Sequential learning – Lesson 2

Pierre Gaillard

January 21, 2022

INRIA

Reminder from last week

Randomized predictions

(Θ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$)

The Gradient Trick and EG

(simplex decision set $\Theta = \Delta_K$, convex and differentiable losses)

Online gradient descent

(convex differentiable Lipschitz losses, convex and compact decision set Θ)

Online Mirrored Descent

Setting of an online learning problem/online convex optimization

At each time step $t = 1, \dots, T$

- the player observes a context $x_t \in \mathcal{X}$ (optional step)
- the player chooses an action $\theta_t \in \Theta$ (compact decision/parameter set);
- the environment chooses a loss function $\ell_t : \Theta \to [0, 1]$;
- the player suffers loss $\ell_t(\theta_t)$ and observes
  - the losses of every actions: $\ell_t(\theta)$ for all $\theta \in \Theta$ $\quad\rightarrow\quad$ full-information feedback
  - the loss of the chosen action only: $\ell_t(\theta_t)$ $\quad\rightarrow\quad$ bandit feedback.

The goal of the player is to minimize his cumulative loss:

$$\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t)\,.$$

Last week

The Exponentially Weighted Average (EWA) forecaster

$$p_t(k) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(k)}}{\sum_{j=1}^{K} e^{-\eta \sum_{s=1}^{t-1} g_s(j)}} \tag{EWA}$$

achieves a cumulative regret $R_T \lesssim \sqrt{T \log K}$ when the set of actions is the K-dimensional simplex and for linear losses $\ell_t(p) = p^\top g_t$ with $g_t \in [-1, 1]^K$.

For prediction with expert advice, the same upper-bound holds for EWA played with loss vectors $g_t = (\ell(x_t(1), y_t), \dots, \ell(x_t(K), y_t))$ for convex loss functions ℓ.

For exp-concave loss functions, we proved a constant regret.

Last week

The Exponentially Weighted Average (EWA) forecaster

$$p_t(k) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(k)}}{\sum_{j=1}^K e^{-\eta \sum_{s=1}^{t-1} g_s(j)}} = \frac{p_{t-1}(k) e^{-\eta g_{t-1}(k)}}{\sum_{j=1}^K p_{t-1}(j) e^{-\eta g_{t-1}(j)}} \tag{EWA}$$

achieves a cumulative regret $R_T \lesssim \sqrt{T \log K}$ when the set of actions is the K-dimensional simplex and for linear losses $\ell_t(p) = p^\top g_t$ with $g_t \in [-1, 1]^K$.

For prediction with expert advice, the same upper-bound holds for EWA played with loss vectors $g_t = (\ell(x_t(1), y_t), \ldots, \ell(x_t(K), y_t))$ for convex loss functions ℓ.

For exp-concave loss functions, we proved a constant regret.

This week

We will generalize the results of last week to **non-linear loss functions** ℓ_t:

1. in the simplest case of finite $\Theta = \{1, \ldots, K\}$ with arbitrary bounded loss functions.
2. in the case of the simplex $\Theta = \Delta_K$ with convex loss functions ℓ_t.
3. for any compact convex set Θ.

Reminder from last week

Randomized predictions

(Θ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$)

The Gradient Trick and EG

(simplex decision set $\Theta = \Delta_K$, convex and differentiable losses)

Online gradient descent

(convex differentiable Lipschitz losses, convex and compact decision set Θ)

Online Mirrored Descent

Need of randomized predictions

Setting: $\Theta \stackrel{\text{def}}{=} \{1, \ldots, K\}$ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$.

The player is restricted to play an action in Θ. The player cannot play convex combinations of the actions as it was done for prediction with expert advice.

Exemple: recommend movies to customers.

Proposition 1 (Need of a random strategy)

Any deterministic algorithm may incur a linear regret. In other words, we can find some sequence of losses ℓ_t such that $R_T \gtrsim \Omega(T)$.

Proof.

