58147 Machine Learning (Spring 2005)

Exercise 11 (Wednesday 20 April)

This is the last set of exercises. The exam will be on Friday 29 April in Auditorium A111.

1. (a) Let X, Y and $f\colon X \times Y \to \mathbf{R}$ be arbitrary. Show that

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

   (b) For $I = \{1,\dots,n\}$, let $\Delta(I)$ be the set of probability measures over I. We represent it as

$$\Delta(I) = \left\{ \boldsymbol{p} \in \mathbf{R}^n \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}.$$

   Show that

$$\sup_{\boldsymbol{p}\in\Delta(I)} \sum_{i=1}^{n} p_i f(i) = \max_{i\in I} f(i)$$

   for any $f\colon I \to \mathbf{R}$.

2. Consider a two-player zero-sum game, where Player A chooses an action from $I = \{1,\dots,n\}$ and Player B chooses an action from $J = \{1,\dots,m\}$. If A plays i and B plays j, then the payoff for A is M_{ij} where $M \in \mathbf{R}^{n\times m}$ is an arbitrary real matrix. The payoff for B is $-M_{ij}$.

   A *strategy* for Player A is a distribution $\boldsymbol{p} \in \Delta(I)$. Similarly, a strategy for B is a distribution $\boldsymbol{q} \in \Delta(J)$. If A plays strategy $\boldsymbol{p}$ and B plays strategy $\boldsymbol{q}$, then the expected payoff for A is

$$V(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_i M_{ij} q_j.$$

   Von Neumann's Minimax Theorem states that

$$\max_{\boldsymbol{p}\in\Delta(I)} \min_{\boldsymbol{q}\in\Delta(J)} V(\boldsymbol{p},\boldsymbol{q}) = \min_{\boldsymbol{q}\in\Delta(J)} \max_{\boldsymbol{p}\in\Delta(I)} V(\boldsymbol{p},\boldsymbol{q}).$$

   (a) Show that the minimax equation on page 322 of the lecture notes follows directly from Von Neumann's theorem. (Use the result from 1(b) above.)

   (b) Prove Von Neumann's theorem by using convex duality. (*Hint:* basically reproduce the derivation on pages 319–322 with appropriate notation.)

3. Show that Optimisation 5.16 (page 328) is indeed the dual of Optimisation 5.15.

4. Consider an optimal solution to Optimisations 5.15 and 5.16. Write $C = 1/(\nu m)$, and (as usual) define the margin of example (x_i, y_i) as $y_i \sum_{j=1}^{m} w_j \tilde{h}_j(x_i)$. Show that (a) at most νm examples have margin less than μ, and (b) at most $(1-\nu)m$ examples have margin greater than μ. (*Hint:* use the KKT conditions; compare with Theorem 4.21.)

5. Fill in the course feedback form at

   `http://ilmo.cs.helsinki.fi/kurssit/servlet/Valinta` .

   (You may wait until the exam to do this, but please do not forget!)