ECE 8803: Online Decision Making in Machine Learning **Fall 2022**

Lecture 4: September 1

Lecturer: Vidya Muthukumar

In this lecture, we will show that the multiplicative weights algorithm, which we introduced last lecture, achieves the optimal regret guarantee of $\mathcal{O}(\sqrt{T})$ on adversarial prediction. We will explicitly call out the terms that trade off exploitation and randomness in the proof. Finally, we will intuitively explain why the $\sqrt{T}$ guarantee is, in fact, the best that we can do.

4.1. Recall: sequential prediction, regret, and multiplicative weights

We begin by recapping relevant notation for binary sequence prediction, regret, and the multiplicative weights algorithm. Last lecture, we introduced the binary sequence prediction paradigm in which we aim to predict the next realization of a sequence, X_t, from the past realizations, denoted by $X^{t-1} := \{X_1, \dots, X_{t-1}\}$. Our prediction is denoted by $\widehat{X}_t$, and our loss function is given by $\ell(\widehat{X}_t, X_t)$. The overall goal is to choose a prediction strategy $f_{\mathsf{predict}}(\cdot)$ to minimize the total loss

$$H_T := \mathbb{E}\left[\sum_{t=1}^{T} \ell(\widehat{X}_t, X_t)\right]. \tag{4.1}$$

We introduced the metric of performance of *regret* with respect to the best fixed predictor in hindsight (who is able to see the entire stream of data at once):

$$R_T(X^T) := \underbrace{H_T}_{\text{our algorithm}} - \underbrace{\min_{x^* \in \{0,1\}} \sum_{t=1}^{T} \ell(x^*; X_t)}_{\text{best fixed prediction}}. \tag{4.2}$$

Henceforth, we denote the loss of the best fixed predictor in hindsight by L_T^* as shorthand. Importantly, we wish to minimize regret on any sequence, even one that could be generated adversarially to our prediction strategy.

We explained why at least some amount of randomization is *required* to avoid being exploited by an adversarial sequence. In other words, successful predictors of an adversarial sequence will generate probabilities $\widehat{P}_t := f_{\mathsf{rand.}}(X^{t-1})$ and then predict $\widehat{X}_t \sim \text{Bernoulli}(\widehat{P}_t)$. On the other hand (as you will see in HW), completely randomizing is not the best strategy either: g $\widehat{P}_t = 1/2$ on all rounds would be very wasteful against the very predictable sequence, $X_t = 1$ for all $t = 1, \dots, T$.

Therefore, we need a strategy that trades off *exploiting* the information in past data and *randomization* to avoid itself being exploited by an adversary. There are many possible ways

that one could achieve this: one is given by the popular *multiplicative weights algorithm* (henceforth abbreviated as MWA). See the previous lecture note for the detailed description of MWA; we only recap the structure of the update. It is convenient to denote the losses incurred by each letter as

$$L_{t,x} := \sum_{s=1}^{t} \mathbb{I}[X_s \neq x] \text{ for each } x \in \{0,1\}.$$

Then, the MWA uses update

$$p_{t,1} = \frac{e^{-\eta L_{t-1,1}}}{e^{-\eta L_{t-1,1}} + e^{-\eta L_{t-1,0}}}.$$

In other words, we have $p_{t,x} \propto e^{-\eta L_{t-1,x}}$ for each $x \in \{0,1\}$: thus, if 1 has incurred less loss until now (which happens if 1 appeared more often in the sequence), then $\widehat{P}_t$ is closer to 1.

We mentioned last lecture that MWA incurs *sublinear regret* of the form $R_T(X^T) = \mathcal{O}(\sqrt{T})$ for any sequence. Now, we will prove it!

4.2. Proof, part 1: Showing low regret with respect to a "mix loss"

The intuition behind the MWA is somewhat mysterious: after all, it is not clear what relationship the Hamming (0-1) loss has with exponentials. Our first step is to introduce a *surrogate benchmark*, that we call the "mix loss", that we will show that MWA does well with respect to. This mix loss will actually be expressed in terms of exponenials. Next, we will show that this "mix loss" in turn has a natural relationship with the loss of the best fixed predictor in hindsight. Putting these together, we will show the requisite regret bound.

We now define this "mix loss", and show that we have low "regret" with respect to it.