Since θ_t is deterministic, the loss function ℓ_t can depend on θ_t. We then choose $\ell_t(\theta_t) = 1$ and $\ell_t(\theta) = 0$ for $\theta \neq \theta_t$. Then one of the chosen actions was picked less then T/K times so that $\max_{1 \leqslant k \leqslant K} \ell_t(k) \leqslant T/K$. Therefore, $R_T \geqslant (1 - 1/K)T$. $\square$

Thus, the strategy of the learner needs to be random $\to$ the player chooses a probability distribution $p_t \in \Delta_K := \{p \in [0,1]^K : \sum_k p_k = 1\}$ and draws $\theta_t \sim p_t$. We recover the setting with actions played in the simplex Δ_K.

A random regret

The regret R_T will be here a random quantity that depends on the randomness of the algorithm (and eventually of the data). We will thus focus on upper-bounding the regret:

- with high-probability: $R_T \leqslant \varepsilon$ with probability at least $1 - \delta$;
- in expectation: $\mathbb{E}[R_T] \leqslant \varepsilon$.

From high-probability bound to expected bound

Since the losses are bounded in $[0,1]$ a bound in high probability entails a bound in expectation.

If $R_T \leqslant \varepsilon$ with probability at least $1-\delta$, then

$$\mathbb{E}[R_T] \leqslant \mathbb{E}\big[R_T | R_T \leqslant \varepsilon\big]\mathbb{P}(R_T \leqslant \varepsilon) + \mathbb{E}\big[R_T | R_T \geqslant \varepsilon\big]\mathbb{P}(R_T \geqslant \varepsilon) \leqslant \varepsilon + T\delta\,. \tag{1}$$

Another useful (and often better) tool to transform a high-probability bound into a bound in expectation is the inequality $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geqslant \varepsilon)d\varepsilon$ for nonnegative random variable X.

From expected bound to high-probability bound

Since the losses are bounded, using Hoeffding-Azuma inequality a bound in expectation entails a bound in high probability at the cost of an additive term of order $\sqrt{T \log(1/\delta)}$ in the regret.

Theorem 1 (Hoeffding-Azuma inequality)

Let $X_1, \ldots, X_T$ be a martingale difference sequence with respect to some filtration $\mathcal{F}_1, \ldots, \mathcal{F}_T$ such that $X_t \in [A_t, A_t + c_t]$ for some $\mathcal{F}_t$-measurable random variable A_t. If $S_T = \sum_{t=1}^T X_t$, then, for any $\varepsilon > 0$

$$\mathbb{P}\big[S_T > \varepsilon\big] \leqslant \exp\left(-\frac{2\varepsilon^2}{\sum_{t=1}^T c_t^2}\right).$$

Exercise: prove Hoeffding-Azuma inequality.

Random EWA

Random EWA

A each $t = 1, \dots, T$

- compute for each $\theta \in \Theta \stackrel{\text{def}}{=} \{1, \dots, K\}$, the weight

$$p_t(\theta) = \frac{e^{-\eta \sum_{s=1}^{t-1} \ell_s(\theta)}}{\sum_{\theta \in \Theta} e^{-\eta \sum_{s=1}^{t-1} \ell_s(\theta)}} .$$

- sample θ_t according to p_t
- observe $\ell_t(\theta) \in [-1, 1]$ for all $\theta \in \Theta$

Proposition 2

Random-EWA satisfies the expected regret

$$\mathbb{E}[R_T] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)\right] \leqslant 2\sqrt{T \log K}$$

for η well tuned.

Exercise: Using Hoeffding-Azuma inequality, provide a bound on the regret R_T with probability $1 - \delta$.

Random EWA

Random EWA

A each $t = 1, \ldots, T$

- compute for each $\theta \in \Theta \stackrel{\text{def}}{=} \{1, \ldots, K\}$, the weight

$$p_t(\theta) = \frac{e^{-\eta \sum_{s=1}^{t-1} \ell_s(\theta)}}{\sum_{\theta \in \Theta} e^{-\eta \sum_{s=1}^{t-1} \ell_s(\theta)}}.$$

- sample θ_t according to p_t
- observe $\ell_t(\theta) \in [-1, 1]$ for all $\theta \in \Theta$

Proposition 2

Random-EWA satisfies the expected regret

$$\mathbb{E}\big[R_T\big] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)\right] \leqslant 2\sqrt{T \log K}$$

for η well tuned.

No assumption on the loss function ℓ_t beside boundedness. In particular, it can be non-convex.

