EN.601.482/682 Deep Learning

Computational Graphs and Backprop

Part II

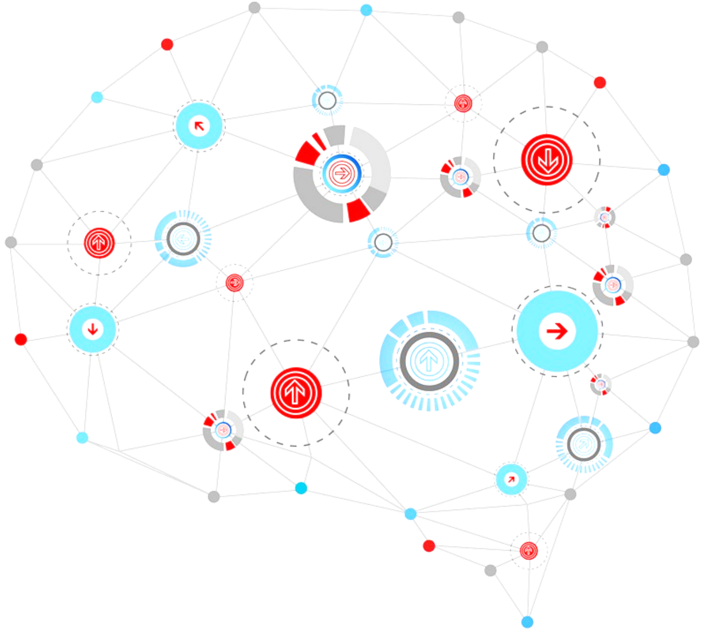
Mathias Unberath, PhD
Assistant Professor
Dept of Computer Science
Johns Hopkins University

Cong Gao, BS
4th-year PhD Student
Dept of Computer Science
Johns Hopkins University

Today’s Lecture

Math of Derivatives

Backpropagation: Matrix Example



Derivatives

Scalar Case:

given a function $f\colon \mathbb{R} \mapsto \mathbb{R}$, the derivative of f at point $x \in \mathbb{R}$ is defined as:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Derivatives

Scalar Case:

given a function $f\colon \mathbb{R} \mapsto \mathbb{R}$, the derivative of f at point $x \in \mathbb{R}$ is defined as:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Measure *change*:

$$f(x+\varepsilon) \approx f(x) + \varepsilon f'(x)$$

Derivatives

Scalar Case:

given a function $f\colon \mathbb{R} \mapsto \mathbb{R}$, the derivative of f at point $x \in \mathbb{R}$ is defined as:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Rephrase $y = f(x)$

$$x \to x + \Delta x \implies y \to \approx y + \frac{\partial y}{\partial x} \Delta x$$

Derivatives

Scalar Case:

- chain rule: how to compute the derivative of the composition of functions

Suppose that $f, g : \mathbb{R} \rightarrow \mathbb{R}$ and $y = f(x),\ z = g(y); \iff z = (g \circ f)(x)$

Derivatives

Scalar Case:

- chain rule: how to compute the derivative of the composition of functions

Suppose that $f, g : \mathbb{R} \to \mathbb{R}$ and $y = f(x),\ z = g(y); \iff z = (g \circ f)(x)$

The (scalar) chain rule tells us that

$$\boxed{\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x}}$$

Derivatives

Scalar Case:

The (scalar) chain rule tells us that

$$\boxed{\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}}$$

$$x \to x + \Delta x \implies y \to\approx y + \frac{\partial y}{\partial x}\Delta x \longrightarrow \boxed{\Delta y = \frac{\partial y}{\partial x}\Delta x}$$

$$y \to y + \Delta y \implies z \to\approx z + \frac{\partial z}{\partial y}\Delta y \longrightarrow \boxed{\frac{\partial z}{\partial y}\Delta y = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}\Delta x}$$

Derivatives

Gradient: Vector in, scalar Out: Binary classification problem

Given a function $f\colon \mathbb{R}^N \mapsto \mathbb{R}$, the derivative of f at the point $x \in \mathbb{R}^N$ is *gradient*:

$$\nabla_x f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{\|h\|}$$

Derivatives

Vector in, scalar Out:

