

Lecture 04. Basics of Analysis

Lecturer: Jie Wang

Date: Sep 26, 2021

The last lecture introduced the linear regression from two perspectives, least squares and maximum likelihood, respectively. The good news is that, we derived the closed form solutions of the estimators $\hat{\mathbf{w}}_{LS}$ and $\hat{\mathbf{w}}_{ML}$, which are indeed the same with each other. The bad news is that, both estimators involve inverses of matrices, which can be computationally intractable as the feature dimension can be huge. To address this challenge, we introduce a suite of powerful tools from **convex analysis** that are widely used in machine learning. The major references of this lecture are [1, 2, 3, 4, 6].

1 Supremum and infimum

Definition 1. A nonempty set $S \subseteq \mathbb{R}$ is **bounded above** if there exists a number $u \in \mathbb{R}$ such that $x \leq u$ for all $x \in S$. The number u is called an **upper bound** for S .

Similarly, the set S is **bounded below** if there exists a number $l \in \mathbb{R}$ such that $l \leq x$ for all $x \in S$. The number l is called a **lower bound** for S .

Definition 2. The real number u is the **least upper bound** for a nonempty set $S \subseteq \mathbb{R}$ if

1. u is an upper bound for S ;
2. if u' is any upper bound for S , then $u \leq u'$.

The least upper bound is called the **supremum** of the set S , which is denoted by

$$u = \sup S.$$

If $u \in S$, then u is called the **maximum** point of S , i.e.,

$$u = \max S.$$

Question 1. For any nonempty subset of real numbers that is bounded above, can we always find it a least upper bound?

The Completeness Axiom. Suppose that S is a nonempty subset of real numbers that is bounded above. Then, the set S has a least upper bound.

Definition 3. The real number l is the **greatest lower bound** for a set $S \subseteq \mathbb{R}$ if

1. l is a lower bound for S ;
2. if l' is any lower bound for S , then $l \geq l'$.

The greatest lower bound is called the **infimum** of the set S , which is denoted by

$$l = \inf S.$$

If $l \in S$, then l is called the **minimum** point of S , i.e.,

$$l = \min S.$$



2 Norms and Inner Products

2.1 Norms

In a vector space, norm measures the “length” of a vector, and thus the “distance” between two vectors. Once we have a distance function defined, we can discuss limits, followed by many important concepts and tools in mathematical analysis, especially differentiation and integration (we can of course discuss these concepts and tools without a distance function defined under a topological space setting, which is out of the scope of this class).

Definition 4. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with $\text{dom } f = \mathbb{R}^n$ is called a **norm** if

- f is nonnegative: $f(x) \geq 0$ for all $x \in \mathbb{R}^n$;
- f is definite: $f(x) = 0$ only if $x = 0$;
- f is homogeneous: $f(tx) = |t|f(x)$, for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$;
- f satisfies the triangle inequality: $f(x + y) \leq f(x) + f(y)$, for all $x, y \in \mathbb{R}^n$.

We often use the notation $f(x) = \|x\|$ to denote the norm function.

Definition 5. The **unit ball** of a given norm $\|\cdot\|$ is the set of vectors with norm less than or equal to one, that is,

$$\mathcal{B} = \{x \in \mathbb{R}^n : \|x\| \leq 1\}.$$

Example 1. For $x \in \mathbb{R}^n$, the commonly seen ℓ_p norm, $p \geq 1$, is defined by

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

The ℓ_1 -norm and ℓ_2 -norm (the Euclidean norm) are commonly-used regularization terms. Moreover, the Chebyshev or ℓ_∞ -norm is given by

$$\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}.$$

Moreover, for any $P \in \mathbb{S}_{++}^n$ —which is the set of $n \times n$ positive definite matrices—we define the P -quadratic norm as

$$\|x\|_P = (\langle x, x \rangle_P)^{1/2} = (\langle x, Px \rangle)^{1/2} = (x^\top Px)^{1/2} = \|P^{1/2}x\|_2.$$

Example 2. Let $A \in \mathbb{R}^{m \times n}$.

