Matrix Calculus

Gradient

Def. Directional Derivative

The directional derivative of function $f : \mathbb{R}^n \to \mathbb{R}$ in the direction u is

$$\nabla_u f(x) = \lim_{h\to 0} \frac{f(x+hu) - f(x)}{h}$$

- When u is the i-th standard unit vector e_i, then $\nabla_u f(x) = f_i'(x) = \frac{\partial f(x)}{\partial x_i}$.
- For any n-dimensional vector u, the directional derivative of f in the direction of u can be represented as $\nabla_u f(x) = \sum_{i=1}^n f_i'(x) \cdot u_i$.
  - Proof.$\Rightarrow$ let $g(h) = f(x+hu)$
    $\nabla_u f(x) = g'(0) = \lim_{h\to 0} \frac{f(x+hu)-g(0)}{h}$
    $\because g'(h) = \sum_{i=1}^n f_i'(x) \frac{d}{dh}(x_i + hu_i) = \sum_{i=1}^n f_i'(x) u_i$
    let $h = 0 \therefore \nabla_u f(x) = \sum_{i=1}^n f_i'(x) u_i$

Def. Gradient

The gradient of f is a vector function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\nabla f(x) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} e_i$$

$$\Rightarrow \nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right]$$

- $\nabla_u f(x) = \nabla f(x) \cdot u = ||\nabla f(x)||\cos a$ Where u is a unit vector.
- When $u = \nabla f(x)$ such that $a = 0$, we have the maximum directional derivative of f.

Matrix Derivatives

- The derivative of $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ with respect to A is defined as:

$$\nabla f(A) = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \cdots & \frac{\partial f}{\partial A_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{m1}} & \cdots & \frac{\partial f}{\partial A_{mn}} \end{bmatrix}$$

trace

Def. $trA = \sum_{i=1}^n A_i i$

- $trABCD = trDABC = trCDAB = trBCDA$
- $trA = trA^T, tr(A+B) = trA + trB, tr(aA) = a \cdot trA$
- $\nabla_A trAB = B^T, \nabla_{A^T} f(A) = (\nabla_A f(A))^T$
- $\nabla_A trABA^TC = CAB + C^TAB^T, \nabla_A |A| = |A|(A^{-1})^T$
- Funky trace derivative $\nabla_{A^T} trABA^TC = B^TA^TC^T + BA^TC$

Jacobian Matrix

$$\mathsf{J} = \begin{bmatrix} \dfrac{\partial \mathsf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathsf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Hesse Matrix

$$G(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{x_0}$$

- $H(f) = J(\nabla f)$