Introduction to Online Learning (CSC 482A/581A) - Lecture 4

Nishant Mehta

1 Stochastic convex optimization

A stochastic convex optimization problem is specified by a probability distribution P over a set $\mathcal{Z}$, a convex set V, and a function $f\colon V \times \mathcal{Z} \to \mathbb{R}$ that is convex in its first argument. The goal is to find some $w \in V$ which minimizes the objective

$$F(w) = \mathsf{E}_{Z\sim P}[f(w, Z)].$$

We will use $w^* \in V$ to denote an arbitrary minimizer of F, so that $F(w^*) = \min_{w\in V} F(w)$. In analogy to statistical learning, we refer to $F(w)$ as the *risk* of w and $F(w) - F(w^*)$ as the *excess risk* of w.

Supervised learning with linear predictors can be recovered by:

- taking $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, so that $Z = (X, Y)$ for feature vector X and label Y;
- defining $f(w, z) = f(w, (x, y)) = \ell(\langle w, x\rangle, y)$ for some loss function $\ell\colon \mathbb{R} \times \mathcal{Y} \to \mathbb{R}$ that is convex in its first argument.

In order to approximately minimize the objective $F(w)$, a learning algorithm will be presented with i.i.d. samples $Z_1, \ldots, Z_T$ distributed according to P, similar to the statistical learning setting.

We will study algorithms for solving the stochastic optimization problem based on online convex optimization (OCO) and a technique known as an *online-to-batch conversion*. The idea will be to:

- first, frame an online version of the above problem as an online convex optimization problem;
- next, use an online learning algorithm (e.g., online gradient descent) to obtain low regret for this problem;
- finally, obtain a single recommended prediction $\hat{w}$ whose excess risk $F(\hat{w}) - F(w^*)$ is approximately bounded by the regret (averaged over rounds) of the online learning algorithm; here, the bound on the excess risk will hold either in expectation or with high probability.

To realize the first step, for each $t \in [T]$, we define the loss function $\ell_t(w) = f(w, Z_t)$. We may then use an OCO algorithm to obtain low regret against any comparator $u \in V$, i.e., to ensure that

$$R_T(u) := \sum_{t=1}^{T} f(w_t, Z_t) - \sum_{t=1}^{T} f(u, Z_t) \tag{1}$$

is not too large.

2 Online-to-batch conversion

Suppose that an online learning algorithm that plays $w_1, \dots, w_T$ against the sequence $Z_1, \dots, Z_T$ obtains regret $R_T(u)$ against action $u \in V$.[1] We will prove that the simple average $\bar{w}_T := \frac{1}{T}\sum_{t=1}^{T}$ obtains low excess risk relative to $u \in V$ whenever $R_T(u)$ is small.

We will derive an in-expectation bound using elementary arguments and then a high probability bound using a more sophisticated martingale-based argument.

2.1 An in-expectation guarantee

To introduce the main ideas in the simplest way possible, in this subsection we assume that Learner is deterministic. That is, given the previous observations $Z_1, Z_2, \dots, Z_{t-1}$, Learner's action w_t is deterministic. Using ideas from the next subsection, Section 2.2, it is not difficult to extend the ideas here to randomized learning strategies.

Theorem 1. *Assume that $Z_1, Z_2, \dots, Z_T$ are i.i.d. according to distribution P. In the setting of OCO, suppose Learner is deterministic and plays actions $w_1, w_2, \dots, w_T$ against loss vectors of the form $\ell_t(w) = f(w, Z_t)$. For any $u \in V$, let $R_T(u)$ be Learner's regret against action u, defined as in (1).*

Then, for all $u \in V$,

$$\mathsf{E}\left[F(\bar{w}_T)\right] \le \mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T} F(w_t)\right] \le F(u) + \frac{\mathsf{E}\left[R_T(u)\right]}{T}. \tag{2}$$

The second inequality actually holds even without any convexity assumptions; of course, we *do* want the regret $R_T(u)$ to be sublinear. The first inequality requires F to be convex, for which it suffices for f to be convex in its first argument.

