Foundations of machine learning

Adversarial online learning

Maximilian Kasy

Department of Economics, University of Oxford

Winter 2026

Outline

- The online learning problem:
  Sequential prediction.

- The adversarial framework:
  Regret guarantees for all possible sequences of outcomes.
  No sampling process is assumed.

- General theory for the case of convex action spaces (e.g. probabilistic forecasts).
  Potentials as a method for proving adversarial regret bounds.

- A very versatile algorithm: Thompson sampling.

Takeaways for this part of class

- Online learning is the most basic sequential decision problem:
  The observable history does not depend on actions.
- We can have performance guarantees
  without any assumptions about the data generating process.
- To do so, our algorithms need to perform well
  whenever there is a "competitor" that performs well.
- How to achieve this?
  Make predictions similar to those of successful competitors.
- Thompson sampling choses actions based on the posterior probability that they are optimal. This principle is successful in a wide variety of settings.
- Bonus slides: Worst case sequences delay learning as long as possible.

Weighted average predictors

Bounding regret

Thompson sampling

References

Setup

- Sequential predictions at times $t = 1, 2, \ldots$
- Outcomes: $Y_t \in \mathcal{Y}$.
- Predictions: $\hat{Y} \in \mathcal{Y}$.
- Experts $h \in \mathcal{H}$, delivering predictions

  $$\hat{Y}_{h,t} \in \mathcal{Y}.$$

  ($\sim$ hypotheses / predictors).
- Any predictive features X_t are left implicit in the expert predictions.
- We assume (for today's discussion)
  1. $\mathcal{H}$ is finite,

Loss and regret

- We want to make a prediction $\hat{Y}_t$, using the expert predictions $\hat{Y}_{h,t}$,
- having observed $\mathcal{S}_{t-1} = (Y_1, \ldots, Y_{t-1})$.
- Loss at time t: $L(\hat{Y}_t, Y_t)$.
- Regret at time t relative to h:

$$r_{h,t} = L(\hat{Y}_t, Y_t) - L(\hat{Y}_{h,t}, Y_t).$$

- Cumulative regret at time t relative to h:

$$R_{h,t} = \sum_{s=1}^{t} r_{h,s}.$$

- Cumulative regret relative to $\mathcal{H}$:

$$R_{\mathcal{H},t} = \max_{h \in \mathcal{H}} R_{h,t}.$$

Successful learning

- Our goal: Find learning algorithms delivering $\hat{Y}_t$
- such that average cumulative regret vanishes
- **for all possible realizations of** $\mathcal{S}_t = (Y_1, \ldots, Y_t)$:

$$\sup_{\mathcal{S}_t} \frac{1}{t} R_{\mathcal{H},t} \to 0.$$

- No probability is involved,
  this is the worst case over all possible realizations of outcomes!!
- How could that even be possible?!?
  The past carries no information about the future?!?!
  There is no stability at all over time?!?!?!

A chaotic, evil world

- **No assumption** is made about how the outcomes Y_t are generated.
- We are interested in worst case behavior over all possible sequences $Y_1, Y_2, \ldots$

> "Imagine another set of results. The first time, the white ball drove the black ball into the pocket. The second time, the black ball bounced away. The third time, the black ball flew onto the ceiling. The fourth time, the black ball shot around the room like a frightened sparrow, finally taking refuge in your jacket pocket. The fifth time, the black ball flew away at nearly the speed of light, breaking the edge of the pool table, shooting through the wall, and leaving the Earth and the Solar System, just like Asimov once described.[13] What would you think then?"
>
> Ding watched Wang. After a long silence, Wang finally said, "This actually happened. Am I right?"

Liu Cixin, The Three Body Problem

Weighted average predictors

Bounding regret

Thompson sampling

References

Weighted average predictors

- We will consider weighted average predictors of the form

$$\hat{Y}_t = \frac{\sum_{h \in \mathcal{H}} w_{h,t-1} \cdot \hat{Y}_{h,t}}{\sum_{h \in \mathcal{H}} w_{h,t-1}},$$

- where the weights of each expert are increasing in the cumulative regret relative to that expert

$$w_{h,t} = \phi'(R_{h,t}),$$

- with ϕ nonnegative, convex, and increasing.

- This gives a larger weight to experts that performed well in the past.