Given a function $f\colon \mathbb{R}^N \mapsto \mathbb{R}$, the derivative of f at the point $x \in \mathbb{R}^N$ is *gradient*:

$$\nabla_x f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{\|h\|}$$

$$\boxed{x} \to \boxed{x} + \boxed{\Delta x} \implies \boxed{y} \to \approx \boxed{y} + \boxed{\frac{\partial y}{\partial x}} \cdot \boxed{\Delta x}$$

$$\frac{\partial y}{\partial x} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \dots, \frac{\partial y}{\partial x_N} \right)$$

dot product

vector

scalar

Derivatives

Jacobian: Vector in, Vector out: Multi-classification problem

Given a function $f\colon \mathbb{R}^N \mapsto \mathbb{R}^M$, the derivative of f at the point $x \in \mathbb{R}^N$ is *Jacobian*:

$$\frac{\partial y}{\partial x} = \underbrace{\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_M}{\partial x_1} & \cdots & \frac{\partial y_M}{\partial x_N} \end{pmatrix}}_{n} \Big] m$$

Derivatives

Jacobian: Vector in, Vector out:

Given a function $f\colon \mathbb{R}^N \mapsto \mathbb{R}^M$, the derivative of f at the point $x \in \mathbb{R}^N$ is *Jacobian*:

$$\frac{\partial y}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_M}{\partial x_1} & \cdots & \frac{\partial y_M}{\partial x_N} \end{pmatrix}$$

$$\boxed{x} \to \underbrace{\boxed{x} + \boxed{\Delta x}}_{n\times 1} \implies \underbrace{\boxed{y}}_{m\times 1} \to \approx \underbrace{\boxed{y}}_{m\times 1} + \underbrace{\boxed{\frac{\partial y}{\partial x}}}_{m\times n} \underbrace{\boxed{\Delta x}}_{n\times 1}$$

$\square$ vector

$\square$ matrix

Matrix-vector multiplication

Derivatives

Jacobian: Vector in, Vector out:

Suppose that $f : \mathbb{R}^N \to \mathbb{R}^M \text{ and } g : \mathbb{R}^M \to \mathbb{R}^K$ $\quad x \in \mathbb{R}^N, \; y \in \mathbb{R}^M, \; z \in \mathbb{R}^K$

$$y = f(x) \text{ and } z = g(y)$$

The (vector) chain rule tells us that

$$\boxed{\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}}$$

Matrix Multiplication

$$\frac{\partial z}{\partial y} : K \times M \text{ matrix}$$

$$\frac{\partial y}{\partial x} : M \times N \text{ matrix}$$

$$\frac{\partial z}{\partial x} : K \times N \text{ matrix}$$

Derivatives

Generalized Jacobian: Tensor* in, Tensor out: Input, output more dimensions

**tensor*: a D-dimensional grid of numbers.

e.g: vector – 1d tensor; matrix – 2d tensor

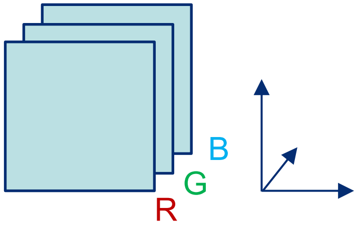
Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

**tensor*: a D-dimensional grid of numbers.

e.g: vector – 1d tensor; matrix – 2d tensor

RGB image – 3d tensor



Derivatives

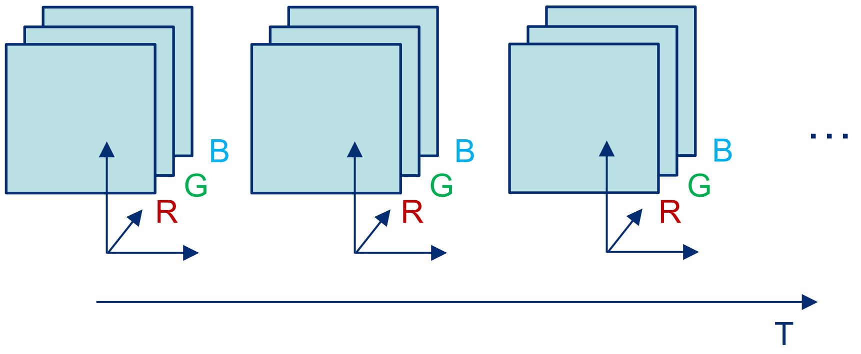
Generalized Jacobian: Tensor* in, Tensor out:

