CSIE5002 Prediction, learning, and games

Lecture 9: Decision theoretic online learning

Yen-Huan Li (yenhuan.li@csie.ntu.edu.tw)

20.05.2019

Department of Computer Science and Information Engineering
National Taiwan University

Abstract

We have studied learning with expert advice with mixable (or exp-concave) losses. What if the loss is not mixable?

Recommended reading

- Y. Freund & R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting.

- S. Arora *et al.* 2012. The multiplicative weights update method: A meta-algorithm and applications.

Table of contents

Decision theoretic online learning

Decision theoretic online learning

Protocol. (Decision theoretic online learning) Let $T \in \mathbb{N}$. Let $\mathcal{A} \coloneqq \{ 1, \ldots, K \}$ for some $K \in \mathbb{N}$. Let the initial cumulative (expected) loss $L_0 = 0$. For every $1 \le t \le T$, the following happen sequentially.

1. $\textsc{Learner}$ announces a probability distribution $\gamma_t \in \Delta(\mathcal{A})$.
2. $\textsc{Reality}$ announces a loss vector $\omega_t \in [0, 1]^K$.
3. Update $L_t \leftarrow L_{t-1} + \lambda(\omega_t, \gamma_t)$, where the *mixture loss* λ is defined as $\lambda(\omega_t, \gamma_t) \coloneqq \langle \omega_t, \gamma_t \rangle$.

Question. How do we interpret the protocol?

Y. Freund & R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1997.

Regret(s)

We define two notions of regrets:

$$R_T(\gamma) \coloneqq \sum_{t=1}^{T} \lambda(\omega_t, \gamma_t) - \sum_{t=1}^{T} \lambda(\omega_t, \gamma), \quad \forall \gamma \in \Delta(\mathcal{A}),$$

$$R_T(i) \coloneqq \sum_{t=1}^{T} \langle \omega_t, \gamma_t \rangle - \sum_{t=1}^{T} \omega_t(i), \quad \forall 1 \le i \le K,$$

where $\omega_t(i)$ denotes the i-th entry of ω_t.

Proposition 1. It holds that

$$\max_{\gamma \in \Delta(\mathcal{A})} R_T(\gamma) = \max_{1 \le i \le K} R_T(i),$$

Proof of Proposition 1

Proof. (Proposition 1) It suffices to prove that

$$\min_{\gamma \in \Delta(\mathcal{A})} \langle s, \gamma \rangle = \min_{1 \leq i \leq K} s(i),$$

where $s \coloneqq \sum_{t=1}^{T} \omega_t$. Notice that

$$\langle s, \gamma \rangle \geq \min_{1 \leq i \leq K} s(i), \quad \forall \gamma \in \Delta(\mathcal{A}).$$

Also notice that

$$\min_{1 \leq i \leq K} s(i) = \langle s, e_{i^\star} \rangle \geq \min_{\gamma \in \Delta(\mathcal{A})} \langle s, \gamma \rangle ,$$

where $i^\star \in \arg\min_{1 \leq i \leq K} s(i)$. The proposition follows.

Perspective of AA

Theorem 1. The mixture loss is (c, η)-mixable, if and only if

$$c \geq \frac{\eta}{K \log \frac{K}{K + \mathrm{e}^{-\eta} - 1}}.$$

Corollary 1. The AA achieves

$$\sum_{t=1}^{T} \lambda(\omega_t, \gamma_t) \leq \frac{\eta}{K \log \frac{K}{K + \mathrm{e}^{-\eta} - 1}} \min_{1 \leq i \leq K} \sum_{t=1}^{T} \omega_t(i) + \frac{\log K}{K \log \frac{K}{K + \mathrm{e}^{-\eta} - 1}}.$$

V. Vovk. A game of prediction with expert advice. 2011.

Algorithm. (Hedge) Let π_1 be the uniform distribution on $\mathcal{A} = \{ 1, \ldots, K \}$. For every $1 \leq t \leq T$, announce $\gamma_t = \pi_t$, and compute

$$\pi_{t+1}(i) \propto \pi_t(i) \mathrm{e}^{-\eta \omega_t(i)}, \quad \forall 1 \leq i \leq K,$$

for some $\eta > 0$.