Proof

1. **Apply the regret bound of EWA with $g_t = \big(\ell_t(1), \dots, \ell_t(K)\big) \in [-1,1]^K$:** from Theorem 1 of last class,

$$\sum_{t=1}^{T} p_t \cdot g_t - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta) \leqslant 2\sqrt{T \log K}\,.$$

2. **Take the expectation over $\theta_t \sim p_t$**

$$\mathbb{E}\big[\ell_t(\theta_t)\big] = \mathbb{E}\big[\mathbb{E}[\ell_t(\theta_t)|p_t]\big] = \mathbb{E}\big[p_t \cdot g_t\big]\,.$$

$\square$

Example: Online classification

Assume that you may want to predict a sequence of labels $y_1, \dots, y_T \in \{0,1\}$ (such as spams) based on expert advice $x_t(k) \in \{0,1\}$ (such as different spam detectors).

Then, using the losses $\ell_t(k) = \mathbb{1}_{x_t(k) \neq y_t}$, Random-EWA ensures

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{\theta_t \neq y_t} - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^{T} \mathbb{1}_{x_t(k) \neq y_t}\right] \leqslant 2\sqrt{T \log K}\,.$$

Hence, the expected number of mistakes of the algorithms will not be much larger than the one of the best expert. This is valid though the loss function is nonconvex.

Reminder from last week

Randomized predictions

(Θ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$)

The Gradient Trick and EG

(simplex decision set $\Theta = \Delta_K$, convex and differentiable losses)

Online gradient descent

(convex differentiable Lipschitz losses, convex and compact decision set Θ)

Online Mirrored Descent

From linear to convex losses

Setting: simplex decision set $\Theta = \Delta_K$, convex and differentiable loss functions

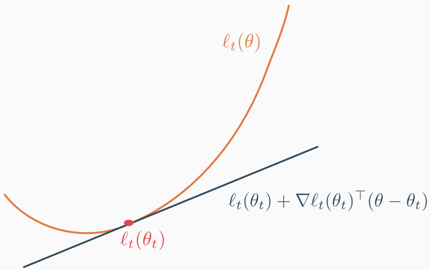
Assumptions and notations: Actions are denoted by p_t (instead of θ_t). The losses are assumed to be convex and Lipschitz

$$\forall p_t \in \Theta, \qquad \left\|\nabla \ell_t(p_t)\right\|_\infty \leqslant G.$$

We will see a simple trick, so-called the gradient trick that allows to extend the results we saw for linear losses to convex losses.

The resulted algorithm is called the Exponentiated Gradient forecaster (EG). It consists in playing EWA with the gradients $g_t = \nabla \ell_t \in [-G, G]^K$ as loss vectors.

The gradient trick



For $g_t = \nabla \ell_t(\theta_t)$, the linear loss $\tilde{\ell}_t(\theta) = g_t^\top \theta$ satisfies for any $\theta \in \Theta$

$$\ell_t(\theta_t) - \ell_t(\theta) \leqslant g_t^\top (\theta_t - \theta) \leqslant \tilde{\ell}_t(\theta_t) - \tilde{\ell}_t(\theta) \,.$$

To prevent infinite regret, need finite $|\tilde{f}_t(\theta)|$ and hence bounds on the dual norms of the domain and gradients

$$|\tilde{\ell}_t(\theta)| \leqslant \|g_t\|_p \|\theta\|_q, \qquad \frac{1}{p} + \frac{1}{q} = 1 \,.$$

Algorithm

The Exponentiated Gradient forecaster (EG)

Parameter: $\eta > 0$

Initialize: $p_1 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$

For $t = 1, \dots, T$

- select p_t; incur loss $\ell_t(p_t)$ and observe $\ell_t : \Theta \to [0, 1]$;
- compute the gradient $g_t = \nabla \ell_t(p_t) \in [-G, G]^K$
- update for all $k \in \{1, \dots, K\}$

$$p_{t+1}(k) = \frac{e^{-\eta \sum_{s=1}^{t} g_s(k)}}{\sum_{j=1}^{K} e^{-\eta \sum_{s=1}^{t} g_s(j)}} .$$

Regret bound of EG

Theorem 2

Let $T \geqslant 1$. For all sequences of convex differentiable losses $\ell_1, \dots, \ell_T : \Theta \to \mathbb{R}$ with bounded gradient $\max_{p\in\Theta} \|\nabla \ell_t(p)\|_\infty \leqslant G$, EWA applied with $g_t = \nabla \ell_t$ achieves the regret bound

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t) - \min_{p\in\Theta} \sum_{t=1}^{T} \ell_t(p) \leqslant \eta G^2 T + \frac{\log K}{\eta}\,. \tag{2}$$

Therefore, for the choice $\eta = \frac{1}{G}\sqrt{\frac{\log K}{T}}$, EWA satisfies the regret bound $R_T \leqslant 2G\sqrt{T \log K}\,$.

