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

Lecturer: Jie Wang | Homework 1
Posted: Sep. 17, 2022 | Due: Oct. 8, 2022

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.

   (c) The closure of C is the union of C' and C; that is $\mathbf{cl}\, C = C' \cup C$. Moreover, $C' \subset C$ if and only if C is closed.

Exercise 2: 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 3: 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. Every bounded sequence in $\mathbb{R}$ has at least one limit point.

3. Every bounded sequence in $\mathbb{R}^n$ has at least one limit point.

Exercise 4: Extreme Value Theorem

1. Show that a set $C \subset \mathbb{R}^n$ is compact if and only if C is closed and bounded.

2. Let C be a compact subset of $\mathbb{R}^n$ and $f : C \to \mathbb{R}$ be continuous. Please show that there exist $\mathbf{a}, \mathbf{b} \in C$ such that

$$f(\mathbf{a}) \leq f(\mathbf{x}) \leq f(\mathbf{b}), \forall \mathbf{x} \in C.$$

(**Hint:** first prove that $f(C)$ is compact, in $\mathbb{R}$.)

3. Let $f : [a, b] \to \mathbb{R}$ be continuous. Show that the range of f is a compact interval $[c, d]$ for some $c, d \in \mathbb{R}$.

Exercise 5: 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 $V = \mathbb{R}^n$ and $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n) = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)\mathbf{P}$, 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 $i \in \{1, \ldots, n\}$.

Exercise 6: 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 $f(\mathbf{X}) = \det(\mathbf{X})$, 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 $f'(\mathbf{X})$.
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. 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.)
6. Define a function $f : \mathbf{S}^n_{++} \to \mathbb{R}$ by $f(\mathbf{X}) = \log\det\mathbf{X}$. Prove $\nabla f(\mathbf{I}) = \mathbf{I}$. Deduce $\nabla f(\mathbf{X}) = \mathbf{X}^{-1}$ for any $\mathbf{X}$ in $\mathbf{S}^n_{++}$.

Exercise 7: 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) $\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top)$;
   (b) $\mathbf{rank}(\mathbf{AB}) \leq \mathbf{rank}(\mathbf{A})$;
   (c) $\mathbf{rank}(\mathbf{AB}) \leq \mathbf{rank}(\mathbf{B})$;
   (d) $\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top\mathbf{A})$.

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

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

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

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

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

   Please show that
   (a) $\mathbf{rank}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A})) = n$;
   (b) $\mathbf{y} = \mathbf{0}$ if and only if $\mathbf{a}_i^\top\mathbf{y} = 0$ for $i = 1, \ldots, m$, where $\mathbf{y} \in \mathbb{R}^m$ and $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ is a basis of $\mathbb{R}^m$.

3. Show that

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

4. Suppose that the first term on the right-hand side (RHS) of Eq. (1) changes to $\mathbf{rank}(\mathbf{A})$. Please find the second term on the RHS of Eq. (1) such that it still holds.

5. Show the results in 1. by Eq. (1) or the one you established in 4.

Exercise 8: Linear Equations

Consider the system of linear equations in $\mathbf{w}$

$$\mathbf{y} = \mathbf{X}\mathbf{w}, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{w} \in \mathbb{R}^d$, and $\mathbf{X} \in \mathbb{R}^{n\times d}$.

1. Give an example for "$\mathbf{X}$" and "$\mathbf{y}$" to satisfy the following three situations respectively:
   (a) there exists one unique solution;
   (b) there does not exist any solution;
   (c) there exists more than one solution.
2. Suppose that $\mathbf{X}$ has full column rank and $\mathbf{rank}((\mathbf{X},\mathbf{y})) = \mathbf{rank}(\mathbf{X})$. Show that the system of linear equations (2) always admits a unique solution.
3. (**Normal equations**) Consider another system of linear equations in $\mathbf{w}$

   $$\mathbf{X}^\top\mathbf{y} = \mathbf{X}^\top\mathbf{X}\mathbf{w}. \tag{3}$$

   Please show that the system (3) always admits a solution. Moreover, does it always admit a unique solution?

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. 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

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

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

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

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

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

Exercise 10: Projection to a Linear Subspace

1. Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ with rank d and $\mathbf{y} \in \mathbb{R}^n$. Consider the optimization problem

$$\min_{\mathbf{w}\in\mathbb{R}^d} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2,$$

   (a) We denote the column space of $\mathbf{X}$ by $\mathcal{C}(\mathbf{X})$. Please show that $\hat{\mathbf{y}} := \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ is the projection of $\mathbf{y}$ on $\mathcal{C}(\mathbf{X})$, i.e. $\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{x}\rangle = 0$ for any $\mathbf{x} \in \mathcal{C}(\mathbf{X})$.

   (b) Please solve the above optimization problem by completing the square.

   (c) Please show that $\min_{\mathbf{w}\in\mathbb{R}^d} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2 \leq \|\mathbf{y}\|_2$. Then find the necessary and sufficient condition where the equality holds and give it a geometric interpretation.

2. Suppose X and Y are both random variables defined in the same sample space Ω with finite second-order moment, i.e. $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$.

   (a) Let $L^2(\Omega) = \{Z : \Omega \to \mathbb{R} \mid \mathbb{E}[Z^2] < \infty\}$ be the set of random variables with finite second-order moment. Please show that $L^2(\Omega)$ is a linear space, and $\langle X, Y\rangle := \mathbb{E}[XY]$ defines an inner product in $L^2(\Omega)$. Then find the projection of Y on the subspace of $L^2(\Omega)$ consisting of all constant variables.

   (b) Please find a real constant $\hat{c}$, such that

$$\hat{c} = \mathop{\mathbf{argmin}}_{c\in\mathbb{R}} \mathbb{E}[(Y - c)^2].$$

   [Hint: you can solve it by completing the square.]

   (c) Please find the necessary and sufficient condition where $\min_{c\in\mathbb{R}} \mathbb{E}[(Y - c)^2] = \mathbb{E}[Y^2]$. Then give it a geometric interpretation using inner product and projection.

3. Suppose X and Y are both random variables defined in the same sample space Ω and all the expectations exist in this problem. Consider the problem

$$\min_{f:\mathbb{R}\to\mathbb{R}} \mathbb{E}[(f(X) - Y)^2].$$

   (a) Please solve the above problem by completing the square.

   (b) We let $\mathcal{C}(X)$ denote the subspace $\{f(X) \mid f(\cdot) : \mathbb{R} \to \mathbb{R}, \mathbb{E}[f(X)^2] < \infty\}$ of $L^2(\Omega)$. Please show that the solution of the above problem is the projection of Y on $\mathcal{C}(X)$.

   (c) Please show that question 2 is a special case of question 3.

References

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