作业一

最优化方法

截止时间：10 月 16 日 23:59（周三晚）

请在规定时间前提交到大夏学堂，超时得分将有折扣。
计算、证明题提交 pdf 电子版，编程题提交 Python 代码、结果及必要的解释。

1 阅读任务

阅读以下阅读材料，并针对每部份简要探讨对该部份的认识。

- Chapter 1. Numerical Optimization by Jorge Nocedal and Stephen J. Wright, Springer, 2006.
- Mazumder et al. (2011), "Spectral regularization algorithms for learning large incomplete matrices" (论文 pdf 见大夏学堂)

2 凸集

a. 证明若函数 f 二次可微，则 f 为凸函数的充要条件为：
   - $\operatorname{dom} f$ 为凸集
   - $\nabla^2 f(x) \succeq 0, \text{ for all } x \in \operatorname{dom} f$

b. 闭集与凸集
   - b.1. 证明多面体 $\{x : Ax \le b\}, A \in \mathbb{R}^{mn}, b \in \mathbb{R}^m$ 为闭的凸集
   - b.2. 举例：$\mathbb{R}^2$ 上闭集的凸包不一定是闭的
   - b.3. 若 $A \in \mathbb{R}^{m\times n}, S \subseteq \mathbb{R}^m$ 为凸集，证明集合 S 在 A 下的原像 $\{x \in \mathbb{R}^n : Ax \in S\}$ 是凸集
   - b.4. 若 $A \in \mathbb{R}^{m\times n}, S \subseteq \mathbb{R}^n$ 为凸集，证明集合 S 在 A 下的像 $\{Ax : x \in S\}$ 是凸集
   - b.5. 举例：存在 $A \in \mathbb{R}^{m\times n}$ 及闭凸集 $S \subseteq \mathbb{R}^n$，使得 $A(S)$ 不是闭集

c. 多面体
   - c.1. 证明：若 $P \subseteq \mathbb{R}^n$ 为多面体，$A \in \mathbb{R}^{m\times n}$，则 $A(P)$ 为多面体，提示：可使用以下事实：
   $$P \subseteq \mathbb{R}^{m+n} \text{ 为多面体} \Rightarrow \{x \in \mathbb{R}^n : (x, y) \in P \text{ for some } y \in \mathbb{R}^m\} \text{ 是多面体}$$
   - c.2. 证明若 $Q \subseteq \mathbb{R}^m$ 为多面体，$A \in \mathbb{R}^{m\times n}$，则 $A^{-1}(Q)$ 为多面体

c. 考虑二次不等式
$$x^\top Ax + b^\top x + c \le 0, \tag{1}$$
其中 A 为 n 阶对称矩阵，设 C 为上述不等式的解集.
   - (a) 证明：当 A 正定时，C 为凸集；
   - (b) 设 C' 是 C 和超平面 $g^\top x + h = 0$ 的交集 $(g \ne 0)$，若存在 $\lambda \in \mathbb{R}$，使得 $A + \lambda gg^\top$ 正定，证明 C' 为凸集.

3 凸函数

a. 证明熵函数：

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i), \quad \operatorname{dom} f = \{x \in \mathbb{R}^n_{++} : \sum_{i=1}^{n} x_i = 1\}$$

是严格凹的。

b. 若 f 为二次可微函数且 $\operatorname{dom} f$ 为凸集，证明 f 为凸函数的充要条件为：

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0, \quad \forall x, y$$

这被称为梯度 ∇f 的单调性。

c. 举例：严格凸函数并不一定能达到其极小值。

d. 函数 $f : \mathbb{R}^n \to \mathbb{R}$ 被称为强制的（coercive），如果当 $\|x\|_2 \to \infty$ 时，有 $f \to \infty$。强制函数的一个关键事实为其可以达到极小值。证明对于一个二次可微的强凸函数是强制的（coercive），并因此可达到极小值。

e. 证明在有界多面体上的凸函数的最大值一定在其中一个顶点上。
（提示：已知一个有界的多面体可以被表示为其顶点的凸组合。）

4 实践：使用 CVXPY 解优化问题

- 确保你已经配置好 Python 环境，并已安装 CVXPY
- CVXPY: `http://www.cvxpy.org/`

a. 给定标签 $y \in \{-1, 1\}^n$，特征矩阵 $X \in \mathbb{R}^{n \times p}, X = (x_1^\top, \ldots, x_n^\top)^\top$，回顾支持向量机问题（SVM）：

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n} \xi_i \\ \text{subject to} \quad & \xi_i \geq 0,\ i = 1, \ldots n \\ & y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i,\ i = 1, \ldots n. \end{aligned}$$

a.1. 从 `xy_train.csv` 中读取训练数据（200×3 的矩阵），前两列为 $p = 2$ 的特征变量，第三列为标签。使用 CVXPY 求解上述 SVM 问题（令 $C = 1$）。给出最优值，最优系数 $\beta \in \mathbb{R}^2$ 和截距项 $\beta_0 \in \mathbb{R}$。

a.2. 回顾 SVM 的解定义了一个超平面

$$\beta_0 + \beta^\top x = 0$$

其被用作 SVM 分类器的决策边界。画出训练数据并注意区别标记两类点。在同一幅图上画出决策边界。

a.3. 定义 $\tilde{X} \in \mathbb{R}^{n \times p}$ 有行向量 $\tilde{x}_i = y_i x_i$，用 CVXPY 求解问题

$$\begin{aligned} \max_{w} \quad & -\frac{1}{2} w^\top \tilde{X} \tilde{X}^\top w + \mathbf{1}^\top w \\ \text{subject to} \quad & 0 \leq w \leq C\mathbf{1},\ w^\top y = 0, \end{aligned}$$

报告最优值，其应该与上述 SVM 问题的最优值相同。（注意这不是巧合，因为这将是我们后面对偶性的一个例子）

a.4. 选择使用不同的参数 $C = 2^a$，令 a 从-4 到 4 变动。对每个值求解 SVM 问题，画出决策边界，在测试集 `xy_test.csv` 中计算错误率。画出错误率（y 轴）和参数 C（x 轴，log-scale）的关系。