Mathematics for Deep Learning

Zheguang Zhao
zheguang.zhao@gmail.com

Last updated: February 1, 2019

1 Single-variable real-valued function

Let $z \in \mathbb{R}$, and $f(z)$ be a scalar map from scalars to scalars, i.e. $\mathbb{R} \to \mathbb{R}$:

$$\frac{\mathrm{d}f(z)}{\mathrm{d}z} := \lim_{\delta_x \to 0} \frac{f(z+\delta_z) - f(z)}{\delta_z} \tag{1}$$

The computation graphs representation of $f(z)$ and its derivative $f'(z)$ are both one-to-one, which reflects that they are both scalar-to-scalar functions:

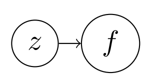

Figure 1: $f(z)$

Figure 2: Single-variable real-valued function.

2 Vector-valued functions

Let $t \in \mathbb{R}$, and $\mathbf{x} = \mathbf{f}(t) := (x_1(t), x_2(t), \cdots, x_3(t))$ be a map of scalars to vectors, i.e. $\mathbb{R} \to \mathbb{R}^3$

The derivative of this vector-valued function w.r.t. its scalar variable is simply the the derivatives of its components w.r.t. this scalar variable:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} := \left(\frac{\mathrm{d}x_1}{\mathrm{d}t}, \frac{\mathrm{d}x_2}{\mathrm{d}t}, \frac{\mathrm{d}x_3}{\mathrm{d}t}\right) \tag{2}$$

We need to use three nodes to denote the components of $\mathbf{x}$ in the computation graph:

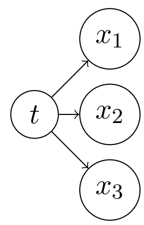

Figure 3: Vector-valued function

3 Multi-variable real-valued functions

Let $z = f(\mathbf{x}) = f(x_1, x_2, x_3)$ be a map from $\mathbb{R}^3 \to \mathbb{R}$.

The derivative of z w.r.t. $\mathbf{x}$ is also called the gradient of z, denoted as $\boldsymbol{\nabla}_{\mathbf{x}} z$:

$$\boldsymbol{\nabla}_{\mathbf{x}} z = \frac{\mathrm{d}z}{\mathrm{d}\mathbf{x}} := \left(\frac{\partial z}{\partial x_1}, \frac{\partial z}{\partial x_2}, \frac{\partial z}{\partial x_3} \right) \tag{3}$$

A multi-variable real-valued function can be represented as many-to-one graph:

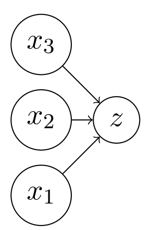

Figure 4: Multi-variable real-valued function

4 Function composition

Let $\mathbf{x} \in \mathbb{R}^3$, $t, z \in \mathbb{R}$, $\mathbf{x} = \mathbf{f}(t)$ be a vector-valued function, $z = g((x))$ be.a multi-variable real-valued function. Then we can compose a map $h = \mathbf{f} \circ g$ from $\mathbb{R} \to \mathbb{R}$ as $z = h(t) = (\mathbf{f} \circ g)(t) = g(\mathbf{f}(t))$.

The computation graph of h is just a concatenation of the subgraphs of a vector-valued function $\mathbf{f}$ (Figure 3) and g (Figure 4):

What is the derivative of z w.r.t. t? Because h is a function composition,

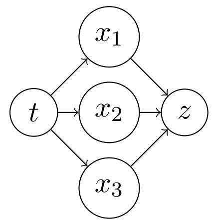

Figure 5: $z = h(t) = (\mathbf{f} \circ g)(t)$

we need to use chain rule.

$$\begin{aligned}
\frac{\mathrm{d}z}{\mathrm{d}t} &= \frac{\mathrm{d}h(t)}{\mathrm{d}t} && (4)\\
&= \frac{\mathrm{d}(\mathbf{f} \circ g)(t)}{\mathrm{d}t} && (5)\\
&= \frac{\mathrm{d}g(\mathbf{x})}{\mathrm{d}\mathbf{x}} \cdot \frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} && (6)\\
&= \frac{\mathrm{d}z}{\mathrm{d}\mathbf{x}} \cdot \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} && (7)\\
&= \left(\frac{\partial z}{\partial x_1}, \frac{\partial z}{\partial x_2}, \frac{\partial z}{\partial x_3}\right) \cdot \left(\frac{\mathrm{d}x_1}{\mathrm{d}t}, \frac{\mathrm{d}x_2}{\mathrm{d}t}, \frac{\mathrm{d}x_3}{\mathrm{d}t}\right) && (8)\\
&= \frac{\partial z}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}t} + \frac{\partial z}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}t} + \frac{\partial z}{\partial x_3}\frac{\mathrm{d}x_3}{\mathrm{d}t} && (9)
\end{aligned}$$

This shows that the derivative is also just the dot product of the two functions of the two subgraphs. This establishes the correspondence between the graph composition and the function composition.

5 License