SDSC6015 Stochastic Optimization and Online Learning

Xiang ZHOU

School of Data Science

City University of Hong Kong

Chapter 2: Convex Function

A ***convex optimization problem*** is of the form:

$$\min_{x \in D} f(x)$$

subject to

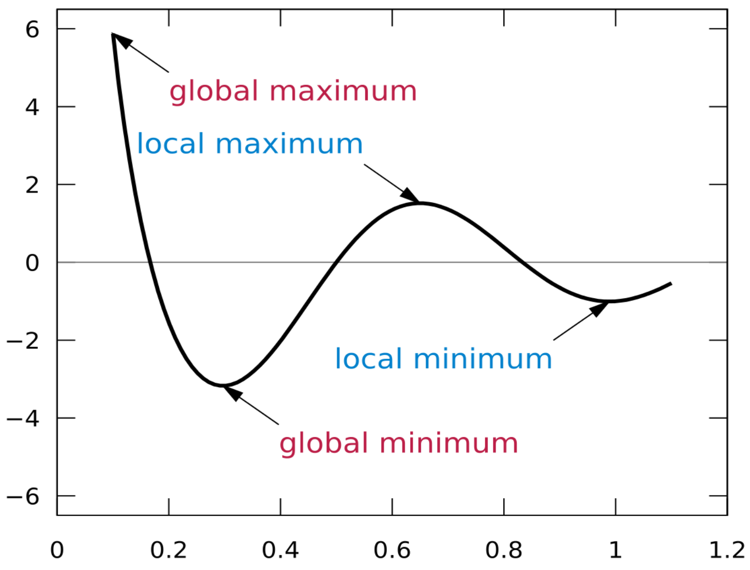
$$\begin{aligned} g_i(x) \le 0, &\quad i = 1, \ldots, m \\ h_j(x) = 0, &\quad j = 1, \ldots, r \end{aligned}$$

where f and g_i are all convex functions, and h_j are affine functions. D is the intersection of the domains of definition of all functions.

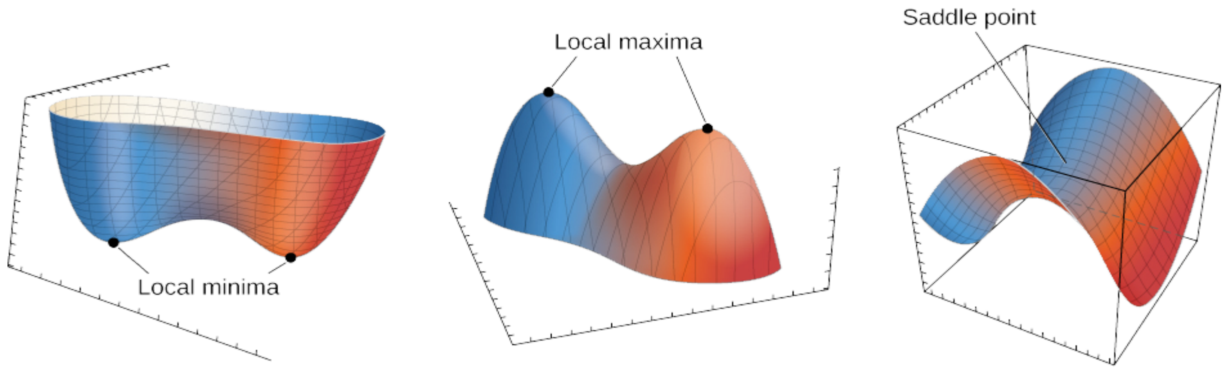
Convex optimization problem always has a solution.
Any local minimizer of a convex optimization problem is a global minimizer.

---

nonconvex optimization : local minimum v.s. global minimum



minima, maxima and saddle points



(a) (b) (c)

Convex Sets

- affine and convex sets
- some important examples
- operations that preserve convexity

Affine sets

- The collection of the lines through any two distinct points in the set
- representation:

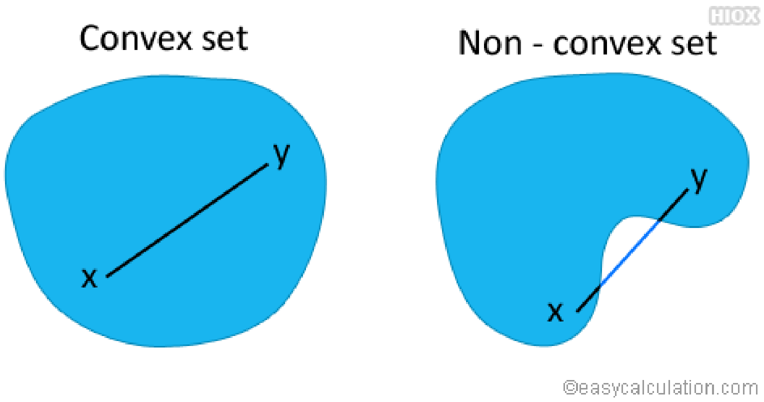
$$x = \sum_i \alpha_i e_i + x_0, \quad \alpha_i \in \mathbb{R}, x_0, e_i \in \mathbb{R}^n$$

- *example* : solution set of the linear equations $\{x|Ax = b\}$
- conversely, every affine set can be expressed as solution set of system of linear equations.
- affine function: $f(x) = Ax + b$ for a matrix A and a vector b.
- The level set of affine function $\{x|Ax + b = 0\}$ is an affine set.