Convex loss functions

Lemma 2.1

- Suppose that the loss function is convex in $\hat{Y}_t$,
- and $\hat{Y}_t$ is given by a weighted average predictor of this form.
- Then

$$\sup_{Y_t} \sum_{h \in \mathcal{H}} r_{h,t} \cdot \phi'(R_{h,t-1}) \leq 0.$$

- *Proof*:
  - By convexity of L, Jensen's inequality, :

$$\sum_{h \in \mathcal{H}} w_{h,t-1} \cdot L(\hat{Y}_t, Y_t) = L(\hat{Y}_t, Y_t) \leq \sum_{h \in \mathcal{H}} w_{h,t-1} \cdot L(\hat{Y}_{h,t}, Y_t).$$

  - Weights are proportional to $\phi'(R_{h,t-1})$.

Potential function

- Use boldface for vectors, with components corresponding to $h \in \mathcal{H}$.
- Potential function (a proof device):

$$\Phi(u) := \psi\left(\sum_{h\in\mathcal{H}} \phi(u_h)\right).$$

- With this notation

$$\hat{Y}_t = \frac{\langle \nabla\Phi(R_{t-1}),\, \hat{Y}_t \rangle}{\langle \nabla\Phi(R_{t-1}),\, 1 \rangle}$$

- The lemma then can be rewritten as the **Blackwell condition**

$$\sup_{Y_t} \langle r_t, \nabla\Phi(R_{t-1}) \rangle \leq 0.$$

- Note that $R_t = R_{t-1} + r_t$.

Illustrating the Blackwell condition

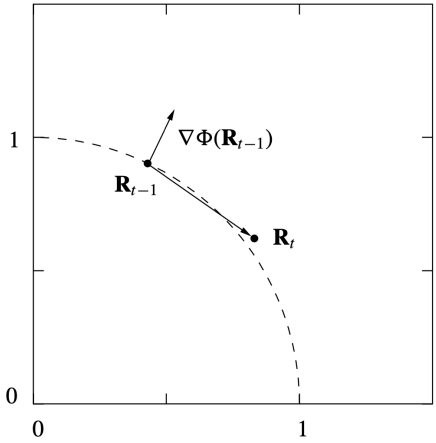

Figure 2.1. An illustration of the Blackwell condition with $N = 2$. The dashed line shows the points in regret space with potential equal to 1. The prediction at time t changed the potential from $\Phi(\mathbf{R}_{t-1}) = 1$ to $\Phi(\mathbf{R}_t) = \Phi(\mathbf{R}_{t-1} + \mathbf{r}_t)$. Though $\Phi(\mathbf{R}_t) > \Phi(\mathbf{R}_{t-1})$, the inner product between $\mathbf{r}_t$ and the gradient $\nabla\Phi(\mathbf{R}_{t-1})$ is negative, and thus the Blackwell condition holds.

Weighted average predictors

Bounding regret

Thompson sampling

References

Bounding the potential

Theorem 2.1.

- Suppose that $\hat{Y}_t$ satisfies the Blackwell condition.
- Then, for all t,

$$\Phi(R_t) \leq \Phi(0) + \frac{1}{2}\sum_{s=1}^{t} C(r_s)$$

- where

$$C(r) = \sup_{u} \psi'\left(\sum_{h\in\mathcal{H}} \phi(u_h)\right) \sum_{h\in\mathcal{H}} \phi''(u_h) r_h^2.$$

- *Proof*:
  - Second order Taylor expansion of $\Phi(R_t) = \Phi(R_{t-1} + r_t)$ in r_t.
  - Bounding the first-order term using the Blackwell condition.

Exponential weighting

- Special case: Exponential weights.
- Potential (with tuning parameter η):

$$\phi(u) = \frac{1}{\eta} \log \left(\sum_{h \in \mathcal{H}} \exp(\eta \cdot u_h) \right).$$

- Corresponding weights:

$$w_{h,t-1} = \frac{\exp\left(\eta \cdot R_{h,t-1}\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\eta \cdot R_{h',t-1}\right)} = \frac{\exp\left(-\eta \cdot \sum_{s=1}^{t-1} L(\hat{Y}_{h,t}, Y_t)\right)}{\sum_{h' \in \mathcal{H}} \exp\left(-\eta \cdot \sum_{s=1}^{t-1} L(\hat{Y}_{h',t}, Y_t)\right)}.$$

- These weights only depend on the loss of each expert, but not on our prediction $\hat{Y}_t$.
- For quadratic error loss, this is Bayesian model averaging, for normal likelihood with variance $2/\eta$, uniform prior over experts.

