Introduction to Machine Learning

Fall 2024

University of Science and Technology of China

Lecturer: Jie Wang | Homework 1

Posted: Sep. 19, 2024 | Due: Oct. 8, 2024

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

Exercise 1: Bolzano-Weierstrass Theorem

The Least Upper Bound Axiom

Any nonempty set of real numbers with an upper bound has a least upper bound. That is, $\sup C$ *always exists for a nonempty bounded above set* $C \subset \mathbb{R}$.

Please show the following statements from **the least upper bound axiom**.

1. Let C be a nonempty subset of $\mathbb{R}$ that is bounded above. Prove that $u = \sup C$ if and only if u is an upper bound of C and

$$\forall\, \epsilon > 0, \exists\, a \in C \text{ such that } a > u - \epsilon.$$

2. Suppose (x_n) be a sequence of real numbers such that $x_n \in [a, b], \forall n$, please show that there exists $c \in [a, b]$ and a subsequence (x_{n_k}) of (x_n) such that $x_{n_k} \to c$ as $k \to \infty$.

Exercise 2: Limit and Limit Points

1. Show that $\{\mathbf{x}_n\}$ in $\mathbb{R}^n$ converges to $\mathbf{x} \in \mathbb{R}^n$ if and only if $\{\mathbf{x}_n\}$ is bounded and has a unique limit point $\mathbf{x}$.

2. (**Limit Points of a Set**). Let C be a subset of $\mathbb{R}^n$. A point $\mathbf{x} \in \mathbb{R}^n$ is called a limit point of C if there is a sequence $\{\mathbf{x}_n\}$ in C such that $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{x}_n \neq \mathbf{x}$ for all positive integers n. If $\mathbf{x} \in C$ and $\mathbf{x}$ is not a limit point of C, then $\mathbf{x}$ is called an isolated point of C. Let C' be the set of limit points of the set C. Please show the following statements.

   (a) If $C = (0, 1) \cup \{2\} \subset \mathbb{R}$, then $C' = [0, 1]$ and $x = 2$ is an isolated point of C.

   (b) The set C' is closed.

Exercise 3: Norms

In this exercise, we will give some examples of norms and a useful theorem related to norms in **finite** dimensional vector space.

1. l_p **norm:** The l_p norm is defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

   where $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $p \geq 1$.

   (a) Please show that the l_p norm is a norm.

   (b) Please show that the following equality.

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty = \max_{1\leq i\leq n} |x_i|.$$

   The l_∞ norm is defined as above.

2. **Operator norms:** Suppose that $\mathbf{A} \in \mathbb{R}^{m\times n}$, which can be viewed as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Please show the following operator norms' equality.

   (a) Let $\|\mathbf{A}\|_1 = \sup_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_1}{\|\mathbf{x}\|_1}$. Please show that

$$\|\mathbf{A}\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|.$$

   (b) Let $\|\mathbf{A}\|_\infty = \sup_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty}$. Please show that

$$\|\mathbf{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|.$$

3. **(Optional) Dual norm:** Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. The dual norm of $\|\cdot\|$ is defined by

$$\|\mathbf{x}\|_* = \sup_{\mathbf{y}\in\mathbb{R}^n,\|\mathbf{y}\|\leq 1} \mathbf{y}^\top\mathbf{x}.$$

   (a) Please show that the dual of the Euclidean norm is the Euclidean norm itself. i.e.,

$$\sup_{\mathbf{y}\in\mathbb{R}^n,\|\mathbf{y}\|_2\leq 1} \mathbf{y}^\top\mathbf{x} = \|\mathbf{x}\|_2.$$

   (b) Please show that the dual of the l_1 norm is the l_∞ norm. i.e.,

$$\sup_{\mathbf{y}\in\mathbb{R}^n,\|\mathbf{y}\|_1\leq 1} \mathbf{y}^\top\mathbf{x} = \|\mathbf{x}\|_\infty.$$

4. **(Optional) Equivalence of norms:**

   (a) Let $\|\cdot\|_a$ and $\|\cdot\|_b$ be two norms on $\mathbb{R}^n$. We say that $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent if there exist two positive constants c_1 and c_2 such that

   $$c_1\|\mathbf{x}\|_a \leq \|\mathbf{x}\|_b \leq c_2\|\mathbf{x}\|_a, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

   Please show that all norms on $\mathbb{R}^n$ are equivalent.

