MS&E 111X & 211X
Introduction to Optimization (Accelerated)
Homework 2
Course Instructor: Yinyu Ye
Due Date: 11:59 pm Oct 21, 2021

Please submit your homework through Gradescope. If you haven't already been added to Gradescope, you can use the entry code **2RJNKV** to join. Please note: late homework will not be accepted. Each problem will be graded out of 10 points.
Some problems allow group work. Groups should be no larger than 4. If you decide to work together, provide the names of those you worked with.

Problem 1 Convex sets and convex functions

Identify whether the following sets are convex

(a) $F := \{x \in R^n : \ Ax = b, \ x \geq 0\}$, where data matrix $A \in R^{m \times n}$ and vector $b \in R^m$.

(b) Fix data matrix A and consider the b-data set for F defined above:

$$B := \{b \in R^m \ : \ F \text{ is not empty}\}$$

(c) $\{x : x^2 \geq 1\}$.

(d) $\{x : x^2 \leq 1\}$.

Identify whether the following functions are convex.

(e) *Negative entropy*

$$h(p) = \sum_{i=1}^{n} p_i log(p_i)$$

On $\{p \in R^n, p_i > 0 \ \forall \ i, e^T p = 1\}$, where $e \in R^n$ is the vector of ones.

(f) *Floor function*

$$f(x) = \lfloor x \rfloor = \max\{k \leq x, k \text{ is an integer}\}$$

(g) *Sum of largest components*

$$f(x) = \sum_{i=1}^{k} x_{(i)}$$

where $x \in R^n$, k is an integer between 1 and n and $x_{(i)}$ denote the i^{th} largest element of the vector x.

Problem 2

Consider the two-variable linear program with 6 inequaltiy constraints:

$$\max\ 3x_1 + 5x_2$$

$$s.t. \quad x_1 \geq 0$$

$$x_2 \geq 0$$

$$-x_1 + x_2 \leq 2.5$$

$$x_1 + 2x_2 \leq 9$$

$$x_1 \leq 4$$

$$x_2 \leq 3$$

(a) Plot the lines $x_j = 0$, $j = 1, \ldots 6$ (all on the same two-dimensional graph of x_1 and x_2) where for $j = 3, 4, 5, 6$, x_j denotes the slack variable in j^{th} constraint in the LP standard form .

(b) Identify the extreme points of the feasible region as intersections of suitable lines $x_j = 0$.

(c) For each pair of adjacent extreme points of the feasible region, describe how each direction of the edge between them can be generated by increasing the value of a single variable chosen from $\{x_1, x_2, x_3, x_4, x_5, x_6\}$.

Problem 3 Phase One Problem

Consider a system of m linear equations in n non-negative variables, say

$$Ax = b, \quad x \geq 0$$

Assume the right-hand side vector b is non-negative. Now consider the related linear program

$$\min \quad e^T y$$

$$s.t. \quad Ax + Iy = b$$

$$x \geq 0, y \geq 0$$

where e is the vector of all ones, and I is the $m \times m$ identity matrix. This linear program is called a Phase One problem.

(a) Write the Lagrange function of the Phase One problem and the dual of Phase One Problem.

(b) Write the complementary slackness conditions for the Phase One problem.

(c) True of False Questions

1. The Phase One problem always has a basic feasible solution.

2. The Phase One Problem always has an optimal solution.

3. $\{x : \ Ax = b, \quad x \geq 0\} \neq \emptyset$ if and only if the optimal value of the objective function in the corresponding Phase One problem is zero.

Problem 4

Consider the following problem for a parameter $\kappa > 0$:

$$\min \ (x_1 - 1)^2 + x_2^2$$

$$s.t. \ - x_1 + \frac{x_2^2}{\kappa} \geq 0$$

(a) Is $x = 0$ a first order KKT solution?

(b) Is $x = 0$ a second order KKT solution for some value of κ?

Problem 5

Consider the Support Vector Machine problem

$$\min \ \beta + \mu \|x\|^2$$

$$s.t. \ a_i^T x + x_0 + \beta \geq 1, \ \forall i$$

$$b_j^T x + x_0 - \beta \leq -1, \ \forall j$$

$$\beta \geq 0$$

with $\beta, x_0 \in R$ and $x = [x_1; \ldots; x_n], a_i, b_j \in R^n$ for all i, j, and $\|x\|^2 = \sum_{i=1}^n x_i^2$. $\|x\|^2$ could be viewed as a L2 regularization.

(a) Write out the Lagrangian function of SVM problem.

(b) Suppose that we have 6 training data in R^2: $a_1 = (0; 0)$, $a_2 = (1; 0)$, $a_3 = (0; 1)$, $b_1 = (0; 0)$, $b_2 = (-1; 0)$, $b_3 = (0; -1)$. Use the optimality conditions to verify whether the following solutions is optimal.

(1) for $\mu = 0$, $x = [1; 1]$, $x_0 = 0$, $\beta = 1$

(2) For $\mu = 10$, $x = [1; 1]$, $x_0 = 0$, $\beta = 1$

(c) Use solver to find the optimal solution for $\mu = 10$.

Bonus question: When $\mu = 0$, is the optimal solution unique? When $\mu = 10$, is the optimal solution unique? Prove your claim.

Problem 6

In logistic regression, we determine x_0 and x by maximizing

$$\left(\Pi_{i,c_i=1} \frac{1}{1 + \exp(-a_i^T x - x_0)} \right) \left(\Pi_{i,c_i=-1} \frac{1}{1 + \exp(b_i^T x + x_0)} \right)$$

which is equivalent to minimize the log-likelihood loss

$$\sum_{i,c_i=1} log(1 + \exp(-a_i^T x_i - x_0)) + \sum_{i,c_i=-1} log(1 + \exp(b_i^T x + x_0))$$

Suppose that we have 6 training data, with $a_1 = (0; 0)$, $a_2 = (1; 0)$, $a_3 = (0; 1)$, $b_1 = (0; 0)$, $b_2 = (-1; 0)$, $b_3 = (0; -1)$, where a_i are points we label as spam and b_i are points we label as non-spam. (You may view a_i has label $c_i = 1$ and b_i has label $c_i = -1$).

(a) What is the KKT Condition for logistic regression?

(b) Use the KKT condition to verify whether the following solution is optimal: $x = [0; 0]$ and $x_0 = 0$.

(c) Typically, we add a regularization to the objective

$$\min \sum_{i,c_i=1} log(1 + \exp(-a_i^T x_i - x_0)) + \sum_{i,c_i=-1} log(1 + \exp(b_i^T x + x_0)) + \mu \sum_{i=0}^{n} x_i^2$$

what is the KKT Condition for the regularized logistic regression?

Bonus Teamwork : solve the problem using any non-linear programming solver for the non-regularized logistic regression.