Proof (of Theorem 1). For the first inequality in (2), use the convexity of F and Jensen's inequality.

We now establish the second inequality. Let $u \in V$ be an arbitrary, fixed action. Then

$$\sum_{t=1}^{T} f(w_t, Z_t) = \sum_{t=1}^{T} f(u, Z_t) + R_T(u).$$

Also, as u is fixed, we have $\mathsf{E}[f(u, Z_t)] = F(u)$.

Next, for any $t \in [T]$, observe that

$$\begin{aligned}\mathsf{E}[f(w_t, Z_t)] &= \mathsf{E}\big[\mathsf{E}[f(w_t, Z_t) \mid Z_1, Z_2, \dots, Z_{t-1}]\big] \\ &= \mathsf{E}[F(w_t)],\end{aligned}$$

where the second equality follows because the action w_t is fixed when conditioning on $Z_1, \dots, Z_{t-1}$.

Therefore,

$$\frac{1}{T}\sum_{t=1}^{T} \mathsf{E}\left[F(w_t)\right] \le F(u) + \frac{\mathsf{E}\left[R_T(u)\right]}{T}.$$

□

[1] Note that $R_T(u)$ is a random variable by way of its dependence on $Z_1, \dots, Z_T$ and Learner's randomization (if any).

2.2 High probability bound

In order to obtain a high probability bound, we will develop some machinery to analyze stochastic processes.

Let $X_1, X_2, \ldots, X_T$ be a stochastic process for which each X_t is deterministic given a history H_t. Informally, the history can be thought of as "everything that has happened until the end of round t."[2]. We call the sequence of histories $(H_t)_{t\in[T]}$ a *filtration*.

Definition 1 (Martingale)**.** Let the sequence $(X_t)_{t\in[T]}$ be as above. We say $X_1, X_2, \ldots, X_T$ is a *martingale* adapted to the filtration $(H_t)_{t\in[T]}$ if for all $t \in [T]$:

- $\mathsf{E}[|X_t|] < +\infty$;
- $\mathsf{E}[X_t \mid H_{t-1}] = X_{t-1}$.

Similar to $X_1, X_2, \ldots, X_T$ above, let $Y_1, Y_2, \ldots, Y_T$ be a stochastic process for which each Y_t is deterministic given a history H_t. We say $Y_1, X_2, \ldots, Y_T$ is a *martingale difference sequence* adapted to the filtration $(H_t)_{t\in[T]}$ if for all $t \in [T]$:

- $\mathsf{E}[|Y_t|] < +\infty$;
- $\mathsf{E}[Y_t \mid H_{t-1}] = 0$.

Gambling offers an excellent way to illustrate martingales.

[2]Formally, each H_t is a σ-algebra, and we have $H_0 \subset H_1 \subset \ldots, \subset H_T$. For our purposes here, the informal notion of H_t as history will be enough to understand the main ideas.

Example 1 (Gambling)**.** Suppose a gambler starts with some positive wealth $X_1 = Y_1$. In each of a sequence of rounds $t = 2, 3, \ldots, T$, the gambler bets all of her wealth X_t on a fair coin coming out Heads, and then the coin is flipped:

- if the outcome is Heads, she wins $Y_t = X_{t-1}$ and hence doubles her wealth;
- if the outcome is Tails, she wins $Y_t = -X_{t-1}$ (i.e., she loses X_{t-1}, which is all of her wealth!).

The gambler's wealth is then updated according to $X_t = X_{t-1} + Y_t$, which may be rewritten as

$$X_t = \sum_{s=1}^{t} Y_s.$$

As we will now confirm, the sequence $(X_t)_{t\in[T]}$ is a martingale, and the sequence $(Y_t)_{t\in[T]}$ is a martingale difference sequence. First, $\mathsf{E}[|X_t|] < +\infty$ since $X_t \in [0, 2^{t-1}X_1]$. In addition, we clearly must have $\mathsf{E}[|Y_t|] < +\infty$ since $|Y_t| = X_{t-1}$. Next, observe that

$$\begin{aligned}
\mathsf{E}[X_t \mid H_{t-1}] &= \mathsf{E}\left[\sum_{s=1}^{t} Y_s \mid H_{t-1}\right] \\
&= \sum_{s=1}^{t-1} Y_s + \mathsf{E}[Y_t \mid H_{t-1}] \\
&= X_{t-1} + \frac{1}{2}\cdot X_{t-1} + \frac{1}{2}\cdot(-X_{t-1}) \\
&= X_{t-1}.
\end{aligned}$$

Hence, $(X_t)_{t\in[T]}$ is a martingale.

