CS260: Machine Learning Theory
Lecture 12: No Regret and the Minimax Theorem of Game Theory

November 2, 2011

Lecturer: Jennifer Wortman Vaughan

1 Regret Bound for Randomized Weighted Majority

In the last class, we proved a general regret bound for Follow the Regularized Leader. In particular, for any arbitrary sequence of losses $\vec{\ell}_1, \cdots, \vec{\ell}_T$ with each $\ell_{i,t} \in [0, 1]$, let $\vec{p}_1, \cdots, \vec{p}_T$ be the distributions chosen by Follow the Regularized Leader with parameter η and regularizer R. Then for any $\vec{p} \in \Delta_n$,

$$\sum_{t=1}^{T} \vec{\ell}_t \cdot \vec{p}_t - \sum_{t=1}^{T} \vec{\ell}_t \cdot \vec{p} \leq \sum_{t=1}^{T} \vec{\ell}_t \cdot (\vec{p}_t - \vec{p}_{t+1}) + \frac{1}{\eta}(R(\vec{p}) - R(\vec{p}_1))$$

which implies that

$$\sum_{t=1}^{T} \vec{\ell}_t \cdot \vec{p}_t - \min_{p \in \Delta_n} \sum_{t=1}^{T} \vec{\ell}_t \cdot \vec{p} \leq \sum_{t=1}^{T} \vec{\ell}_t \cdot (\vec{p}_t - \vec{p}_{t+1}) + \frac{1}{\eta} \left(\max_{\vec{p} \in \Delta_n} R(\vec{p}) - \min_{\vec{p} \in \Delta_n} R(\vec{p}) \right) .$$

Today we will use this result to prove a regret bound of $O(\sqrt{T \log n})$ for Randomized Weighted Majority, which we know is a Follow the Regularized Leader algorithm with

$$R(\vec{p}) = -H(\vec{p}) = -\sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

To do this, we must bound the two terms on the right hand side of the bound above.

Step 1: Bounding the Range of the Regularizer

We begin by deriving upper and lower bounds on the entropy function $H(\vec{p})$. The lower bound is easy. Since for all i, $0 \leq p_i \leq 1$, $p_i \log \frac{1}{p_i} \geq 0$. (Remember that we define $0 \log(1/0)$ to be 0 by convention.) As we discussed before, $H(\vec{p}) = 0$ is achieved when $\vec{p}$ puts all of its weight on a single expert.

To upper bound $H(\vec{p})$, we can use Jensen's inequality. We get

$$H(\vec{p}) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leq \log \sum_{i=1}^{n} p_i \frac{1}{p_i} = \log \sum_{i=1}^{n} 1 = \log n \ .$$

This value is achieved when $\vec{p}$ is uniform.

We have that

$$\frac{1}{\eta} \left(\max_{\vec{p} \in \Delta_n} R(\vec{p}) - \min_{\vec{p} \in \Delta_n} R(\vec{p}) \right) = \frac{1}{\eta}(0 - (-\log n)) = \frac{1}{\eta} \log n \ .$$

All CS260 lecture notes build on the scribes' notes written by UCLA students in the Fall 2010 offering of this course. Although they have been carefully reviewed, it is entirely possible that some of them contain errors. If you spot an error, please email Jenn.

Step 2: Bounding the Stability Term

We will prove the following lemma, which gives a bound on the stability term for RWM. We make use of the fact that we can write the probability that RWM assigns to expert i at time t as

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{n} w_{j,t}}$$

where

$$w_{i,t} = e^{-\eta L_{i,t-1}} \ .$$

The constant 2 in this bound can be improved, but we won't worry about that here.

Lemma 1. *For any sequence of losses $\vec{\ell}_1, \cdots, \vec{\ell}_T$ with each $\ell_{i,t} \in [0,1]$, let $\vec{p}_1, \cdots, \vec{p}_T$ be the distributions chosen by Randomized Weighted Majority with parameter η. For all t,*

$$\vec{\ell}_t \cdot (\vec{p}_t - \vec{p}_{t+1}) \leq 2\eta \ .$$

Proof: First we note that this bound holds trivially for all $\eta \geq 1/2$ since $\vec{p}_t \in \Delta_n$ and $\vec{0} \leq \vec{\ell}_t \leq \vec{1}$ for all t. For the remainder of the proof, we consider the case in which $\eta < 1/2$.

