OTUS

ОНЛАЙН-ОБРАЗОВАНИЕ

Матричные производные

Матричные производные
Дифференциальные уравнения в матрицах

Лоскутов Антон Алексеевич

Меня хорошо слышно && видно?

Напишите в чат, если есть проблемы!

Ставьте + если все хорошо

Знакомство

- Заканчиваю механико-математический факультет МГУ им. Ломоносова
- Учился в Техносфере от Mail.Ru Group
- Являюсь ментором в Техносфере
- Работаю программистом-исследователем в Mail.Ru Group
- Веду лекции открытого курса mlcourse.ai

Правила вебинара

Активно участвуем

Задаем вопросы в чат

Off-topic обсуждаем в Slack

Вопросы вижу в чате, могу ответить не сразу

Градиент

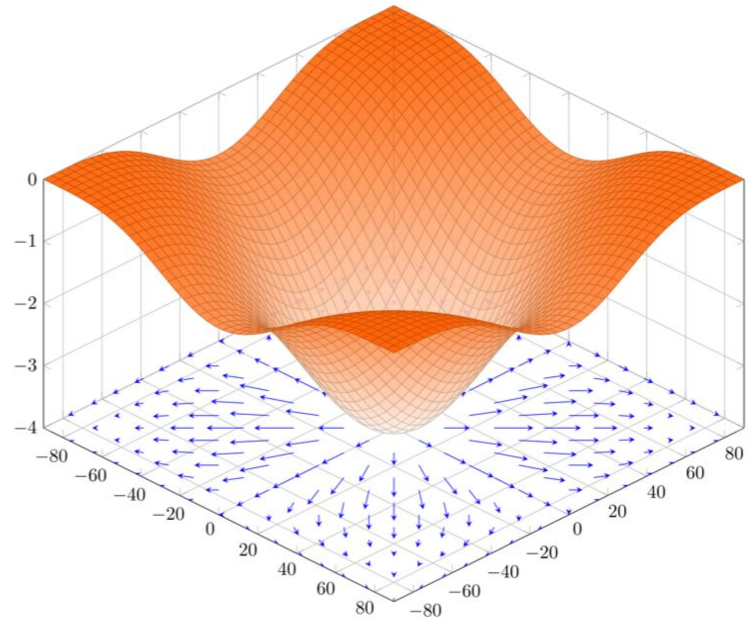
Рассмотрим скалярную функцию трёх переменных: $f(x_1, x_2, x_3)$

Градиент - это ***вектор***:

$\nabla f = \frac{\partial f}{\partial x_1}\hat{x}_1 + \frac{\partial f}{\partial x_2}\hat{x}_2 + \frac{\partial f}{\partial x_3}\hat{x}_3$,где $\hat{x}_1, \hat{x}_2, \hat{x}_3$ - единичные векторы

Геометрическая интерпретация:

- Направление градиента - направление наибольшего возрастания
- Длина градиента - “скорость” этого возрастания



Вычисление градиента

Дана функция

$f(x, y) = x^2 y$

$$\nabla f(x, y) = (2xy, x^2)$$

Найти её градиент в точке (3, 2)

$\nabla f(3, 2)$.

$$\nabla f(3, 2) = (12, 9)$$

$$df = 2xy\,dx + x^2\,dy$$

Гессиан

Рассмотрим функцию с векторным аргументом и скалярным значением:

$$f: \mathbb{R}^n \to \mathbb{R}$$

Если все вторые частные производные существуют, то Гессианом называется матрица:

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

$$\mathbf{H}_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Для дважды непрерывно дифференцируемых функций гессиан является симметричной матрицей.

Якобиан

Рассмотрим функцию с векторным аргументом и векторным значением:

$$\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$$

Якобианом называется матрица:

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

$$\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}$$

Типы производных

Вход/Выход	Скаляр	Вектор	Матрица
Скаляр	$\frac{dy}{dx}$	$\frac{d\boldsymbol{y}}{dx}$	$\frac{d\boldsymbol{Y}}{dx}$
Вектор	$\frac{dy}{d\boldsymbol{x}}$	$\frac{d\boldsymbol{y}}{d\boldsymbol{x}}$	
Матрица	$\frac{dy}{d\boldsymbol{X}}$		

Правила дифференцирования

1. $$\frac{\partial f(\mathbf{w})g(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} g(\mathbf{w}) + f(\mathbf{w}) \frac{\partial g(\mathbf{w})}{\partial \mathbf{w}}$$

2. $$\frac{\partial f(\mathbf{w})/g(\mathbf{w})}{\partial \mathbf{w}} = \frac{\left[\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} g(\mathbf{w}) - f(\mathbf{w}) \frac{\partial g(\mathbf{w})}{\partial \mathbf{w}}\right]}{g^2(\mathbf{w})}$$