- The Frobenius norm is

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{i,j}^2}.$$

- The matrix p -norms are

$$\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p.$$

Specifically,

$$\begin{aligned}\|A\|_1 &= \max_j \sum_i |a_{ij}|, \\ \|A\|_2 &= \sigma_{\max}(A) = (\lambda_{\max}(A^\top A))^{1/2}, \\ \|A\|_\infty &= \max_i \sum_j |a_{ij}|.\end{aligned}$$

- The trace (nuclear/spectral) norm is

$$\|A\|_* = \sum_i \sigma_i(A).$$

2.2 Inner Products

Definition 6. A function $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ with $\text{dom } f = \mathbb{R}^n \times \mathbb{R}^n$ is called an **inner product** if

- f is nonnegative: $f(x, x) \geq 0$ for all $x \in \mathbb{R}^n$;
- f is definite: $f(x, x) = 0$ if and only if $x = 0$;
- f is symmetric: $f(x, y) = f(y, x)$, for all $x, y \in \mathbb{R}^n$;
- f is bilinear: $f(ax + by, z) = af(x, z) + bf(y, z)$ and $f(x, ay + bz) = af(x, y) + bf(x, z)$, for all $x, y, z \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$.

We often use the notation $f(x, y) = \langle x, y \rangle$ to denote the inner product function.

Example 3. Let $x, y \in \mathbb{R}^n$. Then

$$\langle x, y \rangle = x^\top y = x_1 y_1 + \cdots + x_n y_n$$

is an inner product. For any positive definite matrix P , we can also define an inner product as

$$\langle x, y \rangle_P = \langle x, Py \rangle = x^\top Ay.$$

Example 4. Inner product can also be defined as above on a general linear space.

- Let $A, B \in \mathbb{R}^{m \times n}$. Then

$$\langle A, B \rangle = \text{tr}(A^\top B) = \sum_{i,j} a_{ij} b_{ij}.$$

is an inner product.

- Let $l^2(\mathbb{R}) = \{(x_1, x_2, \dots, x_n, \dots), x_i \in \mathbb{R}, \sum_{i=1}^{\infty} |x_i|^2 < \infty\}$. Then

$$\langle x, y \rangle_{l^2} = \sum_{i=1}^{\infty} x_i y_i$$

is an inner product.



- Let $L^2(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R}, \int_{\mathbb{R}} |f(x)|^2 dx < \infty\}$. Then

$$\langle f, g \rangle_{L^2} = \int_{\mathbb{R}} f(x)g(x)dx$$

is an inner product. (Here we view two functions that equals almost everywhere as the same in $L^2(\mathbb{R})$).

Note that for any inner product, we can naturally define a norm, that is, a norm induced by the inner product

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Question 2. Can any norm be induced by an inner product?

Proposition 1 (Cauchy-Schwarz Inequality). For all $x, y \in \mathbb{R}^n$, there holds

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

The equality holds when and only when x and y are linearly dependent, i.e., $\lambda x + \mu y = 0$ for some $\lambda, \mu \in \mathbb{R}$.

Definition 7. For any non-zero vectors $x, y \in \mathbb{R}^n$, the **included angle** is defined as

$$\Theta(x, y) = \arccos \frac{\langle x, y \rangle}{\|x\| \|y\|}.$$

Definition 8. Two vectors x and y are **orthogonal** if $\langle x, y \rangle = 0$.

3 Basic Topology of \mathbb{R}^n

3.1 Open sets

Definition 9. Given $\epsilon > 0$, the **ϵ -neighborhood** of a point $x \in \mathbb{R}^n$ is

$$N_\epsilon(x) = \{y : y \in \mathbb{R}^n, \|y - x\| < \epsilon\}.$$

The number ϵ is called the radius of $N_\epsilon(x)$.

Definition 10. An element $x \in S \subseteq \mathbb{R}^n$ is called an **interior point** of S if there exists an $\epsilon > 0$ such that $N_\epsilon \subseteq S$.

Definition 11. The set of interior points of S is called the **interior** of S , which is denoted by S° or **int** S .

Definition 12. A set $O \subseteq \mathbb{R}^n$ is **open** if every point in O is an interior point of O , i.e., $O = \text{int } O$.

