

Paper Presentation

No-Regret and Incentive Compatible Online Learning

Austin Shouli and Kirill Abrosimov

Faculty of Computer Science

March 6, 2026

No-Regret and Incentive-Compatible Online Learning

Rupert Freeman¹ David M. Pennock² Chara Podimata³ Jennifer Wortman Vaughan⁴

Abstract

We study online learning settings in which experts act strategically to maximize their influence on the learning algorithm's predictions by potentially misreporting their beliefs about a sequence of binary events. Our goal is twofold. First, we want the learning algorithm to be no-regret with respect to the best fixed expert in hindsight. Second, we want incentive compatibility, a guarantee that each expert's best strategy is to report his true beliefs about the realization of each event. To achieve this goal, we build on the literature on wagering mechanisms, a type of multi-agent scoring rule. We provide algorithms that achieve no regret and incentive compatibility for myopic experts for both the full and partial information settings. In general,

But what if the experts that the learner consults are strategic agents, capable of reporting predictions that do not represent their true beliefs? As pointed out by [Roughgarden & Schriivers \(2017\)](#), when the learner is not only making predictions but also (implicitly or explicitly) evaluating the experts, experts might have incentive to misreport. The Good Judgment Project,¹ a competitor in IARPA's Aggregative Contingent Estimation geopolitical forecasting contest, scored individual forecasters and rewarded the top 2%—dubbed “Superforecasters” ([Tetlock & Gardner, 2015](#))—with perks such as paid conference travel; some are now employed by a spinoff company. Similarly, the website FiveThirtyEight² not only predicts election results by aggregating different pollsters, but also publicly scores the pollsters, in a way that correlates with the amount of influence that the pollsters have over the FiveThirtyEight aggregate. It is natural to

Online Learning: Recap

- ▶ A framework for sequential decision making under uncertainty
- ▶ Let's consider a basic protocol for Online Learning, prediction with expert advice

Prediction with Expert Advice

Let's consider a basic Online Learning Problem, Prediction with Expert Advice.

We have:

- ▶ K experts
- ▶ Action space \mathcal{A}
- ▶ Outcome space \mathcal{Y}
- ▶ Loss Function $\ell : \mathcal{A} \rightarrow \mathcal{Y}$

Prediction with Expert Advice

Protocol: At each round $t = 1, 2, \dots, T$:

- ▶ Each expert $i \in [K]$ reveals their advice $f_{j,t}$
- ▶ Learner (Algorithm) plays an action $a_t \in \mathcal{A}$
- ▶ Nature plays and reveals an outcome $y_t \in \mathcal{Y}$ [Could be Stochastic, or Adversarial]
- ▶ Learner suffers loss $\hat{\ell}_t = \ell(a_t, y_t)$ [based on what they played, and the actual outcome of nature]

Prediction with Expert Advice

Regret in round n :

$$R_n = \hat{\ell}_t - \min_{i \in [K]} \ell_{i,t}.$$

Cumulative regret:

$$R_T = \sum_{t=1}^T L_t - \min_{i \in [K]} \sum_{t=1}^T \ell_{i,t}.$$

Goal:

- ▶ We want to design an algorithm that achieves sub-linear regret
- ▶ This means that our algorithm will eventually converge to perform as well as the best expert
- ▶ Sub-linear regret is equivalent to No-regret

Exponentially Weighted Average Forecaster

Let's look at an algorithm for PEA that achieves sub-linear regret.

This algorithm is known as:

- ▶ Exponential Weighted Average Forecaster (EWA)
- ▶ The Hedge Algorithm
- ▶ Multiplicative Weights Update (MWU)

[There are subtle differences but for our case they are functionally equivalent]

Exponentially Weighted Average Forecaster

Algorithm:

Each forecaster (expert) is assigned an initial weight $w_{i,t} = 1$

In each round $t = 1, 2, \dots, T$:

- ▶ Each expert i makes a prediction $f_{i,t}$
- ▶ Probability distribution p among experts, such that
$$p_{j,t} = \frac{w_{j,t}}{W_t}, \quad W_t = \sum_{i=1}^K w_t^{(i)}$$
- ▶ Learner makes a prediction $a_t = \sum_{j=1}^K p_{j,t} f_{j,t}$
- ▶ Learner incurs loss \hat{l}_t
- ▶ Experts incur loss $l_{i,t}$
- ▶ Each expert's weight is updated such that:

$$w_{i,t+1} = \begin{cases} w_{i,t}, & \text{if } f_{i,t} = y_t \quad [\text{ie. experts prediction was correct}] \\ e^{-\eta l_{i,t}}, & \text{if } f_{i,t} \neq y_t \quad [\text{ie. experts prediction was incorrect}] \end{cases}$$

