

# Optimization for Machine Learning

## 机器学习中的优化方法

陈 程

华东师范大学 软件工程学院

chchen@sei.ecnu.edu.cn

# Outline

1 Convex Set

2 Convex Function

# Outline

1 Convex Set

2 Convex Function

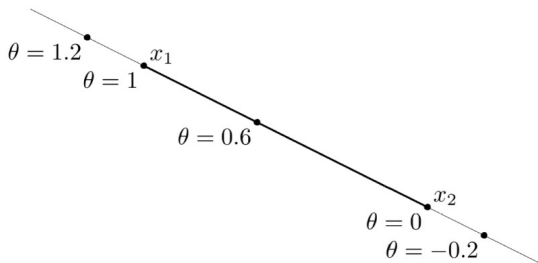
# Lines and Line Segments (直线与线段)

**line** through  $\mathbf{x}_1$  and  $\mathbf{x}_2$ : all points

$$\mathbf{x} = \theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2, \quad \theta \in \mathbb{R}.$$

**line segment** between  $\mathbf{x}_1$  and  $\mathbf{x}_2$ : all points

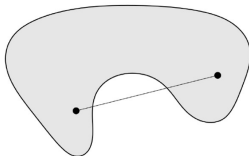
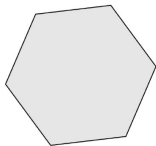
$$\mathbf{x} = \theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2, \quad 0 \leq \theta \leq 1.$$



# Convex Sets (凸集)

A set  $\mathcal{S} \subseteq \mathbb{R}^n$  is **convex** if the line segment between any two points of  $\mathcal{S}$  lies in  $\mathcal{S}$ , i.e., if for any  $\mathbf{x}, \mathbf{y} \in \mathcal{S}$  and  $\theta \in [0, 1]$ , we have

$$\theta \mathbf{x} + (1 - \theta) \mathbf{y} \in \mathcal{S}.$$



Every two points can see each other.

# Properties of Convex Sets

- If  $\mathcal{S}$  is a convex set, then  $k\mathcal{S} = \{k\mathbf{s} | k \in \mathbb{R}, \mathbf{s} \in \mathcal{S}\}$  is convex.
- If  $\mathcal{S}$  and  $\mathcal{T}$  are convex sets, then  $\mathcal{S} + \mathcal{T} = \{\mathbf{s} + \mathbf{t} | \mathbf{s} \in \mathcal{S}, \mathbf{t} \in \mathcal{T}\}$  is convex.
- If  $\mathcal{S}$  and  $\mathcal{T}$  are convex sets, then  $\mathcal{S} \times \mathcal{T} = \{(\mathbf{s}, \mathbf{t}) | \mathbf{s} \in \mathcal{S}, \mathbf{t} \in \mathcal{T}\}$  is convex.
- If  $\mathcal{S}$  and  $\mathcal{T}$  are convex sets, then  $\mathcal{S} \cap \mathcal{T}$  is convex.

# Convex Combination (凸组合)

**Convex combination** of  $\mathbf{x}_1, \dots, \mathbf{x}_k$ : any point  $\mathbf{x}$  of the form

$$\mathbf{x} = \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + \dots + \theta_k \mathbf{x}_k$$

with  $\theta_1 + \dots + \theta_k = 1$ ,  $\theta_i \geq 0$ .

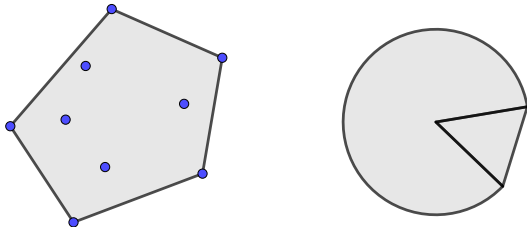
If  $\mathbf{x}_1, \dots, \mathbf{x}_k$  belong to a convex set  $\mathcal{S}$ , then their convex combination  $\mathbf{x}$  also belongs to  $\mathcal{S}$ .

