Machine Learning Theory (CSC 431/531) - Lecture 20

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 $f \colon \mathcal{Z} \times V \to \mathbb{R}$ that is convex in its second argument. The goal is to find some $w \in V$ which minimizes the objective

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

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 $f(z, w) = f((x, y), w) = \ell(y, \langle w, x \rangle)$ for some loss function $\ell \colon \mathcal{Y} \times \mathbb{R} \to \mathbb{R}$ that is convex in its second 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 $\ell_t(w) = f(Z_t, w)$. We may then use an OCO algorithm to obtain low regret against any comparator $u \in V$, i.e., to ensure that

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

is not too large.

2 Online-to-batch conversion

Suppose that an online learning algorithm that plays $w_1, \ldots, w_T$ against the sequence $Z_1, \ldots, Z_T$ obtains regret $R_T(u)$ against action $u \in V$.[1] We will prove that the simple average $\bar{w}_T := \frac{1}{T} \sum_{t=1}^T w_t$ 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 $Z_1, Z_2, \ldots, Z_{t-1}$, 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 $Z_1, Z_2, \ldots, Z_T$ are i.i.d. according to distribution P. In the setting of OCO, suppose Learner is deterministic and plays actions $w_1, w_2, \ldots, w_T$ against loss vectors of the form $\ell_t(w) = f(Z_t, w)$. For any $u \in V$, let $R_T(u)$ be Learner's regret against action u, defined as in (1).*

Then, for all $u \in V$,

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

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 second 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

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

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

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

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

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

Therefore,

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

□

[1] Note that $R_T(u)$ is a random variable by way of its dependence on $Z_1, \ldots, Z_T$ 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 $X_0, X_1, \dots, X_T$ 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."

Definition 1 (Martingale)**.** Let the sequence $(X_t)_{t\in[T]}$ be as above. We say $X_1, X_2, \dots, X_T$ is a *martingale* with respect to $(H_t)_{0\le t\le T}$ if:

- $\mathsf{E}[|X_t|] < +\infty$ for all $t = 0, 1, \dots, T$;
- $\mathsf{E}[X_t \mid H_{t-1}] = X_{t-1}$ for all $t \in [T]$.

Let $Y_1, Y_2, \dots, Y_T$ be a stochastic process for which each Y_t is deterministic given a history H_t. We say $Y_1, Y_2, \dots, Y_T$ is a *martingale difference sequence* with respect to $(H_t)_{t\in[T]}$ if for all $t \in [T]$:

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

It is easy to verify that if $(X_t)_{t\in[T]}$ is a martingale, then the sequence $(Y_t)_{t\in[T]}$ defined by $Y_t = X_t - X_{t-1}$ 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 $Y_1, Y_2, \dots, Y_T$ be a martingale difference sequence with respect to $(H_t)_{t\in[T]}$. Assume that there are stochastic processes $(A_t)_{t\in[T]}$ and $(B_t)_{t\in[T]}$ and positive constants $c_1, c_2, \dots, c_T$ such that, for all $t \in [T]$, with probability 1:*

- *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 $\varepsilon > 0$,

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

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

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

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

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

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

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, including Learner's selection of w_{t+1}). In addition, define

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

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 $f(Z, w) \in [0, b]$ 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},[2] 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 $1 - \delta$,

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

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

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

The result follows by dividing through by T.

□

[2] Remember that H_{t-1} includes any randomization Learner may have used to select w_t.