (14) ensures that the upper bound is less than $1/m$, and hence no errors are made (convince yourself of this fact!). That is precisely what we found by ensuring a margin of $\gamma/2$.

Observe that $k = |\mathcal{F}|$ never enters the picture since we are using best response strategy for Player II.

### 1.4 From Online to Statistical Learning (individual sequence to i.i.d.)

Consider an abstract scenario where the learner chooses $\widehat{y}_t$, observes $z_t$, and incurs a cost $\boldsymbol{\ell}(\widehat{y}_t, z_t)$. We assume $\boldsymbol{\ell}$ is convex in the first argument and that we are able to prove

$$\frac{1}{n}\sum_{t=1}^n \boldsymbol{\ell}(\widehat{y}_t, z_t) \le \min_{v\in K}\frac{1}{n}\sum_{t=1}^n \boldsymbol{\ell}(v, z_t) + \frac{1}{n}\mathbf{Reg}_n. \tag{16}$$

for any sequence $z_1, \ldots, z_n$. We are not specifying the sets where the decisions and outcomes take values, as the technique we are about to describe is quite general.

Now, suppose the data are actually i.i.d. with some distribution $P$. Then

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^n \boldsymbol{\ell}(\widehat{y}_t, Z_t)\right] \le \mathbb{E}\left[\min_{v\in K}\frac{1}{n}\sum_{t=1}^n \boldsymbol{\ell}(v, Z_t)\right] + \frac{1}{n}\mathbf{Reg}_n \tag{17}$$

since the bound holds for each realization $(Z_1, \ldots, Z_n)$. Let

$$\bar{y} = \frac{1}{n}\sum_{t=1}^n \widehat{y}_t.$$

Observe that

$$\mathbb{E}\boldsymbol{\ell}(\bar{y}, Z) \le \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\ell}(\widehat{y}_t, Z)\right] \tag{18}$$

where the expectation is over both $Z$ and $Z_1, \ldots, Z_n$. The expression in (18) can be written as

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\boldsymbol{\ell}(\widehat{y}_t(Z_{1:t-1}), Z)|Z_{1:t-1}\right]\right] \tag{19}$$

which is the same as

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\boldsymbol{\ell}(\widehat{y}_t(Z_{1:t-1}), Z_t)|Z_{1:t-1}\right]\right] = \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\ell}(\widehat{y}_t, Z_t)\right]. \tag{20}$$

On the other hand,

$$\mathbb{E}\left[\min_{v\in K}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\ell}(v, Z_t)\right] \le \min_{v\in K}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\boldsymbol{\ell}(v, Z_t) = \min_{v\in K}\mathbb{E}\boldsymbol{\ell}(v, Z)$$

Putting everything together,

$$\mathbb{E}\boldsymbol{\ell}(\bar{y}, Z) \le \min_{v\in K}\mathbb{E}\boldsymbol{\ell}(v, Z) + \frac{1}{n}\mathbf{Reg}_n \tag{21}$$

Such a bound is a goal of Statistical Learning Theory. Moreover, if $\ell$ is square loss, the above bound also translates into an exact oracle inequality, a subject of interest in Statistics.

We see that a regret guarantee gives rise to a statistical guarantee for the average of the trajectory if we assume that data are i.i.d.

In particular, consider the supervised learning scenario, where at each step we observe side information $x_t$, predict a real value $\widehat{y}_t$, and observe $y_t$. We can think of $\widehat{y}_t$ as a function of $x_t$, since for different $x_t$ values we would have made a different prediction. In other words, an equivalent formulation would be: on round $t$, predict a function $\widehat{y}_t : \mathcal{X} \to \mathbb{R}$ and observe $(x_t, y_t)$. Applying the earlier argument, if $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d., the function

$$\bar{y}(\cdot) = \frac{1}{n}\sum_{t=1}^{n}\widehat{y}_t(\cdot)$$

has the guarantee

$$\mathbb{E}\boldsymbol{\ell}(\bar{y}(X), Y) \le \min_{f\in\mathcal{F}}\mathbb{E}\boldsymbol{\ell}(f(X), Y) + \frac{1}{n}\mathbf{Reg}_n \tag{22}$$

whenever $\boldsymbol{\ell}(\bar{y}(X), Y)$ is convex in $\bar{y}(\cdot)$. While further development of this technique is beyond the scope of this course, let us mention that under certain assumptions, this way of constructing an estimator $\bar{y}(\cdot)$ is optimal for a wide range of problems and loss functions, both in Nonparametric Statistics and Statistical Learning Theory. (ask me if you'd like to know more)