Next, let us confirm that $(Y_t)_{t\in[T]}$ is a martingale difference sequence. We actually already verified that $\mathsf{E}[Y_t \mid H_{t-1}] = 0$ in the previous sequence of equalities. Even so, it is instructive to verify this fact a different way. Observe that $Y_t = X_t - X_{t-1}$. Therefore,

$$\begin{aligned}
\mathsf{E}[Y_t \mid H_{t-1}] &= \mathsf{E}[X_t - X_{t-1} \mid H_{t-1}] \\
&= \mathsf{E}[X_t \mid H_{t-1}] - X_{t-1} \\
&= 0,
\end{aligned}$$

where the second equality uses the fact that $(X_t)_{t\in[T]}$ is a martingale. Therefore, as one might expect, if Y_t is a difference of successive terms of a martingale, then $(Y_t)_{t\in[T]}$ is a martingale difference sequence.

The following concentration inequality is known as Hoeffding-Azuma's inequality, also commonly referred to as Azuma's inequality.

Theorem 2. *Let $Y_1, Y_2, \ldots, Y_T$ be a martingale difference sequence adapted to the filtration $(H_t)_{t\in[T]}$. Assume that there are stochastic processes $(A_t)_{t\in[T]}$ and $(B_t)_{t\in[T]}$ and positive constants $c_1, c_2, \ldots, c_T$ such that, for all $t \in [T]$, with probability 1:*

- *we have A_t and B_t are deterministic given H_{t-1};*
- *$A_t \le Y_t \le B_t$ and $B_t - A_t \le c_t$.*

Then for all $\varepsilon > 0$,

$$\Pr\left(\sum_{t=1}^{T} Y_t \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{t=1}^{T} c_t^2}\right). \tag{3}$$

We will only need to use a specialization of the above theorem for which $c_t = c$ for all $t \in [T]$, in which (3) specializes to

$$\Pr\left(\sum_{t=1}^{T} Y_t \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{Tc^2}\right). \tag{4}$$

All the tools are in place for a high probability online-to-batch conversion.

Theorem 3. *Take the setting of Theorem 1 but with restriction that $f(w, Z) \in [0, b]$ for all $w \in V$ and $Z \in \mathcal{Z}$. Then for all $u \in V$, with probability at least $1 - \delta$,*

$$F(\bar{w}_T) \leq \frac{1}{T}\sum_{t=1}^{T} F(w_t) \leq F(u) + \frac{R_T(u)}{T} + b\sqrt{\frac{2\log\frac{1}{\delta}}{T}}. \tag{5}$$

Proof. Just like in Theorem 1, the first inequality in (5) is from Jensen's inequality. The main work is establishing the second inequality.

For each $t \in [T]$, let H_t denote the history up until time t (which includes $Z_1, \ldots, Z_t$ and any randomization employed by Learner until the end of round t). In addition, define

$$\begin{aligned} Y_t &:= f(u, Z_t) - f(w_t, Z_t) - \mathsf{E}\left[f(u, Z_t) - f(w_t, Z_t) \mid H_{t-1}\right] \\ &= f(u, Z_t) - f(w_t, Z_t) - (F(u) - F(w_t)) \end{aligned}$$

The idea of the proof is to show that $Y_1, Y_2, \ldots, Y_T$ is a martingale difference sequence, to control its sum via Hoeffding-Azuma's inequality, and then to relate this sum to the excess risk.

