Introduction to Machine Learning

Fall 2025

University of Science and Technology of China

Lecturer: Zhihui Li, Xiaojun Chang — Homework 1

Posted: Sep. 28th, 2025 — Due: Oct. 16th, 2025

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

Exercise 1: 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.

Solution 1: Limit and Limit Points

1.

2. (a)

   (b)

■

Exercise 2: 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 $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ 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.,

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

Exercise 3: 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 4: Projection

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

$$\mathbf{P_A}(\mathbf{x}) = \underset{\mathbf{z}\in\mathbb{R}^m}{\mathbf{argmin}}\,\{\|\mathbf{x}-\mathbf{z}\|_2 : \mathbf{z} \in \mathcal{C}(\mathbf{A})\}.$$

We call $\mathbf{P_A}(\mathbf{x})$ the projection of the point $\mathbf{x}$ onto the column space of $\mathbf{A}$.

1. Please show that $\mathbf{P_A}(\mathbf{x})$ is unique for any $\mathbf{x} \in \mathbb{R}^m$.

2. Let $\mathbf{v}_i \in \mathbb{R}^n$, $i = 1, \ldots, d$ with $d \leq n$, which are linearly independent.

   (a) For any $\mathbf{w} \in \mathbb{R}^n$, please find $\mathbf{P}_{\mathbf{v}_1}(\mathbf{w})$, which is the projection of $\mathbf{w}$ onto the subspace spanned by $\mathbf{v}_1$.

   (b) Please show $\mathbf{P}_{\mathbf{v}_1}(\cdot)$ is a linear map, i.e.,

   $$\mathbf{P}_{\mathbf{v}_1}(\alpha\mathbf{u} + \beta\mathbf{w}) = \alpha\mathbf{P}_{\mathbf{v}_1}(\mathbf{u}) + \beta\mathbf{P}_{\mathbf{v}_1}(\mathbf{w}),$$

   where $\alpha, \beta \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n$.

   (c) Please find the projection matrix corresponding to the linear map $\mathbf{P}_{\mathbf{v}_1}(\cdot)$, i.e., find the matrix $\mathbf{H}_1 \in \mathbb{R}^{n\times n}$ such that

   $$\mathbf{P}_{\mathbf{v}_1}(\mathbf{w}) = \mathbf{H}_1\mathbf{w}.$$

   (d) Let $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_d)$.

      i. For any $\mathbf{w} \in \mathbb{R}^n$, please find $\mathbf{P_V}(\mathbf{w})$ and the corresponding projection matrix $\mathbf{H}$.

      ii. Please find $\mathbf{H}$ if we further assume that $\mathbf{v}_i^\top\mathbf{v}_j = 0$, $\forall\, i \neq j$.

3. (a) Suppose that

   $$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

   What are the coordinates of $\mathbf{P_A}(\mathbf{x})$ with respect to the column vectors in $\mathbf{A}$ for any $\mathbf{x} \in \mathbb{R}^2$? Are the coordinates unique?

   (b) Suppose that

   $$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}.$$

   What are the coordinates of $\mathbf{P_A}(\mathbf{x})$ with respect to the column vectors in $\mathbf{A}$ for any $\mathbf{x} \in \mathbb{R}^2$? Are the coordinates unique?

4. A matrix $\mathbf{P}$ is called a projection matrix if $\mathbf{P}\mathbf{x}$ is the projection of $\mathbf{x}$ onto $\mathcal{C}(\mathbf{P})$ for any $\mathbf{x}$.

(a) Let λ be the eigenvalue of $\mathbf{P}$. Show that λ is either 1 or 0. (*Hint: you may want to figure out what the eigenspaces corresponding to* $\lambda = 1$ *and* $\lambda = 0$ *are, respectively.*)

(b) Show that $\mathbf{P}$ is a projection matrix if and only if $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}$ is symmetric.

5. Let $\mathbf{B} \in \mathbb{R}^{m \times s}$ and $\mathcal{C}(\mathbf{B})$ be its column space. Suppose that $\mathcal{C}(\mathbf{B})$ is a proper subspace of $\mathcal{C}(\mathbf{A})$. Is $\mathbf{P_B}(\mathbf{x})$ the same as $\mathbf{P_B}(\mathbf{P_A}(\mathbf{x}))$? Please show your claim rigorously.

