Производные

→ Зачем? Постоянно ищем минимумы разных ф. потерь.

Объекты:

- скаляр
- вектор
- матрица

- то, по чему дифф.
- то, что дифф.
- что получается

Никакого нового матана. Просто берем частные производные. Иногда их будет удаваться записать в красивой матричной форме.

откуда \ куда	скаляр	вектор	матрица
скаляр	$f'(x)dx$	$\nabla f dx$	—
вектор	$\nabla^T f\, dx$	$J\, dx$	—
матрица	$tr(\nabla f^T dX)$	—	—

Договорённости:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$ – вектор столбец

$$X = \begin{pmatrix} x_{11}, \dots, x_{1m} \\ \vdots \quad \ddots \quad \vdots \\ x_{n1}, \dots, x_{nm} \end{pmatrix}$$ – матрица

- Производная столбца – столбец
- Производная по столбцу – столбец.

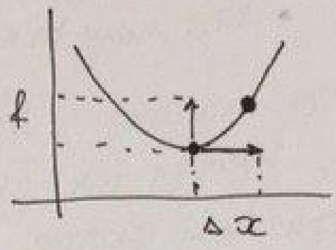
① $f: \mathbb{R} \to \mathbb{R}$

$f(x) = x^2$

$f'(x) = 2x$

$df(x) = 2x \cdot dx$

$$\overbrace{df(x)}^{\text{приращение ф.}} = \underbrace{f'(x)dx}_{\text{представляем его в линейном виде}}$$



Производная — это не просто ф., а какое-то линейное преобраз.

У нас есть

① Таблица производных

② Правила $f \cdot g$ $f+g$ $f(g)$

Хотим это обобщить

② $f: \mathbb{R}^n \to \mathbb{R}$

$f(x) = f(x_1, \ldots, x_n)$

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \underset{\text{градиент}}{\nabla f}$$

$$\left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right) \cdot \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}$$

$$df = \frac{\partial f}{\partial x_1} dx_1 + \ldots + \frac{\partial f}{\partial x_n} dx_n = (\nabla f)^T dx$$

Упражнение

$$\underset{1\times n\ n\times 1}{f(x) = a^T x} \quad \text{скалярное произведение} = (a_1, \ldots, a_n) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \underbrace{a_1}_{\frac{\partial f}{\partial x_1}} x_1 + \ldots + \underbrace{a_n}_{\frac{\partial f}{\partial x_n}} x_n$$

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\nabla f = \frac{\partial f}{\partial x} = a$$

$$df = a^T dx$$

~~[illegible]~~

Упражнение

$$f(x) = \underset{1\times n}{x^T}\ \underset{n\times n}{A}\ \underset{n\times 1}{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j = \dots$$

$$(\dots)\cdot\underbrace{\begin{pmatrix} \vdots & \vdots \\ \vdots & \vdots \end{pmatrix}}_{\text{не обязательно симметричная}}\cdot\begin{pmatrix} \vdots \end{pmatrix}$$

Делаем что-то неестественное. Всё было в компактном матричном виде, но мы решили это испортить...

$$df = (d\, x^T A x) = dx^T \cdot Ax + x^T d(Ax) =$$

$$= \underbrace{\underset{1\times n}{(dx)^T}\ \underset{n\times n}{A}\ \underset{n\times 1}{x}}_{1\times 1 \text{ — скаляр}} + \cancel{x^T dA\, x} + x^T A\, dx = x^T A^T dx + x^T A dx = {}^*$$

$$(dx^T A x)^T = x^T A^T dx$$

$$\boxed{(AB)^T = B^T A}$$

Фокусы со скалярами:

$S_{1\times 1}$ $\quad S^T = S \quad tr(S) = S$

⊕ скаляры можно за скобку выносить

$$^* = \underbrace{\underset{1\times n}{x^T}\ \underset{n\times n}{(A^T + A)}}_{\nabla f^T}\ \underset{n\times 1}{dx}$$

$$\nabla f = \underset{n\times n}{(A + A^T)}\ \underset{n\times 1}{x} \text{ — столбец}$$

df — скаляр

Свойства:

① $A = \begin{pmatrix} 3 & 1 \\ 2 & 7 \end{pmatrix}$ $\quad dA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

② $dX^T = (dX)^T$

③ $d(\underset{n\times m}{X}\underset{m\times k}{Y}) = dX \cdot Y + X \cdot dY$

$\neq$

$Y \cdot dX$

④ $d(\alpha X + \beta Y) = \alpha dX + \beta dY$

Док-во: просто расписываем на бумажке и видим это!

