Exercise 11.1

a) Let $X, Y, f : X \times Y \mapsto \Re$ be arbitrary. Need to show that

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) \leq \inf_{y \in Y} \sup_{x \in X} f(x, y). \tag{212}$$

The reasoning goes quite straightforwardly from the definitions of sup and inf,

$$\begin{aligned} f(x,y) &\leq \sup_{x \in X} f(x,y), \forall x, y && (213) \\ \inf_{y \in Y} f(x,y) &\leq \sup_{x \in X} f(x,y) && (214) \\ \sup_{x \in X} \inf_{y \in Y} f(x,y) &\leq \sup_{x \in X} f(x,y) && (215) \\ \sup_{x \in X} \inf_{y \in Y} f(x,y) &\leq \inf_{y \in Y} \sup_{x \in X} f(x,y). && (216) \end{aligned}$$

On the first line we noted that $f(x, y)$ must always be smaller than its supremum. Since this holds for all x, y, it definitely holds for $\inf_{y \in Y}$, and on the next line, for $\sup_{x \in X}$ too (sup can at most make the bound tight). The last step is reasoned similarly.

b) Let $I = \{1, ...n\}, \Delta(I) = \{p \in \Re^n | p_i \geq 0, \sum p_i = 1\}$. Show that

$$\sup_{p \in \Delta(I)} \sum_i p_i f(i) = \max_{i \in I} f(i), \forall f : I \mapsto \Re. \tag{217}$$

That is, $\Delta(I)$ is the set of all n-dimensional probability distributions and we wish to show that in the above setting the supremum is achieved by a distribution that puts all weight on the single case (this will be useful in the next exercise).

Denote $f(m) = \max_i f(i)$. By selecting $p_m = 1$ and $p_i = 0, \forall i \neq m$, we can instantly see that the maximum can be achieved. It remains to show that the right side upper-bounds the left side.

$$\begin{aligned} \sup_p \sum p_i f_i &\leq \sup_p \sum p_i f(m) && (218) \\ &= f(m) \sup_p \sum p_i && (219) \\ &= f(m) \cdot 1 && (220) \\ &= f(m) = \max_i f(i) && (221) \end{aligned}$$

□

Exercise 11.2

a) Von Neumann's theorem states,

$$\max_p \min_q V(p,q) = \min_q \max_p V(p,q) \tag{222}$$

In this case, we denote the distribution chosen by the max margin player as p and one used by the edge player as q (w and d in the lecture notes). The payoff matrix and the expected payoffs can be written as

$$M(i,j) = h_i(x_j)y_j \tag{223}$$

$$V(p,q) = \sum_i^n \sum_j^m p_i M_{ij} q_j \tag{224}$$

Writing out the Von Neumann's theorem gives

$$\begin{aligned} \max_p \min_q \sum_i \sum_j p_i h_i(x_j) y_j q_j &= \min_q \max_p \sum_i \sum_j p_i h_i(x_j) y_j q_j & (225) \\ \max_p \min_q \sum_j q_j \sum_i h_i(x_j) y_j p_i &= \min_q \max_p \sum_i p_i \sum_j h_i(x_j) y_j q_j & (226) \end{aligned}$$

Lets denote $g(j) = \sum_i h_i(x_j)y_j p_i$ and $f(i) = \sum_j h_i(x_j)y_j q_j$. Note that $f : I \mapsto \Re$ and $g : J \mapsto \Re$, where $I = \{1, ..., n\}$ and $J = \{1, ..., m\}$. This results in

$$\max_p \min_q \sum_j q_j g(j) = \min_q \max_p \sum_i p_i f(i) \tag{227}$$

and we can apply the result from exercise 11.1.b (and the easily seen similar result for inf), this turns into

$$\max_p \min_{j \in J} g(j) = \min_q \max_{i \in I} f(i) \tag{228}$$

and again writing out g and f,

$$\max_p \min_{j \in J} \sum_i h_i(x_j) y_j p_i = \min_q \max_{i \in I} \sum_j h_i(x_j) y_j q_j. \tag{229}$$

Reformatting, we get the wanted result of page 322 (remember that the set of hypothesis H was assumed finite),

$$\min_q \max_{i \in I} \sum_j q_j y_j h_i(x_j) = \max_p \min_{j \in J} y_j \sum_i p_i h_i(x_j). \tag{230}$$

$\square$

b) Interpret player A as the *row-player*. The row-player wishes to select a row of M such that his gains are maximized regardless of which column B (the *col-player*) selects. Note that due to the exercise 11.1, we know that "pure" strategies are optimal (i.e. the optimal distribution puts all weight on a single choice).