Bounding regret for exponential weighting

Corollary 2.2.

- Assume that L is convex in $\hat{Y}$ and bounded by $[0,1]$.
- Then, for all η and for all $\mathcal{S}_t = (Y_1, \ldots, Y_t)$,

$$R_{\mathcal{H},t}(\mathcal{S}_t) \leq \frac{\log(|\mathcal{H}|)}{\eta} + \frac{t\eta}{2}.$$

- For $\eta = \sqrt{2\frac{\log(|\mathcal{H}|)}{t}}$,

$$R_{\mathcal{H},t}(\mathcal{S}_t) \leq \sqrt{2t\log(|\mathcal{H}|)}.$$

Proof

- By assumption, $\phi(x) = \exp(\eta \cdot x)$, $\psi(x) = \phi^{-1}(x) = \log(x)/\eta$.
- For any estimator with weights based on a potential, and $\psi(x) = \phi^{-1}(x)$,

$$\begin{aligned}\max_{h\in\mathcal{H}} R_{h,t} &= \psi\left(\phi\left(\max_{h\in\mathcal{H}} R_{h,t}\right)\right) \\ &\leq \psi\left(\sum_{h\in\mathcal{H}} \phi\left(R_{h,t}\right)\right) = \Phi(R_t).\end{aligned}$$

- Calculation yields $C(r_t) \leq \eta$ (using $|r_{h,t}| \leq 1$), and $\Phi(0) = \log(|\mathcal{H}|)/\eta$.
- The theorem implies

$$\Phi(R_t) \leq \Phi(0) + \frac{1}{2}\sum_{s=1}^{t} C(r_s)$$

$$\log(|\mathcal{H}|) \qquad \eta$$

Discussion

- We can do essentially as well as the best of our experts.
- No matter how the sequence Y_t is generated!
- No stability or invariance in the world is assumed.
- A possible way to address the **induction problem**?
- We are guaranteed to do well *if anyone can do well*.

Is this good enough?

> *The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.*
>
> Bertrand Russell, The Problems of Philosophy.

- Should our regret bound provide consolation to the chicken?

Weighted average predictors

Bounding regret

Thompson sampling

References

Bit prediction

- The simplest special case of online learning.
- Binary outcomes and predictions, $Y_t, \hat{Y}_t \in \{0,1\}$.
- Mis-classification error loss: $L(\hat{Y}_t, Y_t) = 1(\hat{Y}_t \neq Y_t)$.
- No predictors.
- [$\Rightarrow$] Cumulative regret at time t:

$$R_t = \max_{y \in \{0,1\}} \left(\sum_{s=1}^{t} \left[1(\hat{Y}_s \neq Y_s) - 1(y \neq Y_s) \right] \right).$$

- Denote $1_t = \sum_{s=1}^{t} Y_t$, $0_t = t - 1_t$. Then

$$\min_{y \in \{0,1\}} \left(\sum_{s=1}^{t} 1(y \neq Y_s) \right) = \min(0_t, 1_t).$$

A Bayesian model

- Consider the following model, which we will use for the construction of an algorithm
  but *not* for the evaluation of this algorithm!

- i.i.d. draws:
$$Y_t \sim^{i.i.d.} Ber(\theta)$$

- Uniform prior:
$$\theta \sim U[0,1].$$

- Then the time $t+1$ posterior for θ is given by
$$\theta | Y_1, \ldots, Y_t \sim Beta(1+1_t, 1+0_t).$$

- Posterior mean:
$$E[\theta | Y_1, \ldots, Y_{t-1}] = \frac{1+1_t}{2+t}.$$

Thompson sampling

- A very simple, general and successful approach
  for solving online learning and active learning problems.

- Denote by $\mathcal{S}_{t-1}$ the history (information) observed by the beginning of period t.
  Let $p_t(y)$ be the posterior probability that y is the optimal action:

$$p_t(y) = P\left(y = \underset{\tilde{y}}{\operatorname{argmin}}\ E[L(\tilde{y}, Y_t)|\theta] \middle| \mathcal{S}_{t-1} \right).$$

- Thompson sampling chooses $\hat{Y}_t = y$ with probability $p_t(y)$.
  The *sampling probability* is set equal to
  the *posterior probability* that an action is optimal.

