210709 Machine Learning **Spring 2021**

Lecture 2. Elements of Convex Programming I

Lecturer: Jie Wang Date: March 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, 5].

1 Mathematical Background

1.1 Supremum and infimum

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

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

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

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

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

$$u = \sup C.$$

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

$$u = \max C.$$

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

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

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

$$l = \inf C.$$

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

$$l = \min C.$$

1.2 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 **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.

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

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

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

- the Frobenius norm is

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

- the matrix p-norms are

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

1.3 Basic Topology of $\mathbb{R}^n$

1.3.1 Open sets

Definition 5. 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 6. 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 7. The set of interior points of S is called the ***interior*** of S, which is denoted by S° or **int** S.

Definition 8. A set $O \subseteq \mathbb{R}^n$ is ***open*** is every point in O is an interior point of O, i.e., $O =$ **int** O.

Question 1. Is the ϵ-neighborhood an open set?

1.3.2 Closed sets

We first use convergent sequences to characterize topological properties of the so-called ***closed sets***. Then, we will see that we can alternatively define *closed sets* by neighborhoods we introduced in the last section.

Definition 9. 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, $x_k \to x$ or $\lim_{k\to\infty} x_k = x$.

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

Definition 10. 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 $k = 1, 2, \ldots$ such that $x_k \to x$.

Definition 11. A set $F \subseteq \mathbb{R}^n$ is ***closed*** if it contains all of its limit point.

Question 2. 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 2, we give an equivalent definition of limit point by neighborhood.

Definition 12. 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 12, we can easily characterize the points in $S \setminus S'$, which is known as the *isolated points*.

Definition 13. A vector $x \in S \subseteq \mathbb{R}^n$ is a ***isolated (cluster/accumulation) 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 contained 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 14. The ***closure*** of the set S, denoted by **cl** S or $\bar{S}$, is the set $S \cup S'$.

In view of Definitions 11 and 14, 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 $S = \bar{S}$.*

1.3.3 The boundary of a set

Enlightened by the concept, i.e., open sets, introduced in Section 1.3.1, we can characterize the inside of a set $S \subseteq \mathbb{R}^n$ by its interior. This naturally raises two questions.

Question 3.

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 15. 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,*

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

1.3.4 Compact sets

Definition 16. 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 17. 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.*

1.4 Continuous functions on compact sets

Definition 18. Let $f : S \to \mathbb{R}$ 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 $\|x - x_0\| < \delta$.

Question 4. Let $f : \mathbb{N} \to \mathbb{R}$, where $\mathbb{N}$ is the set of all integers. Is f a continous function?

Proposition 1 (Bolzano-Weierstrass Theorem). *Every bounded sequence in $\mathbf{R}^n$ 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*

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

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

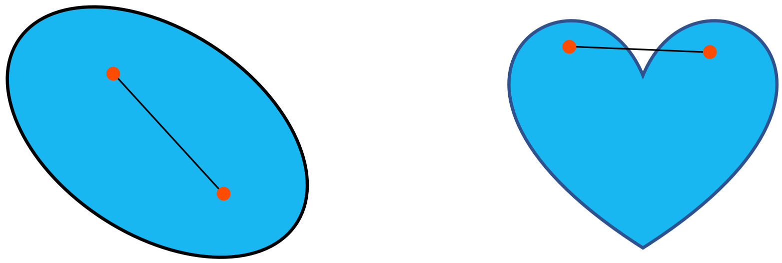

Figure 1: Convex and nonconvex sets.

2 Convex Sets

Definition 19. In $\mathbb{R}^n$, a point x is a ***convex combination*** of the points $\{x_1, \ldots, x_k\}$ if

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k,$$

where $\theta_i \geq 0$ for $i = 1, \ldots, k$ and

$$\theta_1 + \theta_2 + \ldots + \theta_k = 1.$$

Definition 20. The ***convex hull*** of a set $C \subseteq \mathbb{R}^n$, denoted by **conv** C, is the set of all convex combinations of points in C:

$$\mathbf{conv}\ C = \left\{ \sum_{i=1}^{k} \theta_i x_i : x_i \in C, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1 \right\}.$$

Definition 21. A set C is ***convex*** if the line segment between any two points in C lies in C; that is, if $\forall\, x_1, x_2 \in C$ and $\forall\, \theta \in [0, 1]$, we have

$$\theta x_1 + (1 - \theta) x_2 \in C.$$

Example 3. Suppose $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $p(x) \geq 0$ for all $x \in C$ and $\int_C p(x) dx = 1$, where $C \subseteq \mathbb{R}^n$ is convex. Then

$$\int_C p(x) x dx \in C,$$

if the integral exists.

Proposition 2.

1. *The intersection $\cap_{i \in I} C_i$ of any collection $\{C_i : i \in I\}$ of convex sets is convex.*

2. *The closure and the interior of a convex set are convex.*

3. *The image and the inverse image of a convex set under an affine function ($f(x) = Ax + b$) are convex.*

Example 4.

1. hyperplane: $\{x : a^\top x = b\}$, where $a \neq 0$ and $b \in \mathbb{R}$.

2. halfspace: $\{x : a^\top x \le b\}$, where $a \neq 0$ and $b \in \mathbb{R}$.
3. norm ball: $\{x : \|x - x_0\| \le r\}$, where $r > 0$.
4. polyhedron: $\{x : a_i^\top x \le b_i, i = 1, \ldots, m\}$, where $a_i \neq 0$ and $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$.
5. positive semi-definite matrices.

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] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill Education, 1976.