## **Optimization in Machine Learning**

# Mathematical Concepts Convexity

× 0 0 × × ×



#### Learning goals

- Convex sets
- Convex functions

#### **CONVEX SETS**

A set of  $S \subseteq \mathbb{R}^d$  is **convex**, if for all  $\mathbf{x}, \mathbf{y} \in S$  and all  $t \in [0, 1]$  the following holds:

.

$$\mathbf{x} + t(\mathbf{y} - \mathbf{x}) \in \mathcal{S}$$

Intuitively: Connecting line between any  $\mathbf{x}, \mathbf{y} \in S$  lies completely in S.



Left: convex set. Right: not convex. (Source: Wikipedia)



### **CONVEX FUNCTIONS**

Let  $f : S \to \mathbb{R}$ , S convex. f is **convex** if for all  $\mathbf{x}, \mathbf{y} \in S$  and all  $t \in [0, 1]$ 

$$f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) \leq f(\mathbf{x}) + t(f(\mathbf{y}) - f(\mathbf{x})).$$

Intuitively: Connecting line lies above function.



× 0 0 × 0 × ×

Left: Strictly convex function. Right: Convex, but not strictly.

**Strictly convex** if "<" instead of " $\leq$ ". **Concave** (strictly) if the inequality holds with " $\geq$ " (">"), respectively.

**Note:** *f* (strictly) concave  $\Leftrightarrow -f$  (strictly) convex.

#### **EXAMPLES**

#### Convex function: f(x) = |x|Proof: $f(x + t(y - x)) = |x + t(y - x)| = |(1 - t)x + t \cdot y|$ $\leq |(1 - t)x| + |t \cdot y| = (1 - t)|x| + t|y|$ $= |x| + t \cdot (|y| - |x|) = f(x) + t \cdot (f(y) - f(x))$

**Concave function**:  $f(x) = \log(x)$ 

**Neither nor**:  $f(x) = \exp(-x^2)$  (but log-concave)



× × ×

### **OPERATIONS PRESERVING CONVEXITY**

- Nonnegatively weighted summation: Weights w<sub>1</sub>,..., w<sub>n</sub> ≥ 0, convex functions f<sub>1</sub>,..., f<sub>n</sub>: w<sub>1</sub>f<sub>1</sub> + ··· + w<sub>n</sub>f<sub>n</sub> also convex In particular: Sum of convex functions also convex
- **Composition:** g convex, f linear:  $h = g \circ f$  also convex **Proof:**

$$h(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) = g(f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})))$$
  
=  $g(f(\mathbf{x}) + t(f(\mathbf{y}) - f(\mathbf{x})))$   
 $\leq g(f(\mathbf{x})) + t(g(f(\mathbf{y})) - g(f(\mathbf{x})))$   
=  $h(\mathbf{x}) + t(h(\mathbf{y}) - h(\mathbf{x}))$ 

• Elementwise maximization:  $f_1, \ldots, f_n$  convex functions:  $g(\mathbf{x}) = \max \{f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})\}$  also convex × 0 0 × × ×

#### **FIRST ORDER CONDITION**

Prove convexity via **gradient**: Let *f* be differentiable.

f (strictly) convex

$$f(\mathbf{y}) \stackrel{(>)}{\geq} f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$
 for all  $\mathbf{x}, \mathbf{y} \in \mathcal{S}$  (s.t.  $\mathbf{x} \neq \mathbf{y}$ )





### SECOND ORDER CONDITION

Matrix A is **positive (semi)definite** (p.(s.)d.) if  $\mathbf{v}^T A \mathbf{v} \stackrel{(\geq)}{>} 0$  for all  $\mathbf{v} \neq 0$ .

**Notation:** 
$$A \stackrel{(\succ)}{\succ} 0$$
 for  $A$  p.(s.)d. and  $B \stackrel{(\succ)}{\succ} A$  if  $B - A \stackrel{(\succ)}{\succ} 0$ 

Prove convexity via **Hessian**: Let  $f \in C^2$  and  $H(\mathbf{x})$  be its Hessian.

$$f$$
 (strictly) convex  $\iff H(\mathbf{x}) \stackrel{(\succ)}{\succcurlyeq} 0$  for all  $\mathbf{x} \in S$ 

**Alternatively:** Since  $H(\mathbf{x})$  symmetric for  $f \in C^2$ :

 $H(\mathbf{x}) \succcurlyeq 0 \Leftrightarrow$  all eigenvalues of  $H(\mathbf{x}) \ge 0$ 



#### SECOND ORDER CONDITION / 2

Example: 
$$f(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1x_2$$
,  $\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{pmatrix}$ ,  $H(\mathbf{x}) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$ 

× × ×



*f* is convex since  $H(\mathbf{x})$  is p.s.d. for all  $\mathbf{x} \in S$ :

$$\mathbf{v}^{T} \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \mathbf{v} = \mathbf{v}^{T} \begin{pmatrix} 2v_{1} - 2v_{2} \\ -2v_{1} + 2v_{2} \end{pmatrix} = 2v_{1}^{2} - 2v_{1}v_{2} - 2v_{1}v_{2} + 2v_{2}^{2}$$
$$= 2v_{1}^{2} - 4v_{1}v_{2} + 2v_{2}^{2} = 2(v_{1} - v_{2})^{2} \ge 0.$$

### CONVEX FUNCTIONS IN OPTIMIZATION

- For a convex function, every local optimum is also a global one  $\Rightarrow$  No need for involved global optimizers, local ones are enough
- A strictly convex function has at most one optimal point
- $\bullet\,$  Example for strictly convex function without optimum: exp on  ${\rm I\!R}$



Left: Strictly convex; exactly one local minimum, which is also global. Middle: Convex, but not strictly; all local optima are also global ones but not unique. Right: Not convex.

× × 0 × × ×

#### **CONVEX FUNCTIONS IN OPTIMIZATION / 2**

"... in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."

- R. Tyrrell Rockafellar. SIAM Review, 1993.



AMS subject classifications. 49K99, 58C20, 90C99, 49M29

0 0 X X 0 X X