Без этих волшебных св-в и правил нам надо было бы выписывать суммы, что нам делать лениво!!

③ $f: \mathbb{R}^{n\times k} \longrightarrow \mathbb{R}$

$$df = f'_{x_{11}} dx_{11} + \ldots + f'_{x_{nk}} dx_{nk} = tr(\nabla f^T dX)$$

$$\nabla f = \begin{pmatrix} f'_{x_{11}}, \ldots, f'_{x_{1k}} \\ \vdots \\ f'_{x_{n1}}, \ldots, f'_{x_{nk}} \end{pmatrix}$$

Почему это так?

$$f(X) = A^T X$$

$$A_{2\times3} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \qquad \overset{A}{\begin{pmatrix} - \\ - \\ - \end{pmatrix}} \cdot \begin{pmatrix} | & | & | \end{pmatrix} X$$

$$X_{2\times3} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

$$tr(A^T dX) = tr \begin{pmatrix} a_{11}dx_{11} + a_{12}dx_{12} & & * \\ & a_{12}dx_{12} + a_{22}dx_{22} & \\ * & & a_{13}dx_{13} + a_{23}dx_{23} \end{pmatrix} =$$

$$= a_{11}dx_{11} + \ldots + a_{23}dx_{23}$$

Упражнение

$X_{n\times n}$ $A_{n\times n}$ $a_{n\times 1}$

$$\underset{n\times n}{f(X)} = \underset{1\times 1}{a^T X A X a}$$

$$\underset{n\times n}{\nabla F} \qquad \underset{1\times 1}{dF}$$

$$df = tr(\underbrace{\text{нечто}}_{\text{производные}} dX)$$

$$df = d(a^T X) AXa + a^T X\, d(AXa) =$$

$$= \underbrace{a^T dX AXa}_{1\times1} + \underbrace{a^T XA dXa}_{1\times1} =$$

$$= tr(\underbrace{a^T dX}AXa) + tr(\underbrace{a^T XA dX}a) =$$

$tr(ABC) = tr(BCA) = tr(CAB)$
если позволяют размерности

$$= tr(AXaa^T dX + aa^T XA dX) =$$

$$= tr((\underbrace{AXaa^T}_{\substack{n\times n \times n\times 1 \times 1\times n \\ ok}} + aa^T XA) dX)$$

$$\nabla f = [A X a a^T + a a^T X A]^T = a a^T X^T A^T + A^T X a a^T$$

④ $f: \mathbb{R} \to \mathbb{R}^m$

$$x \quad \begin{pmatrix} f_1(x) \\ \cdots \\ f_m(x) \end{pmatrix} \quad \nabla f = \begin{pmatrix} \frac{\partial f_1}{\partial x} \\ \vdots \\ \frac{\partial f_m}{\partial x} \end{pmatrix} \quad df = \nabla f \, dx$$

⑤ $f: \mathbb{R}^n \to \mathbb{R}^m$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} \quad df = \begin{pmatrix} f'_{1x_1} dx_1 + \ldots + f'_{1x_n} dx_n \\ \cdots\cdots \\ f'_{mx_1} dx_1 + \ldots + f'_{mx_n} dx_n \end{pmatrix}$$

$$df = \underbrace{J}_{m \times n} \underbrace{dX}_{n \times 1}$$

Упражнение

$$f(x) = \underset{n\times1}{x} \cdot \underset{1\times n}{x^T} \cdot \underset{n\times1}{x}$$

$$x^T x = \sum x_i^2$$

$$x x^T = \begin{pmatrix} x_1^2 & x_1 x_2 & \ldots \\ x_2 x & x_2^2 & \ldots \\ \vdots & \vdots & \end{pmatrix}$$

$$(\;)_{n\times1} \to (\;)_{n\times1}$$

$$df = d(x x^T x) = d(x x^T) x + x x^T dx =$$

$$= dx \underset{1\times1}{\boxed{x^T x}} + x \underset{1\times1}{\boxed{dx^T x}} + x x^T dx =$$

выносим

транспонируем

$$= \cancel{\underset{1\times1}{x^T x}\, \underset{n\times1}{dx}} + x x^T dx + x x^T dx =$$

$$\underset{1\times1}{x^T x} \cdot I_{n\times n} \cdot \underset{n\times1}{dx}$$

$$= \underset{1\times1}{x^T x} \underset{n\times n}{I} \underset{n\times1}{dx} + 2 \underbrace{x x^T}_{n\times n} \underbrace{dx}_{n\times1} = \overbrace{(x^T x \cdot I + 2 x x^T)}^{J} dx$$

⑥ Другие случаи

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

$f(X)$ X-матрица f-вектор.

