Mathematics reminders for deep learning

Part 2: Differential Calculus

Georges Quénot

Multimedia Information Modeling and Retrieval Group

Laboratory of Informatics of Grenoble

Differential of a function
scalar input and scalar output

- $f:\mathbb{R}\rightarrow\mathbb{R}:x\rightarrow f(x)$ $\quad$ f is differentiable
- $y=f(x)$
- $f(x+h)=f(x)+f'(x)h+o(h)$ $\quad$ $\left(\lim_{h\rightarrow 0}\frac{o(h)}{h}=0\right)$
- $dy=f'(x)dx$ $\quad$ i.e: f is "locally linear"
- $\frac{dy}{dx}\equiv f'(x)$ $\quad$ (notation)
- $dy=\frac{dy}{dx}dx$ $\quad$ ("local scale factor")
- All values are scalar

Differential of a composed function scalar input and scalar output

- $f : \mathbb{R} \to \mathbb{R} : x \to f(x)$ — f is differentiable
- $y = f(x)$
- $g : \mathbb{R} \to \mathbb{R} : y \to g(y)$ — g is differentiable
- $z = g(y)$
- $(g \; o \; f)'(x) = (g' o \; f)(x).f'(x) = g'(y).f'(x)$
- $dy = \frac{dy}{dx} dx$ $\quad dz = \frac{dz}{dy} dy$
- $dz = \frac{dz}{dy} \cdot \frac{dy}{dx} dx$ $\quad \frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$

Differential of a function of a vector vector input and scalar output

- $f : \mathbb{R}^N \rightarrow \mathbb{R} : x \rightarrow f(x)$ $\quad$ f is differentiable
- $y = f(x)$ $\quad$ $x = (x_i)_{(1 \leq i \leq N)}$
- $f(x+h) = f(x) + \text{grad}\, f\,(x).h + o(\|h\|)$
- $dy = \text{grad}\, f\,(x).dx = \sum_{i=1}^{i=n} \frac{\partial f}{\partial x_i}(x).dx_i = \sum_{i=1}^{i=n} \frac{\partial y}{\partial x_i}.dx_i = \frac{\partial y}{\partial x}.dx$
- $\frac{\partial y}{\partial x} \equiv \frac{\partial f}{\partial x}(x) = \text{grad}\, f\,(x)$ $\quad$ $\frac{\partial y}{\partial x_i} \equiv \frac{\partial f}{\partial x_i}(x)$ $\quad$ (notations)
- y, dy and $f(x)$ are scalars;
- x, dx and h are “regular” (column) vectors;
- $\frac{\partial y}{\partial x}$ is a transpose (row) vector.

Differential of a vector function of a vector vector input and vector output

- $f : \mathbb{R}^N \rightarrow \mathbb{R}^P : x \rightarrow f(x)$ f is differentiable
- $y = f(x)$ $x = (x_i)_{(1 \leq i \leq N)}$ $y = (y_j)_{(1 \leq j \leq P)}$ $f = (f_j)_{(1 \leq j \leq P)}$
- $f(x) - f(x + h) = \operatorname{grad} f\,(x).\,h + o(\|h\|)$
- $dy = \operatorname{grad} f\,(x).\,dx = \frac{\partial f}{\partial x}(x).\,dx = \frac{\partial y}{\partial x}.\,dx$ (locally linear)
- $dy_j = \sum_{i=1}^{i=n} \frac{\partial f_j}{\partial x_i}(x).\,dx_i = \sum_{i=1}^{i=n} \frac{\partial y_j}{\partial x_i}.\,dx_i$

- x, dx, y, dy, $f(x)$ and h are all “regular” vectors;
- $\frac{\partial y}{\partial x}$ is a matrix (Jacobian of f: $J_{ij} = \left(\frac{\partial y}{\partial x}\right)_{ij} = \frac{\partial y_j}{\partial x_i} = \frac{\partial f_j}{\partial x_i}(x)$).

Differential of a composed function vector inputs and vector outputs

- $f : \mathbb{R}^N \rightarrow \mathbb{R}^P : x \rightarrow y = f(x)$ $\quad$ f is differentiable
- $g : \mathbb{R}^P \rightarrow \mathbb{R}^Q : y \rightarrow z = g(y)$ $\quad$ g is differentiable
- $x = (x_i)_{(1 \leq i \leq N)}$ $\quad$ $y = (y_j)_{(1 \leq j \leq P)}$ $\quad$ $z = (z_k)_{(1 \leq k \leq Q)}$
- $dz = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x} . dx$
- $\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}$ (matrix multiplication: non commutative!)
- x, dx, y, dy, z, dz, $f(x)$ and $g(y)$ are all regular vectors;
- $\frac{\partial y}{\partial x}$, $\frac{\partial z}{\partial y}$ and $\frac{\partial z}{\partial x}$ are all matrices (f, g and gof Jacobians).

Differential of a composed function vector inputs and scalar output

- $f : \mathbb{R}^N \to \mathbb{R}^P : x \to y = f(x)$ f is differentiable
- $g : \mathbb{R}^P \to \mathbb{R} : y \to z = g(y)$ g is differentiable
- $x = (x_i)_{(1 \leq i \leq N)}$ $y = (y_j)_{(1 \leq j \leq P)}$ $z \in \mathbb{R}$
- $\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}$ (left row vector × matrix mult. → row vector)
- z, dz and $g(y)$ are scalars;
- x, dx, y, dy, and $f(x)$ are regular vectors;
- $\frac{\partial z}{\partial y}$ and $\frac{\partial z}{\partial x}$ are transpose (row) vectors (f and gof gradients);
- $\frac{\partial y}{\partial x}$ is a matrix (f Jacobian).