Let's first think about the relationship between $w_{i,t}$ and $w_{i,t+1}$ for some i and t. We have

$$w_{i,t+1} = e^{-\eta L_{i,t}} = e^{-\eta(L_{i,t-1}+\ell_{i,t})} = e^{-\eta L_{i,t-1}} e^{-\eta \ell_{i,t}} = w_{i,t} e^{-\eta \ell_{i,t}} \ .$$

Since $\ell_{i,t} \in [0,1]$, $e^{-\eta} \leq e^{-\eta \ell_{i,t}} \leq 1$. Combining this with the last equation, we get

$$w_{i,t+1} \leq w_{i,t}$$

and

$$w_{i,t+1} \geq w_{i,t} e^{-\eta} \ .$$

Using these two bounds, we can relate $p_{i,t}$ to $p_{i,t+1}$. In particular, we have

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{n} w_{j,t}} \leq \frac{e^{\eta} w_{i,t+1}}{\sum_{j=1}^{n} w_{i,t+1}} = e^{\eta} p_{i,t+1} \ .$$

Since this holds for all experts i, we have

$$\begin{aligned}
\vec{\ell}_t \cdot (\vec{p}_t - \vec{p}_{t+1}) &\leq \vec{\ell}_t \cdot \big((e^{\eta}\vec{p}_{t+1}) - \vec{p}_{t+1}\big) \\
&= \vec{\ell}_t \cdot \big(e^{\eta} - 1\big)\vec{p}_{t+1} \\
&\leq (e^{\eta} - 1) \\
&\leq (e - 1)\eta \\
&\leq 2\eta
\end{aligned}$$

where the second inequality holds because we are only considering the case in which $\eta \leq 1/2$ (see Figure 1). The point of this step is to get some linear bound on $(e^{\eta} - 1)$; we're not worrying too much about the constants, and clearly this could be made tighter. $\square$

Exercise: During the break, someone in the class pointed out that this bound can be tightened to get $\vec{\ell}_t \cdot (\vec{p}_t - \vec{p}_{t+1}) \leq \eta$ *without* requiring the case analysis at all by using the inequality $1 + x \leq e^x$ that we saw, for example, in Lecture 2. It's true! Try working through this alternative analysis. (The first few steps remain the same.)

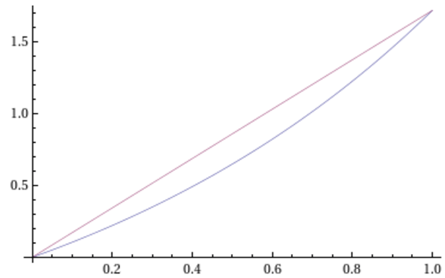

Figure 1: $e^\eta - 1 \le (e-1)\cdot\eta$ for $\eta \le 1$

Step 3: Putting it Together

We can now apply the general FTRL regret bound to get a regret bound for Randomized Weighted Majority:

$$\begin{aligned}\sum_{t=1}^{T} \ell_t \cdot \vec{p}_t - \sum_{t=1}^{T} \ell_t \cdot \vec{p} &\le \sum_{t=1}^{T} \big(\ell_t \cdot \vec{p}_t - \ell_t \cdot \vec{p}_{t-1}\big) + \frac{1}{\eta}\big(R(\vec{p}) - R(\vec{p}_1)\big) \\ &\le 2\eta T + \frac{1}{\eta}\log n \ .\end{aligned}$$

If we know the value T (or an upper bound on T) in advance, we can optimize the value of η to minimize this bound. Setting the derivative of the bound equal to 0, we get that it is minimized when

$$\eta = \sqrt{\frac{\log n}{2T}},$$

giving us a regret bound of $2\sqrt{2T\log n} = O(\sqrt{T\log n})$, which is sublinear as desired. Since the average regret per time step approaches 0 as T gets large, we call RWM a “no-regret” algorithm.

A few comments are in order. First, achieving this bound requires knowing T in advance. However, there are tricks that can be used to get around this (such as the so-called “doubling trick”), yielding bounds that are only slightly worse when T is unknown.

Second, although we only showed it for the particular case of RWM, it turns out that as long as the regularizer R is strongly convex, the upper bound on the regret for FTRL will always be sublinear in T. For example, sublinear regret can be achieved using $R(\vec{p}) = ||\vec{p}||_2$ (which turns out to be equivalent to the well-studied “Online Gradient Descent” algorithm).

2 The Minimax Theorem of Game Theory

We will now see how the existence of no-regret algorithms implies a key result from game theory, Von Neumann’s Minimax Theorem. We begin by reviewing the idea of a two-player zero-sum game.

A two-player zero-sum game is defined by a matrix $M \in R^{n\times m}$. Player 1 (the “row player”) chooses a row $i \in \{1,\cdots,n\}$, and Player 2 (the “column player”) chooses a column $j \in \{1,\cdots,m\}$. Player 1 then suffers a loss of $M_{i,j}$ (or equivalently receives the payoff $-M_{i,j}$), while Player 2 receives the payoff $M_{i,j}$. Table 1 shows the payoff matrix M for the well known game Rock-Paper-Scissors.

	R	P	S
R	0	+1	-1
P	-1	0	+1
S	+1	-1	0

Table 1: Matrix corresponding to the losses for player 1 for Rock-Paper-Scissors.

