MAT 653: Statistical Simulation

Instructor: Dr. Wei Li

2023-09-07

Basics of vector calculus

Gradient. For $f : \mathbb{R}^n \to \mathbb{R}$ a smooth function, its gradient is the $n\times 1$ vector

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

Hessian matrix. The Hessian matrix of f is a $n\times n$ symmetric matrix defined as

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}.$$

Some results from calculus:

(1) $f(x) = a^T x : \mathbb{R}^\ltimes \to \mathbb{R}$, $f(x)$ can also written in $f(x) = \sum_{i=1}^{n} a_i x_i$, the gradient for $f(x)$,

$$\nabla f(x) = a.$$

(2) $f(x) = x^T A x$ (A is square matrix), the gradient for $f(x)$,

$$\nabla f(x) = Ax + A^T x;$$

if A is symmetric matrix,

$$\nabla f(x) = 2Ax.$$

Example: For the linear least square problem $\min_{x} f(x) = \min_{x} \|Ax - b\|^2,$

$$f(x) = \|Ax - b\|^2$$
$$= \langle Ax - b, Ax - b \rangle$$
$$= (Ax - b)^T (Ax - b)$$
$$= x^T A^T Ax - 2b^T Ax + b^T b$$

the gradient of $f(x)$ is

$$\nabla f(x) = \frac{\partial f(x)}{\partial x} = 2A^T Ax - 2b^T A \overset{set}{=} 0$$

Optimization

reference: Numerical Optimization (Nocedal and Wright)

Definition

Given $f : \mathbb{R}^{\times} \to \mathbb{R}$, $S \subseteq \mathbb{R}^{\times}$ is feasible convex set.

1. Convex set: S is a convex set if and only if $x, y \in S$ entails $\alpha x + (1 - \alpha)y \in S$ for all $\alpha \in [0, 1]$

2. Global minimizer: A point $x^* \in S$ is a global minimizer of f in S if and only if $f(x^*) \leq f(x)$ for all $x \in S$.

3. Local minimizer A point $x^* \in S$ is a local minimizer of f in S if and only if there exists a neighborhood of x^*, $N(x^*)$ such that $f(x^*) \leq f(x)$, for all $x \in N(x^*) \cap S$.

4. Strict local minimizer: A point $x^* \in S$ is a local minimizer of f in S if and only if there exists a neighborhood of x^*, $N(x^*)$ such that $f(x^*) < f(x)$, for all $x \in N(x^*) \cap S$.

5. Convex functions:

(1) A "smooth" function f is a convex at the point v if and only if $\nabla^2 f(v) \geq 0$ (positive semidefinite). If $f : \mathbb{R} \to \mathbb{R}$, $(\partial^2 f(x)/\partial x^2)|_{x=v} \geq 0$, noting $\partial f(v)/\partial x$ slope at v = rate of change of f at v; $\partial^2 f(v)/\partial x^2$ = rate of change of the slope of f at v. This is second order definition.

(2) A "differentialbe" function f is convex at v if and only if for all w in the neighborhood of v, $f(w) \geq f(v) + \nabla f(v)^T (w - v)$. This is first order definition.

(3) A function f is convex on S if and only if it's convex at every point of S. A function f is convex on S, if and only if $x, y \in S$ entails $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$ for $\alpha \in [0, 1]$.

6. stationary point and saddle point

A stationary point v of f is any point that satisfies the FOC $\nabla f(v) = 0$. In general, a stationary point could be minimizer, maximizer or saddle point.

A saddle point is a stationary point at which the curvature of the function changes from negative to positive or vice versa (i.e., also a point of inflection). More precise definition is given below.

7. Consequence of convexity

When f is convex on S, then any local minimizer is also a global minimizer. If, in addition, f is differentiable, then any stationary point is a global minimizer.

Necessary and sufficient conditions.

Let's consider the minimization problem $\min_{x \in \mathbb{R}^n} f(x)$ for some smooth $f : \mathbb{R}^n \mapsto \mathbb{R}$.

