冷知识：我没学过 multivariable Calculus.

以下为急救 from D2L 2.4.

2.4.2 Partial Derivatives

$$y = f(x_1, x_2, \dots, x_n)$$

$$\frac{\partial y}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}$$

2.4.3. Gradient

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

无歧义时记为 $\nabla f(x)$.

Partial Differentiation Rule

Let $x \in \mathbb{R}^n$.

$$\Rightarrow \forall A \in \mathbb{R}^{m \times n},\ \boxed{\nabla_x A x = A^T}$$

$$\forall A \in \mathbb{R}^{n \times m},\ \boxed{\nabla_x x^T A = A}$$

$$\forall A \in \mathbb{R}^{n \times n},\ \boxed{\nabla_x x^T A x = (A + A^T)x}$$

$$\nabla_x \|x\|^2 = \boxed{\nabla_x x^T x = 2x}$$

Chain Rule

若 $y = f(u_1, u_2, \dots, u_m)$

$\forall u_i,\ u_i = g(x_1, x_2, \dots, x_n)$

$$\Rightarrow \frac{\partial y}{\partial x_i} = \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x_i} + \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial x_i} + \dots + \frac{\partial y}{\partial u_m} \cdot \frac{\partial u_m}{\partial x_i}$$

$$= \sum_{j=1}^{m} \frac{\partial y}{\partial u_j}\frac{\partial u_j}{\partial x_i}$$

Probability

① Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \left(= \frac{P(AB)}{P(B)}\right)$$

② Marginalization:

$$P(B) = \sum_A P(A, B)$$

③ Expectation:

$$E[X] = \sum_x x P(X = x)$$

if $x \sim P$ (distribution)

$$E_{x \sim P}[f(x)] = \sum_x f(x) P(x)$$

④ To measure the bias between X and its expectation:

$$Var[X] = E[(X - E[X])^2]$$

$$= E[X^2] - E^2[X]$$

To measure the bias between $f(x)$ and its expectation when x is distributed by $f(x)$:

$$Var[f(x)] = E[(f(x) - E[f(x)])^2]$$