COS 511: Theoretical Machine Learning

Lecturer: Rob Schapire
Scribe: Geoffrey Roeder

Lecture #24
May 1, 2019

1 Last Lecture Review

We consider the setting of game theory and draw connections to machine learning in general, especially online learning and boosting. Let M be a matrix with all elements in $[0, 1]$, representing a game. As usual, consider the two players Max and Mindy. Recall that Mindy's goal is to choose a distribution P over rows of M, and plays $i \sim P$, to minimize her expected loss. Max chooses a distribution Q over columns and plays $j \sim Q$ to maximize the same loss. We analyze the loss from the perspective of Mindy, whose expected loss is

$$\sum_{i,j} P(i)M(i,j)Q(j) = P^\top MQ := M(P,Q).$$

Recall the following notation: $M(P, j)$ denotes expected loss when Mindy plays, and Max plays a pure strategy. Similarly, $M(i, Q)$ denotes expected loss when Max plays a mixed strategy, and Mindy plays a pure one. Although each of these is an expected loss, we will often refer to them simply as a "loss".

Also, in what follows, when taking a maximum or minimum, it should be understood that our notation means that we are taking maximum or minimum over all mixed strategies P, Q, or over all pure strategies i, j.

2 Fundamental Theorem of Zero-Sum Games

What does the optimal strategy look like for Mindy? Recall that $M(P, Q)$ denotes expected loss when Mindy plays P and Max plays Q. Let's suppose Max makes the first move. Max *knows* P, and so chooses Q to inflict the maximum loss on Mindy. We can write this loss as

$$\max_Q M(P,Q) = \max_j M(P,j),$$

where the equality holds because, for a fixed P, the maximum over a distribution of all strategies is equivalent to the maximum over the pure strategies since the maximum is always going to be realized by a choice of pure strategy. Then, Mindy will try to minimize this loss with her play:

$$\min_P \max_Q M(P,Q).$$

By similar reasoning, if Max plays first, then the resulting loss would be

$$\max_Q \min_P M(P,Q).$$

This naturally leads to the question, which is better: to play first or last? One might think that to play second is better, because more information will be available. Interestingly, though, it does not matter whether one plays first or second. This has been proved in a theorem due to von Neumann, sometimes called the Minimax Theorem, or the Fundamental Theorem of Zero-Sum Games:

$$\min_P \max_Q M(P,Q) = \max_Q \min_P M(P,Q) := v,$$

where v is the value of game M. A more precise statement of the theorem is $\exists P^*$ s.t. $\forall Q,\ M(P^*,Q) \leq v$, that is, Mindy has some strategy P^* that is the argmin strategy: for any choice of Q by an adversary, the loss is bounded above by the value of the game. P^* is optimal in the sense that Max also has a Q^* that is optimal. Formally, $\exists Q^*$ s.t. $\forall P,\ M(P,Q^*) \geq v$, which means that Max can force Mindy to suffer loss at least v, no matter how she plays. These solutions are also called Nash equilibria.

2.1 Multiplicative Weights Algorithm

According to classical game theory, one should determine P^* through, say, a linear program, and play according to that. There are some problems with that formulation, though. Perhaps we don't know M, or it is prohibitively large, so that we can't form and solve the LP in a reasonable amount of time. We have also made a very strong assumption, that our opponent is both extremely smart, and only wishes to inflict the maximum loss on us (e.g., has no other aims independent of the game that would affect their choice of play). This might be unrealistic.

How can we do a better analysis? Imagine playing over and over again against the same opponent. Then we can learn the characteristics of this opponent, and adapt. Formally, let's consider a game M with n rows, and use the following Multiplicative Weights (MW) algorithm:

```
Initialize $P_1(i) = 1/n$
Fix $\beta \in [0,1]$
For $t = 1, \ldots, T$:
    Mindy (the learner) chooses $P_t$
    Max (the environment or adversary) chooses $Q_t$ [knowing $P_t$]
    Mindy observes $M(i, Q_t)$ for all $i$
    Mindy's loss is $M(P_t, Q_t)$
    $\forall i : P_{t+1}(i) = P_t(i) \cdot \beta^{M(i,Q_t)}/\text{norm}$
```

Consider how this algorithm works: we maintain a distribution P_t that is initially uniform, and then update it multiplicatively according to the loss that would have been incurred by any given pure strategy. Note that in this algorithm, Mindy gets to observe the entire column of results for any choice she would have played. The resulting loss for Mindy over the T iterations is $\sum_{t=1}^{T} M(P_t, Q_t)$.

