

**SDSC6015 (Semester B, 2025)**  
**Stochastic Optimization for Machine Learning**  
**Assignment 2**

All questions are weighted equally. Please submit both your code (e.g. “.py” file if you are using Python) and your results when programming is required. You can submit either a .py or a .txt file. If you are using other formats, you can copy and paste your code to a .txt file for submission.

**Question 1. [2 point]** Consider a dataset  $\{\mathbf{x}_i, y_i\}_{i=1}^m$  where  $\mathbf{x}_i \in \mathbb{R}^n$  and  $y_i \in \{0, 1, \dots, 9\}$ , for  $i \in [m]$ . One can build a classifier via solving the following optimization problem

$$\min_{\beta_0, \dots, \beta_9} \sum_{i=1}^m \left( \log \left( \sum_{k=0}^9 \exp(\mathbf{x}_i^\top \beta_k) \right) - \mathbf{x}_i^\top \beta_{y_i} \right) + \lambda \sum_{k=0}^9 \|\beta_k\|_1, \quad (1)$$

where  $\lambda \in \mathbb{R}_{++}$  is a penalty parameter and  $\beta_k \in \mathbb{R}^n$  is decision variable, for  $k \in \{0, 1, \dots, 9\}$ .

- (a) [0.2 point] Prove that (Log-Sum-Exp) LSE function  $f(\mathbf{y}) = \log(\sum_{k=1}^{10} \exp(y_k))$  is convex. Then, prove that the optimization problem (1) is convex.
- (b) [0.9 point] Design an algorithm that solves problem (1). Your algorithm should not rely on any off-the-shelf (existing) optimization solver. Explain your answer.
- (b) [0.9 point] Write Python code and implement your proposed algorithm. Please provide sufficient details for the implementation of your algorithm.

**Question 2. [1 point]** Derive an **explicit** solution of the following LP

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && \mathbf{c}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \end{aligned}$$

where  $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$  and  $\mathbf{l} \leq \mathbf{u}$ .

**Question 3. [1 point]** Implement the solver of *gradient descent with exact line search* using any programming language (without using build-in optimization solver) and solve the least-square problem

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^n} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2,$$

where  $\mathbf{X} \in \mathbb{R}^{m \times n}$  and  $\mathbf{y} \in \mathbb{R}^m$ . Please simulate different values of  $\mathbf{X}$  and  $\mathbf{y}$  with

(a)  $n = m = 100$

(b)  $n = 100$  and  $m = 10^6$

(c)  $n = 10^6$  and  $m = 100$

What are the numbers of iterations and run times needed to meet the stopping criteria (any)?

**Question 4. [1 point]** Consider the following problem

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}_{++}^n}{\text{minimize}} && \sum_{i=1}^n x_i \log x_i \\ & \text{subject to} && \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & && \sum_{i=1}^n x_i = 1 \end{aligned}$$

What is the corresponding dual problem? Given the optimal solution of the dual problem, what is the optimal solution (if any) of the above primal problem?