Definition 1 *We define the instantaneous mix loss at time step t by*

$$m_t := -\frac{1}{\eta} \log \left(p_{t,0} \cdot e^{-\eta \mathbb{I}[X_t \neq 0]} + p_{t,1} \cdot e^{-\eta \mathbb{I}[X_t \neq 1]} \right) \tag{4.3}$$

and the cumulative mix loss by

$$M_T := \sum_{t=1}^{T} m_t. \tag{4.4}$$

It is interesting to note that the instantaneous mix losses depend not only on the actual losses that result from fixed predictors (i.e. $L_{t,x}$), but also the algorithm itself through the probabilities $p_{t,1}, p_{t,0}$ used for prediction. However, through a coincidence, the cumulative mix loss reduces to a remarkably simple form that is a "soft" form of the best loss in hindsight! We will explore this in the second part of the proof.

At a high level, our proof strategy is as follows: we wlil write

$$R_T = H_T - L_T^* = \underbrace{H_T - M_T}_{\text{mix regret}} + \underbrace{M_T - L_T^*}_{\text{mix loss approximation error}}$$

First, we will show that MWA has low "mix-regret", which is defined by

$$R_{T,\mathsf{mix}} := H_T - M_T \tag{4.5}$$

Importantly, we will show that the more we randomize (i.e. smaller η), the smaller our "mix-regret" is (this is perhaps one reason for the name "mix-loss"!). In particular, we will prove the following lemma.

Lemma 2 *For any sequence, we have*

$$R_{T,\mathsf{mix}} \leq \frac{\eta T}{8}.$$

Before proving the lemma, let us consider two extremes to get some intuition. Suppose we set $\eta \to 0$. Then, one can verify (in a calculus exercise) that $\lim_{\eta\to 0} m_t = p_{t,0}\mathbb{I}[X_t \neq 0] + p_{t,1}\mathbb{I}[X_t \neq 1]$, which is exactly the loss incurred by the actual algorithm at time step t. In this case, we would have $H_T = M_T$ exactly, and the "mix-loss" regret is equal to 0.

This tells us that the mix loss "regret" is in some way quantifying the extent to which the adversary can manipulate the algorithm's decisions. We will now prove the lemma formally.

Proof First, we note that

$$\begin{aligned} R_{T,\mathsf{mix}} &= \sum_{t=1}^{T} \mathbb{E}[\ell(\widehat{X}_t; X_t)] + \frac{1}{\eta}\log \mathbb{E}[e^{-\eta\ell(\widehat{X}_t;X_t)}] \\ &= \frac{1}{\eta}\cdot\left(\sum_{t=1}^{T}\eta\mathbb{E}[\ell(\widehat{X}_t; X_t)] + \log \mathbb{E}[e^{-\eta\ell(\widehat{X}_t;X_t)}]\right) \end{aligned}$$

where the expectation is over the randomness in generating the prediction $\widehat{X}_t$ only. We will now use Hoeffding's lemma (which you saw in the probability review last week) to complete this proof. Consider the random variable $Z_t := -\ell(\widehat{X}_t, X_t)$. Clearly, this is bounded between $[-1, 0]$, so Hoeffding's lemma gives us

$$\begin{aligned} \mathbb{E}[e^{\eta(Z_t - \mathbb{E}[Z_t])}] &\leq e^{\frac{\eta^2}{8}} \\ \implies \log \mathbb{E}[e^{\eta Z_t}] - \eta\mathbb{E}[Z_t] &\leq \frac{\eta^2}{8} \\ \implies \frac{1}{\eta}\left(\log \mathbb{E}[e^{\eta Z_t}] - \eta\mathbb{E}[Z_t]\right) &\leq \frac{\eta}{8}. \end{aligned}$$

Plugging in the definition of Z_t then gives us $R_{T,\mathsf{mix}} \leq \sum_{t=1}^{T}\frac{\eta}{8} = \frac{\eta T}{8}$, which completes the proof. ■