First, for each $t \in [T]$ it holds that $\mathsf{E}\left[Y_t \mid H_{t-1}\right] = 0$. Moreover, since $f(w, Z) \in [0, b]$ for all $w \in V$ and $Z \in \mathcal{Z}$, it holds that $|Y_t| \leq 2b$ and hence $(Y_t)_{t\in[T]}$ is a martingale difference sequence.

In order to apply Theorem 2, recalling that w_t is deterministic given H_{t-1}, observe that we can take $A_t = -b - (F(u) - F(w_t))$ and $B_t = b - (F(u) - F(w_t))$; hence, we can take $c_t = 2b$. Applying Theorem 2, we see that

$$\Pr\left(\sum_{t=1}^{T} Y_t \geq \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2}{2b^2 T}\right).$$

Therefore, with probability at least $1 - \delta$,

$$\sum_{t=1}^{T}(f(u, Z_t) - f(w_t, Z_t)) - \sum_{t=1}^{T}(F(u) - F(w_t)) \leq b\sqrt{2T\log\frac{1}{\delta}}.$$

Rearranging, with probability at least $1 - \delta$,

$$\begin{aligned} \sum_{t=1}^{T} F(w_t) &\leq TF(u) + \sum_{t=1}^{T}(f(w_t, Z_t) - f(u, Z_t)) + b\sqrt{2T\log\frac{1}{\delta}} \\ &= TF(u) + R_T(u) + b\sqrt{2T\log\frac{1}{\delta}}. \end{aligned}$$

The result follows by dividing through by T.

□

Bibliographical notes

1. For the particular form of Hoeffding-Azuma's inequality (Theorem 2) stated here, see Theorem 20.8 of Roch (2018).

2. The first work to provide a general study of online-to-batch conversions for general losses is Cesa-Bianchi et al. (2004); Theorem 3 is essentially from this fundamental paper. Their focus was on converting bounds on the cumulative loss (as compared to regret) of an online learning algorithm to bounds on its risk (as compared to excess risk). A key idea from this work was developed in a more specialized setting by (Blum et al., 1999, see Theorem 4).

3. Zhang (2005) showed, for the setting of regression with squared loss, how to obtain high-probability online-to-batch conversions with faster rates of convergence, paying an additional price in terms of T as low as $O\left(\frac{\log T}{T}\right)$ instead of the price of $O\left(\sqrt{1}T\right)$ present in Theorem 3. A few years later, Kakade and Tewari (2008) demonstrated that this lower price could be achieved generally under the assumption of strongly convex losses that also are Lipschitz (essentially, a bounded gradient assumption). More recently, Mehta (2017), leveraging ideas developed in Van Erven et al. (2015), showed that the lower price can still be enjoyed under exp-concavity.

References

Avrim Blum, Adam Kalai, and John Langford. Beating the hold-out: Bounds for k-fold and progressive cross-validation. In *Proceedings of the twelfth annual conference on Computational learning theory*, pages 203–208, 1999.

Nicolo Cesa-Bianchi, Alex Conconi, and Claudio Gentile. On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057, 2004.

Sham M Kakade and Ambuj Tewari. On the generalization ability of online strongly convex programming algorithms. *Advances in Neural Information Processing Systems*, 21, 2008.

Nishant Mehta. Fast rates with high probability in exp-concave statistical learning. In *Artificial Intelligence and Statistics*, pages 1085–1093. PMLR, 2017.

Sebastien Roch. Math 733-734: Theory of Probability, Notes 20: Azuma's inequality. https://people.math.wisc.edu/~roch/grad-prob/gradprob-notes20.pdf, 2018. [Online; accessed 12-February-2023].

Tim van Erven, Peter D Grünwald, Nishant A Mehta, Mark D Reid, and Robert C Williamson. Fast rates in statistical and online learning. *Journal of Machine Learning Research*, 16:1793–1861, 2015.

Tong Zhang. Data dependent concentration bounds for sequential prediction algorithms. In *Learning Theory: 18th Annual Conference on Learning Theory, COLT 2005, Bertinoro, Italy, June 27-30, 2005. Proceedings 18*, pages 173–187. Springer, 2005.