# Convex Hull (凸包)

**Convex hull**  $\text{conv}\mathcal{S}$ : set of all convex combinations of points in  $\mathcal{S}$ .

$$\text{conv}\mathcal{S} = \{\theta_1\mathbf{x}_1 + \cdots + \theta_k\mathbf{x}_k \mid \mathbf{x}_i \in \mathcal{S}, \theta_i \geq 0, i = 1, \dots, k, \theta_1 + \cdots + \theta_k = 1\}.$$

**Example:** convex hull of  $\{0, 1\}$  is  $[0, 1]$ .





# Affine Sets (仿射集)

A set is called **affine set** if it contains the line through any two distinct points in the set.

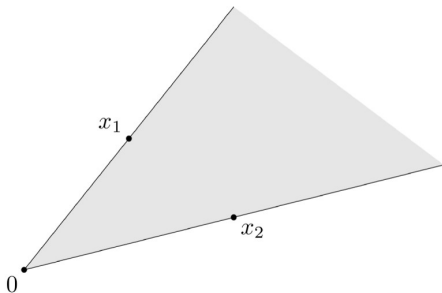
**Example:** solution set of linear equations  $\{\mathbf{x} | \mathbf{Ax} = \mathbf{b}\}$ .

# Cones (锥)

A set  $\mathcal{C}$  is called a **cone** if for every  $\mathbf{x} \in \mathcal{C}$  and  $\theta > 0$  we have  $\theta\mathbf{x} \in \mathcal{C}$ .

A set  $\mathcal{C}$  is called a **convex cone** if it is convex and a cone, which means that for any  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$  and  $\theta_1, \theta_2 > 0$ , we have

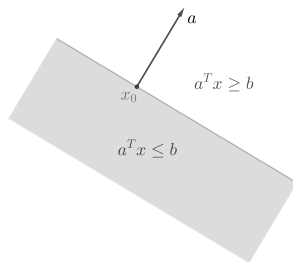
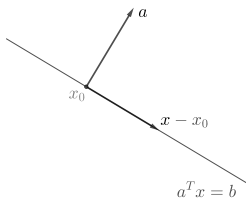
$$\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in \mathcal{C}.$$



# Hyperplanes and Halfspaces (超平面与半平面)

**Hyperplane:** set of the form  $\{\mathbf{x} | \mathbf{a}^\top \mathbf{x} = \mathbf{b}\}$  ( $a \neq 0$ ).

**Halfplane:** set of the form  $\{\mathbf{x} | \mathbf{a}^\top \mathbf{x} \leq \mathbf{b}\}$  ( $a \neq 0$ ).



Hyperplane is affine set.

# Norm Balls (范数球)

**Norm ball** with center  $\mathbf{x}_c$  and radius  $r$ :  $\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$ .



$$p = \infty$$



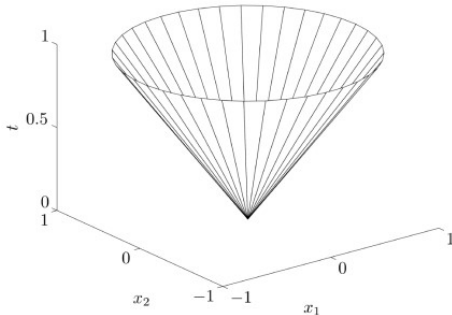
$$p = 2$$



$$p = 1$$

# Norm Cones (范数锥)

**Norm cone:**  $\{(\mathbf{x}, t) \mid \|\mathbf{x}\| \leq t\}$ .



# Operations that preserve convexity (保凸运算)

**Affine functions** (仿射函数).

Suppose  $\mathcal{S}$  is convex and  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is an affine function:

$$f(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}.$$

Then the image of  $\mathcal{S}$  under  $f$ :

$$f(\mathcal{S}) = \{f(\mathbf{x}) | \mathbf{x} \in \mathcal{S}\}$$

is convex. The inverse image:

$$f^{-1}(\mathcal{S}) = \{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) \in \mathcal{S}\}$$

is convex.

