Gradient: The Direction of Steepest Ascent

Feng Li
School of Computer Science and Technology
Shandong University (Qingdao), China
Email: fli@sdu.edu.cn

September 2019

1 Directional Derivative

Consider a differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Definition 1. *The directional derivative of f (at $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$) in the direction of $u \in \mathbb{R}^n$ is*

$$\nabla_u f(x) = \lim_{h \rightarrow 0} \frac{f(x + hu) - f(x)}{h} \tag{1}$$

Intuitively, $\nabla_u f(x)$ can be regarded as the rate at which f is increased as $x \rightarrow \theta$ in direction u. When u is a the i-th standard unit vector e_i [1] (where $i = 1, 2, \cdots, n$), we have $\nabla_u f(x) = f_i'(x)$, where $f_i'(x) = \frac{\partial f(x)}{\partial x_i}$ is the partial derivative of $f(x)$ with respect to x_i.

Theorem 1. *For any n-dimensional vector $u = (u_1, u_2, \cdots, u_n)$, the directional derivative of f in the direction of u can be represented as*

$$\nabla_u f(x) = \sum_{i=1}^{n} f_i'(x) u_i \tag{2}$$

Proof. Let $g(h) = f(x + hu)$. The derivative of g at 0 is

$$\begin{aligned} g'(0) &= \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x + hu) - f(x)}{h} \\ &= \nabla_u f(x) \end{aligned} \tag{3}$$

[1] e_i is a unit vector where the i-the element is 1 while the others are 0's

On the other hand, by the chain rule, we have

$$g'(h) = \sum_{i=1}^{n} f'_i(x) \frac{d}{dh}(x + hu_i) = \sum_{i=1}^{n} f'_i(x) u_i \tag{4}$$

Let $h = 0$, $g'(0) = \sum_{i=1}^{n} f'_i(x)u_i$, substituting which into Eq. (3), we get Eq. (2).
□

2 Gradient

Definition 2. *The gradient of f is a vector function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ defined by*

$$\nabla f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} e_i \tag{5}$$

where e_i is the i-th standard unit vector.

Based on **Definition** 2 and **Theorem** 1, $\nabla_u f(x)$ can be re-written as the dot product of $\nabla f(x)$ and u, i.e.

$$\nabla_u f(x) = \nabla f(x) \cdot u = \|\nabla f(x)\| \ \|u\| \ \cos a \tag{6}$$

where a is the angle between $\nabla f(x)$ and u. Without of loss of generality, we assume u is a unit vector. Then

$$\nabla_u f(x) = \|\nabla f(x)\| \cos a \tag{7}$$

Recall that $\nabla_u f(x)$ represent the rate at which $f(x)$ changes in the direction u, the question is: "in what direction, f changes at the highest rate?" Since $\cos a \in [-1, 1]$, when u is the direction of $\nabla f(x)$ such that $a = 0$, we have the maximum directional derivative of $f(x)$, which implies that $\nabla f(x)$ indicates the direction of steepest ascent of $f(x)$. On the other hand, $-\nabla f(x)$ is the direction of steepest descent of $f(x)$.