   (b) Suppose $\mathbf{X_1} = (\mathbb{R}^n, \|\cdot\|_a)$ and $\mathbf{X_2} = (\mathbb{R}^n, \|\cdot\|_b)$ are two normed vector spaces. Please show that if $(\mathbf{x_n})$ converges to $\mathbf{x}$ in $\mathbf{X_1}$, then $(\mathbf{x_n})$ also converges to $\mathbf{x}$ in $\mathbf{X_2}$, and vice versa.

   (c) The unit ball in $\mathbf{X_1}$ and $\mathbf{X_2}$ may be different. However, the open set in normed vector space $\mathbf{X_1}$ is also open in normed vector space $\mathbf{X_2}$, and vice versa. Please show that by the theorem of equivalence of norms.

   (d) Now we can prove that if f is continuous in normed vector space $\mathbf{X_1}$, then f is also continuous in normed vector space $\mathbf{X_2}$, and vice versa. Please show that by the conclusion in (c).

Exercise 4: Open and Closed Sets

The norm ball $\{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\|_2 < r, \mathbf{x} \in \mathbb{R}^n\}$ is denoted by $B_r(\mathbf{x})$.

1. Given a set $C \subset \mathbb{R}^n$, please show the following are equivalent.
   (a) The set C is closed; that is $\mathbf{cl}\, C = C$.
   (b) The complement of C is open.
   (c) If $B_\epsilon(\mathbf{x}) \cap C \neq \emptyset$ for every $\epsilon > 0$, then $\mathbf{x} \in C$.
2. Given $A \subset \mathbb{R}^n$, a set $C \subset A$ is called open in A if

$$C = \{\mathbf{x} \in C : B_\epsilon(\mathbf{x}) \cap A \subset C \, \text{for some} \, \epsilon > 0\}.$$

   A set C is said to be closed in A if $A \setminus C$ is open in A.
   (a) Let $B = [0, 1] \cup \{2\}$. Please show that $[0, 1]$ is not an open set in $\mathbb{R}$, while it is both open and closed in B.
   (b) Please show that a set $C \subset A$ is open in A if and only if $C = A \cap U$, where U is open in $\mathbb{R}^n$.

Exercise 5: Extreme Value Theorem and Fixed Point

1. Show that any continuous mapping from the closed interval [0,1] to itself has a fixed point, i.e. $\forall f : [0,1] \to [0,1]$, $\exists\ x \in [0,1]$, such that $f(x) = x$.

2. Show that a continuous mapping from the open interval $(0,1)$ to itself may have no fixed point, i.e. $\exists\ f : (0,1) \to (0,1)$, such that for all $x \in (0,1)$, $f(x) \neq x$.

3. If a continuous mapping $f : [0,1] \to [0,1]$ satisfies $f(0) = 0$, $f(1) = 1$, and for some $n \in \mathbb{N}$,

$$f^{(n)}(x) := (f \circ f \circ ... \circ f)(x) \equiv x$$

for all $x \in [0,1]$, show that for all $x \in [0,1]$,

$$f(x) \equiv x.$$

(**Hint:** consider the monotonicity of f.)

4. Show that if a non-decreasing function $f : [0,1] \to \mathbb{R}_+$ is continuous, then there exists a constant $\lambda > 0$, such that for any point $x \in [0,1]$, at least one of the following two cases of the function λf can be realized:

   (a) x is a fixed point of λf.