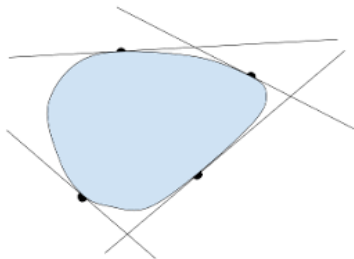
# Operations that preserve convexity (保凸运算)

## Intersection (取交集).

The intersection of (any number of) convex sets is convex, i.e., if  $\mathcal{S}_\alpha$  is convex for any  $\alpha \in \mathcal{A}$ , then  $\cap_{\alpha \in \mathcal{A}} \mathcal{S}_\alpha$  is convex.

A closed convex set  $\mathcal{S}$  is the intersection of all halfspaces contain it:

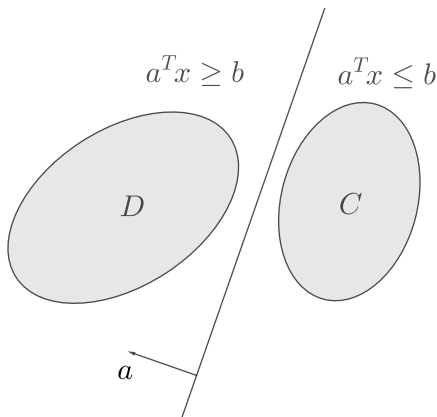
$$\mathcal{S} = \bigcap \{ \mathcal{H} \mid \mathcal{H} \text{ is halfspace, } \mathcal{S} \subseteq \mathcal{H} \}$$



# Hyperplane Separation Theorem

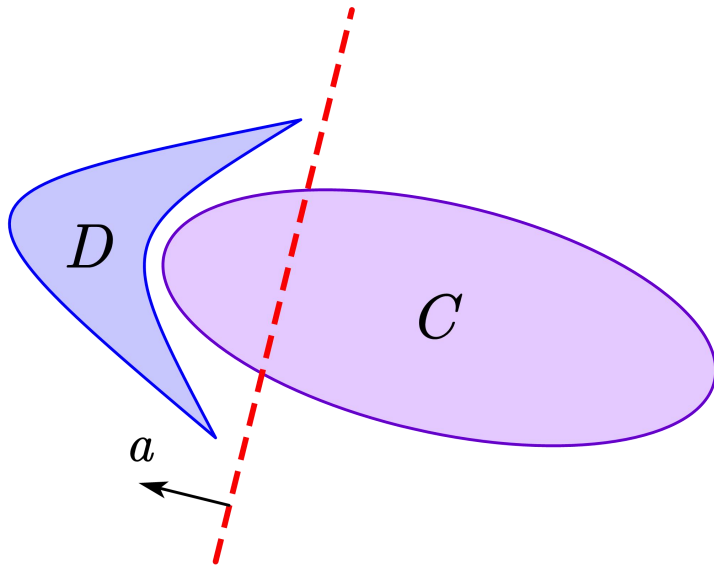
If  $\mathcal{C}$  and  $\mathcal{D}$  are nonempty disjoint convex sets, there exists  $\mathbf{a} \neq 0$  and  $b$  s.t.

$$\mathbf{a}^\top \mathbf{x} \leq b \text{ for } \mathbf{x} \in \mathcal{C}, \quad \mathbf{a}^\top \mathbf{x} \geq b \text{ for } \mathbf{x} \in \mathcal{D}.$$





# Hyperplane Separation Theorem



# Strict Separation Theorem

Suppose  $\mathcal{C}$  and  $\mathcal{D}$  are nonempty disjoint convex sets. If  $\mathcal{C}$  is closed and  $\mathcal{D}$  is compact, there exists  $\mathbf{a} \neq 0$  and  $b$  s.t.

