Introduction to Machine Learning
Spring 2022
University of Science and Technology of China

Lecturer: Jie Wang — Homework 4
Posted: Nov. 3, 2022 — Due: Nov. 15, 2022

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: Convex Functions

1. (Optional) For each of the following functions, determine whether it is convex.
   (a) $f(x) = x^2 \log x$ on $\mathbb{R}_{++}$, where $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$.
   (b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2$.
   (c) $f(x_1, x_2) = \frac{x_1}{x_2}$ on $\mathbb{R}^2_{++}$, where $\mathbb{R}^2_{++} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 > 0\}$.
   (d) $f(x_1, x_2) = \frac{x_1^2}{x_2}$ on $\mathbb{R} \times \mathbb{R}_{++}$.
   (e) $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$ on $\mathbb{R}^2_{++}$, where $0 \leq \alpha \leq 1$.

2. Please show that the following functions are convex.
   (a) $f(\mathbf{x}) = \log \sum_{i=1}^{n} e^{x_i}$ on $\textbf{dom}\ f = \mathbb{R}^n$, where x_i denotes the i^{th} component of $\mathbf{x}$.
   (b) $f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]}$ on $\textbf{dom}\ f = \mathbb{R}^n$, where $1 \leq k \leq n$ and $x_{[i]}$ denotes the i^{th} largest component of $\mathbf{x}$.
   (c) The extended-value extension of the indicator function of a convex set $C \subseteq \mathbb{R}^n$, i.e.,

$$\tilde{I}_c(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in C, \\ \infty, & \mathbf{x} \notin C. \end{cases}$$

   (d) The negative entropy, i.e.,

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

   on $\textbf{dom}\ f = \{\mathbf{p} \in \mathbb{R}^n : 0 < p_i \leq 1, \sum_{i=1}^{n} p_i = 1\}$, where p_i denotes the i^{th} component of $\mathbf{p}$.
   (e) The spectral norm, i.e.,

$$f(\mathbf{X}) = \|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X})$$

   on $\textbf{dom}\ f = \mathbb{R}^{m \times n}$, where $\sigma_{\max}$ denotes the largest singular value of $\mathbf{X}$.
   (f) $f(\mathbf{X}) = \text{tr}(\mathbf{X}^{-1})$ on $\textbf{dom}\ f = \mathbb{S}^n_{++}$, where $\mathbb{S}^n_{++}$ is the space of all $n \times n$ real positive definite matrices.

3. Please show that a continuously differentiable function f is strongly convex with parameter $\mu > 0$ if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|_2^2, \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

4. Suppose that f is twice continuously differentiable and strongly convex with parameter $\mu > 0$. Please show that $\mu \leq \lambda_{\min}(\nabla^2 f(\mathbf{x}))$ for any $\mathbf{x} \in \mathbb{R}^n$, where $\lambda_{\min}(\nabla^2 f(\mathbf{x}))$ is the smallest eigenvalue of $\nabla^2 f(\mathbf{x})$.

5. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, and the gradient of f is Lipschitz continuous, i.e.,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2, \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

where $L > 0$ is the Lipschitz constant. Please show that $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$, where $\lambda_{\max}(\nabla^2 f(\mathbf{x}))$ is the largest eigenvalue of $\nabla^2 f(\mathbf{x})$.

6. Consider the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and convex, and $\textbf{dom}\ f$ is closed.

(a) Please show that the α-sublevel set of f, i.e., $C_\alpha = \{\mathbf{x} \in \textbf{dom}\ f : f(\mathbf{x}) \leq \alpha\}$ is closed.

(b) Please give an example to show that Problem (1) may be unsolvable even if f is strictly convex.

(c) Suppose that f can attain its minimum. Please show that the optimal set $\mathcal{C} = \{\mathbf{y} : f(\mathbf{y}) = \min_{\mathbf{x}} f(\mathbf{x})\}$ is closed and convex. Does this property still hold if $\textbf{dom}\ f$ is not closed?

(d) Suppose that f is strongly convex with parameter $\mu > 0$. Please show that Problem (1) admits a unique solution.

Solution: ■

Exercise 2: Operations that Preserve Convexity

1. (a) Let $f : \mathbb{R}^m \to (-\infty, +\infty]$ be a given convex function, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Please show that

$$F(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b}), \quad \mathbf{x} \in \mathbb{R}^n.$$

is convex.

(b) Let $f_i : \mathbb{R}^n \to (-\infty, +\infty], i = 1, \ldots, m$, be given convex functions. Please show that

$$F(\mathbf{x}) = \sum_{i=1}^{m} w_i f_i(\mathbf{x})$$

is convex, where $w_i \geq 0$, $i = 1, \ldots, m$.

(c) Let $f_i : \mathbb{R}^n \to (-\infty, +\infty]$ be given convex functions for $i \in I$, where I is an arbitrary index set. Please show that the supremum

$$F(\mathbf{x}) = \sup_{i\in I} f_i(\mathbf{x})$$

is convex.