Convex Sets

- $C \subset \mathbb{R}^d$ such that

$$x, y \in C \implies tx + (1-t)y \in C, \quad \text{for all } 0 \leq t \leq 1$$



convex combination and convex hull

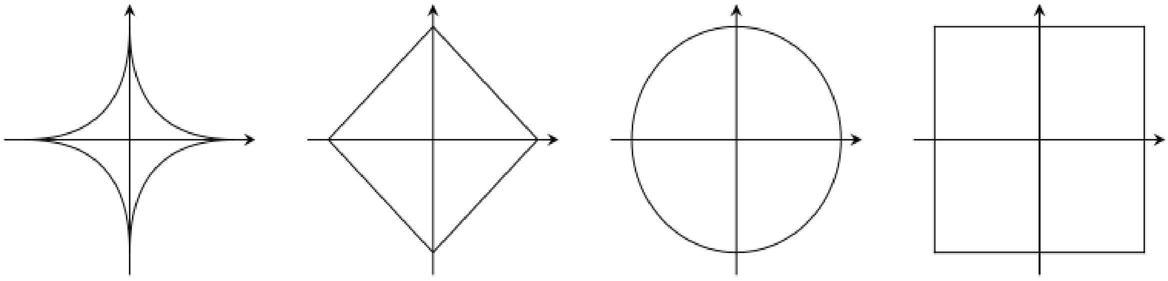
- **Convex combination** of $x_1, \ldots, x_k \in \mathbb{R}^n$: any linear combination

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

  with $\theta_i \geq 0, i = 1, \ldots, k$, and $\sum_{i=1}^k \theta_i = 1$.
- **Convex hull** of a set A , $\operatorname{conv}(A)$, is all convex combinations of points in A.
- convex hull is always convex

Examples of convex sets

- Empty set, point, line.
- Norm ball: $\{x : \|x\|_p \leq r\}$, for $p \geq 1$.



$p = \frac{1}{2}$ $p = 1$ $p = 2$ $p = \infty$

- ellipsoid : $\{x|(x - x_c)^T P^{-1}(x - x_c) \leq r\}$ with the square matrix P positive definite.

Examples of convex sets

- Hyperplane: $\{x : a^T x = b\}$, for given a, b
- Halfspace: $\{x : a^T x \leq b\}$
- Affine space: $\{x : Ax = b\}$, for given A, b
- Polyhedron: $\{x : Ax \leq b\}$, where $\leq$ is interpreted componentwise. The set $\{x : Ax \leq b, Cx = d\}$ is also a polyhedron.
- simplex: special case of polyhedra, given by conv $\{x_0, \ldots, x_k\}$, where these points are affinely independent. The canonical example is the probability simplex,

  $$\text{conv}\{e_1, \ldots, e_n\} = \{w : w \geq 0, 1^T w = 1\}$$

  where e_j is the elementray basis function and 1 is the column vector with one at each entry.

cone and convex cone

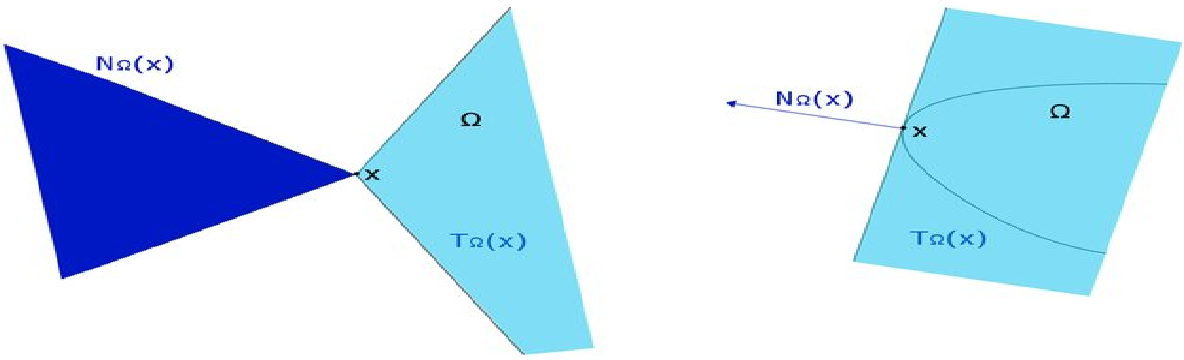
- a set C is a cone if for any $x \in C$, we have $tx \in C$ for all $t \geq 0$.
- convex cone: $x_1, x_2 \in C \Longrightarrow t_1 x_1 + t_2 x_2 \in C$ for all $t_1, t_2 \geq 0$
- *examples* of convex cone:
  - norm cone: $\{(x, r) \in \mathbb{R}^n \times \mathbb{R}_+ : \|x\|_2 \leq r\}$

convex cone

Normal cone: given a set A and a point x on the boundary of A, the normal cone at this point x is

$$\mathcal{N}_A(x) = \{g : g^T x \geq g^T y, \text{ for all } y \in A\}$$

This is always a convex cone, regardless of A.



