Convex Functions

October 9, 2024

1 Convex functions

- The domain of f

$$\mathrm{dom}(f) = \{x : \ f(x) \text{ is defined and finite}\}$$

Definition 1. Convex function: $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathrm{dom}(f) \subset \mathbb{R}^n$ convex, and

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

for $0 \leq t \leq 1$ and all $x, y \in \mathrm{dom}(f)$.

Example 1. Indicator function is a convex function

$$I_C(tx + (1-t)y) \leq tI_C(x) + (1-t)I_C y$$

If $x, y \in C$, then

$$0 \leq 0$$

If $x \notin C$ and $y \notin C$

$$... \leq \infty$$

Proposition 1. *Note: strongly convex* $\implies$ *strictly convex* $\implies$ *convex*

Proposition 2. *If* f *is differentiable, and* $\forall x, y \in \text{dom}(f)$*,*

- f is **convex** $\iff$ $f(y) \geq f(x) + \nabla f(x)(y - x)$
- f is **strictly convex** $\iff$ $f(y) > f(x) + \nabla f(x)(y - x)$
- f is **strongly convex**

$$\iff f(y) \geq f(x) + \nabla f(x)(y - x) + \frac{m}{2}||y - x||_2^2$$

ie,

- if $m = 0$, convex function
- if $m \to 0$, strictly convex
- if $m > 0$, strongly convex

Proposition 3. *If f is twice continuously differentiable,*

- f is **convex** $\iff$

$$f''(x) \geq 0, \quad \forall x \in \text{dom}(f)$$

$$\nabla^2 f(x) \succeq 0, \quad \text{(positive semidefinite)}$$

- f is **strictly convex** $\iff$

$$f''(x) > 0, \quad \forall x \in \text{dom}(f)$$

$$\nabla^2 f(x) \succ 0 \quad \text{(positive definite)}$$

- f is **strongly convex** $\iff$

$$f''(x) \geq m > 0 \quad \forall x \in \mathrm{dom}(f)$$

$$\nabla^2 f(x) \succeq m \succ 0 \quad \text{(bounded)}$$

Example 2. If f is strictly convex with $f''(x_n) = \frac{1}{n}$, then it is not strongly convex since

$$\lim_{n\to\infty} f''(x_n) = \lim_{n\to\infty} \frac{1}{n} = 0$$

Example 3. Least squares loss.

$$\min_{\beta} f(\beta) \iff \min_{\beta} ||y - X\beta||_2^2$$

$$\nabla^2 f(\beta) = X^T X \succeq 0$$

1. $X^T X \succ 0, n \geq p$ full column rank;

2. $X^T X \succeq 0$, otherwise.

Proposition 4. *First-order characterization: If f is differentiable, then f is convex if and only if dom(f) is convex and*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

for all $x, y \in \mathrm{dom}(f)$. *So for a differentiable convex function,*

$$\nabla f(x) = 0 \implies x \textit{ minimizes } f$$

Example 2: Nonnegative linear combination - Logistic Regression

In logistic regression, we wish to solve

$$\min_{\beta} \sum_{i=1}^{n} \log(1 + \exp(-y_i \mathbf{x}_i^T \beta))$$

for $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \pm 1$. To verify that this function is convex, we need to verify if

$$f(t) = \log(1 + \exp(t))$$

is convex. We take the second derivative,

$$f''(t) = \frac{e^t}{(1 - e^t)^2} > 0$$

and conclude that it is convex.

1.1 Lipschitz continuity and strong convexity

Let f be convex and twice differentiable.

Show that the following statements are equivalent.

1. ∇f is Lipschitz with constant $L > 0$;
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;
4. $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|_2^2$ for all x, y.

Show that the following statements are equivalent.

1. f is strongly convex with constant $m > 0$;

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.

1.2 Convex function examples

The distance function to a closed, convex set C

$$dist(x, C) = \min_{y \in C} \|y - x\|_2 = \min_{y \in C} h(x, y) = f(x) \tag{1}$$

is a convex function of x.

Let the solution of the function be $u = P_C(x)$, where $P_C(x)$ is the projection of x onto C. Write $dist(x, C) = \|x - P_C(x)\|_2 = \|x - u\|_2$.

Proposition 5. *It turns out that when* $dist(x, C) > 0$*,*

$$\partial dist(x, C) = \frac{x - u}{\|x - u\|_2}$$

Only has one element, so in fact $dist(x, C)$ is differentiable and this is its gradient.

Proof. We will only show one direction, i.e., that

$$\frac{x - u}{\|x - u\|_2} \in \partial dist(x, C)$$

By first-order optimality condition:

$$\min_x \quad f(x) \quad \text{subject to} \quad x \in C$$

is solved at x, for f convex and differentiable, if and only if

$$\nabla f(x)^T(y-x) \geq 0 \quad \text{for all} \quad y \in C$$

$\square$

for a projection,

$$\nabla f(u)^T(y-u) \geq 0 \quad \text{for all} \quad y \in C$$

which means

$$(u-x)^T(y-u) \geq 0 \quad \text{for all} \quad y \in C$$

Hence

$$C \subseteq H = \{y : (u-x)^T(y-u) \geq 0\} = \{y : (x-u)^T(y-u) \leq 0\}$$

where H is a convex set. Now we claim

$$dist(y, C) \geq \frac{(x-u)^T(y-u)}{\|x-u\|_2} \quad \text{for all} \quad y$$

We can check this is true. First, for $y \in H$, which means that we have $(x-u)^T(y-u) \leq 0$, then the right-hand side is ≤ 0. Now for $y \notin H$, we have $(x-u)^T(y-u) \leq \|x-u\|_2\|y-u\|_2$ by Cauchy–Schwarz inequality. Thus

$$\frac{(x-u)^T(y-u)}{\|x-u\|_2} \leq \frac{\|x-u\|_2\|y-u\|_2}{\|x-u\|_2} = \|y-u\|_2 \leq dist(y, H) \leq dist(y, C) \qquad (2)$$

as desired. Using the claim, we have for any y

$$dist(y, C) \geq \frac{(x-u)^T(y-x+x-u)}{\|x-u\|_2} = \|x-u\|_2 + \left(\frac{x-u}{\|x-u\|_2}\right)^T (y-x) \qquad (3)$$

which is equivalent to

$$f(y) \geq f(x) + \left(\frac{x-u}{\|x-u\|_2}\right)^T (y-x) \qquad (4)$$

Thus by the definition of subgradient $g = \frac{x-u}{\|x-u\|_2}$ is a subgradient of $dist(x, C)$ at x.