3. $$\frac{\partial f(g(\mathbf{w}))}{\partial \mathbf{w}} = f'(g(\mathbf{w})) \frac{\partial g(\mathbf{w})}{\partial \mathbf{w}}$$

Дифференцирование сложной функции

OTUS

$$g(w) = w_1^2 + 3w_2$$

$$f(g) = g^3$$

$$\frac{df}{dw} = 3(w_1^2 + 3w_2)(2w_1 + 3)$$

Матричный градиент

Рассмотрим скалярную функцию *g* матрицы n×m, **W** = {w_{ij}}.
Матричный градиент по **W** - это матрица, состоящая из частных производных *g*(**W**)

$$\frac{\partial g}{\partial \mathbf{W}} = \begin{pmatrix} \frac{\partial g}{\partial w_{11}} & \cdots & \frac{\partial g}{\partial w_{1n}} \\ \vdots & & \vdots \\ \frac{\partial g}{\partial w_{m1}} & \cdots & \frac{\partial g}{\partial w_{mn}} \end{pmatrix}$$

Вычисление градиента по матрице

Пусть y - скалярная функция матрицы **X**, такая что

$$y = \mathrm{tr}(\mathbf{X}) = \sum_{i=1}^{n} x_{ii}\,.$$

Найти производную функции y по матрице **X**

$$\frac{d\,tr(X)}{dX} = E$$

Часто встречающиеся производные

$$\text{(C1)} \qquad \frac{d\mathbf{a}}{dx} = \mathbf{0} \quad \text{(column matrix)}$$

$$\text{(C2)} \qquad \frac{da}{d\mathbf{x}} = \mathbf{0}^\top \quad \text{(row matrix)}$$

$$\text{(C3)} \qquad \frac{da}{d\mathbf{X}} = \mathbf{0}^\top \quad \text{(matrix)}$$

$$\text{(C4)} \qquad \frac{d\mathbf{a}}{d\mathbf{x}} = \mathbf{0} \quad \text{(matrix)}$$

$$\text{(C5)} \qquad \frac{d\mathbf{x}}{d\mathbf{x}} = \mathbf{I}$$

$$\text{(C6)} \qquad \frac{d\,\mathbf{a}^\top\mathbf{x}}{d\mathbf{x}} = \frac{d\,\mathbf{x}^\top\mathbf{a}}{d\mathbf{x}} = \mathbf{a}^\top$$

$$\text{(C7)} \qquad \frac{d\mathbf{x}^\top\mathbf{x}}{d\mathbf{x}} = 2\,\mathbf{x}^\top$$

$$\text{(C8)} \qquad \frac{d(\mathbf{x}^\top\mathbf{a})^2}{d\mathbf{x}} = 2\,\mathbf{x}^\top\mathbf{a}\,\mathbf{a}^\top$$

$$\text{(C9)} \qquad \frac{d\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{A}$$

$$\text{(C10)} \qquad \frac{d\,\mathbf{x}^\top\mathbf{A}}{d\mathbf{x}} = \mathbf{A}^\top$$

$$\text{(C11)} \qquad \frac{d\,\mathbf{x}^\top\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)$$

OTUS

$$E = \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w} - y_i)^2 = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|$$

$$\begin{aligned} E = \|\mathbf{X}\mathbf{w} - \mathbf{y}\| &= (\mathbf{X}\mathbf{w} - \mathbf{y})^T(\mathbf{X}\mathbf{w} - \mathbf{y}) \\ &= (\mathbf{w}^T\mathbf{X}^T - \mathbf{y}^T)(\mathbf{X}\mathbf{w} - \mathbf{y}) \\ &= \mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - \mathbf{w}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\mathbf{w} + \mathbf{y}^T\mathbf{y} \end{aligned}$$

Слагаемое	Формула производной
$\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w}$	$\frac{d\,\mathbf{x}^\top\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)$
$\mathbf{w}^T\mathbf{X}^T\mathbf{y}$	$\frac{d\,\mathbf{a}^\top\mathbf{x}}{d\mathbf{x}} = \frac{d\,\mathbf{x}^\top\mathbf{a}}{d\mathbf{x}} = \mathbf{a}^\top$
$\mathbf{y}^T\mathbf{X}\mathbf{w}$	$\frac{d\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{A}$
$\mathbf{y}^T\mathbf{y}$	$\frac{da}{d\mathbf{x}} = \mathbf{0}^\top \quad \text{(row matrix)}$

Линейная регрессия

$$E = \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w} - y_i)^2 = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|$$

$$\begin{aligned} E = \|\mathbf{X}\mathbf{w} - \mathbf{y}\| &= (\mathbf{X}\mathbf{w} - \mathbf{y})^T(\mathbf{X}\mathbf{w} - \mathbf{y}) \\ &= (\mathbf{w}^T\mathbf{X}^T - \mathbf{y}^T)(\mathbf{X}\mathbf{w} - \mathbf{y}) \\ &= \mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - \mathbf{w}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\mathbf{w} + \mathbf{y}^T\mathbf{y} \end{aligned}$$

