Advanced Course in Machine Learning Spring 2010

Playing games with Hannan, Von-Neumann, and Blackwell

Handouts are jointly prepared by Shie Mannor and Shai Shalev-Shwartz

In this lecture we briefly review several concepts in game theory and present the relation to online learning.

1 Two-person zero-sum game

The game can be described as follows. Let L be an $n \times m$ matrix with $|L_{i,j}| \le 1$ for all i, j. The game is played by two players which we call a row player and a column player. The row player chooses a row index, i, and the column player chooses a column index, $j \in [m]$. The outcome of the game is $L_{i,j}$. The row player thinks on $L_{i,j}$ as its loss, while the column player thinks on $L_{i,j}$ as its gain.[1] The player are allowed to choose a "mixed strategy" instead of a "pure strategy". For the row player, this means that instead of choosing a row i the player can choose a distribution p over $[n]$. Similarly, the column player can choose a distribution q over $[m]$. In such a case, we measure the expected outcome of the round,

$$p^\intercal L q = \sum_{i,j} p_i q_j L_{i,j} \; .$$

The best strategy of the row player is to choose p such that no matter which q the column player will play we will have $p^\intercal L q$ as small as possible. That is, to choose p that minimizes:

$$\min_{p \in \Delta^n} \max_{q \in \Delta^m} p^\intercal L q \; ,$$

where Δ^n is the n-dimensional probability simplex. Similarly, the best strategy of the column player is

$$\max_{q \in \Delta^m} \min_{p \in \Delta^n} p^\intercal L q \; .$$

Von-Neumann proved that the two expressions are equal, namely,

$$\min_{p \in \Delta^n} \max_{q \in \Delta^m} p^\intercal L q \;=\; \max_{q \in \Delta^m} \min_{p \in \Delta^n} p^\intercal L q \; .$$

The common value is called the *value of the game*. There are many ways to prove Von-Neumann's mini-max theorem, e.g. using a strong duality argument. In the next section we will outline another proof that follows from our low-regret strategies for online convex optimization.

2 Playing repeated games

A two-person repeated game is played repeatedly such that at each round t, the row player picks $p_t \in \Delta^n$, the column player picks $q_t \in \Delta^m$, and the outcome of the round is $p_t^\intercal L q_t$. The regret of the row player is defined as

$$\frac{1}{T} \sum_{t=1}^{T} p_t^\intercal L q_t - \min_p \frac{1}{T} \sum_{t=1}^{T} p^\intercal L q_t \; .$$

We say that a strategy is *Hannan consistent* if the regret is $o(1)$, regardless of how the column player behaves. Note that for each t, the function $g_t(p) = p^\intercal L q_t = \langle p, L q_t \rangle$ is a linear (hence convex) function. Therefore, we can apply online convex optimization procedures described in previous lectures to obtain a Hannan consistent hypothesis.

[1] Hence the name "zero-sum", which means that the loss minus gain is zero.

Exercise: Use the existence of Hannan consistent strategies to prove Von-Neumann's minimax theorem.

3 Blackwell's approachability

Blackwell proposed a generalization of the problem of playing repeated two-player zero-sum games. The difference is that now each loss, $L_{i,j}$ is a vector in the unit ℓ_2 ball of $\mathbb{R}^p$ rather than a scalar in $[-1, 1]$. As before, we allow mixed strategies, for which the loss is $\sum_{i,j} p_i q_j L_{i,j} \in \mathbb{R}^p$. We use the notation $L(p, q)$ to denote $\sum_{i,j} p_i q_j L_{i,j}$. For this general game, we define the regret of the row player with respect to a subset S of the unit ball of $\mathbb{R}^p$ to be

$$d\left(\tfrac{1}{T}\sum_{t=1}^{T} L(p_t, q_t)\,,\, S\right) ,$$

where $d(u, S) = \min_{v \in S} \|u - v\|$.

A set S is *approachable* if the row player can guarantee an $o(1)$ regret. Blackwell characterized which convex sets are approachable.