**tensor*: a D-dimensional grid of numbers.

e.g: vector – 1d tensor; matrix – 2d tensor

RGB image – 3d tensor

RGB video – 4d tensor



Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

**tensor*: a D-dimensional grid of numbers. e.g.: matrix – 2d tensor

Given a function $f: \mathbb{R}^{N_1 \times \cdots \times N_{K_x}} \mapsto \mathbb{R}^{M_1 \times \cdots \times M_{K_y}}$, the derivative of f at the point $x \in \mathbb{R}^{N_1 \times \cdots \times N_{K_x}}$ is *generalized Jacobian*:

Shape: $(M_1 \times \ldots \times M_{K_y}) \times (N_1 \times \ldots \times N_{K_x})$

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Given a function $f\colon \mathbb{R}^{N_1 \times \cdots \times N_{K_x}} \mapsto \mathbb{R}^{M_1 \times \cdots \times M_{K_y}}$, the derivative of f at the point $x \in \mathbb{R}^{N_1 \times \cdots \times N_{K_x}}$ is *generalized Jacobian*:

If we let $i \in \mathbb{Z}^{D_y}$ and j $\in \mathbb{Z}^{D_x}$

$$\left(\frac{\partial y}{\partial x}\right)_{i,j} = \frac{\partial y_i}{\partial x_j}$$

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Given a function $f: \mathbb{R}^{N_1 \times \cdots \times N_{K_x}} \mapsto \mathbb{R}^{M_1 \times \cdots \times M_{K_y}}$, the derivative of f at the point $x \in \mathbb{R}^{N_1 \times \cdots \times N_{K_x}}$ is *generalized Jacobian*:

If we let $\underbrace{i \in \mathbb{Z}^{K_x} \text{ and } j \in \mathbb{Z}^{K_y}}_{\text{index}}$

$$\left(\frac{\partial y}{\partial x}\right)_{i,j} = \boxed{\frac{\partial \boxed{y_j}}{\partial \boxed{x_i}}}$$

$\square$ scalar

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Given a function $f\colon \mathbb{R}^{N_1 \times \cdots \times N_{K_x}} \mapsto \mathbb{R}^{M_1 \times \cdots \times M_{K_y}}$, the derivative of f at the point $x \in \mathbb{R}^{N_1 \times \cdots \times N_{K_x}}$ is *generalized Jacobian*:

$$x \to x + \Delta x \implies y \to \approx y + \frac{\partial y}{\partial x} \Delta x$$

Generalized Matrix-vector Multiplication

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Given a function $f\colon \mathbb{R}^{N_1 \times \cdots \times N_{K_x}} \mapsto \mathbb{R}^{M_1 \times \cdots \times M_{K_y}}$, the derivative of f at the point $x \in \mathbb{R}^{N_1 \times \cdots \times N_{K_x}}$ is *generalized Jacobian*:

$$x \to x + \Delta x \implies y \to \approx y + \frac{\partial y}{\partial x} \Delta x$$

Generalized Matrix-vector Multiplication

$$\left(\frac{\partial y}{\partial x} \Delta x\right)_j = \sum_i \left(\frac{\partial y}{\partial x}\right)_{i,j} (\Delta x)_i = \left(\frac{\partial y}{\partial x}\right)_{j,:} \cdot \Delta x$$

Dot product

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Example:
$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \qquad Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$$

2×2 2×2

$$\frac{\partial Y}{\partial X} = \begin{pmatrix} \frac{\partial y_{11}}{\partial X} & \frac{\partial y_{12}}{\partial X} \\ \frac{\partial y_{21}}{\partial X} & \frac{\partial y_{22}}{\partial X} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{12}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{11}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{12}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{12}} \\ \frac{\partial y_{12}}{\partial x_{21}} & \frac{\partial y_{12}}{\partial x_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{12}} \\ \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{22}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{12}} \\ \frac{\partial y_{22}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{22}} \end{pmatrix} \end{pmatrix}$$

