

CSC 446 Notes: Lecture 22

Typed by Shujie Chen

April 19, 2012

1 Introduction

Game theory is not like the reinforcement learning which we have just talked about. In the game theory, the condition that one is playing against someone else is considered. Furthermore, game theory is not about learning, it does not learn from data or previous instances, but tries to figure out what to do given a specific game rule.

A game could be described using a game matrix. The two players are called row player and column player. The value of each element in the matrix denotes the reward for the player. Such a matrix model could be applied to many situations, e.g., board games, bets. A good example would be the rock-paper-scissors game, which has a matrix M like

	Rock	Paper	Scissors
Rock	0	-1	1
Paper	1	0	-1
Scissors	-1	1	0

Note that here, the game has the same set of choices for both players, which is not necessary in general. Moreover, the above matrix satisfies that

$$M = -M^T,$$

which makes the game symmetric. Also, we could find that this is a zero-sum game, where one player gets his/her reward from the other, with the total amount never changed.

2 Definition of the Problem

Define \mathbf{p} as the row player's strategy, and \mathbf{q} to be that of the column player. \mathbf{p} and \mathbf{q} are both probability distributions. Then, the expected reward with respect to the strategies is

$$E(\mathbf{p}, \mathbf{q}) = \sum_i p_i \sum_j q_j m_{ij}.$$

If we consider the problem from the perspective of the row player (which will be the same if we choose the column player), our goal is to maximize the reward E of the row player, i.e., to find a strategy \mathbf{p} that maximizes the row value v_r and

$$v_r = \max_{\mathbf{p}} \min_{\mathbf{q}} E(\mathbf{p}, \mathbf{q}),$$

where the minimization means that our rival (the column player) does whatever that is worst for us, and the maximization chooses the strategy that is best for us given the strategy of the column player. Similarly, the column value v_c is

$$v_c = \min_{\mathbf{q}} \max_{\mathbf{p}} E(\mathbf{p}, \mathbf{q}),$$

which reflects the fact that the row player does whatever that could maximize his/her reward, and that the column player chooses a best strategy to minimize it.

3 Some Simplification

Before solving the optimization problem, we should first do some simplification by showing that, when assuming the column player's strategy, we do not have to consider all the variations of \mathbf{q} .

If we choose one possibility of \mathbf{q} with probability of 1, i.e., the column player only does one action, say, action j , from the available action set, we have

$$\min_{\mathbf{q}} E(\mathbf{p}, \mathbf{q}) \leq \min_j E(\mathbf{p}, j), \quad (1)$$

where $E(\mathbf{p}, j) = \sum_i p_i m_{ij}$ from its definition.

Now, let us consider the strategy of the column player using the structure of v_c , i.e., the column player assumes a \mathbf{p} at first, and then choose an action $l = \operatorname{argmin}_j E(\mathbf{p}, j)$ to make a worst result for the row player. This leads to

$$\forall \mathbf{p}, \mathbf{q}, \quad E(\mathbf{p}, \mathbf{q}) = \sum_j q_j E(\mathbf{p}, j) \geq \sum_j q_j E(\mathbf{p}, l) = E(\mathbf{p}, l),$$

which could be rewritten as

$$\min_{\mathbf{q}} E(\mathbf{p}, \mathbf{q}) \geq \min_j E(\mathbf{p}, j). \quad (2)$$

Combining Equation 1 and Equation 2, we have

$$\min_{\mathbf{q}} E(\mathbf{p}, \mathbf{q}) = \min_j E(\mathbf{p}, j),$$

which means that instead of considering all possibilities of \mathbf{q} , we could consider only one move at a time. In the context of the Rock-Paper-Scissors game, this is to say that $\mathbf{q} = (1, 0, 0)$ or $(0, 1, 0)$ or $(0, 0, 1)$.

Using this equality, the row value v_r that we want to maximize is simplified as

$$\max_{\mathbf{p}} \min_{\mathbf{q}} E(\mathbf{p}, \mathbf{q}) \Rightarrow \max_{\mathbf{p}} \min_j E(\mathbf{p}, j) = \max_{\mathbf{p}} \min_j \sum_i p_i m_{ij}.$$

