Optimization - B. Math, 3rd Semester
Assignment 6 — Odd Semester 2021-2022

Due date: January 02, 2022 (by 11:59 pm)
Note: Total number of points is 60. Plagiarism is prohibited. But after sustained effort, if you cannot find a solution, you may discuss with others and write the solution in your own words **only after** you have understood it.

We denote the set of $m \times n$ matrices with real entries by $M_{m,n}(\mathbb{R})$. When $m = n$, we use the notation $M_m(\mathbb{R})$.

1. (10 points) Let $A \in M_{m,n}(\mathbb{R})$ and $G \in M_{n,m}(\mathbb{R})$ be a generalized inverse of A. Show that the set of all generalized inverses of A is given by

$$\{G + (I_n - GA)U + V(I_m - AG) : U, V \in M_{n,m}(\mathbb{R})\}.$$

2. (10 points) Let

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

   (a) (5 points) Compute the singular value decomposition of A.

   (b) (5 points) Find a vector $\vec{x} \in \mathbb{R}^4$ which minimizes $\|A\vec{x} - \vec{b}\|_2$ (with justification).

Soumyashant Nayak: soumyashant@isibang.ac.in

3. (25 points) (Support vector machines - SVM - for binary classification) 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) (4 points) Let $I := \{i : y_i = +1\}$ be the set of data points in class $+1$ and similarly $J := \{j : y_j = -1\}$ the data points in class -1. Show that there is an affine discriminant function for the data-set if and only if the convex hulls of $\{\vec{x}_i : i \in I\}$ and $\{\vec{x}_j : j \in J\}$ do not intersect.

(b) (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 $\frac{2}{\|\vec{a}\|_2}$. (If we use $\langle \vec{a}, \vec{x} \rangle + b$ as our SVM, then this distance is called the *classification margin.*)

(c) (8 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 $F(\vec{\lambda}, \vec{\mu}) > -\infty$.

(d) (8 points) Show that the DUAL problem of the optimization problem in part (c) is equivalent to,

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

subject to

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

Soumyashant Nayak: soumyashant@isibang.ac.in

4. (15 points) (Piecewise linear convex optimization) For non-zero vectors $\vec{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}$ $(1 \le i \le m)$, consider the piecewise-linear minimization problem,

$$\underset{\vec{x}\in\mathbb{R}^n}{\text{minimize}} \max_{1\le i\le m} (\langle \vec{a}_i, \vec{x}\rangle + b_i).$$

Let A be the $m \times n$ matrix with rows $\vec{a}_i^T$, and $\vec{b}$ be the vector in $\mathbb{R}^m$ with components $b_1, \ldots, b_m$.

(a) (5 points) Derive a dual problem, based on the Lagrange dual of the equivalent problem

$$\underset{\vec{x}\in\mathbb{R}^n, \vec{y}\in\mathbb{R}^m}{\text{minimize}} \max_{1\le i\le m} y_i$$

subject to

$$A\vec{x} + \vec{b} = \vec{y}.$$

(b) (5 points) Let α^* be the optimal value for the above optimization problem. Let $\vec{p} \in \mathbb{R}^m$ be a probability vector, that is, $\vec{p} \ge \vec{0}$ and $\sum_{i=1}^m p_i = 1$, such that $\vec{p}^T A = \vec{0}^T$. Show that $\langle \vec{p}, \vec{b}\rangle \le \alpha^*$.

(c) (5 points) In order to find the best possible lower bound of the form described in part (b), we form the linear programming problem

$$\underset{\vec{p}\in\mathbb{R}^m}{\text{maximize}} \langle \vec{p}, \vec{b}\rangle$$

subject to

$$\vec{p}^T A = \vec{0}^T, \ \sum_{i=1}^m p_i = 1, \vec{p} \ge \vec{0}.$$

Show that the optimal value of this problem is equal to α^*.

Soumyashant Nayak: soumyashant@isibang.ac.in