Remark. Notice the hedge algorithm is not the AA.

Y. Freund & R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1997.

Regret of the hedge (1/2)

Theorem 2. The hedge algorithm achieves

$$\sum_{t=1}^{T} \lambda(\omega_t, \gamma_t) \leq \frac{\eta}{1 - \mathrm{e}^{-\eta}} \min_{1 \leq i \leq K} \sum_{t=1}^{T} \omega_t(i) + \frac{\log K}{1 - \mathrm{e}^{-\eta}}.$$

Lemma 1. It holds that

$$K \log \frac{K}{K + \mathrm{e}^{-\eta} - 1} = 1 - \mathrm{e}^{-\eta}, \quad \forall \eta > 0.$$

Remark. Therefore, the regret bound for the hedge is sub-optimal.

Y. Freund & R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1997.

V. Vovk. A game of prediction with expert advice. 2011.

Proof of Theorem 2 (1/2)

Proof. (Theorem 2) Equivalently, the hedge starts with a weight vector $w_1 = \pi_1$, and computes

$$\pi_t(i) = \frac{w_t(i)}{\sum_{i=1}^K w_t(i)}, \quad \forall 1 \le i \le K,$$
$$w_{t+1}(i) = w_t(i)\mathrm{e}^{-\eta\omega_t(i)}, \quad \forall 1 \le i \le K.$$

Then, we write

$$\begin{aligned}\sum_{i=1}^K w_{T+1}(i) &= \sum_{i=1}^K w_T(i)\mathrm{e}^{-\eta\omega_T(i)} \le \sum_{i=1}^K w_T(i)\left[1 - \left(1 - \mathrm{e}^{-\eta}\right)\omega_T(i)\right] \\ &= \left[\sum_{i=1}^K w_T(i)\right]\left[1 - \left(1 - \mathrm{e}^{-\eta}\right)\lambda(\omega_T, \gamma_T)\right].\end{aligned}$$

Y. Freund & R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1997.

Proof of Theorem 2 (2/2)

Proof continued. (Theorem 2) We obtain

$$\begin{aligned}\sum_{i=1}^{K} w_{T+1}(i) &\leq \prod_{t=1}^{T} \left[1 - \left(1 - \mathrm{e}^{-\eta}\right) \lambda(\omega_t, \gamma_t)\right] \\ &\leq \prod_{t=1}^{T} \mathrm{e}^{-\left(1-\mathrm{e}^{-\eta}\right)\lambda(\omega_t,\gamma_t)} \\ &= \mathrm{e}^{-\left(1-\mathrm{e}^{-\eta}\right)\sum_{t=1}^{T} \lambda(\omega_t,\gamma_t)}.\end{aligned}$$

It remains to notice that

$$w_1(i)\mathrm{e}^{-\eta \sum_{t=1}^{T} \omega_t(i)} \leq \sum_{i=1}^{K} w_{T+1}(i).$$

Y. Freund & R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1997.

Regret of the hedge (2/2)

Corollary 2. The hedge achieves

$$R_T(i) \leq \sqrt{2T \log K} + \log K, \quad \forall 1 \leq i \leq K.$$

Proof sketch. Set

$$\eta = \log\left(1 + \sqrt{\frac{2 \log K}{T}}\right).$$

Y. Freund & R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1997.

Regret of the hedge via the mixability gap

Theorem 3. The hedge achieves

$$R_T(i) \le \sqrt{(1/2)T \log K}, \quad \forall 1 \le i \le K.$$

Lemma 2. It holds that

$$R_T(i) \le \frac{1}{\eta} \log \frac{1}{\pi_1(i)} + \sum_{t=1}^{T} \delta_t,$$

where δ_t denotes the *mixability gap*:

$$\delta_t \coloneqq \lambda(\omega_t, \gamma_t) - \left(\frac{-1}{\eta}\right) \log \sum_{i=1}^{K} \pi_t(i) \mathrm{e}^{-\eta \omega_t(i)}.$$