(2×2)×(2×2)

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Example:

$$X = \boxed{\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}} \qquad Y = \boxed{\begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}}$$

2×2 2×2

$$\frac{\partial Y}{\partial X} = \boxed{\begin{pmatrix} \boxed{\frac{\partial y_{11}}{\partial X}} & \boxed{\frac{\partial y_{12}}{\partial X}} \\ \boxed{\frac{\partial y_{21}}{\partial X}} & \boxed{\frac{\partial y_{22}}{\partial X}} \end{pmatrix}}$$

$$= \boxed{\begin{pmatrix} \boxed{\begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{12}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{11}}{\partial x_{22}} \end{pmatrix}} & \boxed{\begin{pmatrix} \frac{\partial y_{12}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{12}} \\ \frac{\partial y_{12}}{\partial x_{21}} & \frac{\partial y_{12}}{\partial x_{22}} \end{pmatrix}} \\ \boxed{\begin{pmatrix} \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{12}} \\ \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{22}} \end{pmatrix}} & \boxed{\begin{pmatrix} \frac{\partial y_{22}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{12}} \\ \frac{\partial y_{22}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{22}} \end{pmatrix}} \end{pmatrix}}$$

(2×2)×(2×2)

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Example: $\Delta Y = \frac{\partial Y}{\partial X} \Delta X$

Dot product

$$= \begin{pmatrix} \boxed{\begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{12}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{11}}{\partial x_{22}} \end{pmatrix}} & \begin{pmatrix} \frac{\partial y_{12}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{12}} \\ \frac{\partial y_{12}}{\partial x_{21}} & \frac{\partial y_{12}}{\partial x_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{12}} \\ \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{22}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{12}} \\ \frac{\partial y_{22}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{22}} \end{pmatrix} \end{pmatrix} \cdot \boxed{\begin{pmatrix} \Delta x_{11} & \Delta x_{12} \\ \Delta x_{21} & \Delta x_{22} \end{pmatrix}}$$

$$= \begin{pmatrix} \Delta y_{11} & \Delta y_{12} \\ \Delta y_{21} & \Delta y_{22} \end{pmatrix}$$

$$\boxed{\left(\frac{\partial y}{\partial x} \Delta x\right)_j = \sum_i \left(\frac{\partial y}{\partial x}\right)_{i,j} (\Delta x)_i = \left(\frac{\partial y}{\partial x}\right)_{j,:} \cdot \Delta x}$$

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

The (tensor) chain rule tells us that

$$\boxed{\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}}$$

Generalized Matrix-Matrix Multiplication

$$\left(\frac{\partial z}{\partial x}\right)_{i,j} = \sum_k \left(\frac{\partial z}{\partial y}\right)_{i,k}\left(\frac{\partial y}{\partial x}\right)_{k,j} = \left(\frac{\partial z}{\partial y}\right)_{i,:} \cdot \left(\frac{\partial y}{\partial x}\right)_{:,j}$$

Dot product

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Example:

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \qquad Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \qquad Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

$$\Delta Z = \frac{\partial Z}{\partial Y}\frac{\partial Y}{\partial X}\Delta X$$

$$= \begin{pmatrix} \begin{pmatrix} \frac{\partial z_1}{\partial y_{11}} & \frac{\partial z_1}{\partial y_{12}} \\ \frac{\partial z_1}{\partial y_{21}} & \frac{\partial z_1}{\partial y_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial z_2}{\partial y_{11}} & \frac{\partial z_2}{\partial y_{12}} \\ \frac{\partial z_2}{\partial y_{21}} & \frac{\partial z_2}{\partial y_{22}} \end{pmatrix} \end{pmatrix} \cdot \begin{pmatrix} \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{12}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{11}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{12}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{12}} \\ \frac{\partial y_{12}}{\partial x_{21}} & \frac{\partial y_{12}}{\partial x_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{12}} \\ \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{22}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{12}} \\ \frac{\partial y_{22}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{22}} \end{pmatrix} \end{pmatrix} \cdot \begin{pmatrix} \Delta x_{11} & \Delta x_{12} \\ \Delta x_{21} & \Delta x_{22} \end{pmatrix}$$

