Лекция 2.
Матричные производные

Галков Михаил

13 февраля 2017 г.

План лекции

Одномерный случай

$$\lim_{u \to 0} \frac{\phi(c+u) - \phi(c)}{u} = \phi'(c),$$

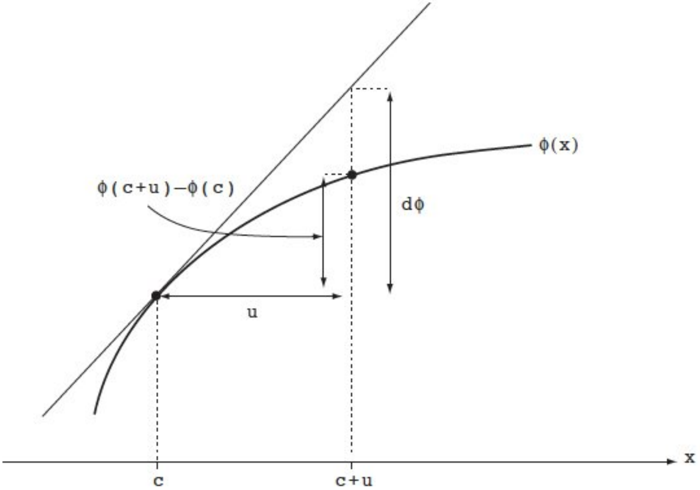
Рис.: Определение производной

$$\phi(c+u) = \phi(c) + u\phi'(c) + r_c(u),$$

Рис.: Эквивалентная формулировка, какое необходимо условие?

Что такое дифференциал и в чем его смысл?

Геометрическая интерпретация



$$\mathrm{d}\phi(c;u) = u\phi'(c)$$

Многомерный случай

$$S \in \mathbb{R}^n, f : S \to \mathbb{R}^m \tag{1}$$

Пусть c внутренняя точка $S, B(c; r)$ - открытый шар

$$u \in \mathbb{R}^n, ||u|| < r, c + u \in B(c; r) \tag{2}$$

Мы хотим, чтобы в многомерном случае все выглядело так же, только с матрицами.

$$f(c + u) = f(c) + A(c)u + r_c(u), \lim_{u \to 0} ||r_c(u)|| \to 0 \tag{3}$$

Структура матрицы A(c)

Рассмотрим вектор e_i заполненый нулями кроме единичной позиции i.

$$f(c + e_j t) = f(c) + tA(c)e_j + r_c(te_j) \tag{4}$$

Если мы разделим на t и устремим $t \to 0$, тогда получим:

$$a_{ij}(c) = \lim_{t \to 0} \frac{f_i(c + e_j t) - f_i(c)}{t} \tag{5}$$

Т.е. матрица A есть матрица частных производных

$$A = [\frac{df}{dx_1}, ..., \frac{df}{dx_n}] \tag{6}$$

Матрица A называется якобианом. A^T - градиент.

Куда указывает градиент?

Рассмотрим случай векторной функции $f : \mathbb{R}^N \to \mathbb{R}$

$$f(x+u) = f(x) + \nabla f(x) u \tag{7}$$

Возьмем u по направлению градиента, т.е. $u = \epsilon \nabla f(x)$
Тогда:

$$f(x+u) = f(x) + \epsilon ||\nabla f(x)||^2 \geq f(x) \tag{8}$$

Почему матрица A^T градиент?

Покажем это!

$$||f(c+u)||^2 = ||f(c) + Au||^2 \tag{9}$$

Возьмем $u = A^T f(c)$, т.е. по направлению градиента. Тогда:

$$||f(c) + AA^T f(c)||^2 = ||(I + AA^T)f(c)||^2 \geq ||f(c)||^2 \tag{10}$$

Chain rule

Пусть f - дифференцируема в точке c.
g - дифференцируема в точке $b = f(c)$,
$h = g \circ f$, тогда

$$dh(c; u) = dg(b; df(c; u)) \tag{11}$$

Аналогично для градиентов!

Gradient types

Type	Scalar	Vector	Matrix
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{Y}}{\partial x}$
Vector	$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\frac{\partial y}{\partial \mathbf{X}}$		

Примеры

Найдем градиент $\frac{da^T x}{dx}$

$$f(x+u) = a^T(x+u) = a^T x + a^T u = f(x) + a^T u \tag{12}$$

Значит, $\frac{da^T x}{dx} = a$
Найдем градиент $\frac{dAx}{dx}$

$$f(x+u) = A(x+u) = Ax + Au = f(x) + Au \tag{13}$$

Значит, $\frac{dAx}{dx} = A^T$
Пусть $Loss$ - функция потерь нашей нейросети (скаляр),
$z_2 = Wz_1 + b$
Проверьте, что все корректно по размерностям:

$$\frac{dLoss}{dz_1} = \frac{dLoss}{dz_2}\frac{dz_2}{dz_1} \tag{14}$$

Задачи

Найдите следующие градиенты:

$$\frac{dx^T x}{dx} = \tag{15}$$

$$\frac{de^{x^T x}}{dx} = \tag{16}$$

$$\frac{dx^T Ax}{dx} = \tag{17}$$

$$\frac{||y - X\beta||^2}{dx} = \tag{18}$$

$$\frac{d \log(N(\mu, \sigma^2))}{d\mu} = \tag{19}$$

Вопросы