Proof

1. **Apply the regret bound of EWA with g_t** (see Theorem 1 of last class):

$$\sum_{t=1}^{T} p_t \cdot g_t - \min_{p \in \Delta_K} \sum_{t=1}^{T} p \cdot g_t \leqslant \eta \sum_{t=1}^{T} \sum_{k=1}^{K} p_t(k) g_t(k)^2 + \frac{\log K}{\eta} \,.$$

Remark that the theorem also holds for loss vectors $g_t \in [-G, G]^K$ as soon as $\eta \leqslant 1/G$.
Upper-bounding $g_t(j)^2 \leqslant \|\nabla \ell_t(p_t)\|_\infty^2 \leqslant G^2$, substituting $g_t = \nabla \ell_t(p_t)$, this yields for all $p \in \Delta_K$

$$\sum_{t=1}^{T} p_t \cdot \nabla \ell_t(p_t) - p \cdot \nabla \ell_t(p_t) \leqslant \eta T G^2 + \frac{\log K}{\eta} \,.$$

2. **Gradient inequality**: by convexity of the losses

$$\ell_t(p_t) - \ell_t(p) \leqslant (p_t - p) \cdot \nabla \ell_t(p_t) \,,$$

which yields

$$\sum_{t=1}^{T} \ell_t(p_t) - \ell_t(p) \leqslant \eta T G^2 + \frac{\log K}{\eta} \,.$$

3. **Optimize η:** $\eta = \frac{1}{G}\sqrt{\frac{\log K}{T}}$. $\square$

Example: Prediction with expert advice (continued)

Setting: A sequence of observations $y_1, \dots, y_T \in [0,1]$ is to be predicted with the help of K expert advice $x_t(k) \in [0,1]$ for $1 \leqslant k \leqslant K$. The learner predict $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$ and suffers a loss $\ell(\widehat{y}_t, y_t)$.

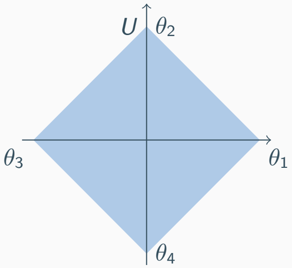
If the loss function is convex and Lipschitz in its first argument, we can apply Theorem 2 with $\ell_t : p \mapsto \ell(p \cdot x_t, y_t)$.

For instance, with the absolute loss, $G = 1$ and EG satisfies:

$$\sum_{t=1}^T |\widehat{y}_t - y_t| - \min_{p \in \Theta} \sum_{t=1}^T \Big| p \cdot x_t - y_t \Big| \leqslant 2\sqrt{T \log K}\,.$$

Hence, on the long run we perform as good as the best convex combination of the experts which may outperform the best expert.

Convex hull of finite point set



The simplex decision set Δ_K can be generalized with any convex hull of a finite point set $S = \{\theta(1), \dots, \theta(K)\}$:

$$\mathrm{Conv}(S) = \left\{ \sum_{i=1}^{K} p_i \theta(i) : \forall i, p_i > 0 \text{ and } \sum_{i=1}^{K} p_i = 1 \right\} .$$

Transforming the loss functions, EG can be applied to compete with such sets.

Such a trick can be used for instance to compete with the ℓ_1-balls using

$$S = \{\theta \in \mathbb{R}^d : \|\theta\|_1 = R, \|x\|_0 = 1\}$$

Since ℓ_p-balls are contained into the ℓ_1-ball (of possibly larger radius depending on p) this can also be used to compete against any ℓ_p-ball for $p \geqslant 1$.

Kivinen and Warmuth, "Exponentiated gradient versus gradient descent for linear predictors", 1997.