From A's point of view, the task is

$$\begin{array}{lll} \text{max.} & \gamma & (== \min -\gamma) \\ \text{s.t.} & \sum_{i=1}^{n} p_i M_{ij} \geq \gamma & , j \in 1, ..., m \\ & \sum p_i = 1 & \\ & p_i \geq 0 & \end{array}$$

Writing out the Lagrangian for the resp. minimization problem gives

$$L(\gamma, p, q, \beta, \lambda) = -\gamma - \sum_{j}^{m} q_j (\sum_{i}^{n} p_i M_{ij} - \gamma) - \sum_{i}^{n} \beta_i p_i + \lambda(1 - \sum_{i}^{n} p_i) \quad (231)$$

Differentiating and setting to zero gives,

$$\begin{aligned} \frac{\delta L}{\delta \gamma} &= -1 + \sum q_j = 0 \Rightarrow \sum q_j = 1 & (232) \\ \frac{\delta L}{\delta p_i} &= \sum_j q_j M_{ij} - \beta_i - \lambda = 0 & (233) \\ &\Rightarrow \sum_j q_j M_{ij} - \lambda = \beta_i & (234) \\ &\Rightarrow \sum_j q_j M_{ij} \geq \lambda & (235) \end{aligned}$$

(since $\beta_i \geq 0$ for all i, from the KKT-cond.). Inserting these back into L leaves us with just

$$G(\lambda) = \lambda \quad (236)$$

Using the derived constraints in the dual maximization problem, we get

$$\begin{array}{lll} \text{max.} & \lambda & \\ \text{s.t.} & \sum_{j=1}^{m} q_j M_{ij} \geq \lambda & , i \in 1, ..., n \\ & \sum_j q_j = 1 & \\ & q_j \geq 0 & \end{array}$$

By changing sign and renaming, the dual problem can be rewritten as

$$\begin{array}{ll} \text{min.} & \lambda \\ \text{s.t.} & \sum_{j=1}^{m} q_j M_{ij} \leq \lambda \quad , i \in 1, ..., n \\ & \sum_j q_j = 1 \\ & q_j \geq 0 \end{array}$$

Consider the optimal values γ^* and λ^*. These are

$$\gamma^* = \max_p \min_j \sum_i p_i M_{ij} \tag{237}$$

$$= \max_p \min_q \sum_i \sum_j p_i M_{ij} q_j \tag{238}$$

$$= \max_p \min_q V(p, q) \tag{239}$$

$$\lambda^* = ... = \min_q \max_p V(p, q) \tag{240}$$

Due to the optimization problem being convex (linear), strong duality holds and the duality gap is zero. Hence $\gamma^* = \lambda^*$ and Von Neumann's theorem holds.

$\square$

Exercise 11.3

We need to show that the optimization problem 5.16 of the lecture notes is the dual of problem 5.15. Lets start with 5.15. Letting $\mu \in \Re, w \in \Re^n, \epsilon \in \Re^m$, the problem was

$$\begin{array}{lll} \max & \mu - C\sum_i^m \epsilon_i & \\ \text{s.t.} & y_i(\sum_j^n w_j h_j(x_i)) \geq \mu - \epsilon_i & , i \in 1,...,m \\ & \epsilon_i \geq 0 & , i \in 1,...,m \\ & w_j \geq 0 & , j \in 1,...,n \\ & \sum_j^n w_j = 1 & \end{array}$$

which equals

$$\begin{array}{lll} \min & -\mu + C\sum_i^m \epsilon_i & \\ \text{s.t.} & y_i(\sum_j^n w_j h_j(x_i)) - \mu + \epsilon_i \leq 0 & , i \in 1,...,m \\ & -\epsilon_i \leq 0 & , i \in 1,...,m \\ & -w_j \leq 0 & , j \in 1,...,n \\ & 1 - \sum_j^n w_j = 0 & \end{array}$$

