DSA3102 Convex Optimization
HW 2

Instruction: Please scan or type your solution into one **single** pdf file. The file name should be **"your-student-number-HW2.pdf"**. Please upload this file to LumiNUS submission folder.

1. In this question, we will work on logistic regression, which is commonly used for data classification. In particular, we will develop a binary classifier to identify spam emails.

   Assume that we independently collect n observations, where each observation has features stored in a vector $\mathbf{x}_i \in R^p$ and a label y_i for $i = 1, ..., n$. Logistic regression is used to construct a model with parameters w such that

$$y_i = \begin{cases} +1 & \text{with probability } \sigma(\mathbf{w}^T\mathbf{x}_i) \\ -1 & \text{with probability } 1 - \sigma(\mathbf{w}^T\mathbf{x}_i) \end{cases}$$

   where $\sigma(a) = \frac{1}{1+e^{-a}}$. The loss function is given by negative likelihood

$$l(\mathbf{w}) = -\frac{1}{n}\sum_{i=1}^{n} \log \sigma(y_i \mathbf{w}^T\mathbf{x}_i)$$

   Data Set: Consider the email spam data set discussed on p. 300 of (Hastie, Tibshirani, Friedman, The Elements of Statistical learning 2nd ed, 2009). This consists of 4601 email messages, from which $p = 57$ features have been extracted. There are multiple ways to preprocess data, in this case, the features are transformed by $\log x_{ij} + 0.1$ and provided to you in tutorial 6 on Canvas (also as HW2data). You will observe that the data is divided into a training set (with n=3065 samples) and a test set (with m=1536 samples). Calculate the best model parameter $\mathbf{w}$ (i.e., find the minimizer of $l(\mathbf{w})$) using the training dataset and report the performance of your classifier by following these steps:

   a) Give an expression for the gradient of l.

   b) Show that l is convex.

   c) Apply the gradient descent algorithm from Assignment 1, Problem 3 to minimize l. Plot the objective function as a function of number of iterations. This is often called the training curve.

   d) Use the iterates obtained (also called trajectory) from (c). However, replace the data in the loss function with the test data. That is:

$$l'(\mathbf{w}) = -\frac{1}{m}\sum_{i=1}^{m=1536} \log \sigma(y_i' \mathbf{w}^T\mathbf{x}_i')$$

where (x_i', y_i') come from the test dataset. Plot the objective function as a function of number of iterations. This is often called the test curve. (Note we are not asking you to minimize l'. We are trying to see how the iterates from (c) work on test data.)

2. Consider the following minimization problem in $\mathbf{R}^3$:

$$\begin{aligned} \min \quad & f(\mathbf{x}) := x_1 - x_2 \\ \text{s.t.} \quad & g_1(\mathbf{x}) := x_1^2 + x_2^2 - 2 = 0 \\ & g_2(\mathbf{x}) := x_2 - x_3^3 = 0 \end{aligned}$$

(a) Let S be the feasible set. Find a positive number $\alpha > 0$ such that

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in S \implies -\alpha \leq x_i \leq \alpha, \ \forall\, i = 1, 2, 3.$$

(b) Does the regularity condition hold at every feasible point? Justify your answer.

(c) Find all the optimal solutions of the minimization problem, assume that they exist.

3. Consider the the scalar minimization problem

$$\begin{aligned} \min \quad & x \\ \text{s.t.} \quad & x^2 \geq 0 \\ & x + 1 \geq 0, \end{aligned}$$

for which the solution is $x^* = -1$.

(a) Let

$$F^< = \{x \in R : x^2 > 0, \ x + 1 > 0\}.$$

Write down an explicit expression of the set $F^<$.

(b) Define $P(\cdot, \mu) : R \to R$ by

$$P(x, \mu) = f(x) + \mu B(x) = x - \mu(\log(x^2) + \log(x + 1))$$

For each $\mu > 0$, find all local minimizers x_μ of $P(\cdot, \mu)$ in $F^<$. Find the limits of all convergent subsequences of $\{x_\mu : \mu \to 0\}$.

4. You are given 20 cm of wire and 16 cm^2 of special paper to make a rectangular box with a bottom and a lid (The paper is to make the 6 faces of the box, and the wire is to stitch the faces together). Find the dimension of the box to maximize its volume. All the wire and paper must be used.

   (a) Formulate the problem as a nonlinear programming problem (NLP).

   (b) Find the optimal solution, assume that it/they exists.