- If the boudary is smooth, then $\mathcal{N}_A(x)$ is just the outter normal vector at x
- If the boundary is cornered at x, then $\mathcal{N}_A(x)$ is a conic area.

Operations that preserve convexity

- **Intersection**: the intersection of convex sets is convex
- Scaling and translation: if C is convex, then

  $$aC + b := \{ax + b : x \in C\}$$

  is convex for any scalars a, b.
- Affine images and preimages: let $f(x) = Ax + b$ and C is convex, then

  $$f(C) = \{f(x) : x \in C\}$$

  is convex, and if D is convex then

  $$f^{-1}(D) = \{x : f(x) \in D\}$$

  is convex

- coordinate projection: if $C \subset \mathbb{R}^m \times \mathbb{R}^n$ is convex, then

$$T = \{x_1 \in \mathbb{R}^m | (x_1, x_2) \in C\}$$

  is convex.
- *try to prove the above*

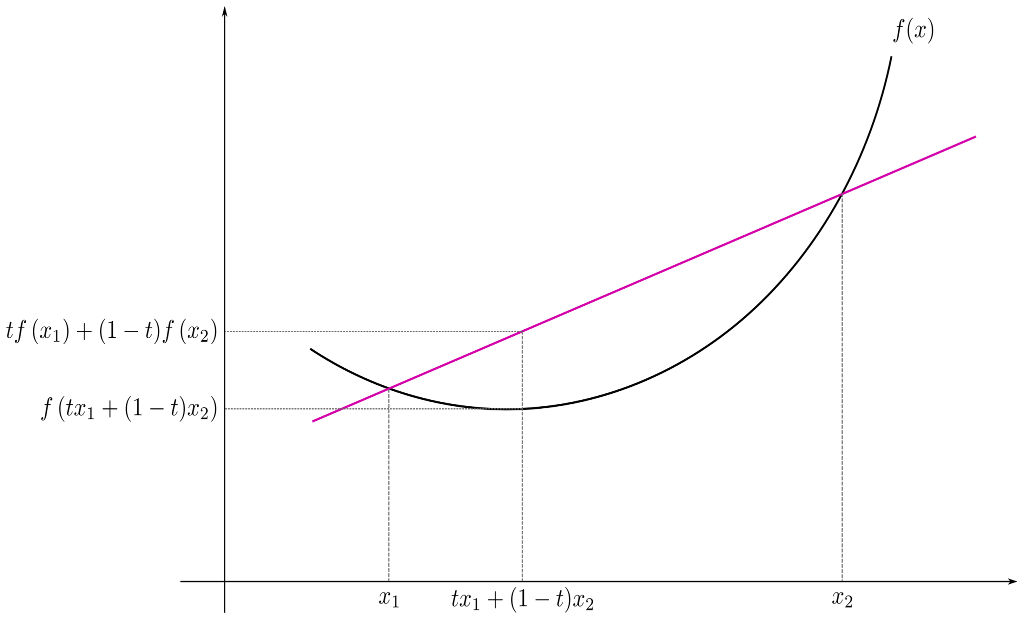
Convex functions

convex function

- **Convex function** $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathrm{dom}(f) \subset \mathbb{R}^n$ is *convex*, and

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \text{ for } 0 \le t \le 1$$

  and all $x, y \in \mathrm{dom}(f)$



- **concave function**: opposite inequality above, so
  f concave if and only if $-f$ convex

strictly convex and strongly convex

- Strictly convex: $f(tx+(1-t)y) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$.
- Strongly convex with parameter $m > 0$: $f - \frac{m}{2}||x||_2^2$ is convex. In words, f is at least as convex as a quadratic function
  Note:
- **strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex**
- *example*:
  - $a^T x + b$ is convex (and concave), not strictly convex;
  - $||x||_1$ is convex, but not strictly convex.
  - $f(x) = x^4$ is strictly convex, but not strongly convex. (see why later)

Examples of convex functions

- Univariate functions:
  - The absolute value function $f(x) = |x|$ is convex , even though it does not have a derivative at the point $x = 0$. It is not strictly convex.
  - Exponential function: e^{ax} is convex for any a over $\mathbb{R}$
  - Power function: x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ (nonnegative reals)
  - Power function: x^a is concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$
  - Logarithmic function: $\log x$ is concave over $\mathbb{R}_{++}$ (positive reals)

- Affine function: $a^T x + b$ is both convex and concave
- Quadratic function: $\frac{1}{2}x^T Q x + b^T x + c$ is convex provided that $Q \succeq 0$ (positive semidefinite)
- Least squares loss: $||y - Ax||_2^2$ is always convex (since $A^T A$ is always positive semidefinite)

Visualization of quadratic function by Python code

Quadratic function: $\frac{1}{2}x^T Q x + b^T x + c$

Norm

- vecotr norm: $||x||_p, p \in [1, \infty]$, is convex as a vector norm; e.g., ℓ_p norms,

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad \text{for } p > 1, \qquad \|x\|_\infty = \max_{i=1,...,n} |x_i|$$

> any norm satisfies triangular inequality : $\|tx + (1-t)y\| \leq t\|x\| + (1-t)\|y\|$ for $t \in [0, 1]$.
> If $p < 1$, $\|x\|_p$ is not a norm.