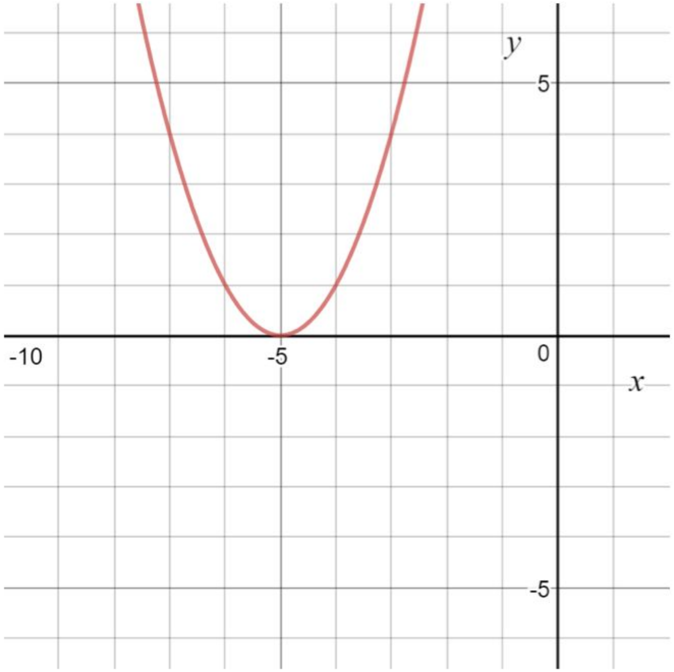
$$\begin{aligned} \frac{dE}{dw} &= \mathbf{w}^T(\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\mathbf{X}) - (\mathbf{X}^T y)^T - \mathbf{y}^T\mathbf{X} + \mathbf{0} \\ &= 2\mathbf{w}^T\mathbf{X}^T\mathbf{X} - 2\mathbf{y}^T\mathbf{X} \end{aligned}$$

$$\begin{aligned} 2\mathbf{w}^T\mathbf{X}^T\mathbf{X} - 2\mathbf{y}^T\mathbf{X} &= 0 \\ \mathbf{w}^T\mathbf{X}^T\mathbf{X} &= \mathbf{y}^T\mathbf{X} \end{aligned}$$

$$\begin{aligned} \mathbf{X}^T\mathbf{X}\mathbf{w} &= \mathbf{X}^T\mathbf{y} \\ \mathbf{w} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \end{aligned}$$

Градиентный спуск

$y=(x+5)^2$



Что такое градиент в случае функции одной переменной?

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma_n \nabla F(\mathbf{x}_n),$$

Начальная точка: $x_0 = 3$

Learning rate = 0.01

$$\frac{dy}{dx} = \frac{d}{dx}(x+5)^2 = 2 * (x+5)$$

$$X_1 = X_0 - (learning\ rate) * (\frac{dy}{dx})$$

$$X_1 = 3 - (0.01) * \big(2 * (3+5)\big) = 2.84$$

Стационарные точки

Определение 2.1 (Локальные экстремумы)**.** Точка $x \in X$ называется *точкой локального минимума* функции f, если существует шар некоторого радиуса $r > 0$, с центром в этой точке: $W = \{z \in U : \|z - x\| < r\}$, и выполнено:

$$f(x) \leq f(z) \quad \text{для любого } z \in W \cap X.$$

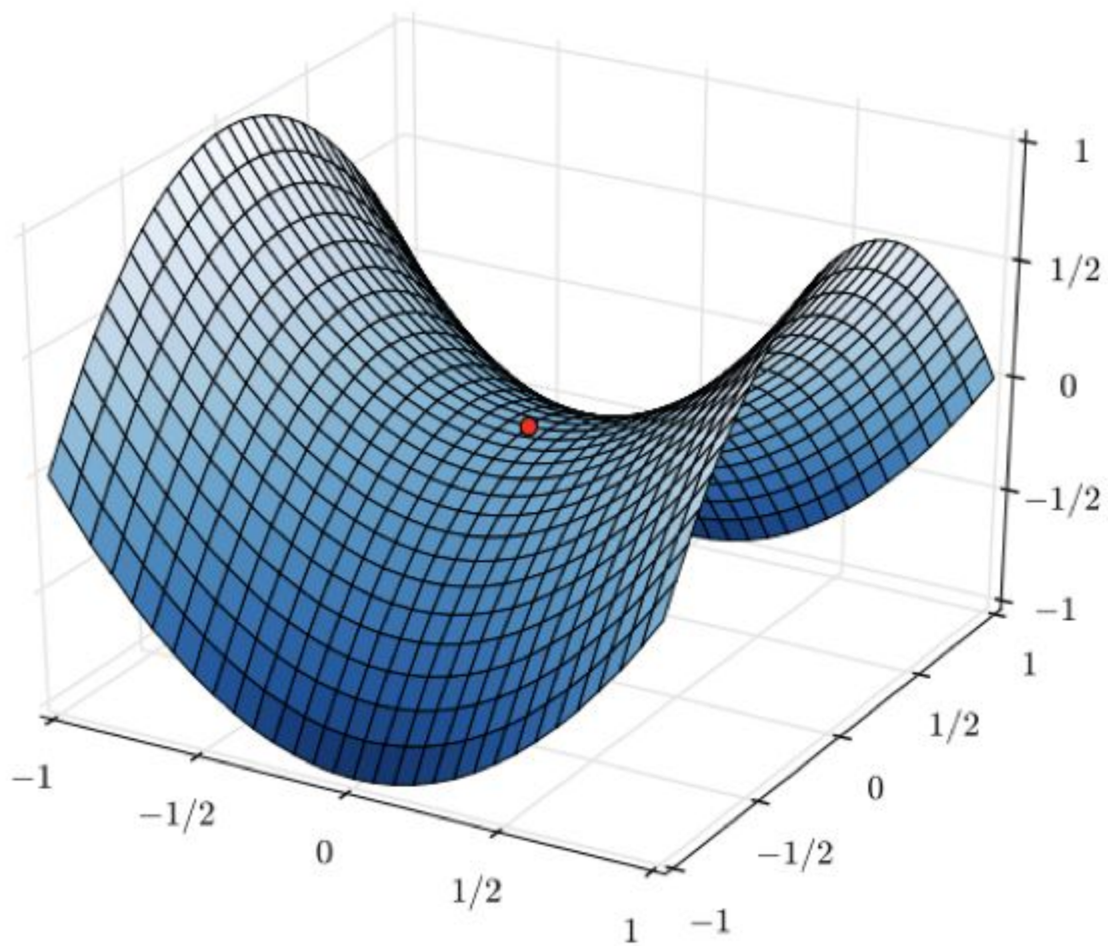
Утверждение 2.1 (условие оптимальности первого порядка)**.** *Пусть для функции $f : X \to \mathbb{R}$ точка x является точкой локального экстремума.*

Тогда если функция непрерывно-дифференцируема в окрестности этой точки, то её производная в этой точке равна нулю:

$$df(x) = 0.$$

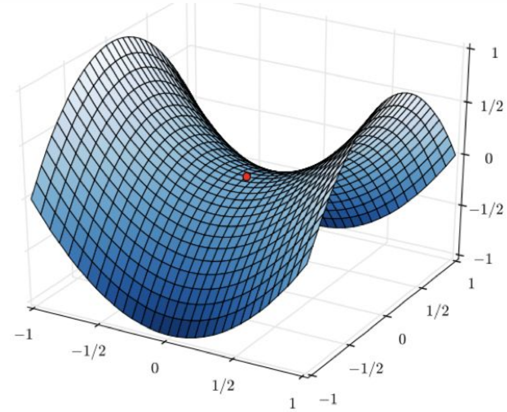
Определение 2.2. Точка x называется *стационарной точкой*, если производная в ней обращается в ноль: $df(x) = 0$. Стационарная точка, которая не является ни локальным минимумом, ни локальным максимумом, называется *седловой точкой*.

Стационарные точки



Стационарные точки

Если f дважды дифференцируемая в окрестности точки x, и вторая производная положительна во всех точках этой окрестности, то x - строгий минимум.
Если отрицательна - строгий максимум.
Если существуют точки h_1 и h_2, такие что в h_1 она положительна, а в h_2 - отрицательна, то x - седловая точка



Симметричная матрица M nxn называется положительно определённой, если $x^T M x > 0$ для любого ненулевого n-мерного вектора x (полуопределённой - если $x^T M x >= 0$) . Если $x^T M x > 0$ всегда меньше нуля - то M называется отрицательно определённой. Если $x^T M x$ принимает как положительные, так и отрицательные значения - то M называется неопределённой.

n=2: $\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + cy^2$

($a>0$ & $ac-b^2$) > 0 <=> M положительно определена

Если для функции n переменных, в точке x_0 её градиент равен нулю, а её Гессиан положительно полуопределён - то в точке x_0 достигается минимум.

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Есть вопросы или замечания?

Напишите в чат свои вопросы и замечания!

Ставьте + если все понятно

Антон Лоскутов

Slack: @LoskutovAnton

Пройдите опрос

Помогите нам стать лучше!
https://otus.ru/polls/8013/

Спасибо за внимание!