The lagrangian is

$$\begin{aligned} L(\mu, w, \epsilon, \alpha, \beta, \gamma, \lambda) &= -\mu + C\sum_i \epsilon_i && (241) \\ & -\sum_i^m \alpha_i (y_i(\sum_j^n w_j h_j(x_i)) - \mu + \epsilon_i) && (242) \\ & -\sum_i^m \beta_i \epsilon_i - \sum_j^n \gamma_j w_j + \lambda(1 - \sum_j^n w_j) && (243) \end{aligned}$$

Differenting L w.r.t. the primal variables and setting to zero yields

$$\begin{aligned} \frac{\delta L}{\delta \mu} &= -1 + \sum \alpha_i = 0 \Rightarrow \sum \alpha_i = 1 && (244) \\ \frac{\delta L}{\delta w_j} &= -\sum_i^m \alpha_i (y_i h_j(x_i))\gamma_j = 0 \Rightarrow \gamma_j = -\sum_i \alpha_i y_i h_j(x_i) - \lambda && (245) \\ \frac{\delta L}{\delta \epsilon_i} &= C - \alpha_i - \beta_i = 0 \Rightarrow \beta_i = C - \alpha_i \Rightarrow \alpha_i \leq C && (246) \end{aligned}$$

As usual, insert these to L to get the dual. The dual simplifies into just

$$G(\lambda) = \lambda \quad (247)$$

Hence, using G and the constraints we got from the differentiation, the dual optimization problem can be written as

$$\begin{array}{lll} \max & \lambda & \\ \text{s.t.} & -\sum_i^m \alpha_i y_i h_j(x_i)) \geq \lambda & , j \in 1, ..., n \\ & \alpha_i \geq 0 & , i \in 1, ..., m \\ & \sum_i \alpha_i = 1 & \\ & \alpha_i \leq C & , i \in 1, ..., m \end{array}$$

The first constraint straightforwardly follows from (245) and the fourth one from (246). The constraints $\alpha_i \geq 0$ follow from α_i being Lagrange multipliers related to inequality constraints in the primal problem. Finally, the dual optimization can be easily written as

$$\begin{array}{lll} \min & \lambda & \\ \text{s.t.} & \sum_i^m \alpha_i y_i h_j(x_i)) \leq \lambda & , j \in 1, ..., n \\ & 0 \leq \alpha_i \leq C & , i \in 1, ..., m \\ & \sum_i \alpha_i = 1, & \end{array}$$

the optimization problem 5.16 of the lecture notes.

$\square$

Exercise 11.4

Consider the optimization problem 5.16 of the lecture notes. This time, let $C = \frac{1}{vm}$ for some $0 < v \leq 1$. Formulated as a minimization problem (see prev. exercise), the problem was

$$\begin{array}{lll} \min & \lambda & \\ \text{s.t.} & \sum_i^m \alpha_i y_i h_j(x_i)) \leq \lambda & , j \in 1, ..., n \\ & 0 \leq \alpha_i \leq C & , i \in 1, ..., m \\ & \sum_i \alpha_i = 1. & \end{array}$$

Note ϵ_i to be the Lagrange multiplier related to the bound $\alpha_i \leq C$. In the primal, ϵ_i are the slack variables. The fourth KKT-condition then states that for the optimal solution

$$\epsilon_i(\alpha_i - C) = 0. \tag{248}$$

Similar to theorem 4.21 of the lecture notes, we can use this condition and the constraints of the optimization problem to reason about the solution.

Suppose first that $\epsilon_i > 0$. Then $(\alpha_i - C) = 0$ and $\alpha_i = C = \frac{1}{vm}$. Since $\sum \alpha_i = 1$, this case can happen at most vm times. Hence at most vm examples violate the margin (the resp. slack variable ϵ_i in primal is active).

We also know that $\alpha_i \leq C = \frac{1}{vm}$, and that $\alpha_i \geq 0$. Hence we need atleast vm examples with $\alpha_i > 0$ to reach $\sum \alpha_i = 1$. As $m = vm + (1 - v)m$, we can have at most $(1 - v)m$ examples left that are allowed to exceed the margin.

□

- Finis -