AMATH 840: Advanced Numerical Methods for Computational and Data Sciences

Giang Tran

Department of Applied Mathematics, University of Waterloo

Feb 13 ~~Jan~~ 26, 2021

Topic 2.2 Back propagation.

Example: Given $\{(x_k, y_k)\}_{k=1}^{m} \subset \mathbb{R}^{d_{in}} \times \mathbb{R}^{d_{out}}$,

find a neural network: $f_{NN}(x_k) \approx y_k$

$$\min_{\theta} \frac{1}{2m} \sum_{k=1}^{m} \| f_{NN}(x_k;\theta) - y_k \|_2^2 = \underline{\mathcal{L}(x_k;\theta)}$$

where $\theta = \{$all weights & bias of the neural network$\}$

For example: x_k = a color image
y_k = "plane", "car", "bike" | Classification

Simplest method

Initialization: θ_0

Update: $$\theta^{new} = \theta^{old} - \tau \frac{\partial \mathcal{L}}{\partial \theta}(\theta^{old})$$

More precisely

$$W_1^{new} = W_1^{old} - \tau \frac{\partial \mathcal{L}}{\partial W_1}(W_1^{old})$$

$$W_2^{new} = W_2^{old} - \tau \frac{\partial \mathcal{L}}{\partial W_2}(W_2^{old})$$

$$\vdots$$

$\hookrightarrow$ Need to comput $\frac{\partial \mathcal{L}}{\partial W_J}$, $\frac{\partial \mathcal{L}}{\partial b_J}$

Einstein

Notation

Jacobian | formulation

$$\left(\frac{\partial y}{\partial x}\right)_{iJ} := \frac{\partial y_i}{\partial x_J}$$

① $y : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

$(x_1, x_2, \ldots, x_n) \longmapsto (y_1, y_2, \ldots, y_m)$

$$\frac{\partial y}{\partial x} := \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Recall
$$\nabla y := \begin{pmatrix} \partial y / \partial x_1 \\ \vdots \\ \partial y / \partial x_n \end{pmatrix}$$

In particular,

ⓐ $m=1$, $y: \mathbb{R}^n \to \mathbb{R}$

$$\frac{\partial y}{\partial x} = \left(\frac{\partial y}{\partial x_1}, \dots, \frac{\partial y}{\partial x_n} \right) \in \mathbb{R}^{1\times n}$$

$$= (\nabla y)^T$$

ⓑ $n=1$, $y: \mathbb{R} \to \mathbb{R}^m$, $\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix} \in \mathbb{R}^{m\times 1}$

ⓒ $y = \overset{m\times n}{A}\,\overset{n\times 1}{x}$, $\frac{\partial y}{\partial x} \in \mathbb{R}^{m\times n} = A$.

② $y: \mathbb{R} \to \mathbb{R}^{m \times n}$

$$\left(\frac{\partial y}{\partial x}\right)_{ij} = \frac{\partial y_{ij}}{\partial x}$$

$$\frac{\partial y}{\partial x} := \begin{pmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & & \ddots & \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

3 $y : \mathbb{R}^{n \times m} \longrightarrow \mathbb{R}$

$$\left(\frac{\partial y}{\partial X}\right)_{ij} = \frac{\partial y}{\partial x_{ji}}$$

$$\frac{\partial y}{\partial X} := \begin{pmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{n1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{n2}} \\ \vdots & \vdots & & \\ \frac{\partial y}{\partial x_{1m}} & \frac{\partial y}{\partial x_{2m}} & \cdots & \frac{\partial y}{\partial x_{nm}} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

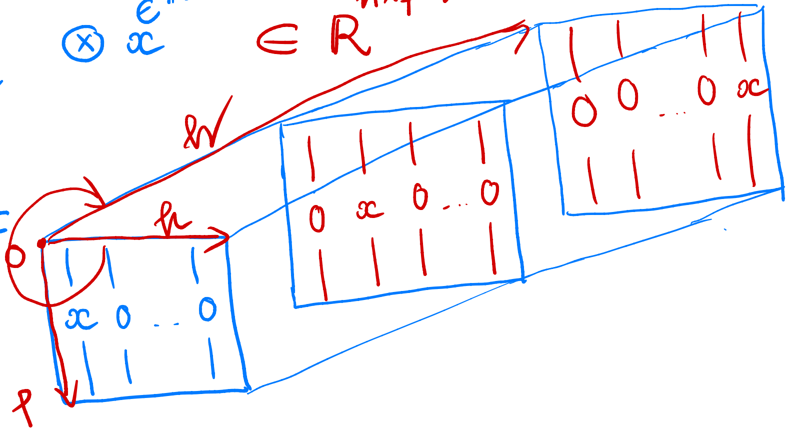
④ $y : \mathbb{R}^{n \times m} \longrightarrow \mathbb{R}^{p}$

$$\frac{\partial y}{\partial X} := \boxed{\frac{\partial y_1}{\partial X}} \quad \boxed{\frac{\partial y_2}{\partial X}} \quad \boxed{\frac{\partial y_p}{\partial X}}$$

$$\left(\frac{\partial y}{\partial X}\right)_{ijk} := \frac{\partial y_i}{\partial x_{kj}}$$

(5) $\underset{\in \mathbb{R}^{h\times h}}{I_h} \otimes \underset{\in \mathbb{R}^{p}}{x} \in \mathbb{R}^{h\times p\times h}$

$I_h \otimes x =$



Example

$$z = \underset{\mathbb{R}^{h\times d}}{W}\,\underset{\mathbb{R}^{d\times 1}}{x} + \underset{\mathbb{R}^{h\times 1}}{b} \in \mathbb{R}^{h\times 1}$$

See 4

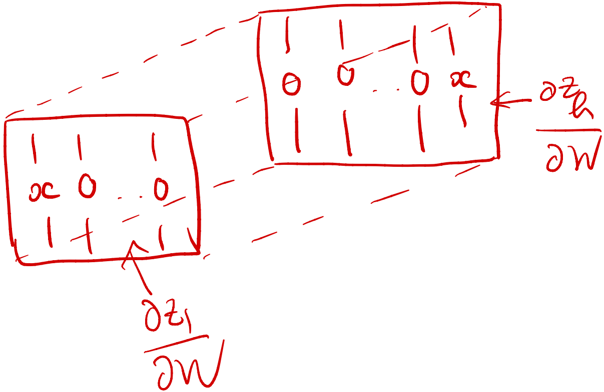
$$\frac{\partial z}{\partial W} := \begin{bmatrix} \dfrac{\partial z_1}{\partial W} & \cdots & \dfrac{\partial z_h}{\partial W} \end{bmatrix}$$

$$W = \begin{bmatrix} - & w_1^T & - \\ - & w_2^T & - \\ & \vdots & \\ - & w_h^T & - \end{bmatrix}$$

$$z_i = (Wx+b)_i = w_i^T x + b_i$$

$$\frac{\partial z_i}{\partial w_{jk}} = \frac{\partial \left(\sum_{t=1}^{d} w_{it} x_t + b_i \right)}{\partial w_{jk}} = \underbrace{\delta_{ij} x_k}$$

$$\frac{\partial(Wx+b)}{\partial W} =$$



$$= I_h \otimes x$$

See note: AMATH840_11_Backpropagation In Neural Networks