MAT 653: Statistical Simulation

Instructor: Dr. Wei Li Scribe: Meng He

Sep 22th , 2021

Basics of vector calculus

Gradient

For $f : \mathbb{R}^n \to \mathbb{R}$, its gradient is $n \times 1$ matrix.

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

Hessian matrix

The Hessian matrix of f if square $n \times n$ matrix, is defined

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

Example: (1) $f(x) = a^T x : \mathbb{R}^n \to \mathbb{R}$, $f(x)$ 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,

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

Proof:

For the linear least square problem$\min_x f(x) = \min_x \|Ax - b\|^2$,

$$\begin{aligned} f(x) &= \|Ax - b\|^2 \\ &= \langle Ax - b, Ax - b \rangle \\ &= (Ax - b)^T (Ax - b) \\ &= x^T A^T A x - 2b^T A x + b^T b \end{aligned}$$

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

1.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 $x \in S$

2.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 $x \in N(x^*) \cap S$

3.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 $x \in N(x^*) \cap S$

4.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 $\alpha \in [0,1]$

5.Convex functions

A function f is convex on S if and only if it's convex at every point of S.

(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 $f:\mathbb{R} \to \mathbb{R}$, $\frac{\partial^2 f(x)}{\partial x^2}\Big|_{x=v} \geq 0$).

$\frac{\partial f(v)}{\partial x}$: slope at v = rate of change of f at v.

$\frac{\partial^2 f(v)}{\partial x^2}$ = rate of change of the slope of f at v.

(2) A "differentialbe" function f is convex at v if and only if for all w, $f(w) \geq f(v) + \nabla f(v)(w-v)$ (when w is neighborhood of v).

(3) 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.First order condition

FOC (first order condition): $\nabla f(x) \overset{set}{=} 0$

7.Saddle point

A saddle point is a stationary point at which the curvature of the function changes from negative to positive or vice versa.

9.Stationary point

A stationary point v of f satisfies the FOC $\nabla f(x)=0$, $f : \mathbb{R}^n \to \mathbb{R}$. A stationary point could be minimizer, maximizer or saddle point.