CSE599s, Spring 2014, Online Learning Lecture 6 - 04/17/2014

Online Learning with Expert Advice

Lecturer: Ofer Dekel *Scribe: Jonathan Bragg*

1 Review

Recall the problem of online learning with expert advice. We have defined our regret in terms of the difference between our loss and the loss of the best fixed expert (if one were to pick the expert in the future). Recall that the expert advice setting proceeds as follows.

For $t = 1, \ldots, T$ (T known in advance):

- d experts (actions) give advice
- Player chooses expert $I_t \in \{1, \ldots, d\}$
- Player receives feedback $l_t \in [0,1]^d$ and incurs loss $l_{t,I_t} \in [0,1]$

Remark: Can I_t be deterministic?

The answer is *no* for the following reason. An adversary (even an oblivious one) knows our code, so he would choose

$$l_{t,i} = \begin{cases} 1 & \text{if } i = I_t \\ 0 & \text{otherwise} \end{cases},$$

resulting in a cumulative loss for the player equal to T. In order to calculate regret, consider how this loss compares to the best fixed strategy. The average loss on each round is $1/d$, so the average cumulative loss of any fixed expert i over T rounds is T/d. Therefore, $\exists i$ s.t. $\sum_{t=1}^{T} l_{t,i} \leq T/d$, since all experts can't be worse than average. Thus,

$$\text{Regret} \geq T - T/d = \frac{d-1}{d} T,$$

yielding linear regret $\mathcal{O}(T)$.

Remark: Does it matter if the adversary can play after the player (i.e., he is non-oblivious)?

The answer is *no* for the same reason. Since the adversary knows our code, which is deterministic, he gains nothing by viewing the outcomes.

2 Achieving Sub-linear Regret with Randomization

Part I: Introduce Randomization

- Instead of choosing a single expert, we choose $p_t \in \Delta_d$, the probability simplex over d outcomes.
- In each round t, we draw expert I_t randomly and independently as $I_t \sim p_t$. More explicitly, $\forall i \in \{1, \ldots d\}, P(I_t = i) = p_{t,i}$.

Importantly, we choose p_t deterministically (the adversary can know our choices), but the adversary does not know the value of the actual sample.

Part II: Measure Expected Regret

Since we are sampling experts, it makes sense to consider our expected loss and regret. We define

$$\begin{aligned}\text{Expected Regret} &= \mathsf{E}\left[\sum_{t=1}^{T} l_{t,I_t}\right] - \min_{i\in\{1,\dots d\}} \sum_{t=1}^{T} l_{t,i} \\ &= \sum_{t=1}^{T} p_t \cdot l_t - \min_{p\in\Delta_d} \sum_{t=1}^{T} p \cdot l_t.\end{aligned}$$

The expected loss of our randomized algorithm can be written as $\sum_{t=1}^{T} p_t \cdot l_t$ by linearity of expectation. To see why we can rewrite loss of the best fixed expert as $\min_{p\in\Delta_d} \sum_{t=1}^{T} p \cdot l_t$, we need the following lemma.

Lemma 1. *The loss of the best fixed expert over T rounds is equal to the expected loss of the best fixed randomized strategy, or*

$$\min_{i\in\{1,\dots,d\}} \sum_{t=1}^{T} l_{t,i} = \min_{p\in\Delta_d} \sum_{t=1}^{T} p \cdot l_t.$$

Proof. In order to prove equality, we will prove inequality in both directions.

$\geq$: Trivial, since we can choose p s.t. $p_i = 1$ for the best fixed expert i, which will result in an expected loss equal to the best fixed strategy.

$\leq$: Choose p. If $\exists i$ s.t. $p_i = 1$, then we are done since the two quantities are equal. Otherwise, there $\exists i, j$ s.t. $p_i > 0$ and $p_j > 0$. We can assume WLOG that $\sum_{t=1}^{T} l_{t,i} \leq \sum_{t=1}^{T} l_{t,j}$. Now, set $p'_i \leftarrow p_i + p_j$ and $p'_j \leftarrow 0$. Thus, $\sum_{t=1}^{T} p' \cdot l_t \leq \sum_{t=1}^{T} p \cdot l_t$.

We can repeat this process until $\exists i$ s.t. $p_i = 1$, and we are done since this condition corresponds to a fixed strategy. □

Conclusion: The expert advice problem is special case of online convex optimization with linear functions parameterized by $p \in \Delta_d$. Thus, Follow the Regularized Leader (FTRL) algorithms should yield better regret bounds, as we will see.

3 Follow the Regularized Leader for the Expert Advice Problem

3.1 Review of First Attempt

First, we will restate a theorem, we have already proved.

