CA684 Jia Yuan Yu

# 6: Online learning

DCU March 21, 2014

These notes are based on [CBL06].

Last lecture, we saw an efficient implementation of a learning algorithm (Kalman filter) for state-estimation when observations are revealed incrementally. This week, we look at other incremental learning algorithms in a different setting.

## 1 Online learning

In this setting, we consider data sequences $\{x_t, y_t\}$ that are not modeled as and i.i.d. sequence: we allow every possible sequence. Online learning describes a number of problems where more data is obtained over time, which can be incorporated in our solution algorithm. Examples of such problem include classification, regression, and prediction with expert advice. We will introduce the prediction problem next.

### 1.1 The setting

We consider the setting of prediction with expert advice. We assume that there are $d$ experts. Define the vector $x_t = (x_t^1, \ldots, x_t^d) \in \mathcal{X}^d$, where $x_t^i \in \mathcal{X}$ is the advice of expert $i$ at time $t$. The set $\mathcal{X}$ is called the decision space and assumed convex. In other words, expert $i$ gives the sequence of advices:

$$x_1^i, x_2^i, \ldots$$

There is an unknown sequence $y_1, y_2, \ldots$, whose elements take values in the outcome space $\mathcal{Y}$. Consider a sequential decision (prediction) problem over rounds $t = 1, 2, \ldots$ At round $t$:

- decision-maker observes $x_t \in \mathcal{X}$,
- decision-maker outputs a prediction $\hat{p}_t$ taking values in a decision space $\mathcal{X}$, which may be equal to $\mathcal{Y}$,
- the true $y_t \in \mathcal{Y}$ is revealed,
- the decision-maker records a loss $\ell(\hat{p}_t, y_t)$, where $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a known loss function.

Observe that the sequence $\{x_t, y_t\}$ can be deterministic, random, or non-oblivious adversarial ($x_t, y_t$ may be functions of $\hat{p}_1, \ldots, \hat{p}_{t-1}$). Such sequences are also called individual sequences. The setting is called adversarial. A decision-making policy is a sequence of functions $\{\hat{p}_1, \hat{p}_2, \ldots\}$, where each $\hat{p}_t$ is a function of past observations.

## 1.2 Performance metric

Previously, we measured the performance of our classifiers and estimators via such metrics as estimation error, approximation error, probability of error, and mean-squared error. In the online learning setting, the performance metric is the regret with respect to expert advices.

First, we define the decision-maker's cumulative loss after $n$ rounds:

$$\hat{L}_n = \sum_{t=1}^{n} \ell(\hat{p}_t, y_t).$$

Is this a random variable? Depends on $\hat{p}_t$, but we consider $\{x_t, y_t\}$ to be an arbitrary deterministic sequence. We also define the loss of each expert $i$:

$$L_{i,n} = \sum_{t=1}^{n} \ell(x_t^i, y_t).$$

The cumulative regret with respect to an expert $i$ at round $n$ is

$$R_{i,n} = \sum_{t=1}^{n} \ell(\hat{p}_t, y_t) - \sum_{t=1}^{n} \ell(x_t^i, y_t),$$

which is the regret for not following expert $i$'s advice at every round until $n$. A policy $\{\hat{p}_t\}$ has vanishing average regret if

$$\lim_{n\to\infty} \frac{1}{n} \left( \hat{L}_n - \min_{i=1,\dots,d} L_{i,n} \right) = 0$$

uniformly over all sequences $\{x_t, y_t\}$.

## 1.3 Weighted average prediction

Consider a policy where each $\hat{p}_t$ is a convex combination of expert advices $x_{1,t}, \dots, x_{d,t}$ at the same round:

$$\hat{p}_t = \sum_{i=1}^{d} \frac{w_{i,t-1}}{\sum_{k=1}^{d} w_{k,t-1}} x_{i,t},$$

where $\{w_{i,t-1}\}$ are parameters to compute from the data. A natural approach is to assign higher weights to an expert $i$ with small loss $L_{i,t-1}$ or high regret $R_{i,t-1}$. For instance, the exponentially weighted average forecaster with parameter $\eta$ is

$$\hat{p}_t = \sum_{i=1}^{d} \frac{e^{-\eta L_{i,t-1}}}{\sum_{k=1}^{d} e^{-\eta L_{k,t-1}}} x_{i,t}, \quad t = 2, 3, \dots,$$

where $\hat{p}_1$ is set arbitrarily.

**Theorem 1.1.** *Suppose that $\ell$ is convex in the first argument and takes values in* $[0,1]$. *Then, for every $n$, every fixed $\eta > 0$, and all $\{x_t, y_t\}$, the regret of the exponentially weighted average forecaster satisfies*

$$\hat{L}_n - \min_{i=1,\dots,d} L_{i,n} \le \frac{\log d}{\eta} + \frac{n}{2}\eta.$$

If we allow $\eta$ to be a function of $n$, we can set $\eta = \sqrt{2\log d/n}$ to optimize the above bound. We can even replace $\eta$ in the algorithm by $\eta_t = \sqrt{8\log d/t}$ to obtain a bound that holds uniformly over time $n$, cf. [CBL06, Chapter 2.3].

