Full-Information Online Learning in Adversarial Environments

Chen-Yu Wei

The Expert Problem

Given: set of experts $\mathcal{A} = \{1, \dots, A\}$

For time $t = 1, 2, \dots, T$:

Learner chooses a distribution over experts $p_t \in \Delta_{\mathcal{A}}$

Environment reveals the reward vector $r_t = (r_t(1), \dots, r_t(A))$

Key difference from before: $r_1(a), \dots, r_T(a)$ do not have the same mean

$$\text{Regret} = \max_{a \in \mathcal{A}} \sum_{t=1}^{T} r_t(a) - \sum_{t=1}^{T} \langle p_t, r_t \rangle$$

Strategies?

- Follow the leader

$$a_t = \max_{a \in \mathcal{A}} \left\{ \sum_{i=1}^{t-1} r_i(a) \right\}$$

Incremental Updates

Exponential Weight Update / Multiplicative Weight Update / Hedge:

$$p_{t+1}(a) = \frac{p_t(a)\, e^{\eta r_t(a)}}{\sum_{a' \in \mathcal{A}} p_t(a')\, e^{\eta r_t(a')}}$$

Equivalent forms:

$$p_{t+1} = \max_{p \in \Delta} \left\{ \langle p, r_t \rangle - \frac{1}{\eta} \mathrm{KL}(p, p_t) \right\} \quad \text{or} \quad p_{t+1} = \max_{p \in \Delta} \left\{ \left\langle p, \sum_{i=1}^{t} r_i \right\rangle + \frac{1}{\eta} H(p) \right\}$$

$$\mathrm{KL}(p, q) \coloneqq \sum_{a=1}^{A} p(a) \ln \frac{p(a)}{q(a)} \qquad H(p) \coloneqq \sum_{a=1}^{A} p(a) \ln \frac{1}{p(a)}$$

Regret Bound for Exponential Weight Updates

Theorem.

Assume that $\eta r_t(a) \leq 1$ for all t, a. Then EWU ensures

$$\text{Regret} = \max_{a^\star} \sum_{t=1}^{T} (r_t(a^\star) - \langle p_t, r_t \rangle) \leq \frac{\ln A}{\eta} + \eta \sum_{t=1}^{T} \sum_{a=1}^{A} p_t(a) r_t(a)^2$$

Online Linear Optimization

Given: Convex feasible set $\Omega \subseteq \mathbb{R}^d$

For time $t = 1, 2, \ldots, T$:

Learner chooses a point $w_t \in \Omega$

Environment reveals the reward vector $r_t \in \mathbb{R}^d$

$$\text{Regret} = \max_{w \in \Omega} \sum_{t=1}^{T} w^\top r_t - \sum_{t=1}^{T} w_t^\top r_t$$

Projected Online Gradient Ascent

Projected Online Gradient Ascent:

$$w_{t+1} = \Pi_{\Omega}[w_t + \eta r_t]$$

Equivalent form: $$w_{t+1} = \max_{w \in \Omega} \left\{ \langle w, r_t \rangle - \frac{1}{2\eta} \|w - w_t\|_2^2 \right\}$$

Regret Bound for Projected Online Gradient Ascent

Theorem. Projected Online Gradient Ascent ensures

$$\text{Regret} = \max_{w^\star \in \Omega} \sum_{t=1}^{T} \langle w^\star - w_t, r_t \rangle \leq \frac{\max_{w \in \Omega} \|w\|_2^2}{\eta} + \eta \sum_{t=1}^{T} \|r_t\|_2^2$$