$\left(\quad\right)_{n\times k}$ $\left(\,\right)_{m\times 1}$ $\frac{\partial f}{\partial X}$ – трёхмерная структура!

Тут обычно ограничиваются просто записью частных производных => пока не пробиваем размерность 2, мы можем жить в мире матриц и размахивать ими во имя удобства!

Упражнения

Упражнение

$$X = \begin{pmatrix} x_{11}, \ldots, x_{1k} \\ \vdots \quad \ddots \quad \vdots \\ x_{n1}, \ldots, x_{nk} \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

$y = X\beta + \varepsilon$

$\sum (y_i - x_i^T\beta)^2 \to \min\limits_{\beta}$

$$(y - X\beta) = \begin{pmatrix} y_1 - x_1^T\beta \\ \cdots \\ y_n - x_n^T\beta \end{pmatrix}$$

$L(\beta) = (y - X\beta)^T (y - X\beta) \to \min\limits_{\beta}$

$dL \quad 1\times 1$ $\quad L: \mathbb{R}^k \to \mathbb{R}$

$\nabla L \quad k\times 1$

$$dL = d(y-X\beta)^T(y-X\beta) + (y-X\beta)^T d(y-X\beta) =$$

$$= \underbrace{\underbrace{d(-X\beta)^T}_{-d\beta^T X^T} \underbrace{(y-X\beta)}_{n\times 1}}_{\text{скаляр}} + (y-X\beta)^T d(-X\beta) =$$

$1\times k$ $k\times n$

$$= -(y-X\beta)^T X\, d\beta - (y-X\beta)^T X\, d\beta =$$

$$= \underbrace{-2(y-X\beta)^T X}_{\nabla^T L} d\beta$$

$$\nabla L = -2X^T(y-X\beta)$$

$$\boxed{\beta_t = \beta_{t-1} - \gamma\cdot\left(-2X^T(y-X\beta_{t-1})\right)}$$ Шаг градиентного спуска.

$-2X^T(y-X\beta) = 0$ ← система уравнений

$$X^T y = X^T X\beta$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty$$

$$f(\beta) = -2X^T(y-X\beta)$$

$$f: (\ )_{k\times 1} \longrightarrow (\ )_{k\times 1}$$

$$df = -2\,d\left[X^T(y-X\beta)\right] = -2X^T d(y-X\beta) = \underbrace{2X^TX}_{H}\,d\beta$$

Если X^TX положительно определена => мы в точке минимума. — Критерий Сильвестра

$$\forall z \neq 0 \quad z^TX^TXz > 0$$

$z^TX^TXz = \|Xz\|^2 > 0\ \forall z\neq 0$ => всё хорошо.

Упражнение

dX^{-1} $\quad X_{n\times n}$

Подсказка: $dI = 0$

$$X^{-1}\cdot X = I$$

$$d(X^{-1}X) = dI$$

$$dX^{-1}X + X^{-1}dX = 0$$

$$\boxed{dX^{-1} = -X^{-1}dX\,X^{-1}}$$

Упражнение

$$f(x) = \ln(\underbrace{x^TAx}_{y})$$

$$df = \frac{1}{y}dy = \frac{1}{x^TAx}\cdot\overbrace{x^T(A^T+A)dx}^{d(x^TAx)}$$

Упражнение

$$f(X) = X$$

$$\frac{\partial X}{\partial x_{ij}} = \begin{array}{cc} & j \\ i & \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix} \end{array} = E_{ij}$$

$$\frac{\partial\, \mathrm{tr}\, X}{X} = I \qquad \frac{\partial x}{\partial x^T} = I$$

Упражнение

$$f(X) = \det X \qquad \det X = \sum_{j=1}^{n} x_{ij}(-1)^{i+j}M_{ij}$$

$$f: (\ )_{n\times n} \to \cdot_{1\times 1} \qquad \frac{\partial \det X}{\partial x_{ij}} = (-1)^{i+j}\cdot M_{ij}$$

$$(\det X) = (-1)^{1+1} M_{11}\, dx_{11} + (-1)^{1+2} M_{12}\, dx_{12} + \ldots$$

$$= \sum_{ij} A_{ij}\, dx_{ij} = tr(A^T dx) = tr(\det X \cdot X^{-T} dX)$$

$$X^{-1} = \frac{A}{\det X} \Rightarrow \det X \cdot X^{-1} = A$$

Выходит, что $\nabla f = \det X \cdot (X^{-1})^T = \det X \cdot X^{-T}$