Foundations of machine learning

Adversarial online learning

Maximilian Kasy

Department of Economics, University of Oxford

Winter 2025

Takeaways

- Thompson sampling choses actions based on the posterior probability that they are optimal. This principle is successful in a wide variety of settings.
- Worst case sequences delay learning as long as possible.

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:

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

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.
  2. Choosing $\hat{Y}_t = \operatorname{argmin}_{\tilde{y}} E[L(\tilde{y}, Y_t)|\theta = \hat{\theta}_t]$.

Expected regret for a given sequence

- For binary bit prediction:

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

  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*