Similarly, v_c could be simplified as

$$\min_{\mathbf{q}} \max_i \sum_j q_j m_{ij}.$$

4 Nash Equilibrium

The objective functions v_r and v_c that have been simplified as above still contain two min/max operations, which makes them too complicated to solve. In order to get rid of one of the min/max, we could introduce a variable v and adding a new constraint for compensation. If we do the same for the case of v_c , we could come up with

$$\begin{array}{ll} \max_{\mathbf{p}, v} & v \\ \text{s.t.} & \sum_i p_i m_{ij} \geq v \quad \forall j \\ & \sum_i p_i = 1 \\ & p_i \geq 0 \quad \forall i \end{array} \qquad \begin{array}{ll} \min_{\mathbf{q}, v} & v \\ \text{s.t.} & \sum_j q_j m_{ij} \leq v \quad \forall i \\ & \sum_j q_j = 1 \\ & q_j \geq 0 \quad \forall j \end{array}$$

where the left one is for v_r and right one v_c . For the left one, we could simplify a little bit more by firstly adding onto the game matrix a constant such that all of its entries are positive, and define

$$y_i = p_i/v, \quad \forall i,$$

then the model is updated as the left one below

$$\begin{array}{ll}
\min_{\mathbf{y}} & \sum_i y_i \\
\text{s.t.} & \sum_i y_i m_{ij} \geq 1 \quad \forall j \\
& y_i \geq 0 \quad \forall i
\end{array}
\qquad
\begin{array}{ll}
\max_{\mathbf{x}} & \sum_j x_j \\
\text{s.t.} & \sum_j x_j m_{ij} \leq 1 \quad \forall i \\
& x_j \geq 0 \quad \forall j
\end{array}$$

The model on the right is obtained by introducing $\forall j, x_j = q_j/v$ for the v_c model. What we just came up with are actually special cases of the following general models

$$\begin{array}{ll}
\min_{\mathbf{y}} & \mathbf{b}^T \mathbf{y} \\
\text{s.t.} & \mathbf{A}^T \mathbf{y} \geq \mathbf{c} \\
& \mathbf{y} \geq 0
\end{array}
\qquad
\begin{array}{ll}
\max_{\mathbf{x}} & \mathbf{c}^T \mathbf{x} \\
\text{s.t.} & \mathbf{A} \mathbf{x} \leq \mathbf{b} \\
& \mathbf{x} \geq 0
\end{array}$$

which are a pair of duals in the linear programming (LP) theory, and thus have the same optimization result. The set of strategy (\mathbf{p}, \mathbf{q}) under such condition is called Nash Equilibrium, which could be mathematically described as

$$\begin{array}{l}
(\mathbf{p}, \mathbf{q}) \\
\text{s.t.} \quad E(\mathbf{p}, \mathbf{q}) = \max_{\mathbf{p}'} E(\mathbf{p}', \mathbf{q}) \\
\quad \quad E(\mathbf{p}, \mathbf{q}) = \max_{\mathbf{q}'} E(\mathbf{p}, \mathbf{q}')
\end{array}$$

When the Nash Equilibrium has been reached, both players know the strategy of each other, and they all get the best reward using their current strategies against those of the rivals. Hence neither of them need to change the current strategy.

5 Proof of the LP Duality

The LP duality that we used to reach the Nash Equilibrium is proved as follows. We will start from the previous linear programming model on the left and will get the other one using the primal-dual method.

Firstly, the Lagrangian of the model is calculated as

$$\begin{aligned}
L(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \sum_i y_i + \sum_j \lambda_j (1 - \sum_i y_i m_{ij}) - \sum_i \mu_i y_i \\
&= \sum_j \lambda_j + \sum_i y_i (1 - \sum_j \lambda_j m_{ij} - \mu_i)
\end{aligned}$$

Then by defining

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{y}} L(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\infty, & \text{if } \exists i, 1 - \sum_j \lambda_j m_{ij} - \mu_i \neq 0 \\ \sum_j \lambda_j, & \text{o.w.} \end{cases}$$

we will have the dual problem as below

$$\begin{array}{ll}
\max & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_j \lambda_j \\
\text{s.t.} & 1 - \sum_j \lambda_j m_{ij} - \mu_i = 0, \quad \forall i \\
& \lambda_j \geq 0, \quad \forall j \\
& \mu_i \geq 0, \quad \forall i
\end{array}$$

Due to the constraint $\forall i, \mu_i \geq 0$, we could remove μ_i in the first constraint and update it as

$$1 - \sum_j \lambda_j m_{ij} \geq 0 \Rightarrow \sum_j \lambda_j m_{ij} \leq 1$$

If we replace $\boldsymbol{\lambda}$ with \boldsymbol{x} , this dual is exactly the same as the linear programming model for v_c , which proves that the previous two LP models are dualities. Note that here, we did not do the proof for general cases, which is easy to get using the variables of \boldsymbol{b} and \boldsymbol{c} , which is set to $\mathbf{1}$ in our proof.

Recall that we also applied the primal-dual in the SVM, the difference is that in SVM, we go to the domain of the dual to make the problem easier to solve, while here, we generally use the duality to prove that the two model are duals, which gives us the Nash Equilibrium.