Theorem 1 *Let S be a closed and convex subset of the unit ball of $\mathbb{R}^p$. Then, S is approachable if and only if for all unit vector $a \in \mathbb{R}^p$ and scalar $c \in \mathbb{R}$ such that the halfspace $H = \{x : \langle a, x\rangle \le c\}$ contains S we have*

$$\min_{p \in \Delta^n} \max_{q \in \Delta^m} \langle a, L(p, q)\rangle \le c .$$

Proof First, assume that S is approachable. Then, exists $v \in S$ such that $\|\frac{1}{T}\sum_{t=1}^{T} L(p_t, q_t) - v\| \le \epsilon$, where ϵ is arbitrarily small (for large T). Take any H s.t. $S \subset H$. Then, $\langle a, v\rangle \le c$. It follows that

$$c - \epsilon \ge \langle a, \tfrac{1}{T}\sum_{t=1}^{T} L(p_t, q_t)\rangle = \tfrac{1}{T}\sum_{t=1}^{T}\sum_{i,j} p_t^{\intercal} \tilde{L} q_t$$

where $\tilde{L}$ is a matrix with $\tilde{L}_{i,j} = \langle a, L_{i,j}\rangle$. Since the above holds for any $\epsilon > 0$ we get that there exists a strategy for the row player in the scalar game such that its asymptotic average loss is bounded by c. It follows that c upper bounds the value of the game, that is,

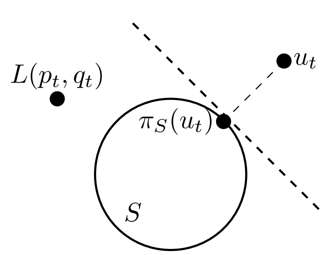
$$c \ge \min_{p \in \Delta^n} \max_{q \in \Delta^m} p^{\intercal} \tilde{L} q ,$$

which implies the desired result by the definition of $\tilde{L}$.

Now, assume that for any unit vector $a \in \mathbb{R}^p$ and scalar $c \in \mathbb{R}$ such that the halfspace $H = \{x : \langle a, x\rangle \le c\}$ contains S we have

$$\min_{p \in \Delta^n} \max_{q \in \Delta^m} \langle a, L(p, q)\rangle \le c .$$

The row player will play the following strategy. Let $u_t = \frac{1}{t-1}\sum_{\tau < t} L(p_\tau, q_\tau)$ and let $\pi_S(u_t)$ be the point in S closest to u_t. If $u_t \notin S$ then the hyperplane defined by $a = \frac{u_t - \pi_S(u_t)}{\|u_t - \pi_S(u_t)\|}$ and $c = \langle a, \pi_S(u_t)\rangle$ contains S. See illustration below:



So, if the player will play p_t in

$$\operatorname*{argmin}_{p\in\Delta^n}\max_{q\in\Delta^m}\langle a, L(p,q)\rangle\ ,$$

then no matter what the value of q_t is, we will have $\langle a, L(p_t,q_t)\rangle \le c$. Thus,

$$\begin{aligned}
d(u_{t+1},S)^2 &= \|u_{t+1}-\pi_S(u_{t+1})\|^2\\
&\le \|u_{t+1}-\pi_S(u_t)\|^2\\
&= \|\tfrac{t-1}{t}u_t + \tfrac{1}{t}L(p_t,q_t)-\pi_S(u_t)\|^2\\
&= \|\tfrac{t-1}{t}(u_t-\pi_S(u_t)) + \tfrac{1}{t}(L(p_t,q_t)-\pi_S(u_t))\|^2\\
&= (\tfrac{t-1}{t})^2 d(u_t,S)^2 + \tfrac{1}{t^2}\|L(p_t,q_t)-\pi_S(u_t)\|^2 + 2\tfrac{t-1}{t^2}\langle u_t-\pi_S(u_t), L(p_t,q_t)-\pi_S(u_t)\rangle\\
&\le (\tfrac{t-1}{t})^2 d(u_t,S)^2 + \tfrac{1}{t^2}\|L(p_t,q_t)-\pi_S(u_t)\|^2\ .
\end{aligned}$$

Additionally, since everything is assumed to be in the unit ball we get that

$$d(u_{t+1},S)^2 \ \le\ (\tfrac{t-1}{t})^2 d(u_t,S)^2 + \tfrac{4}{t^2}\ .$$

The above inequality also holds if $u_t \in S$ because then,

$$d(u_{t+1},S)^2 \le \|u_{t+1}-u_t\|^2 = \tfrac{1}{t^2}\|-u_t + L(p_t,q_t)\|^2 \le \tfrac{4}{t^2}\ .$$

Multiplying by t^2, summing over t, and rearranging, we obtain

$$\sum_t \left(t^2 d(u_{t+1},S)^2 - (t-1)^2 d(u_t,S)^2\right) \le 4T\ .$$

The sum on the left side telescopes and becomes $T^2 d(u_{T+1},S)^2$. Thus,

$$d(u_{T+1},S)^2 \le 4/T\ ,$$

which concludes our proof. ■

4 Exercises

1. Show that Von-Neumann's minimax theorem follows from Blackwell's approachability theorem.
2. Show that the existence of Hannan consistent strategies follows from Blackwell's approachability theorem.