- (optional) matrix norm: **operator** (*spectral*) and **trace** (*nuclear*) **norms** of the matrix,

$$\|X\|_{\text{op}} = \sigma_1(X), \quad \|X\|_{\text{tr}} = \sum_{i=1}^r \sigma_r(X) = \text{ trace}(\sqrt{X^T X})$$

  where $\sigma_1(X) \geq \ldots \geq \sigma_r(X) \geq 0$ are the singular values of the matrix X.

> singular values of X are the square root of (non-negagtive) eigenvalues of the positive semidefinite matrix $X^T X$.

convex envelope (optional)

- The **convex envelope** of a (possibly nonconvex) function f is the largest convex function g such that $g(x) \leq f(x)$.
- The nuclear norm $\|A\|_*$ is the convex envelope of the rank function $\text{rank}(A)$ on the convex set $\{X : \|X\|_{\text{op}} \leq 1\}$.

Examples of convex functions

- **Indicator function**: if C is convex, then its indicator function

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

  is convex.

- **Support function**: for any set A (convex or not), its support function

$$I_A^*(x) = \max_{y \in A} x^T y$$

  is convex.

- **Max function**: $f(x) = \max\{x_1, \ldots, x_n\}$ where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, is convex.

> Proof: Homework

Examples of convex loss functions in machine learning: linear regession

Given $y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}, p$ is the dimension of the input x.

- mean square error in ordinary linear regession is a quadratic function

$$\min_{\beta\in\mathbb{R}^p} L(\beta) = \|y - X\beta\|_2^2$$

- $\lambda > 0$, the objective function in ridge regression:

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2$$

- $\lambda > 0$, the objective function in lasso regression (regularized form):

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Examples of loss functions in machine learning: binary classification

We use t to refer to the product $yf(x)$ for the label $y \in \{-1, +1\}$ and the discriminant function $f(x)$. The classifier is thus $\mathrm{sign}(f(x))$.

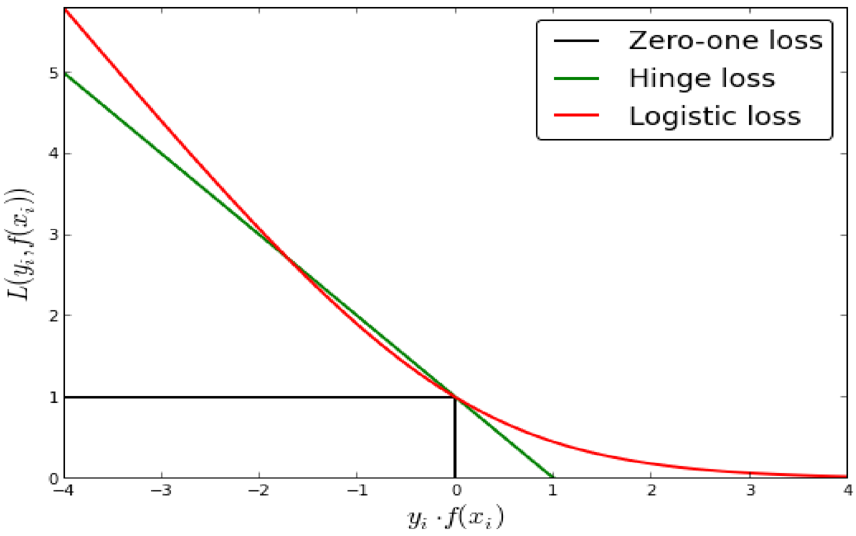
- 0-1 (misclassification) loss is *non-convex*:

$$l_{01}(t) = \begin{cases} 1, & t < 0 \\ 0, & t \geq 0 \end{cases}$$

- **hinge loss** in SVM

$$h(t) = \max(1 - t, 0)$$

  is convex.



- **Logistic loss** in logistic regression

$$l(t) = \frac{1}{\log 2} \log(1 + e^{-t})$$

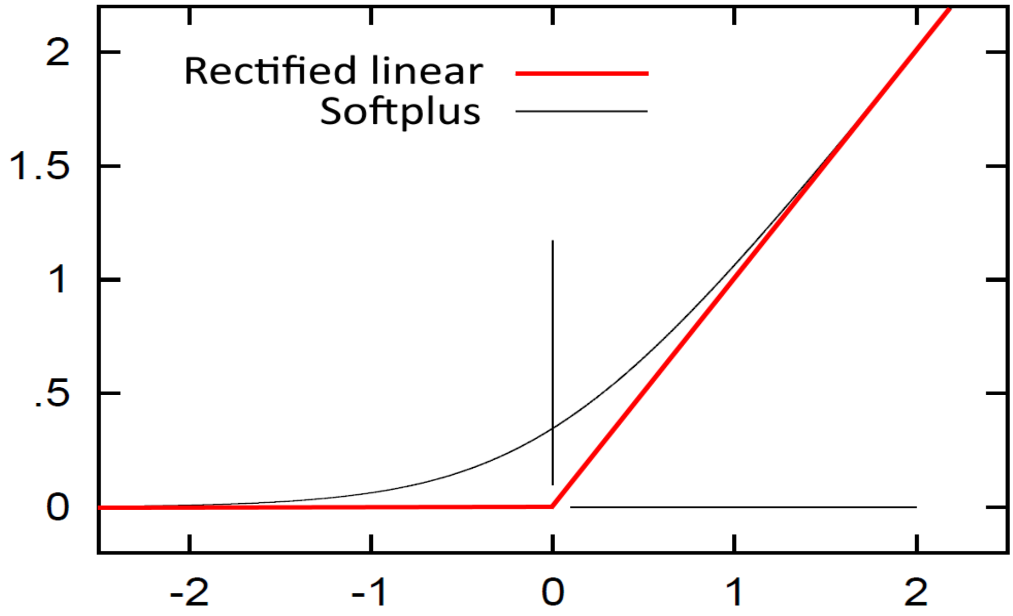
where t refers to the product $yf(x)$ for the label $y \in \{-1, +1\}$ and the discriminant function $f(x)$.

