Lecture 01. Basics of Analysis

Lecturer: Jie Wang Date: Sep 1, 2022

In machine learning area, we model many problems as optimization problems, i.e., finding the maxima and minima of functions. To this end, we need some mathematical tools to analyze the properties of the functions. In this lecture, we introduce a suite of powerful tools from **mathematical analysis** that are widely used in machine learning. The major references of this lecture are [1, 2, 3, 4, 6].

1 Supremum and Infimum

We begin from some basic definitions, which characterize the properties of real numbers.

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 $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbf{dom}\ f = \mathbb{R}^n$ is called a **norm** if

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

We often use the notation $f(\mathbf{x}) = \|\mathbf{x}\|$ 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,

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

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

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

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

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

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

$$\|\mathbf{x}\|_{\mathbf{P}} = (\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{P}})^{1/2} = (\langle \mathbf{x}, \mathbf{P}\mathbf{x} \rangle)^{1/2} = (\mathbf{x}^\top \mathbf{P}\mathbf{x})^{1/2} = \|\mathbf{P}^{1/2}\mathbf{x}\|_2.$$

Example 2. Norm can also be defined on an other space such as a matrix space. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$.

- The Frobenius norm is

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

- The matrix p-norms are

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

Specifically,

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

- The trace (nuclear/spectral) norm is

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

2.2 Inner Products

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

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

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

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

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

is an inner product. For any positive definite matrix $\mathbf{P}$, we can also define an inner product as

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{P}} = \langle \mathbf{x}, \mathbf{P}\mathbf{y} \rangle = \mathbf{x}^\top \mathbf{P} \mathbf{y}.$$

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

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

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

  is an inner product.

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

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

  is an inner product.

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

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

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

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

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

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

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|.$$

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

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

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

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

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

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

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

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

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., $O = \textbf{int } O$.

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,

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

is open.

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

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

$$\|\mathbf{x}_k - \mathbf{x}\|_2 < \epsilon, \, \forall \, k \geq N.$$

Symbolically, $\mathbf{x}_k \to \mathbf{x}$ or $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$.

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

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

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

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

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 $\textbf{cl}\, S$ 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 $\mathbf{x}$ is a **boundary point** of a set $S \subseteq \mathbb{R}^n$ if every ϵ-neighborhood of $\mathbf{x}$ 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

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

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 $f : S \to \mathbb{R}$ where $S \subseteq \mathbb{R}^n$. We say f is **continuous** at $\mathbf{x}_0 \in S$ if for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| < \epsilon,$$

for all $\mathbf{x} \in S$ and $\|\mathbf{x} - \mathbf{x}_0\| < \delta$. A function is continuous if it is continuous at every point in its domain.

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

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

$$\lim_{k\to\infty} f(\mathbf{x}_k) = f\left(\lim_{k\to\infty} \mathbf{x}_k\right) = f(\mathbf{x}).$$

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 $f : C \to \mathbb{R}$ be continuous. Then, there exist $a, b \in C$ such that*

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

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 $S \subset \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and

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

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.

5 Useful Tools

5.1 Taylor Formula

A useful technique to analyze a function's local properties is to approximate it with well-studied functions such as polynomials.

Definition 24. We write $h(x) = o(x^n)$ when $x \to 0$, if

$$\lim_{x\to 0} \frac{h(x)}{|x|^n} = 0.$$

Theorem 6 (**Taylor Formula with Peano Remainder**)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a function that has derivatives up to order n in an interval $(x_0 - \delta, x_0 + \delta)$. Then we have*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + o((x - x_0)^n), \quad x \to x_0.$$

For multivariable functions, the commonly used formulations are the first-order and second-order Taylor formulas.

Proposition 3. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a function that is differentiable in an open neighbourhood $N_\delta(\mathbf{x}_0)$. Then we have*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|), \quad \mathbf{x} \to \mathbf{x}_0,$$

where $\nabla f(\mathbf{x}) = (\frac{\partial}{\partial x_1} f(\mathbf{x}), \cdots, \frac{\partial}{\partial x_m} f(\mathbf{x}))^\top$ denotes the first derivative.

Proposition 4. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a function that has continuous partial derivatives up to order 2 in an open neighbourhood $N_\delta(\mathbf{x}_0)$. Then we have*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \mathbf{H} f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2), \quad \mathbf{x} \to \mathbf{x}_0,$$

where $\nabla f(\mathbf{x}) = (\frac{\partial}{\partial x_1} f(\mathbf{x}), \cdots, \frac{\partial}{\partial x_m} f(\mathbf{x}))^\top$ denotes the first derivative, and

$$\mathbf{H} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(\mathbf{x}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_m} f(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_m \partial x_1} f(\mathbf{x}) & \cdots & \frac{\partial^2}{\partial x_m \partial x_m} f(\mathbf{x}) \end{bmatrix}$$

denotes the Hessian matrix.

5.2 The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus links the two important concepts in calculus—differentiation and integration, and it states that the two operations, in a certain sense, are inverse operations.

Theorem 7 (**The Fundamental Theorem of Calculus**)**.**

1. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and F be the function defined by*

$$F(x) = \int_a^x f(t)\mathrm{d}t.$$

Then F is differentiable on (a, b), and $\frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x)$ for all $x \in (a, b)$.

2. *(Newton–Leibniz Formula) Let* $f : [a, b] \to \mathbb{R}$ *be a Riemann integrable function, where* $a, b \in \mathbb{R}$*, and suppose there is a differentiable function* $F : [a, b] \to \mathbb{R}$ *such that* $f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x)$*. Then*

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a).$$

Another important role of the Fundamental Theorem of Calculus is that it relates the concepts of antiderivative (also refferred to as indefinite integral) and definite integral. This theorem also provides a technique to analyze a function by leveraging its derivative, which we will use in the follow-up lectures.

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.