Minimax strategy for prediction with expert advice under stochastic assumptions

Wojciech Kotłowski

Poznań University of Technology, Poland

Learning Faster from Easy Data II
NIPS 2015 Workshop

Prediction with expert advice

In trials $t = 1, 2, \dots, T$:

- Algorithm predicts with $\boldsymbol{w}_t \in \Delta^K$.
- Loss vector $\boldsymbol{\ell}_t \in [0, 1]^K$ is revealed.
- Algorithm incurs loss $\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t$.

Regret of a strategy $\boldsymbol{\omega} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_T)$:

$$R = \sum_t \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \min_k \underbrace{\sum_{t \le T} \ell_{t,k}}_{L_{T,k}}$$

Goal: find $\boldsymbol{\omega}$ minimizing the worst-case regret over all sequences.

Prediction with expert advice

In trials $t = 1, 2, \ldots, T$:

- Algorithm predicts with $\boldsymbol{w}_t \in \Delta^K$.
- Loss vector $\boldsymbol{\ell}_t \in [0,1]^K$ is revealed.
- Algorithm incurs loss $\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t$.

Regret of a strategy $\boldsymbol{\omega} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_T)$:

$$R = \sum_t \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \min_k \underbrace{\sum_{t \le T} \ell_{t,k}}_{L_{T,k}}$$

Goal: find $\boldsymbol{\omega}$ minimizing the worst-case regret over all sequences.

Too pessimistic: minimax $\boldsymbol{\omega}$ has the same regret on all sequences!

Prediction with expert advice

In trials $t = 1, 2, \ldots, T$:

- Algorithm predicts with $\boldsymbol{w}_t \in \Delta^K$.
- Loss vector $\boldsymbol{\ell}_t \in [0, 1]^K$ is revealed.
- Algorithm incurs loss $\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t$.

Regret of a s

Drop the minimax principle?

Drop the worst-case assumptions?

$L_{T,k}$

Goal: find $\boldsymbol{\omega}$ minimizing the worst-case regret over all sequences.

Too pessimistic: minimax $\boldsymbol{\omega}$ has the same regret on all sequences!

Prediction with expert advice

In trials $t = 1, 2, \ldots, T$:

- Algorithm predicts with $\boldsymbol{w}_t \in \Delta^K$.
- Loss vector $\boldsymbol{\ell}_t \in [0,1]^K$ is revealed.
- Algorithm incurs loss $\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t$.

Regret of a s

Drop the minimax principle?

Drop the worst-case assumptions?

$L_{T,k}$

Goal: find $\boldsymbol{\omega}$ minimizing the worst-case regret over all sequences.

Too pessimistic: minimax $\boldsymbol{\omega}$ has the same regret on all sequences!

Stochastic setting: optimal strategy for “easy” data

Assumption: Each expert $k = 1, \ldots, K$ generates losses i.i.d. from a fixed distribution P_k.

Goal: Find the minimax strategy $\boldsymbol{\omega}$ w.r.t. all choices of distributions $\mathcal{P} = (P_1, \ldots, P_k)$.

Stochastic setting: optimal strategy for "easy" data

Assumption: Each expert $k = 1, \ldots, K$ generates losses i.i.d. from a fixed distribution P_k.

Goal: Find the minimax strategy $\boldsymbol{\omega}$ w.r.t. all choices of distributions $\mathcal{P} = (P_1, \ldots, P_k)$.

Minimax in terms of what?

Stochastic setting: optimal strategy for "easy" data

Assumption: Each expert $k = 1, \ldots, K$ generates losses i.i.d. from a fixed distribution P_k.

Goal: Find the minimax strategy $\boldsymbol{\omega}$ w.r.t. all choices of distributions $\mathcal{P} = (P_1, \ldots, P_k)$.

Minimax in terms of what?

Expected regret: $R_{\text{eg}}(\boldsymbol{\omega}, \mathcal{P}) \;=\; \mathbb{E}\big[\sum_t \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \min_k L_{T,k}\big]$

Redundancy: $R_{\text{ed}}(\boldsymbol{\omega}, \mathcal{P}) \;=\; \mathbb{E}\big[\sum_t \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t\big] - \min_k \mathbb{E}\left[L_{T,k}\right]$

Excess risk: $R_{\text{isk}}(\boldsymbol{\omega}, \mathcal{P}) \;=\; \mathbb{E}\left[\boldsymbol{w}_T \cdot \boldsymbol{\ell}_T\right] - \min_k \mathbb{E}\left[\ell_{T,k}\right]$

Stochastic setting: optimal strategy for "easy" data

Assumption: Each expert $k = 1, \ldots, K$ generates losses i.i.d. from a fixed distribution P_k.

We give a strategy $\boldsymbol{\omega}^*$, which is minimax with respect to all three measures **simultaneously**:

$$\sup_{\mathcal{P}} R(\boldsymbol{\omega}^*, \mathcal{P}) = \inf_{\boldsymbol{\omega}} \sup_{\mathcal{P}} R(\boldsymbol{\omega}, \mathcal{P}),$$

where R is either R_{eg}, R_{ed}, or R_{isk}.

Excess risk: $R_{\mathrm{isk}}(\boldsymbol{\omega}, \mathcal{P}) = \mathbb{E}[\boldsymbol{w}_T \cdot \boldsymbol{\ell}_T] - \min_k \mathbb{E}[\ell_{T,k}]$