## 1.4 What about nonconvex loss $\ell$?

In classification, we have $\mathcal{Y} = \{0,1\}$ and use a loss function of the form

$$\ell(p(x), y) = 1_{[p(x)\neq y]}.$$

This is not convex. In such situations, for every deterministic policy $\{p_t\}$, we can construct a sequence $\{y_t\}$ such that the policy errs at every time instant $t$. Can we achieve vanishing average regret in this case? Yes: with randomization.

Consider a policy where the decision-maker

- computes a probability distribution $\vec{v}_t = (v_{1,t}, \dots, v_{d,t})$ over the set of experts,
- follows a random expert $I_t$, where $\mathbb{P}(I_t = i) = v_{i,t}$ for all $i = 1, \dots, d$.

Hence, we have a random $\hat{L}_n$ and a random cumulative regret wrt the best expert:

$$\sum_{t=1}^{n} \ell(x_t^{I_t}, y_t) - \min_{i=1,\dots,d} \sum_{t=1}^{n} \ell(x_t^i, y_t),$$

Note that if $y_t$ depends on $I_1, \dots, I_{t-1}$ (i.e., a nonoblivious opponent), then it is also random.

## 1.5 Follow the perturbed leader

The exponentially weighted average forecaster works in the nonconvex setting if the normalized weights $\{\hat{w}_t\}$ are taken as the probabilities $v_t$. However, another simple policy works too.

Let $\nu_1, \nu_2, \dots$ be independent random variables. The follow-the-perturbed-leader policy is:

$$I_t \in \arg\min_{i=1,\dots,d} L_{i,t-1} + \nu_t^i, \quad t = 2, 3, \dots,$$

with an arbitrary $I_1$.

**Theorem 1.2.** *Suppose that the opponent is oblivious. If each $\nu_t^i$ uniformly distributed on $[0, \Delta_t]$ and $\Delta_t = \sqrt{td}$, then for all $n$ and uniformly over all individual sequences $\{x_t, y_t\}$, we have*

$$\sum_{t=1}^{n} \mathbb{E}_t \ell(x_t^{I_t}, y_t) - \min_{i=1,\ldots,d} \sum_{t=1}^{n} \ell(x_t^i, y_t) \le 2\sqrt{2dn},$$

*where $\mathbb{E}_t$ is the expectation over $I_t$.*

This result is another notion of consistency called (weak) Hannan consistency.

We can tweak this policy to work even against a nonoblivious opponent [CBL06, Remark 4.2].

# 2 Bandit problems

There are a number of versions of the so-called bandit problem: Markovian (Gittins), stochastic (Lai and Robbins), and adversarial (or nonstochastic). Imagine $d$ slot machines and a player who plays sequentially one coin at a time. What policy should he or she follow to maximize profit? This policy should certainly tradeoff exploration and exploitation.

## 2.1 Adversarial bandits

The bandit problem is similar to the problem of individual sequence prediction. The difference is that after making a prediction $I_t$, the decision-maker only observes $\ell(x_t^{I_t}, y_t)$ instead of $y_t$.

Let $\eta$ and $\{\gamma_t\}$ denote fixed parameters. Consider a version of the exponentially weighted average predictor that picks $I_t$ randomly according to a probability distribution $\vec{v}_t = (v_{1,t}, \ldots, v_{d,t})$ over the set of experts defined as follows (for $t = 2, 3, \ldots$, and arbitrary $\vec{v}_1$):

$$\begin{aligned}
v_{i,t} &= (1-\gamma_t)\frac{\exp(-\eta \tilde{L}_{i,t-1})}{\sum_{k=1}^{d} \exp(-\eta \tilde{L}_{k,t-1})} + \gamma_t/d, \quad i = 1, \ldots, d, \\
\tilde{L}_{k,t-1} &= \sum_{m=1}^{t-1} \tilde{\ell}(x_m^k, y_m), \quad k = 1, \ldots, d, \\
\tilde{\ell}(x_m^k, y_m) &= \begin{cases} \ell(x_m^k, y_m)/v_{k,m}, & \text{if } I_m = k, \\ 0, \quad \text{otherwise.} \end{cases} \\
&= \frac{\ell(x_m^k, y_m)}{v_{k,m}} 1_{[I_m = k]}, \quad m = 1, \ldots, t-1.
\end{aligned}$$

Observe that $\tilde{\ell}(x_m^k, y_m)$ is an unbiased estimator of $\ell(x_m^k, y_m)$, which is potentially unseen.

**Theorem 2.1.** *Suppose that* $\sum_{t=1}^{n} 1/\gamma_t^2 \in o(n^2/\log n)$*. For every fixed* $\eta > 0$ *and* $n$*, we have*

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{1}{n}\left[\sum_{t=1}^{n} \ell(x_t^{I_t}, y_t) - \min_{i=1,\ldots,d} \sum_{t=1}^{n} \ell(x_t^i, y_t)\right] = 0\right) = 1.$$

In particular, the sequence $\gamma_t = t^{-1/3}$ works for the above theorem. This property is called (strong) Hannan consistency.

# 3 What to do during reading week?

"An overview of statistical learning theory" by Vapnik. This will connect SVMs to VC theory and neural nets seen in class.

# References

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