Exercise 5: Derivatives with matrices

Definition 1 (Differentiability)**.** [?] 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) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{x}$.

2. Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$. The **Jacobian Matrix with denominator layout** is defined by:

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_1} \\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \frac{\partial f_2(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

   Please show that

$$L(\mathbf{x}-\mathbf{x}_0) = \left(\frac{\partial f}{\partial \mathbf{x}}\right)^\top (\mathbf{x}-\mathbf{x}_0),$$

   where $L : \mathbb{R}^n \to \mathbb{R}^m$ is the derivative in Definition 1.

3. Please follow Definition 1 and give the definition of the differentiability of the functions $f : \mathbb{R}^{n\times n} \to \mathbb{R}$.

4. 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.

5. (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 $f'(\mathbf{X})$.

6. (Optional) Let $\mathbf{S}^n_{++}$ be the space of all positive definite $n \times n$ matrices. Please show 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 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. A real symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is called *positive definite*, written $\mathbf{A} \succ \mathbf{0}$, if for all $\mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$,
$$\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0.$$
Let the set of all positive definite matrices be
$$\mathbb{S}^n_{++} := \left\{\mathbf{A} \in \mathbb{R}^{n\times n} : \mathbf{A} = \mathbf{A}^\top,\ \mathbf{x}^\top \mathbf{A}\mathbf{x} > 0 \text{ for all } \mathbf{x} \neq \mathbf{0}\right\}.$$
Is $\mathbb{S}^n_{++}$ a linear subspace of $\mathbb{R}^{n\times n}$? Please show your conclusion in detail.

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: 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

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

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

Exercise 9: 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

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

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

   (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

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

Exercise 10: Matrix SVD Decomposition and Pseudoinverse

1. For any real matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, the **Moore-Penrose generalized inverse** (or pseudoinverse) of $\mathbf{A}$, denoted by $\mathbf{A}^{+} \in \mathbb{R}^{m\times n}$, is a matrix that satisfies the following four conditions:

   (a) $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$ (Consistency condition)

   (b) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$ (Reflexivity condition)

   (c) $(\mathbf{A}\mathbf{A}^{+})^{\top} = \mathbf{A}\mathbf{A}^{+}$ (Symmetry condition 1)

   (d) $(\mathbf{A}^{+}\mathbf{A})^{\top} = \mathbf{A}^{+}\mathbf{A}$ (Symmetry condition 2)

   Suppose that the matrix $\mathbf{A}$ can be decomposed via Singular Value Decomposition (SVD) as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$, Please show that $\mathbf{A}^{+} = \mathbf{V}\mathbf{\Sigma}^{+}\mathbf{U}^{\top}$, where $\mathbf{\Sigma}^{+} \in \mathbb{R}^{m\times n}$ is defined by:

$$\mathbf{\Sigma}^{+}_{ij} = \begin{cases} \frac{1}{\mathbf{\Sigma}_{ii}} & \text{if } i = j \text{ and } \mathbf{\Sigma}_{ii} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

2. **(Optional)** Please show that $\mathbf{A}^{+}$ is unique for any matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$.

3. Consider the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ where $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{x} \in \mathbb{R}^{m}$, and $\mathbf{b} \in \mathbb{R}^{n}$. Please show that if the system has no solution (i.e., $\mathbf{b}$ is not in the column space of $\mathbf{A}$), the least squares solution to the system

$$\arg\min_{\mathbf{x}\in\mathbb{R}^{m}} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \,,$$

   is given by $\mathbf{x} = \mathbf{A}^{+}\mathbf{b}$, where $\mathbf{A}^{+} \in \mathbb{R}^{m\times n}$ is the Moore-Penrose generalized inverse of matrix $\mathbf{A}$ defined above.

   (**Hint**: For any orthogonal matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ and vector $\mathbf{x} \in \mathbb{R}^{n}$, then $\|\mathbf{U}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$)