Reminder from last week

Randomized predictions

(Θ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$)

The Gradient Trick and EG

(simplex decision set $\Theta = \Delta_K$, convex and differentiable losses)

Online gradient descent

(convex differentiable Lipschitz losses, convex and compact decision set Θ)

Online Mirrored Descent

Online Gradient Descent (OGD) [2]

Setting: convex differentiable Lipschitz loss function, convex and compact decision set Θ

Online Gradient Descent (OGD)

Parameter: $\eta > 0$

Initialize: $\theta_1 \in \Theta$ arbitrarily chosen

For $t = 1, \ldots, T$

- select θ_t; incur loss $\ell_t(\theta_t)$ and observe $\ell_t : \Theta \to [0, 1]$;
- compute the gradient $\nabla \ell_t(\theta_t)$
- update

$$\theta_{t+1} = \mathrm{Proj}_\Theta \Big(\theta_t - \eta \nabla \ell_t(\theta_t) \Big) .$$

where Proj_Θ is the Euclidean projection onto Θ.

Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent", 2003.

Regret bound for OGD

Online Gradient Descent

$$\boxed{\theta_{t+1} \leftarrow \mathrm{Proj}_\Theta\left(\theta_t - \eta \nabla \ell_t(\theta_t)\right)}$$

Theorem 3 (Regret of OGD)

Let $D, G, \eta > 0$. Assume that $\max_{\theta,\theta' \in \Theta} \|\theta - \theta'\| \leqslant D$ and. Then for any sequence $\ell_1, \dots, \ell_T$ of convex differentiable loss functions such that $\max_{\theta \in \Theta} \|\nabla \ell_t(\theta)\| \leqslant G$, the regret of OGD satisfies

$$\sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta) \leqslant \frac{D^2}{2\eta} + \frac{\eta}{2} G^2 T \,.$$

In particular, for $\eta = \frac{D}{G\sqrt{T}}$, we have $R_T \leqslant DG\sqrt{T}$.

Comparison of EG and OGD

Assume that $\Theta = \Delta_K$ is the simplex and the loss functions are sub-differentiable convex functions with $\|\nabla \ell_t\|_\infty \leqslant G_\infty$. Then both EG and OGD are possible algorithms (see Theorems 2 and 3).

We saw in Theorem 2 that EG has a regret bound $R_T \leqslant 2G_\infty\sqrt{T \log K}$. In this case, for all $p, p' \in \Delta_K$

$$\|p - p'\| = \sum_{k=1}^{K} \big(p(i) - p'(i)\big)^2 \leqslant \sum_{i=1}^{K} \big|p(i) - p'(i)\big| \leqslant \sum_{i=1}^{K} p(i) + p'(i) = 2\,,$$

and $\|\nabla \ell_t(p)\| \leqslant \sqrt{K}\|\nabla \ell_t(p)\|_\infty \leqslant \sqrt{K}G_\infty$. Therefore, the regret of OGD is upper-bounded by $R_t \leqslant G_\infty\sqrt{2KT}$. Thus

$$\text{EG:} \quad R_T \leqslant 2G_\infty\sqrt{T \log K} \qquad \text{and} \qquad \text{OGD:} \quad R_T \leqslant \sqrt{2KT}\,.$$

The dependence on K of OGD is suboptimal in this case. This is solved by OMD, a generalization of both algorithms.

Regret bound for OGD

Online Gradient Descent

$$\theta_{t+1} \leftarrow \mathrm{Proj}_\Theta\left(\theta_t - \eta \nabla \ell_t(\theta_t)\right)$$

Theorem 3 (Regret of OGD)

Let $D, G, \eta > 0$*. Assume that* $\max_{\theta,\theta' \in \Theta} \|\theta - \theta'\| \leqslant D$ *and. Then for any sequence* $\ell_1, \ldots, \ell_T$ *of convex differentiable loss functions such that* $\max_{\theta \in \Theta} \|\nabla \ell_t(\theta)\| \leqslant G$*, the regret of OGD satisfies*

$$\sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta) \leqslant \frac{D^2}{2\eta} + \frac{\eta}{2} G^2 T\,.$$

In particular, for $\eta = \frac{D}{G\sqrt{T}}$*, we have* $R_T \leqslant DG\sqrt{T}$*.*

Proof (Step 1)