4.3. Proof, part 2: Is the "mix loss" close to the best loss in hindsight?

Thus, we have shown that for a sufficiently large amount of randomization (small η), we incur sufficiently low "mix loss" regret. But what does the mix loss tell us about the best loss in hindsight? We will now see some magic: *for the MWA*, the mix loss and L_T^* enjoy

a very special relationship. To see this, we first plug in the definition of $p_{t,1}$ and $p_{t,0}$ from MWA to get

$$\begin{aligned} m_t &= -\frac{1}{\eta}\log\left(p_{t,0}\cdot e^{-\eta\mathbb{I}[X_t\neq 0]} + p_{t,1}\cdot e^{-\eta\mathbb{I}[X_t\neq 1]}\right) \\ &= -\frac{1}{\eta}\log\left(\frac{e^{-\eta L_{t-1,0}}\cdot e^{-\eta\mathbb{I}[X_t\neq 0]} + e^{-\eta L_{t-1,1}}\cdot e^{-\eta\mathbb{I}[X_t\neq 1]}}{e^{-\eta L_{t-1,0}} + e^{-\eta L_{t-1,1}}}\right) \\ &= -\frac{1}{\eta}\log\left(\frac{e^{-\eta L_{t,0}} + e^{-\eta L_{t,1}}}{e^{-\eta L_{t-1,0}} + e^{-\eta L_{t-1,1}}}\right) \\ &= -\frac{1}{\eta}\log\left(e^{-\eta L_{t,0}} + e^{-\eta L_{t,1}}\right) + \frac{1}{\eta}\log\left(e^{-\eta L_{t-1,0}} + e^{-\eta L_{t-1,1}}\right) \end{aligned}$$

for all $t > 1$. For $t = 1$, we get $m_1 = -\frac{1}{\eta}\log\left(\frac{e^{-\eta L_{t,0}}+e^{-\eta L_{t,1}}}{2}\right)$ as the updates start with $p_{t,0} = p_{t,1} = 1/2$.

Summing the m_t's up over $t = 1, \ldots, T$, we see that all the terms cancel (this is also often called a telescoping sum!), and we get

$$M_T = -\frac{1}{\eta}\log\left(\frac{e^{-\eta L_{T,0}} + e^{-\eta L_{T,1}}}{2}\right).$$

Remarkably, although the *instantaneous* mix loss terms depend on the algorithm, the *cumulative* mix loss term does not. This is a very special property of the multiplicative weights update.

In the previous part of the proof, we showed that $H_T - M_T$ was small. It remains now to show that $M_T - L_T^*$ is small, i.e. the cumulative mix loss is a good approximation to the best loss in hindsight. This turns out to be more the case the less randomization we use (i.e. the larger the value of η!). To see this, note that when η is very large, $e^{-\eta L_{T,0}} + e^{-\eta L_{T,1}}$ is dominated by $e^{-\eta L_T^*}$ and so we get $M_T \approxeq -\frac{1}{\eta}\log(e^{-\eta L_T^*}) = L_T^*$. More precisely, we have

$$\begin{aligned} M_T &= -\frac{1}{\eta}\log\left(\frac{e^{-\eta L_{T,0}} + e^{-\eta L_{T,1}}}{2}\right) \leq -\frac{1}{\eta}\log\left(\frac{e^{-\eta L_T^*}}{2}\right) \\ &= L_T^* + \frac{\log 2}{\eta}. \end{aligned}$$

It is interesting to note that this bound may not needed for some special cases, e.g. where $L_{T,0} = L_{T,1}$ (such as the periodic sequence). *In these cases, it may make sense to randomize fully and the best regret guarantee may be obtained with maximal randomization.*

4.4. The role of the learning rate: Exploitation-randomization tradeoff

Putting the bounds on $H_T - M_T$ and $M_T - L_T^*$ together, we get

$$R_T \leq \frac{\eta T}{8} + \frac{\log 2}{\eta}. \tag{4.6}$$

Thus, selecting $\eta = 1/\sqrt{T}$ gives us a $\mathcal{O}(\sqrt{T})$ regret guarantee! (It may be a useful exercise to pick the specific value of η that minimizes the upper bound, and see what constant you will get.)

Intuitively, the first term encourages randomization: the more we "mix", the less an adversary is able to exploit our decisions. This is measured by the mix loss regret quantitatively.

On the other hand, the second term discourages randomization: the more different the cumulative losses are from one another, the more we may want to be able to exploit this and eventually converge to the right decision. The bound on $M_T - L_T^*$ tells us that if we randomize too much, we would not be able to do that. Thus, Equation (4.6) mathematizes the tradeoff between randomization and exploitation.