Question 3.

- 3.1. Is the ϵ -neighborhood an open set?
- 3.2. Is $(0, 1) \subset \mathbb{R}$ an open set?
- 3.3. Is $(0, 1) \subset \mathbb{R}^2$ an open set?



3.2 Closed sets

Another type of sets that is closely related to the open sets is the so-called **closed sets**. We can easily define the closed sets by using open sets.

Definition 13. A set $F \subseteq \mathbb{R}^n$ is **closed** if its complement set, that is,

$$\mathbb{R}^n \setminus F = \{x : x \in \mathbb{R}^n, x \notin F\},$$

is open.

Definition 13 implies that, if $F \subseteq \mathbb{R}^n$ is closed in \mathbb{R}^n , we can find each $x \notin F$ a neighborhood $N_\epsilon(x) \cap F = \emptyset$, where ϵ may depend on x . Another useful approach to characterize the topological properties of closed sets is by **convergent sequences**.

Definition 14. A **sequence** (x_k) of vectors in \mathbb{R}^n is said to **converge** to $x \in \mathbb{R}^n$ if for any $\epsilon > 0$, there exists a positive integer N such that

$$\|x_k - x\|_2 < \epsilon, \forall k \geq N.$$

Symbolically, $x_k \rightarrow x$ or $\lim_{k \rightarrow \infty} x_k = x$.

Theorem 1. *If the limit of a sequence exists, it must be unique.*

Definition 15. A vector $x \in \mathbb{R}^n$ is a **limit (cluster/accumulation) point** of a set $S \subseteq \mathbb{R}^n$ if there exists a sequence $(x_k) \subseteq S$ and $x_k \neq x$ for $k = 1, 2, \dots$ such that $x_k \rightarrow x$.

Definition 16. A set $F \subseteq \mathbb{R}^n$ is **closed** if it contains all of its limit point.

Question 4. Let S' be the set of all limit points of S . How to characterize the points left in S after we remove all the points in S' ? In other words, what can we say about the points in $S \setminus S'$?

To Question 4, we give an equivalent definition of limit point by neighborhood.

Definition 17. A vector $x \in \mathbb{R}^n$ is a **limit (cluster/accumulation) point** of a set $S \subseteq \mathbb{R}^n$ if every neighborhood of x contains a point $x' \neq x$ such that $x' \in S$.

In view of Definition 17, we can easily characterize the points in $S \setminus S'$, which is known as the isolated points.

Definition 18. A vector $x \in S \subseteq \mathbb{R}^n$ is an **isolated point** of S if it is not a limit point of S , that is, there exists a neighbor of x that contains no other points in S other than x .

Remark 1. Notice that, for a nonempty set $S \subseteq \mathbb{R}^n$, its limit points may not belong to S , while its isolated points must be one of the elements in S . However, either S' or $S \setminus S'$ can be empty (when?), but not both under the nonempty assumption of S .

When a set S does not contain all of its limit points, we may say that S is not closed to the limit operations of the sequences in S . This may lead to practical difficulties. For example, what is the length of the diagonal of the unit square if you only know rational numbers? Thus, expanding the set such that it contains all its limit points becomes desirable.

Definition 19. The **closure** of the set S , denoted by $\text{cl } S$ or \bar{S} , is the set $S \cup S'$.

In view of Definitions 16 and 19, we immediately have the result as follows.



Theorem 2. Let $S \subseteq \mathbb{R}^n$.

1. The set \bar{S} is closed.
2. The set S is closed if and only if $S = \bar{S}$.

Question 5.

1. Is $(0, 1]$ closed in \mathbb{R} ?
2. Is $(0, 1]$ closed in $(0, \infty)$?

Remark 2. When we discuss the openness or closedness of a given set S , we always refers to another set Ω that includes S . Specifically, even for the same set S , it can be open with respect to a set Ω with $S \subseteq \Omega$, and it can also be closed with respect to another set Ω' with $S \subseteq \Omega'$ as well (please see Questions 3 and 5). Thus, a rigorous way to claim that “the set S is open or closed” is to say that “the set S is open or closed in Ω (with $S \subseteq \Omega$)”.