   (b) $(\lambda f)^{(n)}(x)$ tends to a fixed point of λf, i.e. $\lim_{n\to\infty} (\lambda f)((\lambda f)^{(n)}(x)) = \lim_{n\to\infty} (\lambda f)^{(n)}(x)$,

   where $(\lambda f)(x) = \lambda * (f(x))$.

   (**Hint:** first consider using the extreme value theorem.)

Exercise 6: Linear Space

1. Let $P_n[x]$ be the set of all polynomials on $\mathbb{R}$ with degree at most n. Show that $P_n[x]$ is a linear space.

2. Let V be a linear space, $\lambda \in \mathbb{F}$ be an arbitrary number, $v \in V$ be an arbitrary vector. Show that:

   (a) The zero vectors of V is unique.

   (b) The additive inverse of v is unique, i.e. $\forall\ v \in V$, there exists only one $v' \in V$, such that $v + v' = \mathbf{0}$. Moreover, $v' = (-1) \cdot v \triangleq -v$.

   (c) $\lambda \cdot \mathbf{0} = 0 \cdot v = \mathbf{0}$, $\forall\ \lambda \in \mathbb{F},\ v \in V$.

   (d) If $\lambda \cdot v = \mathbf{0}$, then $\lambda = 0$ or $v = \mathbf{0}$.

3. How many vectors can a linear space have? How many subspaces does a linear space have at least?

Exercise 7: Basis and Coordinates

Suppose that $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is a basis of an n-dimensional vector space V.

1. Show that $\{\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n\}$ is also a basis of V for nonzero scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$.

2. Let $V = \mathbb{R}^n$, $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) \in \mathbb{R}^{n\times n}$ and $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n) \in \mathbb{R}^{n\times n}$. $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n) = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)\mathbf{P}$, where $\mathbf{P} \in \mathbb{R}^{n\times n}$ and $\mathbf{b}_i \in \mathbb{R}^n$, for any $i \in \{1, \ldots, n\}$. Show that $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is also a basis of V for any invertible matrix $\mathbf{P}$.

3. Suppose that the coordinate of a vector $\mathbf{v}$ under the basis $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is $\mathbf{x} = (x_1, x_2, \ldots x_n)$.

   (a) What is the coordinate of $\mathbf{v}$ under $\{\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n\}$?

   (b) What are the coordinates of $\mathbf{w} = \mathbf{a}_1 + \cdots + \mathbf{a}_n$ under $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ and $\{\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n\}$? Note that $\lambda_i \neq 0$ for any $i \in \{1, \ldots, n\}$.

4. Suppose $\mathbf{a} = (1, 0)$, $\mathbf{b} = (0, 1)$ and $\mathbf{c} = (-1, 0)$ are three unit vectors in two-dimensional space. $\mathbf{v} = (x, y)$ is a vector in two-dimensional space.

   (a) Please find the coordinate of $\mathbf{v}$ under basis $\{\mathbf{c}, \mathbf{b}\}$? Is the coordinate unique?

   (b) Please find all the possible combination coefficients of $\mathbf{v}$ under vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, i.e., $\mathbf{v} = x'\mathbf{a} + y'\mathbf{b} + z'\mathbf{c}$.

   (c) (**Bonus**) Each set of combination coefficients (x', y', z') in (b) forms a vector in $\mathbb{R}^3$. Please find the combination coefficients with minimum ℓ_1-norm.

Exercise 8: Derivatives with matrices

Definition 1 (Differentiability)**.** [1] Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function, $\mathbf{x}_0 \in \mathbb{R}^n$ be a point, and let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say that f is *differentiable at* $\mathbf{x}_0$ *with derivative* L if we have

$$\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\neq\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x}-\mathbf{x}_0)\|_2}{\|\mathbf{x}-\mathbf{x}_0\|_2} = 0.$$

We denote this derivative by $f'(\mathbf{x}_0)$.