Necessary condition   First order result: suppose f is continuously differentiable. If x^* is a minimizer of f then $\nabla f(x^*) = 0$.

Second order result: suppose f is twice continuously differentiable. If x^* is a minimizer of f, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \geq 0$.

Sufficient condition   Second order result: suppose f is twice continuously differentiable, for some $x^* \in \mathbb{R}^n$, $\nabla^2 f(x^*) > 0$ and $\nabla f(x^*) = 0$. Then x^* is a strict local minimizer.

Hessian, eigenvalues and eigenvectors

Hessian matrix contains important geometric information about the surface of the function f (in particular the concavity of the surface).

Recall that the Taylor approximation to $z = f(x, y)$ at (x_0, y_0) is

$$f(x, y) \approx f(x_0, y_0) + \nabla f(x_0, y_0) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x - x_0 & y - y_0 \end{bmatrix} \nabla^2 f(x_0, y_0) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

for (x, y) close to (x_0, y_0).

More generally, $f : \mathbb{R}^k \to \mathbb{R}$ has continuous second partial derivatives such that $\nabla f(x^*) = 0$, Hessian matrix at x^* provides the following information:

1. If $\nabla^2 f(x^*) > 0$, then any local minimizer x^* is a strict local minimizer.
2. If $\nabla^2 f(x^*) < 0$, then local maximizer x^* is a strict local maximizer.
3. If $\nabla^2 f(x^*)$ is indefinite (i.e. neither positive semi-definite nor negative semi-definite), then the local minimizer x^* is a **saddle point**.
4. If those cases not listed above, then the test is inconclusive.

Note f has four second partial derivatives-four measures of concavity (in $\nabla^2 f$). In fact, eigenvalues of the Hessian $\nabla^2 f$ measures concavity in the corresponding eigendirections (take for example, $v^T H v = \lambda v^T v = \lambda$ for an eigenvector v).

In view of the classical results:

- A square (symmetric) matrix is positive definite if and only if all its eigenvalues are positive.
- A square (symmetric) matrix is negative definite if and only if all its eigenvalues are negative.
- A square (symmetric) matrix is indefinite (i.e., neither p.s.d. nor n.s.d.) if and only if it has both positive and negative eigenvalues.

Theorem: Suppose the function $f(x, y)$ has continuous second partial derivatives. Let (x_0, y_0) be a stationary point of f, and let λ_1 and λ_2 be the eigenvalues of $\nabla^2 f(x_0, y_0)$.

a. If $\lambda_1 < 0$ and $\lambda_2 < 0$, then (x_0, y_0) is a strict local maximum.
b. If $\lambda_1 < 0$ and $\lambda_2 > 0$ or if $\lambda_1 > 0$ and $\lambda_2 < 0$, then (x_0, y_0) is a **saddle point**.
c. If $\lambda_1 > 0$ and $\lambda_2 > 0$, then (x_0, y_0) is a strict local minimum.

d. If either $\lambda_1 = 0$ or $\lambda_2 = 0$ (or both), then no conclusion can be drawn without further information.

For a function f of three or more variables, there is a generalization of the rule above. In this context, instead of examining the determinant of the Hessian matrix, one must look at the eigenvalues of the Hessian matrix at the stationary point. The following test can be applied at any stationary point x^*:

- If the Hessian is positive definite (equivalently, has all eigenvalues positive) at x^*, then f attains a strict local minimum at x^*.
- If the Hessian is negative definite (equivalently, has all eigenvalues negative) at x^*, then f attains a strict local maximum at x^*.
- If the Hessian has both positive and negative eigenvalues (i.e., indefinite matrix) then x^* is a **saddle point** for f (and in fact this is true even if the Hessian is degenerate).
- In those cases not listed above (e.g., p.s.d but not p.d, n.s.d. but not n.d., or all entries are zeros), the test is inconclusive (i.e., cannot tell if local max/min or saddle point)

Note: in the definition of a saddle point, some authors would require eigenvalues to be non-zero.

Back   **Updated:** 2023-09-07   Statistical Simulation, Wei Li