$$\mathbf{a}^\top \mathbf{x} < b \text{ for } \mathbf{x} \in \mathcal{C}, \quad \mathbf{a}^\top \mathbf{x} > b \text{ for } \mathbf{x} \in \mathcal{D}.$$

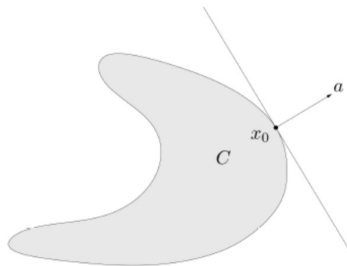
Example: a point and a closed convex set.

# Supporting Hyperplane Theorem

**supporting hyperplane** to set  $\mathcal{C}$  at boundary point  $\mathbf{x}_0$ :

$$\{\mathbf{a}^\top \mathbf{x} = \mathbf{a}^\top \mathbf{x}_0\}$$

where  $\mathbf{a} \neq 0$  and  $\mathbf{a}^\top \mathbf{x} \leq \mathbf{a}^\top \mathbf{x}_0$  for all  $\mathbf{x} \in \mathcal{C}$ .



**Supporting hyperplane theorem:** if  $\mathcal{C}$  is convex, then there exists a supporting hyperplane at every boundary point of  $\mathcal{C}$ .

# Outline

1 Convex Set

2 Convex Function

# Convex Function (凸函数)

A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex if  $\text{dom } f$  is a convex set and

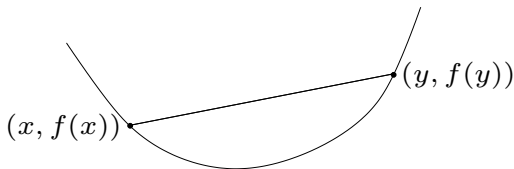
$$f(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y})$$

for all  $\mathbf{x}, \mathbf{y} \in \text{dom } f$ ,  $\theta \in [0, 1]$ .

$f$  is concave if  $-f$  is convex.

**Strict convex function:**

$$f(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) < \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y}), \quad t \in (0, 1), \quad \mathbf{x} \neq \mathbf{y}$$



# Examples

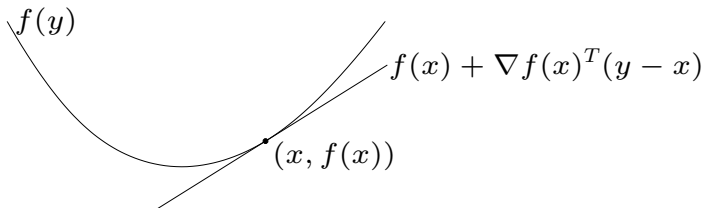
- exponential:  $e^{ax}$ .
- power:  $x^\alpha$  ( $x > 0, \alpha \geq 1$ ).
- logarithm:  $\log_a x$  ( $0 < a < 1$ ).
- negative entropy:  $x \log x$
- affine:  $\mathbf{a}^\top \mathbf{x} + b$ .
- norms:  $\|\mathbf{x}\|$ .

# First-Order Condition

Suppose  $f$  is differentiable and has convex domain, then  $f$  is convex if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$

holds for all  $\mathbf{x}, \mathbf{y} \in \text{dom } f$ .



# First-Order Condition

If  $\nabla f(\mathbf{x}) = 0$ , then for all  $\mathbf{y} \in \text{dom } f$ ,  $f(\mathbf{y}) \geq f(\mathbf{x})$ , i.e.,  $\mathbf{x}$  is a global minimizer of  $f$ .