We would like some kind of regret bound on this quantity, one that compares the actual loss received with the loss on the best possible fixed strategy if Mindy had actually known each Q_t ahead of time. In other words, we want

$$\sum_{t=1}^{T} M(P_t, Q_t) \leq \min_P \sum_{t=1}^{T} M(P, Q_t) + \text{small amount}$$

In fact, we do achieve this using the MW algorithm, and can prove the following using similar techniques as for RWMA:

$$\sum_{t=1}^{T} M(P_t, Q_t) \leq a_\beta \min_P \sum_{t=1}^{T} M(P, Q_t) + c_\beta \ln n,$$

where the constants a_β, c_β are the same as in RWMA. As a corollary, if we set $\beta = 1/(1 + \sqrt{2 \ln n/T})$, and let $\Delta_T = \mathcal{O}(\sqrt{\ln n/T})$, then

$$\sum_{t=1}^{T} M(P_t, Q_t) \leq \min_P \sum_{t=1}^{T} M(P, Q_t) + \Delta_T$$

Note that $M(P^*, Q_t) \leq v$ always, and so the first term on the right is at most v. This shows MW never does much worse than what we would get using the classical approach to game theory. But, it's also possible that the algorithm will do much better if playing against an opponent who is not fully optimal or adversarial.

2.2 Proof of Fundamental Theorem of Zero-Sum Games

We now proceed to prove von Neumann's minmax theorem using MW and its analysis. For each round t, we assume that Mindy picks P_t using the MW algorithm just described, and we further assume that Max picks Q_T as

$$\operatorname*{argmax}_Q M(P_t, Q).$$

This is called the "best response". For the analysis that follows, let

$$\overline{P} = \frac{1}{T} \sum_{t=1}^{T} P_t$$
$$\overline{Q} = \frac{1}{T} \sum_{t=1}^{T} Q_t.$$

We need to prove that $\max\min \leq \min\max$ and $\min\max \geq \max\min$. The result that $\max\min \leq \min\max$ is much more straightforward based on our intuition that playing second is always better (or at least not worse) than playing first. We focus here on the second, less straightforward inequality:

$$\begin{aligned}
\min_P \max_Q P^\top MQ &\le \max_Q \overline{P}^\top MQ && \text{(viewing the LHS as a function of } P\text{)}\\
&= \max_Q \left(\frac{1}{T}\sum_T P_t\right)^\top MQ\\
&\le \frac{1}{T}\sum_t \max_Q P_t^\top MQ && \text{(taking max at each } t \text{ could only be a bigger sum)}\\
&= \frac{1}{T}\sum_t P_t^\top MQ_t && \text{(by assumption of how } Q_t \text{ was picked)}\\
&\le \min_P \frac{1}{T}\sum_t P^\top MQ_t + \Delta_T && \text{(using regret bound for MW)}\\
&= \min_P P^\top M\overline{Q} + \Delta_T\\
&\le \max_Q \min_P P^\top MQ + \Delta_T && \text{(viewing LHS as function of Q).}
\end{aligned}$$

As defined, $\Delta_T \to 0$ as $T \to \infty$, concluding the proof. This proof additionally tells us (from the chain of inequalities above) that

$$\max_Q M(\overline{P}, Q) \le \max_Q \min_P M(P,Q) + \Delta_T = v + \Delta_T.$$

This means that if Mindy plays with $\overline{P}$, the loss is no more than $v + \Delta_T$. This means that for any strategy Q that Max might play, if Mindy plays $\overline{P}$ then she will get within Δ_T of v, which is what she would get if she played using P^*. So, it is in that sense that $\overline{P}$ is an approximate min-max strategy. Similarly, $\overline{Q}$ is an approximate max-min strategy.

3 Connections to Online Learning

In this section we draw connections between game theory and the online learning setting studied throughout the course. Recall the Online Learning problem:

```
For $t = 1, \dots, T$:
    Observe $x_t \in \mathcal{X}$
    Predict $\hat{y}_t \in \{0,1\}$
    Observe true label $c(x_t) \in \{0,1\}$
```

Note that a mistake is defined as $\hat{y}_t \neq c(x_t)$. We want our algorithm to do almost as well as the best $h \in \mathcal{H}$, that is

$$\#\text{mistakes} \le \min_{h\in\mathcal{H}}(\#\text{mistakes of } h) + \text{small amount}$$

We define M in this case to have rows indexed by $h \in \mathcal{H}$ and columns indexed by $x \in \mathcal{X}$. Then, the loss $M(h,x) = 1\{h(x) \neq c(x)\}$. Using this game, we can apply MW as follows:

```
For $t = 1, \ldots, T$:
    Use MW on matrix $M$ to get $P_t$
    Pick random $h \sim P_t$
    Let $\hat{y}_t = h(x_t)$
    Take $Q_t$ to be concentrated on $x_t$ (1 on $x_t$, 0 otherwise)
```

