Incentives and Machine Learning (CSC 482A/581A)
Lectures 14 and 15

Nishant Mehta

1 Prediction with expert advice

In the game of prediction with expert advice, there is an action space $\mathcal{A}$, an outcome space $\mathcal{Y}$, a loss function $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$ mapping each a given action $a \in \mathcal{A}$ and outcome $y \in \mathcal{Y}$ to a loss $\ell(a, y)$. At the start of each round, each of K experts provides Learner with advice in the form of a suggested action from $\mathcal{A}$. Learner then aggregates these actions in some way, producing its own action in $\mathcal{A}$. Finally, Nature selects an outcome, and Learner and each expert suffer loss according to their respective actions and the outcome. The goal of Learner is to ensure that its *regret* over T rounds is small, where the regret is defined as the amount by which Learner's cumulative loss exceeds the cumulative loss of the best action in hindsight. The game protocol is given below.

Algorithm 1: Prediction with Expert Advice

```
for $t = 1 \to T$ do
    Nature plays expert advice $f_{j,t} \in \mathcal{A}$ for each expert $j \in [K]$ and reveals the advice
      $(f_{1,t}, f_{2,t}, \dots, f_{K,t})$ to Learner
    Learner plays action $a_t \in \mathcal{A}$
    Nature plays outcome $y_t \in \mathcal{Y}$ and reveals it to Learner
    Each expert $j \in [K]$ suffers loss $\ell(f_{j,t}, y_t)$ and Learner suffers loss $\ell(a_t, y_t)$
end
```

Note that Nature controls both the experts and the outcomes. In sequential prediction problems, the strength of the adversary (Nature) can vary; the adversary can be either:

- *oblivious* - Nature knows which algorithm Learner is using, but Nature must commit to its entire sequence of expert advice and outcomes before Learner takes its first action.
- *non-oblivious* or *adaptive* - At any point in time, Nature can make its choice (whether expert advice or outcome) based on all of Learner's previous actions.

We will make no assumptions about Nature: Nature will be a non-oblivious adversary. Anytime Nature makes a selection, it can do so using all the information revealed thus far. To be clear, the expert advice $f_{j,t}$ for any expert j can be selected with knowledge of $a_1, a_2, \dots a_{t-1}$, while outcome y_t can be selected with the same knowledge as well as a_t.

The regret takes the form

$$\mathcal{R}_T := \sum_{t=1}^{T} \ell(a_t, y_t) - \min_{j \in [K]} \sum_{t=1}^{T} \ell(f_{j,t}, y_t).$$

That is, the regret is simply Learner's cumulative loss minus the cumulative loss of the best expert in hindsight. Intuitively, regret quantifies how much sadness Learner feels for not having always played the best action in hindsight. A basic demand in online learning is to have a learning algorithm that is *no-regret*, meaning that the time-average of the regret $\frac{1}{T}\mathcal{R}_T$ goes to zero as T approaches infinity, or, equivalently, the regret $\mathcal{R}_T$ is sublinear in T. To see why such a requirement is sensible, consider the case where the loss function takes values in some bounded range. Then if Learner has regret growing linearly in T, up to a multiplicative constant it is doing no better than constantly playing the *worst* action in hindsight! Another perspective comes from a stochastic interpretation of the data. If we were in the statistical learning setting, then at a high level we may think of the time-average of the regret as a form of risk, and the no-regret property is the analogue of the excess risk decaying to zero as the sample size n approaches infinity.

Given Nature's ability to adapt to the previous plays of Learner (and, in particular, to select y_t with knowledge of a_t), one wonders if any algorithm can always (i.e., against any strategy of Nature) obtain sublinear regret for this problem. Without further assumptions on the action space and loss function, Nature actually can ensure that any algorithm is forced to suffer linear regret. However, if we assume the action space is convex, the loss function is convex with respect to its second argument, and the losses are in some bounded range (e.g., the unit interval $[0, 1]$), then we can show that the worst-case regret is sublinear.

Two common examples satisfying these assumptions are:

- Classification with absolute loss: Here, we take $\mathcal{A} = [0, 1]$, $\mathcal{Y} = \{0, 1\}$, and $\ell(a, y) = |a - y|$.
- Classification with squared loss: We take $\mathcal{A}$ and $\mathcal{Y}$ as before and now set $\ell(a, y) = (a - y)^2$.