Recall the update of OGD:

$$\text{OGD}: \quad \theta_{t+1} \leftarrow \text{Proj}_\Theta \big(\underbrace{\theta_t - \eta \nabla \ell_t(\theta_t)}_{z_t} \big)$$

1. Upper-bound the regret with gradient inequality: by convexity

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t) - \ell_t(\theta^*) \quad \overset{\text{Convexity}}{\leqslant} \quad \sum_{t=1}^{T} \langle \nabla \ell_t(\theta_t), \theta_t - \theta^* \rangle$$

Proof (Step 2)

2. Get a telescoping sum:

$$
\begin{aligned}
\left\|\theta_{t+1}-\theta^*\right\|^2 &\overset{\text{Projection}}{\leqslant} \left\|z_t-\theta^*\right\|^2 \\
&= \left\|\theta_t-\eta\nabla\ell_t(\theta_t)-\theta^*\right\|^2 \\
&= \left\|\theta_t-\theta^*\right\|^2+\eta^2\left\|\nabla\ell_t(\theta_t)\right\|^2-2\eta\langle\nabla\ell_t(\theta_t),\theta_t-\theta^*\rangle
\end{aligned}
$$

$$\theta_{t+1} \leftarrow \mathrm{Proj}_\Theta\left(\underbrace{\theta_t-\eta\nabla\ell_t(\theta_t)}_{z_t}\right)$$

Thus,

$$\langle\nabla\ell_t(\theta_t),\theta_t-\theta^*\rangle \leqslant \frac{1}{2\eta}\left(\left\|\theta_t-\theta^*\right\|^2-\left\|\theta_{t+1}-\theta^*\right\|^2\right)+\frac{\eta}{2}\left\|\nabla\ell_t(\theta_t)\right\|^2$$

Summing over $t=1,\ldots,T$ and it telescopes

$$
\begin{aligned}
R_T &\leqslant \frac{1}{2\eta}\left(\left\|\theta_1-\theta^*\right\|^2-\cancel{\left\|\theta_{T+1}-\theta^*\right\|^2}\right)+\frac{\eta}{2}G^2T \\
&\leqslant \frac{D^2}{2\eta}+\frac{\eta G^2T}{2}
\end{aligned}
$$

Exercises

Exercise: Prove an upper-bound on the regret of OGD

a) when η is calibrated with a doubling trick.
b) when η is calibrated using a time-varying parameter $\eta_t = D/(G\sqrt{t})$

Exercise: Prove an upper-bound on the regret of OGD with respect to any sequence of points $\theta_1^*, \ldots, \theta_t^* \in \Theta$ such that $\sum_{t=2}^{T} \|\theta_t^* - \theta_{t-1}^*\| \leqslant X$

$$\sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(\theta_t^*) \leqslant \quad \ldots$$

Logarithmic regret under strong-convexity

Online Gradient Descent:

$$\boxed{\theta_{t+1} \leftarrow \mathrm{Proj}_\Theta\left(\theta_t - \eta_t \nabla \ell_t(\theta_t)\right)}$$

Theorem 4 (Regret of OGD under strong-convexity)

Let $D, G, \gamma > 0$. Assume that $\max_{\theta,\theta' \in \Theta} \|\theta - \theta'\| \leqslant D$ and. Then for any sequence $\ell_1, \dots, \ell_T$ of γ-strongly convex differentiable loss functions such that $\max_{\theta \in \Theta} \|\nabla \ell_t(\theta)\| \leqslant G$, the regret of OGD with $\eta_t = 1/(\gamma t)$ satisfies

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta) \leqslant \frac{G^2}{2\gamma}(1 + \log T)\,.$$

Proof

1. **Upper-bound the regret with strong convexity:**

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t) - \ell_t(\theta^*) \quad \stackrel{\text{Strong Convexity}}{\leqslant} \quad \sum_{t=1}^{T} \langle \nabla \ell_t(\theta_t), \theta_t - \theta^* \rangle - \frac{\gamma}{2} \left\| \theta_t - \theta^* \right\|^2$$

