Taking Derivatives

Andisheh Amrollahi

November 2019

1 Notation

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a differentiable function mapping $x = (x_1, \dots, x_m)$ to $(f_1(x), \dots, f_n(x))$. We define the derivative of this function to be an $m \times n$ matrix where the i, j'th element $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$ is equal to $\frac{\partial f_j}{\partial x_i}$. We denote the derivative by Df. As a special case $f : \mathbb{R}^m \to \mathbb{R}$ will have as derivative a $m \times 1$ column vector which we also call the "gradient".

$$Df = \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \frac{\partial f}{\partial x_m} \end{bmatrix}_{m \times 1}$$

2 Basic building blocks

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be defined as $f(x) \triangleq \|x\|_2^2$. Then:

$$Df = \nabla f = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \dots \\ 2x_m \end{bmatrix}_{m \times 1} = 2x \tag{1}$$

Let $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$, $f : \mathbb{R}^m \to \mathbb{R}^n$, be defined as $f(x) = Ax + b$. Then:

$$Df = A^\top \tag{2}$$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as $f((x_1, x_2)) = x_1 \cdot x_2$. Then:

$$Df = \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} = x_2 \\ \frac{\partial f}{\partial x_2} = x_1 \end{bmatrix}_{2 \times 1} \tag{3}$$

Let $f : \mathbb{R}^m \to \mathbb{R}$ and $A \in \mathbb{R}^{n \times n}$ be defined as $f(x) = x^\top Ax$. Then:

$$Df = \nabla f = (A + A^\top)x \tag{4}$$

3 Differentiation rules

The first rule is that the derivative is just a linear operator. Let $f, g : \mathbb{R}^m \to \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then:

$$D(\alpha f + g) = \alpha Df + g$$

The second rule is the chain rule. Let $f : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^k$. Define $gof : \mathbb{R}^m \to \mathbb{R}^k$ as $(gof)(x) \triangleq g(f(x))$. Then:

$$D(gof)(x) = (Df)(x) \cdot (Dg)(f(x))$$

Where $(Df)(x)$ means the derivative of f evaluated at point x.

4 Examples

4.1 Derivative of a Gaussian

Let $\mu \in \mathbb{R}^m$ and $\Sigma \in \mathbb{R}^{m \times m}$ a symmetric invertible matrix. Define $f : \mathbb{R}^m \to \mathbb{R}$ as follows:

$$f(x) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{(x-\mu)^\top \Sigma^{-1}(x-\mu)}{2}}$$

We decompose f into three functions.

$$g_1 : \mathbb{R}^d \to \mathbb{R}^d, \;\; g(x) = x - \mu$$

$$g_2 : \mathbb{R}^d \to \mathbb{R}, \;\; g_2(y) = -\frac{1}{2} y^\top \Sigma^{-1} y$$

$$g_3 : \mathbb{R} \to \mathbb{R}, \;\; g_3(z) = \frac{1}{2\pi} e^z$$

Thus $f = g_3 o g_2 o g_1$. We know from (2) that $D(g_1) = I_{m \times m}$, by (4) and linearity of the derivative $D(g_2) = -\Sigma^{-1} y$ and finally $D(g_3) = \frac{1}{2\pi} e^z$. Hence by the chain rule the total derivative is equal to:

$$I_{m \times m} \cdot -\Sigma^{-1}(x - \mu) \cdot \frac{1}{2\pi} e^{-\frac{(x-\mu)^\top \Sigma^{-1}(x-\mu)}{2}}$$

4.2 Product rule

Let $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}$. Let us find the derivative of $f : \mathbb{R}^m \to \mathbb{R}$ defined as $f \triangleq f_1(x) \cdot f_2(x)$. We can decompose this function into the following two functions:

$$g_1 : \mathbb{R}^d \to \mathbb{R}^2, \;\; g_1(x) = (f_1(x), f_2(x))$$

$$g_2 : \mathbb{R}^2 \to \mathbb{R}, \;\; g_2(y_1, y_2) = y_1 \cdot y_2$$

Then by definition of the derivative:

$$D(g_1) = [\nabla f_1 | \nabla f_2]_{m \times 2}$$

And by (3):

$$D(g_2)(y_1, y_2) = \begin{bmatrix} y_2 \\ y_1 \end{bmatrix}$$

Hence by the chain rule $Df = \nabla f_1 \cdot f_2 + \nabla f_2 \cdot f_1$