Матрично-векторное дифференцирование.

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

$\nabla f(x)$ - ?

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

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

$$= \sum_{i \neq k}{}' a_{ik} x_i + \sum_{j \neq k}{}'' a_{kj} x_j = (Ax)_k + (A^T x)_k$$

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

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:

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

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

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

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

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

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

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

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

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

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

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.

$$= d w^T \cdot \frac{1}{N} \cdot \Big[\underbrace{-2X^T y + 2X^T X w}_{\nabla F(w)} \Big]$$

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

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

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:

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

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

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

$$g(x) = x^Tx \Rightarrow dg = 2x^Tdx$$

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

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

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

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

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

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

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