Proof of Lemma 2

Proof. (Lemma 2) We write

$$\lambda(\omega_t, \gamma_t) = \delta + \left(\frac{-1}{\eta}\right) \log \sum_{i=1}^{K} \pi_t(i) \mathrm{e}^{-\eta \omega_t(i)}.$$

It remains to notice

$$\begin{aligned} &\sum_{t=1}^{T} \left[\left(\frac{-1}{\eta}\right) \log \sum_{i=1}^{K} \pi_t(i) \mathrm{e}^{-\eta \omega_t(i)} \right] \\ &\quad = \frac{-1}{\eta} \log \sum_{i=1}^{K} \pi_1(i) \mathrm{e}^{-\eta \sum_{t=1}^{T} \omega_t(i)} \\ &\quad \le \frac{-1}{\eta} \log \left[\pi_1(i) \mathrm{e}^{-\eta \sum_{t=1}^{T} \omega_t(i)} \right], \quad \forall 1 \le i \le K. \end{aligned}$$

Proof of Theorem 3

Proof. (Theorem 3) Notice that $\omega_t(i) \in [0, 1]$ for all i and t. By Hoeffding's lemma, we have

$$\delta_t \leq \frac{\eta}{8}.$$

Then, we obtain

$$R_T(i) \leq \frac{1}{\eta} \log K + \frac{\eta T}{8}.$$

Choosing

$$\eta = 2\sqrt{\frac{2 \log K}{T}}$$

competes the proof.

Exponentiated gradient method

Online convex optimization on the probability simplex

Protocol. (Online convex optimization on the probability simplex)
Let $T \in \mathbb{N}$. Let the initial cumulative loss $L_0 = 0$. For every $1 \leq t \leq T$, the following happen sequentially.

1. $\textsc{Learner}$ announces $x_t \in \Delta \subset \mathbb{R}^n$.
2. $\textsc{Reality}$ announces a convex function $f_t : \Delta \to \mathbb{R}$.
3. Update the cumulative loss as $L_t \leftarrow L_{t-1} + f_t(x_t)$.

Definition. (Regret) The regret is standard:

$$R_T(x) := \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x), \quad \forall x \in \Delta.$$

Online gradient descent

Suppose for every $1 \le t \le T$ and $x \in \Delta$, there exists some $g_t(x) \in \mathbb{R}^n$ such that

$$g_t(x) \in \partial f_t(x), \quad \|g_t(x)\|_2 \le L_2.$$

Suppose we adopt the online gradient descent:

$$\tilde{x}_t \leftarrow x_{t-1} - \eta g_{t-1}(x_{t-1}),$$
$$x_t \leftarrow \arg\min_{x \in \Delta} \|x - \tilde{x}_t\|_2^2.$$

Recall the online gradient descent achieves

$$R_T(\gamma) \le L_2 \max_{x \in \Delta} \|x - x_1\|_2 \sqrt{T}.$$

Exponentiated gradient method (1/2)

Recall online convex optimization can be reduced to online linear optimization, as

$$R_T(x) \leq \sum_{t=1}^{T} \langle g_t(x_t), x_t \rangle - \sum_{t=1}^{T} \langle g_t(x_t), x \rangle, \quad \forall x \in \Delta.$$

Notice online linear optimization on the probability simplex is equivalent to decision theoretic online learning. Therefore, we can adopt the hedge.

J. Kivinen & M. K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. 1997.

Regret of the exponentiated gradient method (2/2)

Algorithm. (Exponentiated gradient method) Let x_1 be the uniform distribution. For every $2 \leq t \leq T$, announce $x_t = (x_t(i))_{1 \leq i \leq n}$ such that

$$x_t(i) \propto x_{t-1}(i) \mathrm{e}^{-\eta [g_{t-1}(x_{t-1})]_i}, \quad \forall 1 \leq i \leq n,$$

where $[g_{t-1}(x_{t-1})]_i$ denotes the i-th entry of $[g_{t-1}(x_{t-1})]$.