2. **Upper-bound the gradient term as for OGD analysis**

$$\langle \nabla \ell_t(\theta_t), \theta_t - \theta^* \rangle \leqslant \frac{1}{2\eta_t} \left(\left\| \theta_t - \theta^* \right\|^2 - \left\| \theta_{t+1} - \theta^* \right\|^2 \right) + \frac{\eta_t}{2} \left\| \nabla \ell_t(\theta_t) \right\|^2$$

3. **Substitute in the previous inequality and conclude**

$$\begin{aligned} R_T &\leqslant \sum_{t=1}^{T} \frac{1}{2\eta_t} \left(\left\| \theta_t - \theta^* \right\|^2 - \left\| \theta_{t+1} - \theta^* \right\|^2 \right) + \frac{\eta_t G^2}{2} - \frac{\gamma}{2} \left\| \theta_t - \theta^* \right\|^2 \\ &= \frac{1}{2} \sum_{t=1}^{T} \cancel{\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} - \gamma \right)} \left\| \theta_t - \theta^* \right\|^2 + \frac{G^2}{2} \sum_{t=1}^{T} \frac{1}{\gamma t} \\ &\leqslant \frac{G^2}{\gamma} (1 + \log T) \end{aligned}$$

Reminder from last week

Randomized predictions

(Θ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$)

The Gradient Trick and EG

(simplex decision set $\Theta = \Delta_K$, convex and differentiable losses)

Online gradient descent

(convex differentiable Lipschitz losses, convex and compact decision set Θ)

Online Mirrored Descent

Online Mirrored Descent (OMD)

Generalization of OGD to better exploit the geometry of the decision space Θ.

OMD is the online counterpart of the Mirrored Descent algorithm from convex optimization.

Updates are performed into a dual space defined by a convex differentiable function $R : \Theta \to \mathbb{R}$.

Definition (Bregman divergence)

For any continuously differentiable convex function R, the Bregman divergence with respect to R is defined as

$$D_R(x||y) \leqslant R(x) - R(y) - \nabla R(y) \cdot (x - y) \quad \forall x, y \in \Theta\,.$$

It is the difference between the value of the regularization function at x and the value of its first order Taylor approximation.

Online Mirrored Descent (OMD)

Online Mirrored Descent (OMD)

Parameters: $\eta > 0$, regularization function R

Initialize: $z_1 \in \mathbb{R}^d$ such that $\nabla R(z_1) = 0$ and $\theta_1 = \arg\min_{\theta \in \Theta} B_R(\theta || y_1)$

For $t = 1, \ldots, T$

- select θ_t; incur loss $\ell_t(\theta_t)$ and observe $\ell_t : \Theta \to [0, 1]$;
- compute the gradient $\nabla \ell_t(\theta_t)$
- update z_t such that

$$\nabla R(z_{t+1}) = \nabla R(\theta_t) - \eta \nabla \ell_t(\theta_t)\,.$$

- project according to the Bregman divergence

$$\theta_{t+1} \in \arg\min_{\theta \in \Theta} D_R(\theta || z_{t+1})\,.$$

Regret of OMD

Theorem 5

Let $t \geqslant 1$. Let Θ be a compact and convex set. Then, for all sequences of convex subdifferentiable loss functions $\ell_1, \ldots, \ell_T : \Theta \to [0,1]$, the regret of OMD is upper-bounded as

$$\sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta) \leqslant \frac{D}{\eta} + \frac{1}{\eta} \sum_{t=1}^{T} D_{R^*}\left(\nabla R(\theta_t) - \eta \nabla \ell_t(\theta_t) || \nabla R(\theta_t)\right)$$

where $D \geqslant \max_{\theta \in \Theta} |R(\theta)|$ and R^* is the Fenchel conjugate of R defined as $R^*(z) \stackrel{\text{def}}{=} \max_{\theta \in \Theta} \left\{\theta \cdot z - R(\theta)\right\}$.

The proof can be found for instance in Bubeck, Cesa-Bianchi, et al., "Regret analysis of stochastic and nonstochastic multi-armed bandit problems", 2012. EG and OGD are two particular cases of Online Mirror Descent.