To simplify the presentation, we set up some notation. For any $j \in [K]$ and $t \in [T]$, let $\ell_{j,t} = \ell(f_{j,t}, y_t)$ denote the loss of expert j in round t, and let $L_{j,t} = \sum_{s=1}^{t} \ell_{j,s}$ denote expert j's cumulative loss at the end of round t. Also, for any $t \in [T]$, denote Learner's loss in round t $\hat{\ell}_t = \ell(a_t, y_t)$, and denote Learner's cumulative loss at the end of round t by $\hat{L}_t = \sum_{s=1}^{t} \hat{\ell}_s$.

2 Exponentially Weighted Average Forecaster

The algorithm that we study for this setting is called the exponentially weighted average forecaster (EWA forecaster); it works as follows. In each round, the algorithm maintains weights over the experts, with $w_{j,t} = e^{-\eta L_{j,t}}$ indicating the weight of the j^{th} expert at the end of round t. In round t, the forecaster predicts according to the following weighted average of the experts' actions:

$$a_t = \frac{\sum_{j=1}^{K} w_{j,t-1} \, f_{j,t}}{\sum_{j=1}^{K} w_{j,t-1}}.$$

It will be convenient to rewrite the above using normalized weights; to this end, we introduce the probability vector $p_t \in \Delta_K$, defined as

$$p_{j,t} = \frac{w_{j,t-1}}{\sum_{i=1}^{K} w_{i,t-1}}.$$

Using p_t, we can re-express a_t as $a_t = \mathsf{E}_{j \sim p_t}[a_{j,t}]$.

In later lectures, we may consider a time-varying learning rate. However, since the learning rate η is currently constant throughout the rounds, we also can write an incremental update for the weights for all $j \in [K]$ by:

- initializing via $w_{j,0} = 1$;
- at the end of round t, incrementally updating via $w_{j,t} = w_{j,t-1} \cdot e^{-\eta \ell_{j,t}}$.

I emphasize that if the learning rate is not constant, then we *should not using the incremental form of the update*; doing so can result in linear regret.

We are now ready to see an upper bound on the worst-case regret of the EWA forecaster.

Theorem 1. *Assume $\mathcal{A}$ is convex, the loss function is convex in its second argument, and the losses are in the range $[0, 1]$. For any learning rate $\eta > 0$, any sequence of expert advice $(f_{j,t})_{j\in[K],t\in[T]}$, and any sequence of outcomes $y_1, \ldots, y_T$, the regret of the EWA forecaster satisfies*

$$\hat{L}_T - \min_{j\in[K]} L_{j,T} \;\leq\; \sum_{t=1}^{T} \langle p_t, \ell_t \rangle - \min_{j\in[K]} L_{j,T} \;\leq\; \frac{\log K}{\eta} + \eta \sum_{t=1}^{T} \sum_{j=1}^{K} p_{j,t} \ell_{j,t}^2 \;\leq\; \frac{\log K}{\eta} + T\eta.$$

In particular, setting $\eta = \sqrt{\frac{\log K}{T}}$ makes the upper bound $2\sqrt{T \log K}$.

Proof. For the first inequality, observe that from the convexity of the loss, Jensen's inequality implies that

$$\hat{\ell}_t = \ell\left(\sum_{j=1}^{K} p_{j,t} f_{j,t},\, y_t\right) \leq \sum_{j=1}^{K} p_{j,t} \ell(f_{j,t}, y_t) = \sum_{j=1}^{K} p_{j,t} \ell_{j,t}.$$

The main part of the proof is the second inequality. The proof centers around the sum W_t of the experts' weights in any round t. To this end, for $t \in \{0, 1, \ldots, T\}$, we define $W_t = \sum_{j=1}^{K} w_{j,t}$. Note that $\hat{L}_{j,0} = 0$ and hence $w_{j,0} = 1$ for all j.