Theorem 4. Suppose that

$$\| g_t(x_t) \|_\infty \leq L_\infty, \quad \forall 1 \leq t \leq T.$$

Then, the exponentiated gradient method achieves

$$R_T(x) \leq L_\infty \sqrt{\frac{T \log n}{2}}, \quad \forall x \in \Delta.$$

Proof of Theorem 4

Proof. (Theorem 4) Recall that

$$R_T(x) \leq \frac{1}{\eta} \log n + \sum_{t=1}^{T} \delta_t,$$

where δ_t denotes the mixability gap in the t-th round. By Hoeffding's lemma, we have

$$\delta_t \leq \frac{\eta L^2}{8}, \quad \forall 1 \leq t \leq T.$$

Therefore, we write

$$R_T(x) \leq \frac{1}{\eta} \log n + \frac{\eta L^2 T}{8}, \quad \forall x \in \Delta.$$

Optimizing η completes the proof.

Online gradient descent vs. exponentiated gradient method (1/2)

Online gradient descent. Requires $\|g_t(x_t)\|_2 \leq L_2$, and achieves

$$R_T = O\left(L_2\sqrt{T\gamma}\right),$$

where $\gamma = O\left(1\right)$.

Exponentiated gradient method. Requires $\|g_t(x_t)\|_\infty \leq L_\infty$, and achieves

$$R_T = O\left(L_\infty\sqrt{T\log n}\right).$$

Online gradient descent vs. exponentiated gradient method (2/2)

Notice that

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty, \quad \forall x \in \mathbb{R}^n.$$

Then, we have

$$\frac{1}{\sqrt{\log n}} \leq \frac{L_2\sqrt{T}}{L_\infty\sqrt{T\log n}} \leq \sqrt{\frac{n}{\log n}}.$$

Observation. In terms of dependence on the dimension, the exponentiated gradient method can be significantly better or slightly worse than the online gradient descent.

Application: Online portfolio selection (1/)

Recall the online portfolio selection problem. We formulate it as online convex optimization on the probability simplex.

Protocol. (Online portfolio selection) Let $T \in \mathbb{N}$. Let the initial cumulative loss $L_0 = 0$. For every $1 \leq t \leq T$, the following happen sequentially.

1. $\textsc{Learner}$ announces $x_t \in \Delta \subset \mathbb{R}^n$.
2. $\textsc{Reality}$ announces $f_t : x \mapsto -\log \langle y_t, x \rangle$ for some $y_t \in \mathbb{R}^n$.
3. Update the cumulative regret: $L_t \leftarrow L_{t-1} + f_t(x_t)$.

Application: Online portfolio selection (2/)

Notice the loss function is convex, as

$$\nabla^2 f_t(x) = \frac{y_t y_t^{\mathrm{T}}}{\langle y_t, x \rangle^2} \geq 0, \quad \forall x \in \{ z \in \Delta \mid \langle y_t, z \rangle > 0 \}, 1 \leq t \leq T.$$

Then, we can reduce online portfolio selection to online linear optimization on the probability simplex, with the linear losses defined by

$$\nabla f_t(x_t) = -\frac{y_t}{\langle y_t, x_t \rangle}.$$

Exponentiated gradient method for online portfolio selection

Algorithm. (Exponentiated gradient method for online portfolio selection) Let x_1 be the uniform distribution. For every $1 \le t \le T$, announce x_t such that

$$x_t(i) \propto x_{t-1}(i) \mathrm{e}^{\frac{\eta y_{t-1}(i)}{\langle y_{t-1}, x_{t-1} \rangle}}, \quad \forall 1 \le i \le n.$$

Question. Does the exponentiated gradient method achieve a $O\left(\sqrt{T \log n}\right)$ regret?

D. P. Helmbold *et al.* On-line portfolio selection using multiplicative updates. 1998.

EG for OPS with the market variability condition (1/2)

Observation. Notice that

$$\|\nabla f_t(x)\|_\infty = \left\| \frac{y_t}{\langle y_t, x \rangle} \right\|_\infty$$

can be arbitrarily large on $\{ z \in \Delta \mid \langle y_t, z \rangle > 0 \}$. Therefore, the standard regret guarantee for the EG method does not apply.