Example: OMD with Balls in $\mathbb{R}^d$ = OGD

Recall the update of OGD and OMD:

OGD : $\theta_{t+1} \leftarrow \mathrm{Proj}_\Theta\left(\theta_t - \eta \nabla \ell_t(\theta_t)\right)$ **OMD** : $\begin{array}{l} \nabla R(z_{t+1}) = \nabla R(\theta_t) - \eta \nabla \ell_t(\theta_t) \\ \theta_{t+1} \in \arg\min_{\theta \in \Theta} D_R(\theta \| z_{t+1}) \end{array}$

If $\Theta \subset \mathbb{R}^d$, we can choose $R(x) = \frac{1}{2}\|x\|^2$.

Then

$$\nabla R(x) = x \quad \text{and} \quad D_R(x\|y) = \frac{1}{2}\|x - y\|^2 .$$

Therefore, the update of OMD becomes $z_{t+1} = \theta_t - \eta \nabla \ell_t(\theta_t)$ and $\theta_{t+1} = \mathrm{Proj}_\Theta(z_{t+1})$.

We recover the online gradient descent algorithm.

OMD in the Simplex = EG

Recall the update of EG and OMD:

$$\textbf{EG}: \quad \begin{aligned} g_t &= \nabla \ell_t(\theta_t) \\ \theta_{t+1}(k) &= \frac{\theta_t(k) e^{-\eta g_t(k)}}{\sum_{j=1}^K \theta_t(j) e^{-\eta g_t(j)}} \end{aligned} \qquad \textbf{OMD}: \quad \begin{aligned} \nabla R(z_{t+1}) &= \nabla R(\theta_t) - \eta \nabla \ell_t(\theta_t) \\ \theta_{t+1} &\in \arg\min_{\theta \in \Theta} D_R(\theta \| z_{t+1}) \end{aligned}$$

If $\Theta = \Delta_K$. We can choose the negative entropy

$$R(x) = \sum_{i=1}^K x(i) \log x(i) \,.$$

In this case, $\nabla R(x)_i = 1 + \log x(i)$ and the Bregman Divergence is $D_R(x\|y) = \sum_{i=1}^K x(i) \log(x(i)/y(i))$ also known as the Kullback-Leibler divergence. The update of OMD is then

$$1 + \log(z_{t+1}(i)) = 1 + \log \theta_t(i) - \eta g_t(i) \,,$$

where $g_t = \nabla \ell_t(\theta_t) \in \mathbb{R}^K$. This can be rewritten

$$z_{t+1}(i) = \theta_t(i) e^{-\eta g_t(i)} .$$

The projection to the simplex is a simple renormalization (exercise), we thus recover EG.

Next class

We will see what we can do with bandit feedback.

At each time step $t = 1, \dots, T$

- the player observes a context $x_t \in \mathcal{X}$ (optional step)
- the player chooses an action $\theta_t \in \Theta$ (compact decision/parameter set);
- the environment chooses a loss function $\ell_t : \Theta \to [0, 1]$;
- the player suffers loss $\ell_t(\theta_t)$ and observes
  - the losses of every actions: $\ell_t(\theta)$ for all $\theta \in \Theta$ $\quad \to \quad$ full-information feedback
  - the loss of the chosen action only: $\ell_t(\theta_t)$ $\quad \to \quad$ bandit feedback.

The goal of the player is to minimize his cumulative loss:

$$\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t)\,.$$

References

Thank you!

Bubeck, Sébastien, Nicolo Cesa-Bianchi, et al. "Regret analysis of stochastic and nonstochastic multi-armed bandit problems". In: Foundations and Trends® in Machine Learning 5.1 (2012), pp. 1–122.

Cesa-Bianchi, Nicolo and Gábor Lugosi. Prediction, learning, and games. Cambridge university press, 2006.

Hazan, Elad et al. "Introduction to online convex optimization". In: Foundations and Trends® in Optimization 2.3-4 (2016), pp. 157–325.

Kivinen, Jyrki and Manfred K Warmuth. "Exponentiated gradient versus gradient descent for linear predictors". In: information and computation 132.1 (1997), pp. 1–63.

Lattimore, Tor and Csaba Szepesvári. "Bandit algorithms". In: preprint (2019).

Shalev-Shwartz, Shai et al. "Online learning and online convex optimization". In: Foundations and Trends® in Machine Learning 4.2 (2012), pp. 107–194.

Zinkevich, Martin. "Online convex programming and generalized infinitesimal gradient ascent". In: Proceedings of the 20th International Conference on Machine Learning (ICML-03). 2003, pp. 928–936.