1. Let $\mathbf{x}, \mathbf{a} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Consider the functions as follows. Please show that they are differentiable and find $f'(\mathbf{x})$.
   (a) $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$.
   (b) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{x}$.
   (c) $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$.
2. Please follow Definition 1 and give the definition of the differentiability of the functions $f : \mathbb{R}^{n\times n} \to \mathbb{R}$.
3. Let $f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}^\top \mathbf{X})$, where $\mathbf{A}, \mathbf{X} \in \mathbb{R}^{n\times m}$, and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Please discuss the differentiability of f and find f' if it is differentiable.
4. (Optional)Let $f(\mathbf{X}) = \det(\mathbf{X})$, where $\det(\mathbf{X})$ is the determinant of $\mathbf{X} \in \mathbb{R}^{n\times n}$. Please discuss the differentiability of f rigorously according to your definition in the last part. If f is differentiable, please find $f'(\mathbf{X})$.
5. (Optional)Let $\mathbf{S}_{++}^n$ be the space of all positive definite $n \times n$ matrices. Prove the function $f : \mathbf{S}_{++}^n \to \mathbb{R}$ defined by $f(\mathbf{X}) = \mathrm{tr}\,\mathbf{X}^{-1}$ is differentiable on $\mathbf{S}_{++}^n$. (Hint: Expand the expression $(\mathbf{X} + t\mathbf{Y})^{-1}$ as a power series.)

Exercise 9: Rank of Matrices

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$.

1. Please show that
   (a) $\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top) = \mathbf{rank}(\mathbf{A}^\top\mathbf{A}) = \mathbf{rank}(\mathbf{A}\mathbf{A}^\top)$;
   (b) $\mathbf{rank}(\mathbf{AB}) \leq \mathbf{rank}(\mathbf{A})$; (please give an example when the equality holds)

2. The *column space* of $\mathbf{A}$ is defined by

$$\mathcal{C}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x},\, \mathbf{x} \in \mathbb{R}^n\}.$$

   The *null space* of $\mathbf{A}$ is defined by

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = 0\}.$$

   Notice that, the rank of $\mathbf{A}$ is the dimension of the column space of $\mathbf{A}$.

   Please show that
   (a) $\mathbf{rank}(\mathbf{A}) = \dim(\mathcal{C}(\mathbf{A})$;
   (b) $\mathbf{rank}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A})) = n$.

3. Given that

$$\mathbf{rank}(\mathbf{AB}) = \mathbf{rank}(\mathbf{B}) - \dim(\mathcal{C}(\mathbf{B}) \cap \mathcal{N}(\mathbf{A})). \tag{1}$$

   Please show the results in 1.(b) by Eq. (1).

Exercise 10: Properties of Eigenvalues and Singular Values

1. Suppose the maximum eigenvalue, minimum eigenvalue and maximum singular value of a given symmetric matrix $\mathbf{A} \in S^n$ are denoted by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$, respectively. Please show that

$$\lambda_{\max}(\mathbf{A}) = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\mathbf{x}^\top\mathbf{x}}, \quad \lambda_{\min}(\mathbf{A}) = \inf_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\mathbf{x}^\top\mathbf{x}}.$$

(**Hint:** considering the orthogonal decomposition of **A**.)

2. (Optional) Suppose $\mathbf{B} = (b_{ij}) \in \mathbb{R}^{m\times n}$ with maximum singular value $\sigma_{\max}(\mathbf{B})$.

   (a) Let $\|\mathbf{B}\|_2 := \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{B}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$. Please show that

$$\sigma_{\max}(\mathbf{B}) = \|\mathbf{B}\|_2.$$

   (b) Please show that

$$\sigma_{\max}(\mathbf{B}) = \sup_{\mathbf{x}\in\mathbb{R}^m, \mathbf{y}\in\mathbb{R}^n, \mathbf{x},\mathbf{y}\neq 0} \frac{\mathbf{x}^\top\mathbf{B}\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}.$$

References

[1] T. Tao. *Analysis II*. Springer, 2015.