> logistic regression usually encodes the labels by $v \in \{0, 1\}$, not by $\{-1, +1\}$. By transforming $y = 2v - 1$, the loss function takes the conventional form
>
> $$\log(1 + e^{(1-2v)f(x)}) = \begin{cases} \log(1 + e^{f(x)}) = \log(1 + e^{-f(x)}) + f(x), & v = 0 \\ \log(1 + e^{-f(x)}), & v = 1 \end{cases}$$
>
> is convex in θ as the discriminant function takes the linear form $f(x) = \theta \cdot x$, regardless $v = 0, 1$.

- **softplus** function

$$\log(1 + e^{x})$$

is convex (note $\log(1 + e^{-x})$ is convex too).

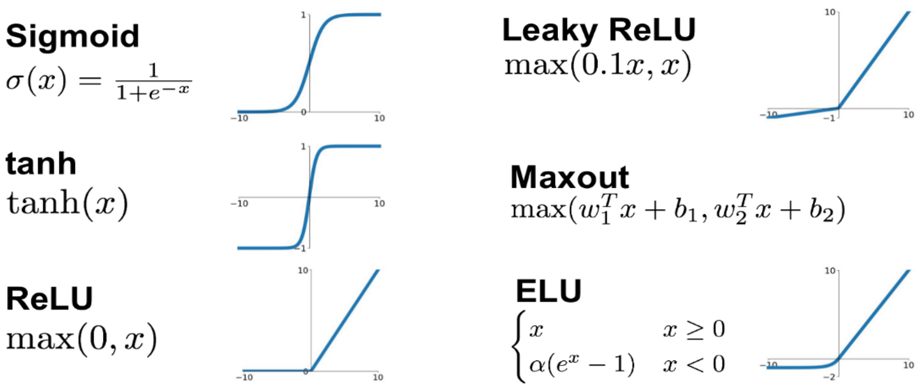


- Exponential loss used in AdaBoost: e^{-t}

---

activation functions

which are convex and which are not in the following figures?



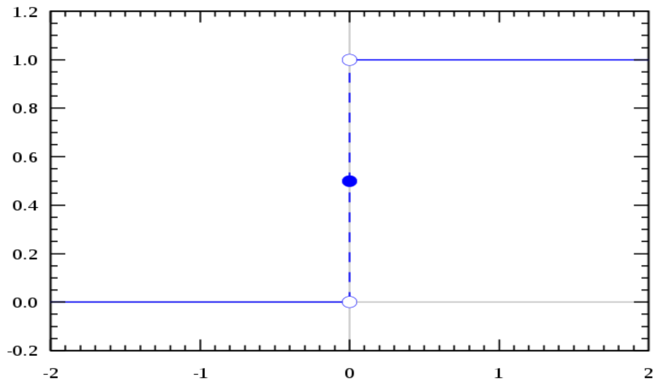
> ELU (Exponential Linear Unit) is convex iff $\alpha \in [0, 1]$

Examples of non-convex functions

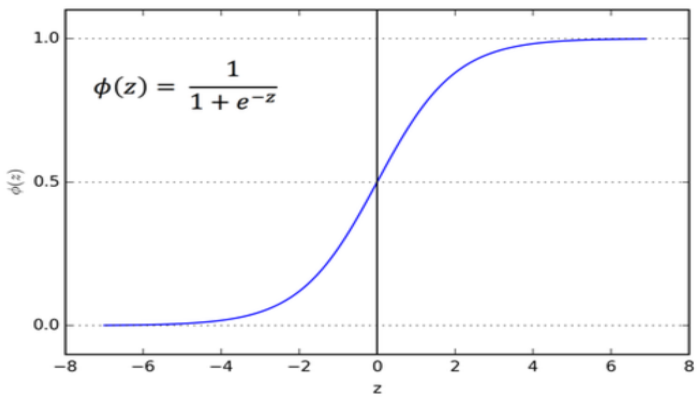
- 0-1 (misclassification) loss is *non-convex*:

$$l_{01}(t) = \begin{cases} 1, & t < 0 \\ 0, & t \geq 0 \end{cases}$$

- Heaviside step function /sign function



- sigmoid function/logistic function



- Gaussian function

$$f(x) = a \cdot \exp\left(-\frac{(x-b)^2}{2c^2}\right)$$

Key properties of convex functions

- A function $f(x)$ is convex if and only if its restriction to any line (a straight line is represented by $t \mapsto x_0 + tv$),

$$F(t) = f(x_0 + tv)$$

is convex, where x_0 and v are given and $\mathrm{dom}(F) = \{t | x_0 + tv \in \mathrm{dom}(f)\}$

