Section 05: Convexity and Subgradients

1. Convexity

Convexity is defined for both sets and functions. For today we'll focus on discussing the convexity of functions.

Definition 1 (convex functions). *A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** on a set A if for all $x, y \in A$ and $\lambda \in [0, 1]$:*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

When the this definition holds with the inequality being reversed, then f is said to be concave. From the definition, it is clear that a function f is convex if and only if $-f$ is concave.

(a) Why do we care whether a function is convex or not?

(b) Which of the following functions are convex? (Hint: draw a picture)
(i) $|x|$ (ii) $\cos(x)$ (iii) $x^T x$

(c) Can a function be both convex and concave on the same set? If so, give an example. If not, describe why not.

2. Practical Methods for Checking Convexity

Using the definition to check whether a function is convex or not can be a tedious task in many situations. Some basic methods that can help us achieve the task in an efficient way are introduced below:

- for differentiable function, examine $f(y) \geq f(x) + \nabla f(x)^T(y - x)$ for any x, y in the domain of f.
- for twice differentiable functions, examine $\nabla^2 f(x) \succeq 0$ (i.e., the Hessian matrix is positive semidefinite).
- nonnegative weighted sum
- composition with affine function
- composition with convex function
- pointwise maximum and supremum

Note: there are even more such methods, which are covered in a convex optimization course or textbook.

(a) If f is differentiable, then f is convex if and only if $f(y) \geq f(x) + \nabla f(x)^T(y - x)$ for any x, y in the domain of f. A geometric interpretation of this characterization is that any tangent plane of a convex function f must lie entirely below f. One interesting application of this characterization is one of the most important inequalities in probability and statistics: the Jensen's inequality, which states that $\mathbb{E}f(X) \geq f(\mathbb{E}(X))$ when f is convex.

(b) Let $\alpha \geq 0$ and $\beta \geq 0$, and if f and g are convex, then αf, $f + g$, $\alpha f + \beta g$ are all convex. One application: When a (possibly complicated) objective function can be expressed as a sum (e.g., the negative log-likelihood function), then showing the convexity of each individual term is typically easier.

(c) Suppose $f(\cdot)$ is convex, then $g(x) := f(Ax + b)$ is convex. Use this method to show that $||Ax + b||_1$ is convex (in x), where $||z||_1 = \sum_i |z_i|$.

(d) Suppose $f(\cdot)$, $g(\cdot)$ are both convex and f is non-decreasing. Show that $(f \circ g)(\cdot)$ is also convex.

(e) Suppose you know that f_1 and f_2 are convex functions on a set A. The function $g(x) := \max\{f_1(x), f_2(x)\}$ is also convex on A. In general, if $f(x, y)$ is convex in x for each y, then $g(x) := \sup_y f(x, y)$ is convex. Use this method to show that the largest eigenvalue of a matrix X, $\lambda_{\text{Max}}(X)$, is convex in X (Using the definition of convexity would make this question quite difficult).

(f) Does the same result hold for $h(x) := \min\{f_1(x), f_2(x)\}$? If so, give a proof. If not, provide convex functions f_1, f_2 such that h is not convex.

(g) Show that the objective function in linear regression $\|Y - Xw\|_2^2$ is convex in w.

(h) Show that the objectives for LASSO $||Y - X\omega||_2^2 + \lambda||\omega||_1$ is convex in ω.

(i) If f is twice differentiable with convex domain, then f is convex if and only if

$$\nabla^2 f(x) \succeq 0,$$

for any x in the domain of f.

Recall that the objective for logistic regression is

$$\arg\min_w \sum_{i=1}^{n} \log(1 + \exp(-y_i x_i^T w)) + \lambda\|w\|_2^2$$

Use this method to show that the objective function above is convex in w.

3. Subgradients

We start with the definition of subgradients before discussing the motivation and its usefulness.

Definition 2 (subgradients). *A vector $g \in \mathbb{R}^d$ is a subgradient of a convex function $f : D \longrightarrow \mathbb{R}$ at $x \in D \subseteq \mathbb{R}^d$ if*

$$f(y) \geq f(x) + g^T(y - x) \quad \textit{for all } y \in D.$$

One interpretation of subgradient g is that the affine function (of y) $f(x) + g^T(y - x)$ is a global underestimator of f. Note that if a convex function f is differentiable at x (i.e., $\nabla f(x)$ exists), then $f(y) \geq f(x) + \nabla f(x)^T(y - x)$ is true for all $y \in D$, meaning that $\nabla f(x)$ is a subgradient of f at x. But a subgradient can exist even when f is not differentiable at x.

(a) Why are subgradients useful in optimization? If $g = 0$ is a subgradient of a function f at x^*, what does it imply?

(b) What are the subgradients of $f(x) = |x|$ at 0? (Hint: draw a picture and note that subgradients at a point might not be unique)

(c) Some important results about subgradients are

- If f is convex, then a subgradient of f at $x \in \text{int}(D)$ (interior of the domain of f) always exists.
- If f is convex, then f is differentiable at x if and only if $\nabla f(x)$ is the only subgradient of f at x.
- A point x^* is a global minimizer of a function f (not necessarily convex) if and only if $g = 0$ is a subgradient of f at x^*.