2. (Optional) Let $\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{x}_0 \in \mathbb{R}^n$. The restriction of $f : \mathbb{R}^n \to \mathbb{R}$ to the affine set $\{\mathbf{A}\mathbf{z} + \mathbf{x}_0 | \mathbf{z} \in \mathbb{R}^m\}$ is defined as the function $F : \mathbb{R}^m \to \mathbb{R}$ with

$$F(\mathbf{z}) = f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)$$

on $\mathbf{dom}\ F = \{\mathbf{z}|\mathbf{A}\mathbf{z} + \mathbf{x}_0 \in \mathbf{dom}\ f\}$. Suppose f is twice differentiable with a convex domain.

(a) Show that F is convex if and only if for all $\mathbf{z} \in \mathbf{dom}\ F$, we have

$$\mathbf{A}^\top \nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)\mathbf{A} \succeq 0.$$

(b) Suppose $\mathbf{B} \in \mathbb{R}^{p\times n}$ is a matrix whose nullspace is equal to the range of $\mathbf{A}$, i.e., $\mathbf{A}\mathbf{B} = \mathbf{0}$ and $\text{rank}(\mathbf{B}) = n - \text{rank}(\mathbf{A})$. Show that F is convex if for all $\mathbf{z} \in \mathbf{dom}\ F$, there exists a $\lambda \in \mathbb{R}$ such that

$$\nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0) + \lambda \mathbf{B}^\top \mathbf{B} \succeq 0.$$

(**Hint:** you can use the result as follows. If $\mathbf{C} \in \mathbb{S}^n$ and $\mathbf{D} \in \mathbb{R}^{p\times n}$, then $\mathbf{x}^\top \mathbf{C}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathcal{N}(\mathbf{D})$ if there exists a λ such that $\mathbf{C} + \lambda \mathbf{D}^\top \mathbf{D} \succeq 0$.)

3. (Optional)

(a) Consider the function $f(\mathbf{X}) = \lambda_{\max}(\mathbf{X})$, with $\mathbf{dom}\ f = \mathbb{S}^n$, where $\lambda_{\max}(\mathbf{X})$ is the largest eigenvalue of $\mathbf{X}$ and $\mathbb{S}^n$ is the set of $n \times n$ real symmetric matrices. Show that f is a convex function.

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, with $\textbf{dom}\ f = \mathbb{R}^n$. Show that it can be represented as the pointwise supremum of a family of affine functions, i.e.,

$$f(\mathbf{x}) = \sup\{g(\mathbf{x}) : g \text{ is affine}, g(\mathbf{z}) \leq f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n\}.$$

4. Suppose that the training set is $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the i^{th} data instance and $y_i \in \mathbb{R}$ is the corresponding label. Recall that Lasso is the regression problem:

$$\min_{\mathbf{w}} \frac{1}{2n}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_1,$$

where $\mathbf{X} \in \mathbb{R}^{n\times d}$ with its i^{th} row being $\mathbf{x}_i^\top$, $\mathbf{w} \in \mathbb{R}^d$, and $\lambda > 0$ is the regularization parameter. Show that the objective function in the above problem is convex.

Solution: ■

Exercise 3: Subdifferentials

1. Calculation of subdifferentials.

   (a) Let $H \subset \mathbb{R}^n$ be a hyperplane. The extended-value extension of its indicator function I_H is

   $$\tilde{I}_H(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in H, \\ \infty, & \mathbf{x} \notin H. \end{cases}$$

   Find $\partial \tilde{I}_H(\mathbf{x})$.

   (b) Let $f(\mathbf{x}) = \exp \|\mathbf{x}\|_1$, $\mathbf{x} \in \mathbb{R}^n$. Find $\partial f(\mathbf{x})$.

   (c) Let $f(x) = \max\{0, x\}$, $x \in \mathbb{R}$. Find $\partial f(x)$.

   (d) For $\mathbf{x} \in \mathbb{R}^n$, let $x_{[i]}$ be the i^{th} largest component of $\mathbf{x}$. Find the subdifferential of

   $$f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]}.$$

   (e) Let $f(\mathbf{x}) = \max\limits_{1 \le i \le n} x_i$, $\mathbf{x} \in \mathbb{R}^n$. Find $\partial f(\mathbf{x})$.

   (f) Let $f(\mathbf{x}) = \|\mathbf{x}\|_\infty$, $\mathbf{x} \in \mathbb{R}^n$. Find $\partial f(\mathbf{x})$.

   (g) Let $f(X) = \max\limits_{1 \le i \le n} |\lambda_i|$, where $X \in \mathbb{S}^n$ and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of X. Find $\partial f(X)$.