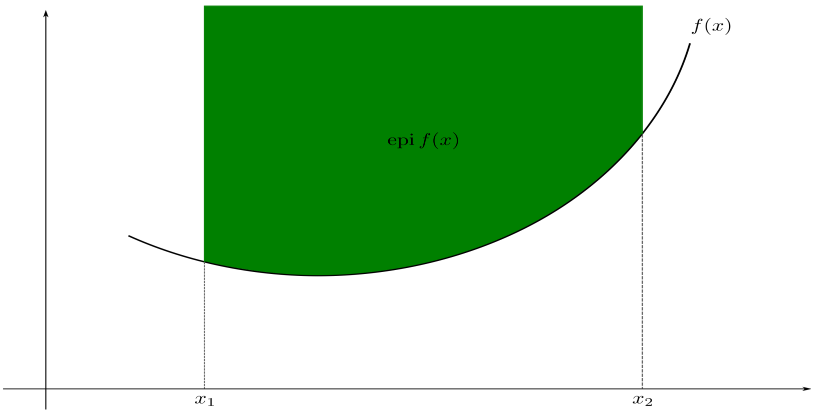
> Proof: Homework

- Epigraph characterization: a function f is convex **if and only if** its ***epigraph***

$$\mathrm{epi}(f) = \{(x,t) \in \mathrm{dom}(f) \times \mathbb{R} : f(x) \leq t\} \subset \mathbb{R}^{n+1}$$

is a convex set.

> Proof: If f is convex, then let $z_i = (x_i, t_i), i = 1, 2$ be two different points in $\mathrm{epi}(f)$. Then for any $\alpha \in (0, 1)$,
>
> $$\alpha z_1 + (1-\alpha)z_2 = (\alpha x_1 + (1-\alpha)x_2, \alpha t_1 + (1-\alpha)t_2).$$
>
> Since $\mathrm{dom}(f)$ is convex, then $\alpha x_1 + (1-\alpha)x_2 \in \mathrm{dom}(f)$ and since f is convex, then $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) \leq \alpha t_1 + (1-\alpha)t_2$. So, $\alpha z_1 + (1-\alpha)z_2 \in \mathrm{epi}(f)$,i.e., $\mathrm{epi}(f)$ is convex.
> If $\mathrm{epi}(f)$ is convex, then $\mathrm{dom}(f)$ is convex and let $t_1 = f(x_1), t_2 = f(x_2)$, and consider (x_1, t_1) and (x_2, t_2) in the convex set $\mathrm{epi}(f)$, we then have $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha t_1 + (1-\alpha)t_2 = \alpha f(x_1) + (1-\alpha)f(x_2)$. So f is a convex function.



- Convex sublevel sets: if f is convex, then its **sublevel sets**

$$L_t := \{x \in \mathrm{dom}(f) : f(x) \leq t\}$$

are convex for all $t \in \mathbb{R}$. The converse is *not* true.

> Proof: Homework.

Differentiable convex function

characterization of differentiable convex function: gradient

- **First-order characterization**: if f is differentiable, then f is convex *if and only if*

- $\mathrm{dom}(f)$ is convex, and
- $$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

  for all $x, y \in \mathrm{dom}(f)$.

> Theorem 2.1.2 in [Nesterov (2004)]
> Proof: Let $x_t = tx + (1-t)y$ for any $t \in [0, 1)$.
> If f is convex, then $f(x_t) \leq (1-t)f(y) + tf(x)$, i.e. $f(y) \geq \frac{1}{1-t}f(x_t) - \frac{t}{1-t}f(x) = \frac{1}{1-t}(f(x_t) - f(x)) + f(x)$. Let $t \to 1$, then the conclusion is proved by noting $f'(x_t) = \nabla f(x_t) \cdot (x - y)$.
> On the other hand, if $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ for any x, y, then $f(y) \geq f(x_t) + \nabla f(x_t)^T(y - x_t)$ and $f(x) \geq f(x_t) + \nabla f(x_t)^T(x - x_t)$.
> Then adding them together: $tf(x) + (1-t)f(y) \geq f(x_t) + \nabla f(x_t)^T\Big((1-t)(y - x_t) + t(x - x_t)\Big) = f(x_t)$.

- Therefore for a *differentiable convex function*

$$\nabla f(x^*) = 0 \Longleftrightarrow x^* \text{ minimizes } f$$

> Proof: " $\Longrightarrow$ " : $f(y) \geq f(x)$ so x is a minimizer; " $\Longleftarrow$ ": first order optimality condition

- If f is differentiable, then f is convex *if and only if*
  - $\mathrm{dom}(f)$ is convex, and
  - $$(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0$$

  for all $x, y \in \mathrm{dom}(f)$. This property is called *monotonicity* of the gradient.