From $-\frac{1}{\eta} \log W_T$, we will "extract" the cumulative loss of the best expert in hindsight, and from $-\frac{1}{\eta} \log \frac{W_t}{W_{t-1}}$, we will extract Learner's loss in round t. We then use the relation (from telescoping)

$$\sum_{t=1}^{T} \log \frac{W_t}{W_{t-1}} = \log W_T - \log W_0 \tag{1}$$

to relate Learner's cumulative loss to the cumulative loss of the best expert in hindsight.

Step 1: We show that

$$-\frac{1}{\eta} \log W_T \leq \min_{j\in[K]} L_{j,T}.$$

Since the weights are nonnegative, we have for all $i \in [K]$ (including the best expert in hindsight)

$$W_T = \sum_{j=1}^{K} w_{j,T} \geq w_{i,T}.$$

Hence,

$$-\frac{1}{\eta} \log W_T \leq -\frac{1}{\eta} \log e^{-\eta L_{i,T}} = L_{i,T}.$$

Step 2: We claim that

$$-\frac{1}{\eta}\log\frac{W_t}{W_{t-1}} \geq \sum_{t=1}^{T}\langle \ell_t, p_t\rangle - \eta\sum_{t=1}^{T} p_{j,t}\ell_{j,t}^2.$$

We establish the claim as follows:

$$\begin{aligned}\frac{W_t}{W_{t-1}} &= \frac{\sum_{j=1}^{K} w_{j,t}}{\sum_{j=1}^{K} w_{j,t-1}} \\ &= \frac{\sum_{j=1}^{K} w_{j,t-1}e^{-\eta\ell_{j,t}}}{\sum_{j=1}^{K} w_{j,t-1}} \\ &= \sum_{j=1}^{K} p_{j,t}e^{-\eta\ell_{j,t}}.\end{aligned}$$

Now, we use that for all $x \geq -1$, it holds that $e^x \leq 1 - x + x^2$, so that the above is at most

$$\sum_{j=1}^{K} p_{j,t}\left(1 - \eta\ell_{j,t} + \eta^2\ell_{j,t}^2\right) = 1 - \eta\langle \ell_t, p_t\rangle + \eta^2\sum_{j=1}^{K} p_{j,t}\ell_{j,t}^2.$$

Finally, we use $\log(1+x) \leq x$ for $x \geq -1$ to get

$$\begin{aligned}-\frac{1}{\eta}\log\frac{W_t}{W_{t-1}} &= -\frac{1}{\eta}\log\left(1 - \eta\langle \ell_t, p_t\rangle + \eta^2\sum_{j=1}^{K} p_{j,t}\ell_{j,t}^2\right) \\ &\geq \langle \ell_t, p_t\rangle - \eta\sum_{j=1}^{K} p_{j,t}\ell_{j,t}^2.\end{aligned}$$

Step 3: Combining everything Using (1) and the previous steps, we have

$$\begin{aligned}\sum_{t=1}^{T}\langle \ell_t, p_t\rangle - \eta\sum_{t=1}^{T}\sum_{j=1}^{K} p_{j,t}\ell_{j,t}^2 &\leq \sum_{t=1}^{T} -\frac{1}{\eta}\log\frac{W_t}{W_{t-1}} \\ &= -\frac{1}{\eta}\log W_T + \frac{1}{\eta}\log W_0 \\ &= -\frac{1}{\eta}\log W_T + \frac{\log K}{\eta} \\ &\leq \min_{j\in[K]} L_{j,T} + \frac{\log K}{\eta}.\end{aligned}$$

Rearranging gives the second inequality in the theorem statement:

$$\sum_{t=1}^{T}\langle \ell_t, p_t\rangle \leq \min_{j\in[K]} L_{j,T} + \frac{\log K}{\eta} + \eta\sum_{t=1}^{T}\sum_{j=1}^{K} p_{j,t}\ell_{j,t}^2.$$

The third inequality is immediate from the losses being in $[0,1]$. □

Against a worst-case adversary who seeks to maximize the regret, the bound $2\sqrt{T\log K}$ has the optimal rate; there is a matching lower bound (matching in terms of the rate). The constant 2 is not the best possible. With a little more work, one can use a result known as Hoeffding's Lemma — instead of the inequalities $e^x \leq 1 - x + x^2$ and $\log(1+x) \leq x$ — to improve the constant 2 to $\frac{1}{\sqrt{2}}$. As shown by a matching lower bound, the latter constant is the optimal constant as $T \to \infty$.