For a game like Rock-Paper-Scissors, there is an advantage to using a randomized strategy for each player. We can therefore think of Player 1 as choosing a distribution $\vec{p}$ over actions (rows) $1, \cdots, n$, and Player 2 choosing a distribution $\vec{q}$ over actions (columns) $1, \cdots, m$.

Consider the following two scenarios:

1. Player 1 announces a distribution $\vec{p}$, after which Player 2 announces a distribution $\vec{q}$. Player 1 receives the expected payoff $-\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j M_{i,j}$, while Player 2 receives $\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j M_{i,j}$.

2. Player 2 announces a distribution $\vec{q}$, after which Player 1 announces a distribution $\vec{p}$. Again, Player 1 receives the expected payoff $-\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j M_{i,j}$, while Player 2 receives $\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j M_{i,j}$.

Intuitively, it seems that the person who can choose their strategy second should have an advantage. However, one can show mathematically that neither player has an advantage if both play optimally. This is formalized in the following theorem, which we can prove using the existence of no-regret algorithms.

Theorem 1 (Von Neumann Minimax Theorem)**.** *For any $n \times m$ matrix M,*

$$\min_{\vec{p} \in \Delta_n} \max_{\vec{q} \in \Delta_m} \sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j M_{i,j} = \max_{\vec{q} \in \Delta_m} \min_{\vec{p} \in \Delta_n} \sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j M_{i,j} \ .$$

Proof: We will prove a special case of this theorem for matrices M with entries in $[0, 1]$, but this result can be generalized easily. (Exercise: Show how it can be generalized to any arbitrary matrix M.)

Suppose that two players play the zero-sum game defined by the matrix M repeated for a sequence of rounds $t = 1, \cdots, T$. At each round t, Player 1 chooses a distribution $\vec{p}_t$ and Player 2 chooses a distribution $\vec{q}_t$.

For every i and t, define $\ell_{i,t} = \sum_{j=1}^{m} q_{j,t} M_{i,j}$. This is the loss (negative payoff) that Player 1 would suffer for choosing the action i. We can then write the expected loss of Player 1 on round t as $\sum_{i=1}^{n} p_{i,t} \ell_{i,t} = \vec{p}_t \cdot \vec{\ell}_t$.

Suppose that Player 1 chooses $\vec{p}_t$ at each round by running a no-regret algorithm on this sequence of losses, and suppose that Player 2, knowing this, chooses the distribution $\vec{q}_t$ on each round that will hurt Player 1 the most. Since the average per time step regret of Player 1's algorithm goes to 0, we know that for any $\epsilon > 0$, it is possible to make T large enough so that

$$\frac{1}{T} \left(\sum_{t=1}^{T} \vec{p}_t \cdot \vec{\ell}_t - \min_{\vec{p} \in \Delta_n} \sum_{t=1}^{T} \vec{p} \cdot \vec{\ell}_t \right) \leq \epsilon \ .$$

Suppose we make T sufficiently large.

For any t,

$$\begin{aligned}
\min_{\vec{p}\in\Delta_n} \max_{\vec{q}\in\Delta_m} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} &\leq \max_{\vec{q}\in\Delta_m} \sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,t} q_j M_{i,j} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,t} q_{j,t} M_{i,j} \ .
\end{aligned}$$

This implies that

$$\begin{aligned}
\min_{\vec{p}\in\Delta_n} \max_{\vec{q}\in\Delta_m} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,t} q_{j,t} M_{i,j} \\
&= \frac{1}{T}\sum_{t=1}^{T} \vec{p}_t \cdot \vec{\ell}_t \\
&\leq \frac{1}{T} \min_{\vec{p}\in\Delta_n} \sum_{t=1}^{T} \vec{p} \cdot \vec{\ell}_t + \epsilon \\
&\leq \max_{\vec{q}\in\Delta_m} \min_{\vec{p}\in\Delta_n} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} + \epsilon \ .
\end{aligned}$$

We can drive ϵ arbitrarily close to 0 and this bound will still hold, so we must have that

$$\min_{\vec{p}\in\Delta_n} \max_{\vec{q}\in\Delta_m} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} \leq \max_{\vec{q}\in\Delta_m} \min_{\vec{p}\in\Delta_n} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} \ .$$

It remains to show the opposite, that is, that

$$\max_{\vec{p}\in\Delta_n} \min_{\vec{q}\in\Delta_m} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} \leq \min_{\vec{q}\in\Delta_m} \max_{\vec{p}\in\Delta_n} \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j M_{i,j} \ .$$

This is the more intuitive direction that says that the loss that Player 1 can guarantee himself when playing second is no more than the loss that Player 1 can guarantee himself when playing first. The proof is left as an exercise. $\square$