Question 6.

1. Can you find a set that is **open-and-closed**?
2. Can you find a set that is neither open nor closed?

3.3 The boundary of a set

Enlightened by the definition of open sets introduced in Section 3.1, we can characterize the inside of a set $S \subseteq \mathbb{R}^n$ by its interior. This naturally raises two questions.

Question 7.

1. How to characterize the outside of a set $S \subseteq \mathbb{R}^n$?
2. How to characterize the boundary of a set $S \subseteq \mathbb{R}^n$?

As long as we know how to characterize the inside of a set, we can easily characterize its outside (how?). Thus, given a set S , the points left by removing the inside and outside of S naturally belong to the boundary of S . We formalize this idea by the definition as follows.

Definition 20. A point x is a **boundary point** of a set $S \subseteq \mathbb{R}^n$ if every ϵ -neighborhood of x contains both points belonging to S and points not belonging to S .

We can further characterize the boundary points by the results as follows.

Theorem 3. Let ∂S (also denoted by **bd** S) be the boundary of a set $S \subseteq \mathbb{R}^n$. Then,

$$\partial S = \bar{S} \setminus S^\circ.$$

Question 8.

1. Can we claim that **bd** $S \subseteq S$?
2. Is that possible **bd** $S = S$?



3.4 Compact sets

Definition 21. A set $S \subseteq \mathbb{R}^n$ is **bounded** if there exists a scalar M such that

$$\|x\|_2 \leq M, \forall x \in S.$$

Definition 22. A set $S \subseteq \mathbb{R}^n$ is **compact** if every sequence in S has a subsequence that converges to a point in S .

Theorem 4. A set $S \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Remark 3. Different from the “openness” and “closedness”, the property of “compactness” is intrinsic [5]—that is, if $A \subseteq B \subseteq C$, then A is compact in B if and only if A is compact in C , while the property of being closed (or open) is not intrinsic (see Question 5).

4 Continuous functions

Definition 23. Let $f : S \rightarrow \mathbb{R}$ where $S \subseteq \mathbb{R}^n$. We say f is **continuous** at $x_0 \in S$ if for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x) - f(x_0)| < \epsilon,$$

for all $x \in S$ and $\|x - x_0\| < \delta$. A function is continuous if it is continuous at every point in its domain.

Question 9. Let $f : \mathbb{N} \rightarrow \mathbb{R}$, where \mathbb{N} is the set of all integers. Is f a continuous function?

A handy property of continuous functions f is that, for any sequence $(x_k) \subset \mathbf{dom} f$ that converges to $x \in \mathbf{dom} f$, we have

$$\lim_{k \rightarrow \infty} f(x_k) = f\left(\lim_{k \rightarrow \infty} x_k\right) = f(x).$$

Proposition 2 (Bolzano-Weierstrass Theorem). Every bounded sequence in \mathbb{R}^n has a convergent subsequence.

Theorem 5 (Extreme Value Theorem). Let C be a compact subset of \mathbb{R}^n and $f : C \rightarrow \mathbb{R}$ be continuous. Then, there exist $a, b \in C$ such that

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

In other words, f attains maximum and minimum values in C .

Theorem 5 is one of the most important results in calculus, as it provides a method to show the existence of the optimum of optimization problems.

Question 10.

Which property will be preserved by a continuous function, openness, closedness, or compactness? Specifically, let $S \subset \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function, and

$$I = \{y : y \in \mathbb{R}, \exists x \in S, \text{ such that } f(x) = y\}.$$

If S is open/closed/compact, will I be open/closed/compact?

If you can show that the compactness is preserved by continuous functions, you can immediately obtain the result in Theorem 5.



References

- [1] S. Abbott. *Understanding Analysis, 2ed.* Springer, 2015.
- [2] D. Bertsekas. *Convex Optimization Theory.* Athena Scientific, 2009.
- [3] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.
- [4] R. Courant and F. John. *Introduction to Calculus and Analysis.* Springer, 1989.
- [5] J. M. Erdman. *A Problems Based Course in Advanced Calculus.* AMS, 2018.
- [6] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill Education, 1976.