For those who know the analysis of ERM (empirical risk minimization) in Statistical Learning may ask whether we have miraculously circumvented the issue of uniform convergence of means to expectations, a necessary step in the analysis of ERM. The answer is that we haven't — it is "hidden" in the regret bound $\mathbf{Reg}_n = \mathbf{Reg}_n(\mathcal{F})$. More precisely, it can be shown that this term is controlled by a martingale version of uniform convergence [RST15]. In many cases of interest, this convergence is equivalent to i.i.d. convergence. Hence, by passing through online methods, we might be gaining in terms of algorithmic understanding, and (under certain conditions) not losing in terms of accuracy for i.i.d. data.

### 1.5 Equivalence of deterministic regret bounds and martingale inequalities

The connection between online regret bounds and Statistics and Probability are even more intriguing than outlined above. Let us sketch one example.

Let $z_1, \ldots, z_n \in B_2$ be an arbitrary sequence in the unit Euclidean ball. We construct the sequence of predictions by $\widehat{y}_0 = 0$ and

$$\widehat{y}_{t+1} = \Pi_{B_2}\left(\widehat{y}_t - \frac{1}{\sqrt{n}} z_t\right), \tag{23}$$

which we can all recognize as gradient descent with projections onto $B_2$ and step size $\eta = \frac{1}{\sqrt{n}}$. Earlier in the course we have proved a regret bound of

$$\sum_{t=1}^{n} \langle \widehat{y}_t - v, z_t \rangle \leq \sqrt{n} \tag{24}$$

for any $z_1, \ldots, z_n \in B_2$ and $v \in B_2$. Let us choose the optimal $v$ and rearrange terms:

$$\forall z_1, \ldots, z_n \in B_2, \qquad \left\| \sum_{t=1}^{n} z_t \right\| - \sqrt{n} \leq \sum_{t=1}^{n} \langle \widehat{y}_t, -z_t \rangle . \tag{25}$$

Now, apply this inequality to a realization of a martingale difference sequence $Z_1, \ldots, Z_n$ (that is, $\mathbb{E}[Z_{t+1}|Z_{1:t}] = 0$). Since the inequality holds for any realization,

$$\mathbf{P}\left( \left\| \sum_{t=1}^{n} Z_t \right\| - \sqrt{n} \geq u \right) \leq P\left( \sum_{t=1}^{n} \langle \widehat{y}_t, -Z_t \rangle \geq u \right). \tag{26}$$

for any $u > 0$. Since $\widehat{y}_t = \widehat{y}_t(Z_1, \ldots, Z_{t-1})$, the right-hand side is a sum of real-valued martingale differences with values in $[-1, 1]$, and so the right-hand side is upper bounded by $\exp\{-u^2/(2n)\}$ via Hoeffding-Azuma inequality. Therefore,

$$\mathbf{P}\left( \left\| \sum_{t=1}^{n} Z_t \right\| - \sqrt{n} \geq u \right) \leq \exp\left\{-u^2/(2n)\right\}. \tag{27}$$

One may pass from this statement in probability to an expected-value bound

$$\mathbb{E}\left\| \sum_{t=1}^{n} Z_t \right\| \leq c\sqrt{n} \tag{28}$$

by integrating the tails in (27). Note that the inequality holds for all martingale difference sequences with values in $B_2$.

Using the minimax theorem, one can show (we basically have done this already) that there exists a strategy with (25) for any sequence because (28) holds. That is, (25) implies (27) which implies (28) which implies (25) (with a worse constant). Hence, the statements are, in a certain sense, equivalent. One can, in fact (see the relevant paper [RS15]) improve tail bounds (27) by taking a smarter gradient descent method with an adaptive step size, and, conversely, it is possible to derive new optimization methods by proving tail bounds for martingales.

This lecture was meant to give a brief overview of some interesting connections between online methods and the near-by areas of game theory, statistics, probability, and optimization.

## References


[FS95] Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[RS15] Alexander Rakhlin and Karthik Sridharan. On equivalence of martingale tail bounds and deterministic regret inequalities. *arXiv preprint arXiv:1510.03925*, 2015.

[RST15] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Sequential complexities and uniform martingale laws of large numbers. *Probability Theory and Related Fields*, 161(1-2):111–153, 2015.

[SF12] Robert E Schapire and Yoav Freund. *Boosting: Foundations and algorithms*. MIT press, 2012.