Theorem 2. *Suppose that $f_1, \dots, f_T$ are convex functions with regularizer $R(w)$, and we use online gradient descent to choose $w_1, \dots, w_T$ with $g_t \in \partial f_t(w_t)$ according to the online gradient descent algorithm. For any norm $\|\cdot\|$, let σ, G be constants s.t.*

1. *R is σ-strongly convex w.r.t. $\|\cdot\|$ and*
2. *$\forall t$, $\|g_t\|_* \leq G$.*

Then, $\forall u$,

$$\text{Regret}(u) \leq R(u) + \frac{TG^2}{\sigma}.$$

Plugging in $R(p) = \frac{1}{2\eta}\|p\|_2^2 + I_{\Delta_d}(p)$ and setting $\eta = \frac{1}{\sqrt{2dT}}$, we have already shown that $\text{Regret}(u) \leq \sqrt{2dT} = \mathcal{O}(\sqrt{T})$. Last time, we observed that we may be able to provide a better regret bound by using a norm that will result in a tighter bound for our dual norm.

3.2 FTRL with Negative Entropy Function

It turns out that a better choice for regularization is

$$R(p) = \frac{1}{\eta}\sum_{i=1}^{d} p_i \log p_i + \log d + I_{\Delta_d}(p),$$

which uses the negative entropy function and guarantees that our solution is in the probability simplex. How, though, should we choose our norm? We must make sure that it is still strongly convex, or our regret bound will be vacuous. Recall that last time, we showed $\frac{1}{2}\|w\|_2^2$ is 1-strongly convex w.r.t. $\|\cdot\|_2$. We will use the following fact to improve our regret bound.

Lemma 3. *$R(p) = \sum_{i=1}^{d} p_i \log p_i$ is 1-strongly convex w.r.t. $\|\cdot\|_1$ in Δ_d.*

Proof. It is sufficient to establish that $R(p)$ satisfies the following inequality, which follows from the definition of strong convexity: $R(q) \geq R(P) + \nabla R(P)\cdot(q-p) + \frac{1}{2}\|q-p\|_1^2$. Substituting, we have

$$\begin{aligned}\sum_{i=1}^{d} q_i \log q_i &\overset{?}{\geq} \sum_{i=1}^{d} p_i \log p_i + \sum_{i=1}^{d}(1+\log p_i)\cdot(q_i - p_i) + \frac{1}{2}\|p-q\|_1^2 \\ &\geq \sum_{i=1}^{d} q_i \log p_i + \frac{1}{2}\|p-q\|_1^2,\end{aligned}$$

since $\sum_{i=1}^{d} q_i = 1$ and $\sum_{i=1}^{d} p_i = 1$ (they are on the simplex). Rewriting,

$$\sum_{i=1}^{d} q_i \log \frac{q_i}{p_i} \overset{?}{\geq} \frac{1}{2}\|p-q\|_1^2.$$

We know that this inequality holds by Pinsker's inequality, a result from information theory that bounds the difference in terms of the Kullback-Leibler divergence, so we are done. □

We can use this result to apply the online convex optimization theorem with the $\|\cdot\|_1$. Since $\sum_{i=1}^{d} p_i \log p_i$ is 1-strongly convex w.r.t. $\|\cdot\|_1$ in Δ_d, $R(p) = \frac{1}{\eta}\sum_{i=1}^{d} p_i \log p_i + \log d + I_{\Delta_d}(p)$ is $\frac{1}{\eta}$-strongly convex w.r.t $\|\cdot\|_1$. Also, by the following theorem, we know $\|\cdot\|_1$ is dual to $\|\cdot\|_\infty$:

Theorem 4.

$$\|\cdot\|_p \text{ is dual to } \|\cdot\|_q \iff \frac{1}{p} + \frac{1}{q} = 1.$$

The gradient is simply the vector of losses $g_t = l_t$, due to $f_t = p_t \cdot l_t$. Since we are interested in the $\|\cdot\|_1$, in order to apply our theorem we consider the dual norm of the gradient $\|g_t\|_\infty \leq 1,\ \forall t$. (This inequality holds since we assume that losses are bounded by 1).

Applying our theorem, we have that

$$\begin{aligned}\text{Regret}(p) &\leq R(p) + \eta T \\ &= \frac{1}{\eta}\sum_{i=1}^{d} p_i \log p_i + \log d + I_{\Delta_d} + \eta T.\end{aligned}$$

Using the fact that $p \in \Delta_d$, we can bound $\sum_{i=1}^{d} p_i \log p_i \leq \log d$. Thus, we have

$$\text{Regret}(p) \leq \frac{1}{\eta}\log d + \log d + I_{\Delta_d} + \eta T.$$

Setting $\eta = \sqrt{\frac{\log d}{2T}}$, our final regret bound is

$$\text{Regret}(p) \leq 2\sqrt{T \log d}.$$

This result improves over our previous bound by reducing the factor of d under the square root to $\log d$.

4 Looking Ahead

Next time, we will present a closed form solution, so that we don't have to run gradient descent minimization at each step of the FTRL algorithm.