**Strict convex:**

$$f(\mathbf{y}) > f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \text{ if } \mathbf{y} \neq \mathbf{x}.$$



# Second-Order Condition

Suppose  $f$  is twice differentiable and has convex domain, then  $f$  is convex if and only if

$$\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}.$$

Strict convex:

$$\nabla^2 f(\mathbf{x}) \succ \mathbf{0}.$$

# Examples

- least-square:  $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2$
- quadratic-over-linear:  $f(x, y) = x^2/y$
- log-sum-exp:  $f(x) = \log \sum_{i=1}^n \exp(x_i)$

The  $\alpha$ -sublevel set of a function  $f$  is defined as

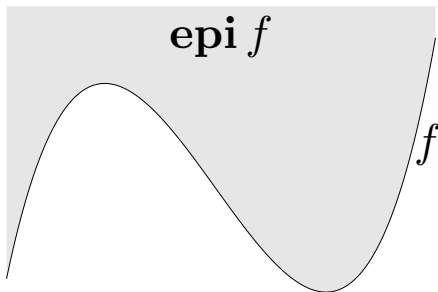
$$\mathcal{C}_\alpha = \{\mathbf{x} \in \text{dom } f \mid f(\mathbf{x}) \leq \alpha\}$$

Sublevel sets of convex functions are convex (converse is false).

## Epigraph (上方图)

The epigraph of a function  $f : \mathcal{S} \rightarrow \mathbb{R}$  is defined as the set

$$\text{epi } f \triangleq \{(\mathbf{x}, u) \in \mathcal{S} \times \mathbb{R} : f(\mathbf{x}) \leq u\}.$$



**Theorem.** A function  $f(\mathbf{x})$  is convex if and only if its epigraph is a convex set.

# Jensen Inequality

Jensen Inequality and extensions:

$$f(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y}), \quad \theta \in [0, 1]$$

$$f(\theta_1 \mathbf{x}_1 + \cdots + \theta_k \mathbf{x}_k) \leq \theta_1 f(\mathbf{x}_1) + \cdots + \theta_k f(\mathbf{x}_k), \quad \theta_1 + \cdots + \theta_k = 1$$

$$f\left(\int_S p(\mathbf{x}) \mathbf{x} d\mathbf{x}\right) \leq \int_S f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

$$f(\mathbb{E}[\mathbf{x}]) \leq \mathbb{E}[f(\mathbf{x})], \text{ for any random variable } \mathbf{x}$$

# Operations that preserve convexity

## **Nonnegative weighted sums:**

A nonnegative weighted sum of convex functions

$$f = w_1 f_1 + \cdots + w_m f_m$$

is convex.

## **Composition with affine function:**

If  $f$  is convex, then  $f(\mathbf{Ax} + \mathbf{b})$  is convex.

# Operations that preserve convexity

## Pointwise maximum:

If  $f_1, \dots, f_m$  are convex, then  $f(x) = \max\{f_1(x), \dots, f_m(x)\}$  is convex.

## Example:

- piecewise-linear function:  $f(x) = \max_{i=1, \dots, m} (\mathbf{a}_i^\top \mathbf{x} + \mathbf{b}_i)$  is convex
- sum of  $r$  largest components of  $\mathbf{x} \in \mathbb{R}^n$ :

$$f(\mathbf{x}) = x_{[1]} + \dots + x_{[r]}$$

is convex. ( $x_{[i]}$  is  $i$ -th largest component of  $\mathbf{x}$ )

# Operations that preserve convexity

## Pointwise supremum:

If  $f(x, y)$  is convex in  $x$  for each  $y \in \mathcal{A}$ , then

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$

is convex.

## Example:

- distance to farthest point in a set  $\mathcal{C}$ :

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\|$$



# Operations that preserve convexity

## Minimization:

If  $f(x, y)$  is convex in  $(x, y)$  and  $\mathcal{C}$  is a convex set, then

$$g(x) = \inf_{y \in \mathcal{C}} f(x, y)$$

is convex.

**Example:** distance to a set:  $\text{dist}(\mathbf{x}, \mathcal{S}) = \inf_{\mathbf{y} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\|$  is convex if  $\mathcal{S}$  is convex.

# Questions