Definition. (Market variability condition) We say the *market variability condition* holds, if for some $\varepsilon > 0$, we have

$$\varepsilon \leq y_t(i) \leq 1, \quad \forall 1 \leq t \leq t, 1 \leq i \leq n.$$

D. P. Helmbold *et al.* On-line portfolio selection using multiplicative updates. 1998.
E. Hazan & S. Kale. An online portfolio selection algorithm with regret logarithmic in price variation. 2015.

EG for OPS with the market variability condition (2/2)

Lemma 3. If the market variability condition holds for some $\varepsilon > 0$, then we have

$$\|\nabla f_t(x)\|_\infty \leq \frac{1}{\varepsilon}, \quad \forall x \in \Delta, 1 \leq t \leq T.$$

Remark. With the market variability condition, the standard guarantee for the exponentiated gradient method applies.

Theorem 5. Suppose the market variability condition holds for some $\varepsilon > 0$. Then, the exponentiated gradient method achieves an $O\left(\varepsilon^{-1}\sqrt{T \log n}\right)$ regret for online portfolio selection.

Some state-of-the-art results for online portfolio selection

MVC = market variability condition

1. Universal portfolio selection: Logarithmic regret, high computational complexity, no need for MVC.
2. Exponentiated gradient method: $O(\sqrt{T})$ regret, low computational complexity, requiring MVC.
3. Online Newton step: Logarithmic regret, moderate computational complexity, requiring MVC.

T. M. Cover & E. Ordentlich. Universal portfolios with side information. 1996.
D. P. Helmbold *et al.* On-line portfolio selection using multiplicative updates. 1998.
E. Hazan *et al.* Logarithmic regret algorithms for online convex optimization. 2007.
H. Luo *et al.* Efficient online portfolio with logarithmic regret. 2018.

Matrix exponentiated gradient method? (1/2)

Define the *spectraplex* as

$$\mathcal{D} \coloneqq \left\{ \rho \in \mathbb{R}^{d \times d} \;\middle|\; \rho \geq 0, \rho = \rho^{\mathrm{T}}, \mathrm{Tr}\rho = 1 \right\}.$$

Protocol. (Online convex optimization on the spectraplex) Let $T \in \mathbb{N}$. Let the initial cumulative loss $L_0 = 0$. For every $1 \leq t \leq T$, the following happen sequentially.

1. $\textsc{Learner}$ announces $\rho_t \in \mathcal{D}$.
2. $\textsc{Reality}$ announces a convex function $f_t : \mathcal{D} \to \mathbb{R} \cup \{ +\infty \}$.
3. Update the cumulative loss: $L_t \leftarrow L_{t-1} + f_t(\rho_t)$.

Matrix exponentiated gradient method? (2/2)

Observation. The spectraplex is a matrix counterpart of the probability simplex. In particular, for any density matrix, its vector of eigenvalues is a probability vector.

Question. How do we extend the exponentiated gradient method for the matrix case?

Matrix exponentiated gradient method

Algorithm. (Matrix exponentiated gradient method) Let $\rho_1 \in \mathcal{D}$ be non-singular. For every $1 \leq t \leq T$, announce

$$\rho_{t+1} \leftarrow C_t^{-1} \exp\left[\log\left(\rho_t\right) - \eta_t \nabla f_t\left(\rho_t\right)\right].$$

Idea.

1. Hedge is equivalent to *entropic mirror descent*.
2. Entropic mirror descent can be directly extended for the matrix case using the *quantum relative entropy*.

K. Tsuda *et al.* Matrix exponentiated gradient updates for on-line learning and Bregman projection. 2005.

Conclusions

Conclusions

- We have studied the problem of decision theoretic online learning and the hedge algorithm.

- The regret analysis for the hedge algorithm is similar to that for the AA. The key difference is for the former, the mixability gaps are non-zero.

- Online convex optimization on the probability simplex can be solved via the exponentiated gradient method, or the hedge for the associated online linear optimization problem.

Next lecture

- Second-order & quantile regret bounds for decision theoretic online learning.

- Shifting regret.