> Theorem 2.1.3 in [Nesterov (2004)]
> Proof: If f is convex, then $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ as well as $f(x) \geq f(y) + \nabla f(y)^T(x-y)$. Adding together, $0 \geq (\nabla f(y) - \nabla f(x))^T(x-y)$.
> On the other hand, if $(\nabla f(x) - \nabla f(y))^T(x-y) \geq 0$, then let $x_t = x + t(y-x)$, then $\frac{d}{dt}f(x_t) = \nabla f(x_t) \cdot (y - x)$ and $f(y) = f(x) + \int_0^1 \nabla f(x_t) \cdot (y-x)dt = f(x) + \nabla f(x) \cdot (y-x) + \int_0^1 (\nabla f(x_t) -$

> $\nabla f(x)) \cdot \frac{1}{t}(x_t - x)dt \geq f(x) + \nabla f(x) \cdot (y-x)$. So, $f(y) \geq f(x) + \nabla f(x) \cdot (y-x)$, then f is convex.

- If $x \in \mathbb{R}^1$ and $f(x)$ is differentiable and $f'(x)$ is nondecreasing, then f is convex.
- If $x \in \mathbb{R}^1$ and $f(x)$ is differentiable and $f'(x)$ is strictly increasing, then f is stirctly convex.

> $f(x) = x^4$ is strictly convex since x^3 is strictly increasing.

Bregmann divergence

Bregman divergence or *Bregman distance* for continuously-differentiable and strictly convex function:

$$D_f(y,x) := f(y) - f(x) - \nabla f(x)^T (y-x) \geq 0$$

> **Exercise**: if $f(x) = \|x\|_2^2$, show that $D_f(y,x) = \|y-x\|_2^2$.
> **Exercise** Show the (non-negative) linearity of Bregmann divergence: $D_{f+\lambda g}(y,x) = D_f(y,x) + \lambda D_g(y,x)$ for f and g strictly convex and differentiable, $\lambda \geq 0$

---

characterization of twice-differentiable convex function: Hessian matrix

- Second-order characterization: if f is *twice differentiable*, then f is convex **if and only if** $\mathrm{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$
- $\nabla^2 f(x) \succ 0$ (strictly positive definite) for all $x \Longrightarrow$ strictly convex
(the converse is not true)
- **strongly convex**: there exists $m > 0$ such that $\nabla^2 f(x) \succeq m$

Exercise : show that $-\log(1-x) > x$ for all $x \in (0,1)$

Examples of (differentiable) convex functions

- $f(x) = x^4$ is strictly convex, but $f''(0) = 0$ and not strongly convex.

- $f(x) = \exp(-x)$ is strictly convex, but $f''(x) = \exp(-x)$ can be arbitrarily small, so it is not strongly convex.
- x^2 is strongly convex.
- $f(x) = 1/x^2$, with $\operatorname{dom} f = \{x \in \mathbb{R} \mid x \neq 0\}$, satisfies $f''(x) > 0$ for all $x \in \operatorname{dom}(f)$, but is not a convex function.

> Why? Is $\operatorname{dom}(f)$ convex?

- Quadratic-over-linear function. The function $f(x, y) = x^2/y$, with

$$\operatorname{dom} f = \mathbf{R} \times \mathbf{R}_{++} = \left\{(x, y) \in \mathbf{R}^2 \mid y > 0\right\}$$

is convex (check Hessian matrix is positive semidefinite. $H = \begin{bmatrix} 2/y, & -2x/y^2 \\ -2x/y^2, & 2x^2/y^3 \end{bmatrix}$. $\det H = 4x^2/y^4 - 4x^2/y^4 = 0$).
- Log-sum-exp. The function $f(x) = \log\left(e^{x_1} + \cdots + e^{x_n}\right)$ is convex on $\mathbb{R}^n$. This function can be interpreted as a differentiable approximation of the max function, since

$$\max\left\{x_1, \ldots, x_n\right\} \leq f(x) \leq \max\left\{x_1, \ldots, x_n\right\} + \log n$$

---

characterization of differentiable strictly convex function:

If f is differentiable and $\operatorname{dom}(f)$ is convex, then the following statement

1. f is strictly convex.
2. $$f(y) > f(x) + \nabla f(x)^T (y - x)$$
   for all $x \neq y$ in $\operatorname{dom}(f)$.
3. $$(\nabla f(x) - \nabla f(y))^T (x - y) > 0$$
   for all $x \neq y$ in $\operatorname{dom}(f)$.

Lipschitz gradients and strong convexity

Let f be convex and twice continuously differentiable. Then the following statements are equivalent.

1. $\nabla f(x)$ is Lipschitz with constant L, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

   for any x, y.
2. $(\nabla f(x) - \nabla f(y))^T(x - y) \leq L\|x - y\|_2^2$ for all x, y
3. $\nabla^2 f(x) \preceq LI$ for all x where I is the identity matrix
4. $f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|_2^2$ for all x, y.

> Some literature call f is L-smooth.

equivalent condition of strong convexity

The following statements are equivalent.

