Матрично-векторное дифференцирование.

$f(x) = x^T A x, \quad x \in R^n, \; A \in R^{n \times n}.$

$\nabla f(x) - ?$

$$f(x) = \sum_{i,j} a_{ij} x_i x_j.$$

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i \neq k} a_{ik} x_i + \sum_{j \neq k} a_{kj} x_j + 2a_{kk} x_k$$

$$= \sum_{i \neq k} a_{ik} x_i + \sum_{j \neq k} a_{kj} x_j = (Ax)_k + (A^T x)_k$$

$$\Rightarrow \nabla f(x) = (A + A^T)\, x.$$

Метод неудобный.

Нужно разбивать матричное выражение на суммы, а потом снова собирать.

Дифф. исчисление

Таблица стандартных дифф.

$$d(c^T x) = c^T dx$$

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

$$d \det X = \det X \, \mathrm{tr}(X^{-1} dX), \quad \det X \neq 0$$

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

$$d X^{-1} = -X^{-1} dX \cdot X^{-1}$$

Правила дифф.

$$dA = 0$$

$$d(X + Y) = dX + dY$$

$$d(X \cdot Y) = dX \cdot Y + Y \cdot dX$$

$$d\left(\frac{X}{\varphi}\right) = \frac{dX \cdot \varphi - X d\varphi}{\varphi^2}, \quad \varphi \in R.$$

$$d(AXB) = A \, dX \, B.$$

Пример:

1) Лин. регрессия:

$\{x_i, y_i\}_{i=1}^{N}$, $x_i \in R^D$, $y_i \in R$

Решающее правило: $g(x) = w^T \circ x$.

$$F(w) = \frac{1}{N} \sum_{i=1}^{N} (y_i - w^T \circ X)^2 =$$

$$= \frac{1}{N} \| y_i - Xw \|_2^2 \to \min_{w \in R^D}$$

$$X \in R^{N \times D}$$

$$F(w) = \frac{1}{N} (y - Xw)^T (y - Xw) =$$

$$= \frac{1}{N} [\underbrace{y^T y}_{R} - \underbrace{y^T X w}_{R} - \underbrace{w^T X^T y}_{R} + \underbrace{w^T X^T X w}_{R}] =$$

$$= \frac{1}{N} [y^T y - 2w^T X^T y + w^T X^T X w]$$

$$dF = \frac{1}{N} [0 - 2dw^T X^T y + w^T (X^T X + (X^T X)^T) dw]$$

$$dF = \frac{1}{N} [-2dw^T X^T y + 2w^T X^T X dw] =$$

Каноническая форма дифференциала

$U \backslash V$	$\mathbb{R}$	$\mathbb{R}^p$	$\mathbb{R}^{p \times q}$
$\mathbb{R}$	$f'(x)dx$	вектор $\nabla f(x)dx$	матрица $\nabla f(X)dx$
$\mathbb{R}^m$	$\nabla f(x)^T dx$	якобиан $J(x)dx$	
$\mathbb{R}^{m \times n}$	$\langle \nabla f(X), dX \rangle$		

В нашем случае, мы из вектора получаем число.

$$\ominus\; dw^T \cdot \frac{1}{N} \cdot [\underbrace{-2X^Ty + 2X^TXw}_{\nabla F(w)}]$$

$$\nabla F(w) = \frac{2}{N} \cdot X^T(Xw - y) = 0$$

$$\Rightarrow w_{opt} = (X^T \cdot X)^{-1} \cdot X^T \cdot y.$$

Дифф. сложной функции.

$h(x) := f(g(x))$.

$h(x)\cdot[dx] = df(y)\cdot[dg] = df(g(x))\cdot[dg(x)\cdot[dx]]$

Пример:

$h(x) = \frac{1}{3}\|x\|^3 : \mathbb{R}^n \to \mathbb{R}$.

$\frac{1}{3}(x^T\cdot x)^{3/2}$

$h(x) = f(g(x))$.

$g(x) = x^T x \Rightarrow dg = 2x^T dx$

$f(y) = \frac{1}{3}y^{3/2} \Rightarrow df = \frac{1}{3}\cdot\frac{3}{2}y^{\frac{1}{2}}dy$.

$dh(x)[dx] = \cancel{\frac{1}{2}}(x^T x)^{1/2}\cancel{2}x^T dx = \|x\|\cdot x^T dx$.

$\nabla h(x) = \|x\|\cdot x$.

$$f(X) = tr(AX^{-1}B)$$

$$df = tr(A\,dX^{-1}B) = tr(A\cdot(-X^{-1}dX\cdot X^{-1})B)$$

$$= -tr(X^{-1}B\cdot AX^{-1}dX)$$

$$\Rightarrow \nabla f(X) = X^{-T}A^{T}B^{T}X^{-T}$$

Матричное дифференцирование Яндекс.

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

$f(x)$ дифф. в т. x_0, если:

$$f(x_0+h) = f(x_0) + [D_{x_0} f](h) + \bar{o}(\|h\|).$$

$D_{x_0} f$ – дифф. функции f: линейное отображение $x \to f$. $\boxed{h = \Delta x \to \Delta f}$.

$$f(x_0+h) - f(x_0) \approx [D_{x_0} f](h)$$

Дифф. зависит от точки x_0, в которой он берется.

Примеры.

① $f: \mathbb{R} \to \mathbb{R}$.

$$f(x_0+h) - f(x_0) \approx f'(x_0) \cdot h$$

$$[D_{x_0} f](h) = f'(x_0) \cdot h$$

(2) $f(x): \mathbb{R} \to \mathbb{R}^n$.

$$f(x_0 + h) - f(x_0) \approx \sum_i \frac{\partial f}{\partial x_i}\Big|_{x = x_0} \cdot h_i$$

$$[D_{x_0} f](h) = (\nabla_{x_0} f)^T \cdot h = \langle \nabla_{x_0} f, h \rangle$$

(3) $f(X): \mathbb{R} \to \mathbb{R}^{n \times m}$

$$f(X_0 + H) - f(X_0) = \sum_{i,j} \frac{\partial f}{\partial X_{i,j}}\Big|_{X = X_0} \cdot H_{ij}.$$

$$[D_{X_0} f](H) = tr\left(\left[\frac{\partial f}{\partial X_{ij}}\Big|_{X = X_0} \right]^T H \right)$$

Заметим, что:

$$[D_{x_0} f](x - x_0) = \langle \nabla_{x_0} f, x - x_0 \rangle$$