   (h) (Optional) Let

   $$f(\mathbf{x}) = \left(\sum_{i=1}^{k} x_i^2 \right)^{\frac{1}{2}} + \left(\sum_{i=k+1}^{n} x_i^2 \right)^{\frac{1}{2}}, \quad \mathbf{x} \in \mathbb{R}^n,$$

   where $1 \le k \le n-1$. Find $\partial f(\mathbf{x})$.

   (i) (Optional) Let $f(\mathbf{X}) = \|\mathbf{X}\|_*$ be the trace norm of $\mathbf{X} \in \mathbb{R}^{m \times n}$. Find $\partial f(\mathbf{X})$.

Solution: ■

Exercise 4: Problems from the Lecture Notes

1. **Mean Value Theorem in Vector Functions.** The mean value theorem is generally not holds in vector valued functions. In the proof of Theorem 5 in Lecture 06, we use some techniques to avoid applying mean value theorem directly on vector valued functions.

   Theorem 5 in Lecture 06. Suppose that f is twice continuously differentiable. Then, f is convex if and only if **dom** f is convex and $\nabla^2 f(\mathbf{x}) \succeq 0$.

   In the proof of necessity, we let $\mathbf{x}_t = \mathbf{x} + t\mathbf{s}, t > 0$.

   (a) We write down the following formula without proof in class,

   $$0 \leq \frac{1}{t^2}\langle \nabla f\left(\mathbf{x}_t\right) - \nabla f(\mathbf{x}), \mathbf{x}_t - \mathbf{x}\rangle = \frac{1}{t}\langle \nabla f\left(\mathbf{x}_t\right) - \nabla f(\mathbf{x}), \mathbf{s}\rangle$$
   $$= \frac{1}{t}\int_0^t \langle \nabla^2 f(\mathbf{x} + \tau\mathbf{s})\mathbf{s}, \mathbf{s}\rangle d\tau.$$

   Please show why the second equality holds.

   (b) By the mean value theorem, we can find an $\alpha \in (0, t)$ such that

   $$\int_0^t \left\langle \nabla^2 f(\mathbf{x} + \tau\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle d\tau = t\left\langle \nabla^2 f(\mathbf{x} + \alpha\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle.$$

   Please explain how we use the mean value theorem in detail.

2. **Log-determinant Function.** Recall Example 3 in Lecture06, the log-determinant function $f(X) = -\log\det X$ with **dom** $f = \mathbb{S}^n_{++}$. Let $X_0 \in \mathbb{S}^n_{++}$ and $V \in \mathbb{S}^n$. We define

   $$g(t) = f\left(X_0 + tV\right)$$

   with **dom** $g = \left\{t : X_0 + tV \in \mathbb{S}^n_{++}\right\}$.

   (a) Please show that **dom** g is nonempty.

   (b) (**Optional**) Please find **dom** g.

3. **Subdifferential.** Recall the Example 5 in Lecture 07. Let $f : \mathbb{S}^n \to \mathbb{R}$ be defined by $f(X) = \lambda_{\max}(X)$. We want to Find $\partial f(X)$.

   By eigen-decomposition, a symmetric matrix can be written as $X = U\Lambda U^\top$, where $U^\top U = I$ and $\Lambda = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_n\right)$ with $\lambda_1 \geq \cdots \geq \lambda_n$. Let $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, i.e., $\mathbf{u}_i$ is the eigenvector corresponding to λ_i. We then write $f(X)$ as

   $$f(X) = \max\{\langle \mathbf{s}, X\mathbf{s}\rangle : \|\mathbf{s}\| = 1\} = \max\{\langle \mathbf{s}\mathbf{s}^\top, X\rangle : \|\mathbf{s}\| = 1\}.$$

   Assume that $\lambda_{\max} = \lambda_1 = \cdots = \lambda_r$, where $1 \leq r \leq n$. Let $U^r = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$,

   $$S^* := \mathop{\mathbf{argmax}}_{\|\mathbf{s}\|=1} \langle \mathbf{s}\mathbf{s}^\top, X\rangle$$

(a) Please show that

$$S^* = \{\mathbf{v} : \mathbf{v} \in \textbf{span } U^r, \|\mathbf{v}\| = 1\} = \{\mathbf{v} : \mathbf{v} = U^r\mathbf{q}, \mathbf{q} \in \mathbb{R}^r, \|\mathbf{q}\| = 1\}.$$

(b) Please find $\frac{d}{dX}\langle \mathbf{s}\mathbf{s}^\top, X\rangle$, then show that

$$\partial f(X) = \textbf{conv}\left\{\mathbf{v}\mathbf{v}^\top : \mathbf{v} \in S^*\right\} = \left\{U^r G \left(U^r\right)^\top : G \succeq 0,\ \text{trace } G = 1\right\}.$$

(c) Suppose $n = 3$, please find $\partial f(X)$ at $X = \text{diag}(2, 4, 4)$ and $X = \text{diag}(1, 2, 4)$.

Solution: ■

References