Final Exam - Optimization
B. Math II

07 January, 2022

(i) Duration of the exam is 3 hours.

(ii) The maximum number of points you can score in the exam is 100.

(iii) You are not allowed to consult any notes or external sources for the exam.

Name: ____

Roll Number: ____________________________________

1. Let

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 2 & 1 & 0 \end{bmatrix}, \vec{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

   (a) (5 points) Compute the pseudoinverse of A.

   (b) (5 points) Find a vector $\vec{x} \in \mathbb{R}^3$ which minimizes $\|A\vec{x} - \vec{b}\|_2$ (with justification).

Total for Question 1: 10

2. (10 points) Every matrix $A = (a_{ij})$ defines the following game. In each round, the *row player* selects one of the rows $i = 1, 2, \ldots, m$, and the *column player* selects one of the columns $j = 1, 2, \ldots, n$, the resulting *payoff* to the row player is a_{ij}; that is to say, the column player gives the row player a_{ij} monetary units. (If a_{ij} is negative, then it is the row player who pays $|a_{ij}|$ to the column player.)

   In the game with the payoff matrix:

$$\begin{bmatrix} 3 & 2 & 0 & -1 & 5 & -2 \\ -2 & -3 & 2 & 4 & 0 & 4 \\ 5 & -3 & 4 & 0 & 4 & 7 \\ 1 & 3 & 3 & 2 & -6 & 5 \end{bmatrix}$$

the row player's mixed strategy $[\frac{1}{2}, \frac{1}{4}, 0, \frac{1}{4}]$ is optimal. Describe all the optimal strategies of the column player.

Total for Question 2: 10

3. Consider the linear program:

$$\text{Maximize } \; 10x_1 + 14x_2 + 20x_3$$

subject to

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &\leq 220 \\ 4x_1 + 2x_2 - x_3 &\leq 385 \\ x_1 + 4x_3 &\leq 160 \\ x_1, x_2, x_3 &\geq 0. \end{aligned}$$

   (a) (5 points) Write down the **dual** linear programming problem.

   (b) (10 points) Assuming $x_1 = 60, x_2 = 0, x_3 = 25$ is an optimal solution to the LP problem mentioned above, find an optimal solution to the dual problem (with justification).

   (c) (5 points) Show that $x_1 = 60, x_2 = 0, x_3 = 25$ is indeed an optimal solution to the LP problem mentioned above.

Total for Question 3: 20

4. The following tableaux corresponds to an *optimal basis* for a linear programming problem, where x_1, x_2, x_3 are the original primal variables, and s_1 and s_2 are the slack variables corresponding to the two constraints.

	x_1	x_2	x_3	s_1	s_2
$\frac{65}{2}$	0	2	0	3	2
$\frac{5}{2}$	0	$\frac{1}{4}$	1	$\frac{1}{2}$	0
$\frac{5}{2}$	1	$-\frac{1}{2}$	0	$-\frac{1}{6}$	$\frac{1}{3}$

   (a) (10 points) Deduce the original linear program with justification.

   (b) (4 points) Write the dual of the original program.

   (c) (6 points) Explain why the reduced costs of the slack variables in the tableau are equal to the negative of values of the dual optimal variables for the corresponding constraints.

   (d) (4 points) From the given tableax, determine B^{-1}, where B is the given optimal basis.

(e) (6 points) Consider adding a new variable x_4 to the original problem with corresponding column $\begin{bmatrix}1\\1\end{bmatrix}$ in the constraint matrix, and (unknown) cost coefficient c_4. Calculate the (expanded) tableau corresponding to B, the given basis, by calculating the entries for the new column corresponding to x_4 (the reduced cost in row zero must be expressed as a function of c_4, the unknown cost coefficient).

(f) (5 points) Determine for what values of c_4 the current basis B remains optimal.

Total for Question 4: 35

5. In a classification problem, the statistician is given a sample of correctly classified data:

$$\{(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)\}.$$

For each data point i, $\vec{x}_i \in \mathbb{R}^n$ is the vector of features and $y_i \in \{\pm 1\}$ is the class of the data point. The goal of the statistician is to learn a *discriminant* function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\begin{cases} f(\vec{x}_i) > 0 & \text{if } y_i = +1 \\ f(\vec{x}_i) < 0 & \text{if } y_i = -1 \end{cases}$$

Such a function f can be used to classify new data in the following manner: Given a new feature vector $\vec{x} \in \mathbb{R}^n$, if $f(\vec{x})$ is positive, assign $\vec{x}$ to class $+1$, and if $f(\vec{x})$ is negative, assign $\vec{x}$ to class -1. *Affine discriminant functions* are said to be **support vector machines**.

(a) (5 points) For a non-zero vector $\vec{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, show that the distance between the two hyperplanes $\{\vec{x} \in \mathbb{R}^n : \langle \vec{a}, \vec{x} \rangle + b = -1\}$ and $\{\vec{x} \in \mathbb{R}^n : \langle \vec{a}, \vec{x} \rangle + b = 1\}$ is given by $\frac{2}{\|\vec{a}\|_2}$.

(b) (10 points) We want to select a hyperplane which separate the data points "the most"; in other words, we want a non-zero vector $\vec{a} \in \mathbb{R}^n$ with maximum classification margin. We phrase this in terms of the following optimization problem:

$$\underset{\vec{a},b}{\text{minimize}} \;\; \langle \vec{a}, \vec{a} \rangle = \|\vec{a}\|_2^2$$

subject to

$$\langle \vec{a}, \vec{x}_i \rangle + b \geq 1, i \in I$$
$$\langle \vec{a}, \vec{x}_j \rangle + b \leq -1, j \in J.$$

Compute the dual function $F(\vec{\lambda}, \vec{\mu})$ of the above optimization problem, and specify for what values of $(\vec{\lambda}, \vec{\mu})$, we have $F(\vec{\lambda}, \vec{\mu}) > -\infty$.

(c) (10 points) Show that the DUAL problem of the optimization problem in part (c) is equivalent to,

$$\underset{\vec{\lambda},\vec{\mu}}{\text{minimize}} \;\; \left\| \sum_{i \in I} \lambda_i \vec{x}_i - \sum_{j \in J} \mu_j \vec{x}_j \right\|_2$$

subject to

$$\sum_{i \in I} \lambda_i = \sum_{j \in J} \mu_j = 1, \vec{\lambda} \geq \vec{0}, \vec{\mu} \geq \vec{0}.$$

Total for Question 5: 25