Exponentially Weighted Average Forecaster

EWA Regret:

- ▶ K Experts
- ▶ T Rounds
- ▶ $R_T =$ Cumulative Regret at Round T

The EWA Algorithm's regret is defined as

$$R_T = \hat{L} - L^* = \sum_t^T \sum_i^K p_t^{(i)} \ell_t^{(i)} - \ell_t^{(i^*)}$$

The EWA Algorithm achieves a regret of

$$R_T \leq \sqrt{2T \ln K}.$$

No-Regret Dynamics

Why do we care about sub-linear or no-regret dynamics?

- ▶ If regret is sub-linear, our algorithm will eventually converge on the performance of the best expert
- ▶ In other words, our algorithm will achieve optimal performance for the given set of experts

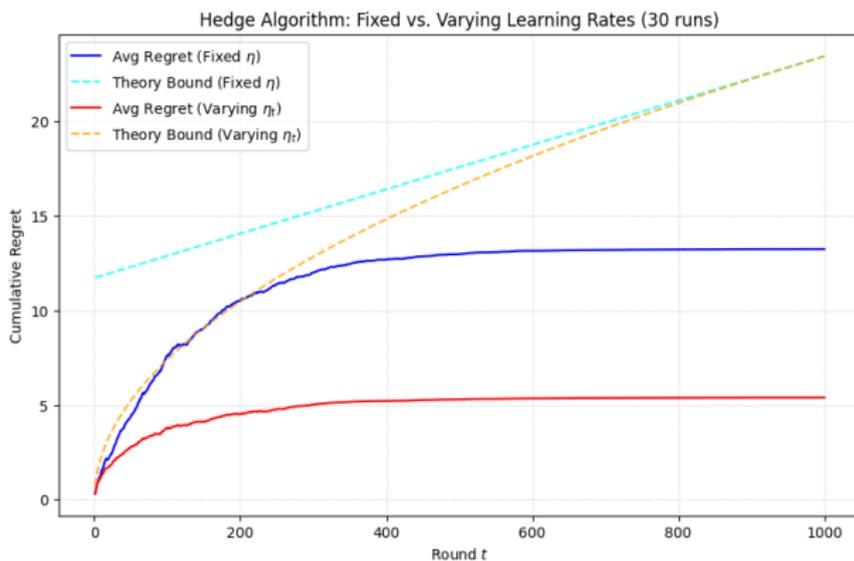


Figure: Hedge Algorithm Experimental Sub-linear Regret

Incentive Compatibility / Lying

We have discussed Dominant Strategy Incentive-Compatible (DSIC) mechanisms

- ▶ This means that it is in an agent's best interest to report truthfully
- ▶ The same concept of Incentive Compatibility can be applied to PEA
- ▶ Can we design a mechanism that encourages the experts to report their true belief?

Suppose in round t an expert has a true belief $v_{i,t}$. We wish to design a mechanism for which every recommendation $f_{i,t}$ a rational agent i makes, $f_{i,t} = v_i$

- ▶ In the paper we are covering, the author's do this by introducing a mechanism for agents called wagering.

Wagering Mechanism

Budget-balanced Wagering Mechanisms (WMs):

A mechanism Γ that redistributes "money" between agents, depending on the quality of their forecasts/ predictions.

Each agent $i \in [K]$:

- ▶ has a belief $b_i \in [0, 1]$
- ▶ reports a prediction $p_i \in [0, 1]$
- ▶ wagers an amount $w_i \in W$
- ▶ receives a payout; depends on realization r

$$\text{payout}_i = \Gamma_i(\vec{p}, \vec{w}, r) \quad (1)$$

$$\sum_{i \in [K]} \text{payout}_i = 0 \implies W_{\text{after}} = W_{\text{before}} \implies \sum_{i \in [K]} \Gamma_i(\vec{p}, \vec{w}, r) = \sum_{i \in [K]} w_i^{(i)} \quad (2)$$

We can think of budget-balanced wagering as a type of zero-sum game. No money is created or destroyed in only moves between agents.