We can now apply the analyses developed earlier to see that

$$\begin{aligned}\sum_t M(P_t, x_t) &\leq \min_P \sum_t M(P, x_t) + \mathcal{O}\left(\sqrt{T \ln |\mathcal{H}|}\right) \\ &= \min_{h \in \mathcal{H}} \sum_t M(h, x_t) + \mathcal{O}\left(\sqrt{T \ln |\mathcal{H}|}\right) \quad \text{(min will be realized by some } h\text{).}\end{aligned}$$

Notice that

$$\begin{aligned}M(P_t, x_t) &= \mathbb{E}_{h \sim P_t}\left[M(h, x_t)\right] \\ &= \Pr_{h \sim P_t}\left[h(x_t) \neq c(x_t)\right] = \Pr\left[\hat{y}_t \neq c(x_t)\right].\end{aligned}$$

We can make the connection more explicit by noticing that $\sum_t M(h, x_t)$ is the number of mistakes that h makes, so we can rewrite the regret bound from MW as

$$\mathbb{E}[\text{\# of mistakes learner makes}] \leq \min_{h \in \mathcal{H}}(\text{\# mistakes } h \text{ makes}) + \mathcal{O}(\sqrt{T \ln |\mathcal{H}|})$$

This is exactly the statement about the expected number of mistakes of the learner that we wanted to show above.

3.1 Connection to Boosting

We can also draw connections between boosting and the game theoretic setting discussed at the start of these notes. Let us take $\mathcal{H}$ to be the weak hypothesis space, and $\mathcal{X}$ to be the training set. (It is unusual for us to write the training set in this way, but the reason will become clear shortly.) Recall the boosting problem:

```
For $t = 1, \ldots, T$:
    Booster picks $D_t$ over $\mathcal{X}$
    Weak learner picks $h_t \in \mathcal{H}$ s.t. $\Pr_{x \sim D_t}[h_t(x) \neq c(x)] \leq \frac{1}{2} - \gamma$
    Output $H = \text{MAJORITY}(h_1, \ldots, h_T)$
```

To fit this into the game theoretic framework, we first note that the game M must be transformed, because we want to have a distribution over the $\mathcal{X}$ samples rather than hypotheses, as in online learning. Hence, we define

$$M' = 1 - M^\top,$$

which is a matrix whose entries represent exactly the function $1\{h(x) = c(x)\}$. This is exactly the same game as M, but with the roles of the row and column players reversed. Specifically, we take the transpose to reverse rows and columns, then we negate the matrix to reverse what is being minimized or maximized (since the row player always wants to minimize loss), and finally we add 1 so that entries of the new matrix will be in $[0, 1]$.

This leads to the following game-theoretic style of algorithm for solving the boosting problem:

```
For $t = 1, \ldots, T$:
    Use MW on $M'$ to compute $P_t$
    Let $D_t = P_t$
    Get $h_t$ from the weak learner
    Set $Q_t$ to be the distribution concentrated on $h_t$
```

Then, we have that

$$\begin{aligned} M'(P_t, h_t) &= \mathbb{E}_{x \sim P_t}\left[M'(x_t, h_t)\right] \\ &= \Pr_{x \sim D_t}\left[h_t(x) = c(x)\right] && \text{(because of how } M' \text{ is defined)} \\ &\geq \frac{1}{2} + \gamma && \text{(by weak learning assumption)} \end{aligned}$$

Further, we know that the following relationship holds from our regret bound on MW:

$$\frac{1}{2} + \gamma \leq \frac{1}{T}\sum_t M'(P_t, h_t) \leq \min_{x \in \mathcal{X}} \frac{1}{T}\sum_t M'(x, h_t) + \Delta_T$$

That is, $\forall x \in \mathcal{X}$:

$$\frac{1}{T}\sum_t M'(x, h_t) \geq \frac{1}{2} + \gamma - \Delta_T > \frac{1}{2},$$

where $M'(x, h_t)$ is exactly $1\{h_t(x) = c(x)\}$, making the left-hand side exactly the fraction of weak hypotheses h_t that are correct on x. The rightmost inequality follows for Δ_T smaller than γ (in particular, for $T = \Omega(\ln|\mathcal{X}|/\gamma^2)$). This implies that more than half of the weak hypotheses h_t are correct on x, and so the majority vote itself will be correct. Then, $H(x) = c(x) \forall x \in \mathcal{X}$, as desired. In other words, after T rounds of boosting, the training error of H will be zero.

We have seen, in the preceding notes, the connections between a game theoretic setting, online learning, and boosting. And we have seen that these last two are in fact duals, in that they use the same game matrix.