3 Decision-theoretic online learning

We now introduce the setting of decision-theoretic online learning (DTOL). This setting is a very important special case of prediction with expert advice. The DTOL protocol unfolds as follows.

Algorithm 2: Decision-Theoretic Online Learning

```
for $t = 1 \to T$ do
    Learner plays probability distribution $p_t \in \Delta_K$
    Nature plays loss vector $\ell_t = (\ell_{1,t}, \ldots, \ell_{K,t})^\top \in [0,1]^K$ and reveals it to Learner
    Each expert $j \in [K]$ suffers loss $\ell_{j,t}$ and Learner suffers loss $\langle \ell_t, p_t \rangle$
end
```

This setting could be thought of as "prediction with expert advice *without the expert advice*". Learner never gets to observe each expert's action (advice), but it still gets to observe each expert's loss. As such, Learner cannot mix the advice of the experts — it cannot mix inside the loss — but it *can* randomize over the experts — it can mix outside the loss. The interpretation of $\langle \ell_t, p_t \rangle = \mathsf{E}_{j \sim p_t}[\ell_{j,t}]$ is then Learner's expected loss if it randomly selects expert j with probability $p_{j,t}$. Alternatively, if Learner is allocating a fixed sum of money across the experts, then p_t specifies how Learner divides its money among the experts, and $\langle \ell_t, p_t \rangle$ is Learner's actual (not expected) loss.

Let us see how DTOL is a special case of prediction with expert advice. In prediction with expert advice, we take action space $\mathcal{A} = \Delta_K$ and outcome space $\mathcal{Y} = [0,1]^K$. Next, for given action $a_t \in \mathcal{A}$ and outcome $y_t \in \mathcal{Y}$ — which we may view as probability vector $p_t \in \Delta_K$ and loss vector $\ell_t \in [0,1]^K$ respectively — the loss is $\ell(a_t, y_t) = \langle y_t, a_t \rangle = \langle \ell_t, p_t \rangle$. That is, for a fixed outcome, the loss function is linear in the action. Finally, each expert $j \in [K]$ uses the *constant* strategy of setting (for all t) $f_{j,t}$ equal to e_j; here, e_j is the j^{th} standard basis vector in $\mathbb{R}^K$. This choice satisfies $\ell(f_{j,t}, y_t) = \ell(e_j, \ell_t) = \langle \ell_t, e_j \rangle = \ell_{j,t}$, as desired.

What regret can Learner hope to acheve in DTOL? Since the losses are in $[0,1]$, the loss function is convex in its second argument (recall that linear functions are convex), and the action space is convex, Learner can use the EWA forecaster. Our regret bound Theorem 1 applies, and so for all sequences of loss vectors, the regret is at most $2\sqrt{T \log K}$.

It is a bit clunky to view the algorithm as the EWA forecaster for prediction with expert advice in the special case of DTOL. For simplicity, we directly present the algorithm, which is known as Hedge, below.

Algorithm 3: Hedge

```
Input: $\eta > 0$
Set $w_{j,0} = 1$ for $j = 1, \ldots, K$
for $t = 1 \to T$ do
    Set $p_{j,t} = \frac{w_{j,t-1}}{\sum_{i=1}^K w_{i,t-1}}$ for $j = 1, \ldots, K$
    Observe loss vector $\ell_t$ from Nature
    Suffer loss $\langle \ell_t, p_t \rangle$
    Set $w_{j,t} = w_{j,t-1} e^{-\eta \ell_{j,t}}$ for $j = 1, \ldots, K$
end
```

For convenience, we also summarize Hedge's regret guarantee.

Theorem 2. *Let Hedge be run with learning rate $\eta = \sqrt{\frac{\log K}{T}}$. Then, for all sequences of loss vectors $\ell_1, \ldots, \ell_T$,*

$$\hat{L}_T \leq \min_{j \in [K]} L_{j,T} + 2\sqrt{T \log K}.$$

Precisely the same comment about improving the constant applies to Hedge as well. Without any change to the algorithm, the constant 2 can be improved to $\frac{1}{\sqrt{2}}$.