- Thompson sampling can be implemented by

  1. Sampling one draw $\hat{\theta}_t$ from the posterior for θ.

Expected regret for a given sequence

- For binary bit prediction:

$$\operatorname*{argmin}_{\tilde{y}} E[L(\tilde{y}, Y_t)|\theta] = 1(\theta > \tfrac{1}{2})$$

and thus

$$p_t(0) = P(\theta < \tfrac{1}{2}|\mathcal{S}_{t-1}) = F_{Beta(1+1_{t-1},1+0_{t-1})}(\tfrac{1}{2}).$$
$$p_t(1) = 1 - F_{Beta(1+1_{t-1},1+0_{t-1})}(\tfrac{1}{2}).$$

- Fix the sequence $Y_1, \ldots, Y_T$ and assume wlog that $1_T > T/2 > 0_T$.

- Consider two sequences (Y_t) and (Y'_t), which are the same, except the order of Y_s and Y_{s+1} is swapped in sequence (Y'_t).

Swapping

- Suppose wlog $(Y_s, Y_{s+1}) = (0,1)$.
  Let $1_s = k$, $0_s = s-k$.

- Then the difference in expected regret between the two sequences equals

$$\begin{aligned} R'_t - R_t &= \left[P(\hat{Y}'_s = 0) + P(\hat{Y}'_{s+1} = 1)\right] \\ &\quad - \left[P(\hat{Y}_s = 1) + P(\hat{Y}_{s+1} = 0)\right] \\ &= \left[F_{Beta(1+k,1+s-k)}(\tfrac{1}{2}) + (1 - F_{Beta(2+k,1+s-k)}(\tfrac{1}{2}))\right] \\ &\quad - \left[(1 - F_{Beta(1+k,1+s-k)}(\tfrac{1}{2})) + F_{Beta(1+k,2+s-k)}(\tfrac{1}{2}))\right] \\ &= 2F_{Beta(1+k,2+s-k)}(\tfrac{1}{2})) \\ &\quad - \left[F_{Beta(2+k,1+s-k)}(\tfrac{1}{2})) + F_{Beta(1+k,2+s-k)}(\tfrac{1}{2}))\right]. \end{aligned}$$

- By the properties of the Beta distribution (*Fact 2*), we can rewrite this as

$$R'_t - R_t = \frac{1}{2^s \cdot B(1+k, 1+s-k)} \cdot \left[\frac{1}{1+k} - \frac{1}{1+s-k}\right]$$

Swapping continued

- It follows that the difference $R'_t - R_t$ is negative iff $k > s/2$.
  (cf. *Lemma 4* in the paper).

- In words: If there were more 1s than 0s thus far,
  it is worse if the "unexpected" observation $Y_s = 0$
  comes before the "expected" $Y_{s+1} = 1$.

- We can use this observation to figure out the worst case sequence $(Y_1, \ldots, Y_T)$,
  among all sequences with $1_T = k > T/2$.

- *Theorem 5* in the paper does exactly that:
  The worst-case sequences are exactly the sequences such that

  1. The sequence ends with $2k - T$ 1s.

  2. Before that, all pairs (Y_s, Y_{s+1}) (for s odd) are equal to either $(0,1)$ or $(1,0)$.

Practice problem

- Consider any sequence with $1_T = k$ that is not of this form.
- Show that for such a sequence there exists a swap which increases regret.

Intuition and implications

- The algorithm tries to learn whether $1_T > 0_T$, or the other way around.
- The worst case sequence delays learning as much as possible, by alternating 0s and 1s.
- One can calculate / bound regret for such a worst-case sequence. By *Theorem 6* in the paper:

$$R_T = O\left(\sqrt{\min(1_T, 0_T)}\right) = O(\sqrt{T}).$$

References

- Adversarial online learning:

  Cesa-Bianchi, N. and Lugosi, G. (2006). Prediction, learning, and games. *Cambridge University Press, chapter 2.*

- Thompson sampling:

  Lewi, Y., Kaplan, H., and Mansour, Y. (2020). Thompson sampling for adversarial bit prediction. In Algorithmic Learning Theory*, pages 518–553. PMLR*