Machine Learning **Fall 2023**

Lecture 7: Minimax Theorem and Duality

Lecturer: Liwei Wang *Scribe: Weijie Guo, Yutong Wu, Weicheng Lin, Xuezhi Huang, Hao Yang, Jiaqi Pan, Zhaobin Li*

7.1 Game Theory and Minimax Theorem

To recap, we introduced the case of a zero-sum, one-shot, two player matrix game. The game is described by the payoff matix M, whose element m_{ij} denotes the value player A sends to player B when actions i and j are chosen respectively.

Today, we continue our discussion on the result of such games,

7.1.1 Pure Strategy

When executing a **pure strategy**, a player only makes deterministic choices. Here's how such a game would unfold:

1. Alice chooses a row i.
2. Bob, after observing Alice's strategy (in this case, row i), chooses a column j.
3. Alice pays M_{ij} to Bob.

Since we assume both players are rational agents, the result is simple:

- When Alice goes first, $\min_i \max_j M_{ij}$ is payed.
- When Bob goes first, $\max_j \min_i M_{ij}$ is payed.

It can be proved that the second player always have the upper hand. In mathematical terms,

$$\min_i \max_j M_{ij} \geq \max_j \min_i M_{ij}$$

7.1.2 Mixed Strategy

A **mixed strategy** can be viewed as a probabilistic combination of pure strategies. A mixed strategy game would proceed as follows:

1. Alice chooses a probability distribution p over the rows.

2. Bob, after observing Alice's strategy (i.e. p), chooses probability distribution q over the columns.
3. Alice pays Bob $p^\top Mq$.

Since the choices are probabilistic, $p^\top Mq$ is the expectation of final results. Again, assuming Alice and Bob are rational, we get:

- When Alice goes first, an expected $\min_p \max_q p^\top Mq$ is payed.
- When Bob goes first, an expected $\max_q \min_p p^\top Mq$ is payed.

Note that p, q cannot be any arbitrary vector, but are rather probability vectors with non-negative entries that add up to one.

The eminent question is: what's the relationship between these values? Does the second player still hold an advantage? John von Neumann answered this is his 1928 paper[VN28].

Theorem 7.1 (John von Neumann Minimax Theorem)

1. $\min_p \max_q p^\top Mq = \max_q \min_p p^\top Mq$
2. *Equivalently,* $\exists(p^*, q^*)$ *s.t.* $\forall p, q,\ p^*Mq \le p^*Mq^* \le pMq^*$*, and* (p^*, q^*) *is the equilibrium.*

The original proof was given via a generalization of the Brouwer fixed-point theorem. Although topology is beyond the scope of this course, a proof using ML theory will be given in future lectures.

We also consider a generalization of this theorem, given by Maurice Sion[S58], which would soon come in handy in our following discussion on Lagrange duality.

Theorem 7.2 (Sion's Minimax Theorem)

Let $f(x, y)$ *be a function. If for any fixed* y*,* $f(x, y)$ *is convex in* x*, and for any fixed* x*,* $f(x, y)$ *is concave in* y*, Then:*

1. $\min_x \max_y f(x, y) = \max_y \min_x f(x, y)$
2. $\exists(x^*, y^*)$ *s.t.* $\forall x, y,\ f(x^*, y) \le f(x^*, y^*) \le f(x, y^*)$

7.2 Lagrange Duality

In optimization, **duality** allows optimization problems to be viewed from two perspectives: the primal form and the dual form. Adopting the dual form allows for new insight, while often preserving the optimal value.

Let's consider the following **primal** optimization problem:

$$
\begin{aligned}
(P) \qquad & \min_x \quad f(x) \\
& \text{s.t.} \quad g_i(x) \le 0, \quad i \in [m] \\
& \qquad\;\; h_i(x) = 0, \quad i \in [n].
\end{aligned}
$$

Where f and g_i's are convex functions, and h_i's are linear. The P here denotes primal form.

We now transform this problem to its dual form.

Step 1. It can be shown that the following optimization problem is equivalent to the primal problem,

$$\min_x \max_{\lambda,\mu} f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^n \mu_i h_i(x)$$
$$s.t.\ \lambda \geq 0$$

as when one of the constraints is not satisfied, the corresponding λ_i or μ_i can make the function value arbitrarily large. We call this new objective function the **Lagrange function**, denoted as $L(x;\lambda,\mu)$.

$$L(x;\lambda,\mu) := f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^n \mu_i h_i(x)$$

Step 2. We now apply the Sion's Minimax Theorem on this min-max optimization problem. The theorem constraints are satisfied ($L(x;\lambda_0,\mu_0)$ is the non-negative weighted sum of convex functions; $L(x_0;\lambda,\mu)$ is linear, therefore concave in λ,μ). Thus

$$\min_x \max_{\lambda,\mu:\lambda\geq 0} L(x;\lambda,\mu) = \max_{\lambda,\mu:\lambda\geq 0} \min_x L(x;\lambda,\mu)$$

Combining steps 1 and 2, we now consider the problem $\max_{\lambda,\mu:\lambda\geq 0} \min_x L(x;\lambda,\mu)$. Solving $\nabla_x L(x;\lambda,\mu) = 0$, we get $x^* = \varphi(\lambda,\mu)$.

Step 3. Substituting x with $\varphi(\lambda,\mu)$, we get the dual problem:

$$(D) \qquad \max_{\lambda,\mu} f(\varphi(\lambda,\mu)) + \sum_{i=1}^m \lambda_i g_i(\varphi(\lambda,\mu)) + \sum_{i=1}^n \mu_i h_i(\varphi(\lambda,\mu))$$
$$s.t.\ \lambda \geq 0$$

7.3 Example: Optimal Margin Classifier

Previously, we posed the optimization problem for finding the optimal margin classifier:

$$(P) \qquad \min_{w,b} \quad \frac{1}{2}||w||_2^2$$
$$\text{s.t.} \quad y_i(w^T x_i + b) \geq 1, \forall i \in [n],$$

We now develop its dual form. The problem can be rewritten as:

$$\max_\lambda \min_{w,b} \frac{1}{2}||w||_2^2 + \sum_{i=1}^n \lambda_i(1 - y_i(w^T x_i + b))$$
$$s.t.\ \lambda \geq 0$$

Setting the derivative of L with respect to w, b to zero, we get

$$w = \sum_{i=1}^n \lambda_i y_i x_i$$
$$\sum_{i=1}^n \lambda_i y_i = 0$$

Plugging this back to the Lagragian function, we get the dual form

$$
\begin{aligned}
(D) \qquad & \min_{\lambda} \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^\top x_j - \sum_i \lambda_i \\
& s.t.\ \lambda_i \geq 0, \quad i \in [n] \\
& \quad \sum_i \lambda_i y_i = 0
\end{aligned}
$$

Note that the dual problem is still a quadratic programming problem: the first term is a quadratic form derived from a Gram matrix, which is positive-semidefinite.

References

[VN28] J. v. Neumann, Zur Theorie der Gesellschaftsspiele. *Math. Ann.* **100** , 295–320 (1928). https://doi.org/10.1007/BF01448847

[S58] M. Sion, On General Minimax Theorems. *Pacific Journal of Mathematics.* **8** : 171–176(1958). doi:10.2140/pjm.1958.8.171