1. f is strongly convex with constant m
2. $(\nabla f(x) - \nabla f(y))^T(x - y) \geq m\|x - y\|_2^2$ for all x, y
3. $\nabla^2 f(x) \succeq mI$ for all x
4. $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|_2^2$ for all x, y

> Proof : skip

Jensen's inequality

Jensen's inequality: if f is convex, and X is a random variable supported on $\operatorname{dom}(f)$, then

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

> Example: $f(x) = x^2$. then Jensen's inequalty is $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$, which is equivalent to the familiar nonnegativity of variance$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$.

Operations that preserve convexity

- **nonnegative linear combination**: $f_1, \ldots, f_m$ convex implies

$$a_1 f_1 + \ldots + a_m f_m$$

  are also convex for any $a_1, \ldots, a_m \geq 0$.
  - ***example***: If each loss $\ell_i(\theta)$ from each data sample is convex, then the empirical risk from all data samples

$$L(\theta) = \sum_i \ell_i(\theta)$$

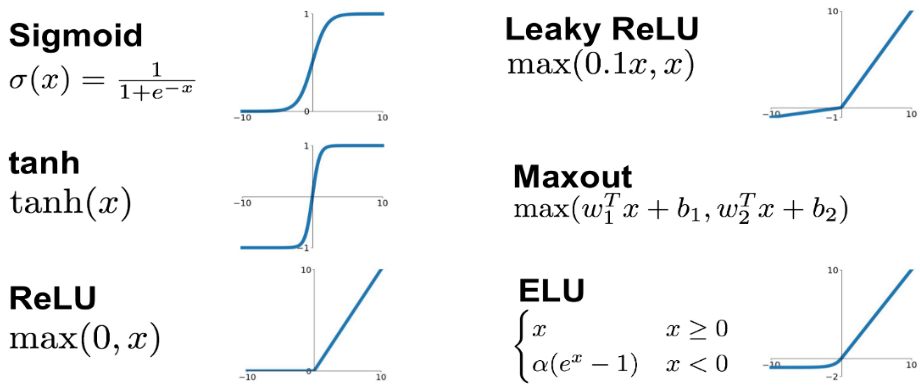
    is also convex.
  - The objective functions in ridge regression and lasso regression are both convex.
  - In general, convex risk function + convex regularization is convex.
- **pointwise maximization**: if $f(x; s)$ is convex in x for each s, then

$$f(x) = \max_s f(x; s)$$

  is also convex
  - support function $f(x) = \max_{y \in A} y^T x$ is a special case
  - if $f_1, \ldots, f_m$ are all convex, then

$$f(x) = \max(f_1(x), \ldots, f_m(x))$$

    is also convex
  - Maxout/Leaky RELU function is convex:



- **affine composition**: if f is convex, then

$$g(x) = f(Ax + b)$$

  is convex.
  - $||\beta||_2^2$ is convex , so $||y - X\beta||_2^2$ is also convex.
  - $f(x)$ is convex, then $f(-x)$ is also convex.
- **partial minimization**: if $g(x, y)$ is convex in (x, y), and the set C is convex, then

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

is convex.

- *Example*: the distance function to a convex set C $d(x, C) = \min_{y \in C} \|x - y\|_2$ is convex.

> Proof : For two poitns x_0, x_1, let $x_t = tx_1 + (1-t)x_0$ and define y_0, y_1 such that $f(x_0) = g(x_0, y_0)$ and $f(x_1) = g(x_1, y_1)$. Let $y_t = ty_1 + (1-t)y_0$. Then $f(x_t) = \min_y g(x_t, y) \leq g(x_t, y_t) \leq tg(x_1, y_1) + (1 - t)g(x_0, y_0) = tf(x_1) + (1-t)f(x_0)$.

- reader: try to prove the above.

Homework

Reference of convex function

- S. Boyd and L. Vandenberghe (2004), “Convex optimization”, Chapters 2 and 3
- J.P. Hiriart-Urruty and C. Lemarechal (1993), “Fundamentals of convex analysis”, Chapters A and B
- R. T. Rockafellar (1970), “Convex analysis”, Chapters 1–10
- Yurii Nesterov (2004), “Introductory Lectures on Convex Optimization”. Chapter 2.1
- S. Bubeck et al. Convex optimization: Algorithms and complexity. Foundations and Trends in Machine Learning, 8(3-4):231–357, 2015.

Homework

Prove the following statements by the definition of the convex function

1. **Max function**: $f(x) = \max\{x_1, \ldots, x_n\}$ where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, is convex.
2. A function $f(x)$ is convex if and only if its restriction to any line,

$$F(t) = f(x_0 + tv)$$

is convex, where x_0 and v are given and $\mathrm{dom}(F) = \{t | x_0 + tv \in \mathrm{dom}(f)\}$.

3. If f is convex, then its **sublevel sets**

$$L_t := \{x \in \mathrm{dom}(f) : f(x) \leq t\}$$

are convex for all $t \in \mathbb{R}$. Show the converse is *not* true.

4. Prove that the **entropy function**, defined as

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i)$$

with $\mathrm{dom}(f) = \left\{x \in \mathbb{R}^n_{++} : \sum_{i=1}^n x_i = 1\right\}$, is **strictly concave**.