Weighted Score Wagering Mechanism

Weighted Score Wagering Mechanism: a wagering mechanism that uses proper scoring rules for payout;

$$\Gamma_i^{WSWM}(\vec{p}, \vec{w}, r) = w^{(i)} \cdot (S(p_i, r) - \sum_{j \in [K]} \frac{w_j}{W} \cdot S(p_j, r)) \quad (3)$$

Is WSWM incentive compatible?

Proof: Show that

$$\mathbb{E}_{r \sim \text{Bern}(b)} [\Gamma_i(p_i = b, p_{-i}), w, r] \leq \mathbb{E}_{r \sim \text{Bern}(b)} [\Gamma_i(p_i, p_{-i}), w, r] \quad (4)$$

Weighted-Score Update (WSU) Algorithm

Suppose:

- ▶ Each agent $i \in [K]$ has a wager (weight) $\pi_{i,t}$
- ▶ Each wager $\pi_{i,1}$ is initialized to $1/K$
- ▶ We define an update rule of the form $\pi_{t+1} = \Gamma(p_t, \pi_t, r_t)$
- ▶ Each agent's is allowed to wager $\eta \in (0, 0.5]$

By defining an update rule of the form $\pi_{t+1} = \Gamma(p_t, \pi_t, r_t)$, and budget-balanced wagering mechanism Γ can be transformed into an online learning algorithm.

Weighted-Score Update (WSU) Algorithm

Protocol: At each round $t = 1, 2, \dots, T$:

- ▶ Each expert i provides a prediction p_t and a wager $\pi_{i,t}$
- ▶ Nature plays and reveals an outcome r_t
- ▶ Each expert receives a payout:

$$\pi_{i,t+1} = \eta \Gamma_i^{\text{WSWM}}(p_t, \pi_t, r_t) + (1 - \eta)\pi_{i,t}$$

Regret: Let i^* be the best-performing expert. The regret of the algorithm at round t is bounded by

$$R_T = \sum_t \sum_i \pi_t^{(i)} \ell_t^{(i)} - \sum_t \ell_t^{(i^*)}$$

WSU Proof

WSU Proof: inequality

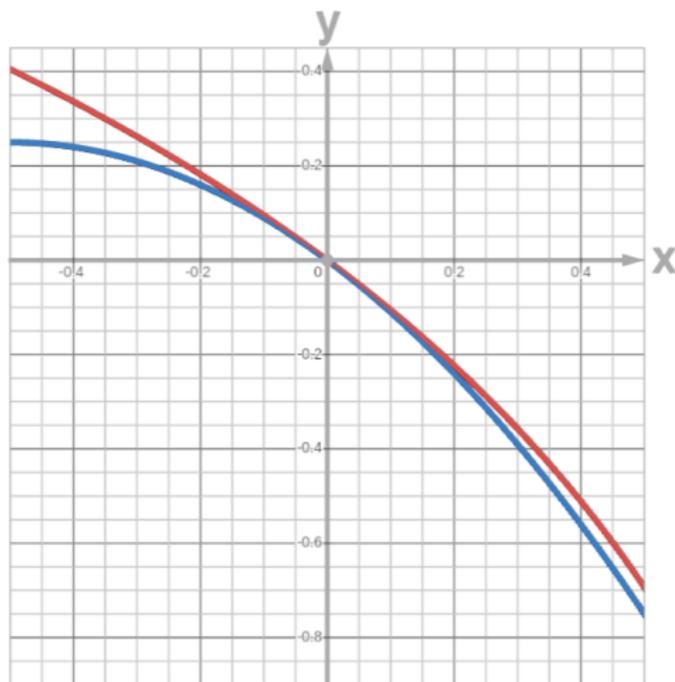


Figure: $\ln(1-x) \geq -x - x^2; \forall x \leq \frac{1}{2}$

Reading Questions and Discussion

- ▶ When using the Exponentially Weighted Average Forecaster, why might experts/forecasters be incentivized to lie, even when a proper scoring rule is used? What role does the normalization of weights play in enabling strategic behavior?
- ▶ In the WSU algorithm, experts are only allowed to "wager" a fraction $\eta \in (0, 0.5)$ of their current probability at each timestep. What happens to an expert's weight when they are allowed to wager $\eta \in (0.5, 1]$? What happens to the algorithm's regret?
- ▶ Excluding Section 5, the authors consider the case of myopic experts: these are experts that select their report to maximize their probability in the next round. Is the assumption of myopic experts reasonable? Why or why not?