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Example: $X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$ $\quad Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$ $\quad Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$

$$\Delta Z = \frac{\partial Z}{\partial Y}\frac{\partial Y}{\partial X}\Delta X$$

$$= \begin{pmatrix} \begin{pmatrix} \frac{\partial z_1}{\partial y_{11}} & \frac{\partial z_1}{\partial y_{12}} \\ \frac{\partial z_1}{\partial y_{21}} & \frac{\partial z_1}{\partial y_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial z_2}{\partial y_{11}} & \frac{\partial z_2}{\partial y_{12}} \\ \frac{\partial z_2}{\partial y_{21}} & \frac{\partial z_2}{\partial y_{22}} \end{pmatrix} \end{pmatrix} \cdot \begin{pmatrix} \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{12}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{11}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{12}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{12}} \\ \frac{\partial y_{12}}{\partial x_{21}} & \frac{\partial y_{12}}{\partial x_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{12}} \\ \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{22}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{12}} \\ \frac{\partial y_{22}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{22}} \end{pmatrix} \end{pmatrix} \cdot \begin{pmatrix} \Delta x_{11} & \Delta x_{12} \\ \Delta x_{21} & \Delta x_{22} \end{pmatrix}$$

Derivatives

Generalized Jacobian: Tensor* in, Tensor out:

Example: $X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \qquad Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \qquad Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$

$$\Delta Z = \frac{\partial Z}{\partial Y}\frac{\partial Y}{\partial X}\Delta X$$

$$= \begin{pmatrix} \begin{pmatrix} \frac{\partial z_1}{\partial y_{11}} & \frac{\partial z_1}{\partial y_{12}} \\ \frac{\partial z_1}{\partial y_{21}} & \frac{\partial z_1}{\partial y_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial z_2}{\partial y_{11}} & \frac{\partial z_2}{\partial y_{12}} \\ \frac{\partial z_2}{\partial y_{21}} & \frac{\partial z_2}{\partial y_{22}} \end{pmatrix} \end{pmatrix} \cdot \begin{pmatrix} \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{12}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{11}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{12}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{12}} \\ \frac{\partial y_{12}}{\partial x_{21}} & \frac{\partial y_{12}}{\partial x_{22}} \end{pmatrix} \\ \begin{pmatrix} \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{12}} \\ \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{22}} \end{pmatrix} & \begin{pmatrix} \frac{\partial y_{22}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{12}} \\ \frac{\partial y_{22}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{22}} \end{pmatrix} \end{pmatrix} \cdot \begin{pmatrix} \Delta x_{11} & \Delta x_{12} \\ \Delta x_{21} & \Delta x_{22} \end{pmatrix}$$

$$\left(\frac{\partial z}{\partial x}\right)_{i,j} = \sum_k \left(\frac{\partial z}{\partial y}\right)_{i,k} \left(\frac{\partial y}{\partial x}\right)_{k,j} = \left(\frac{\partial z}{\partial y}\right)_{i,:} \cdot \left(\frac{\partial y}{\partial x}\right)_{:,j}$$

Dot product

Summary

Input → f(·) → Output

Scalar: a — Scalar: b

Vector: $\vec{a}$ — Vector: $\vec{b}$

Tensor: A — Tensor: B

Backpropagation → Derivative: $\frac{\partial f(*)}{\partial *}$

Summary

$$y = f(x),\ z = g(y);$$

Scalar in	Scalar out	$\Delta y = \frac{\partial y}{\partial x}\Delta x$	Scalar product	$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$	Scalar product
Vector in	Scalar out	$\Delta y = \frac{\partial y}{\partial x}\Delta x$	Dot product		
Vector in	Vector out	$\Delta y = J\Delta x$	Matrix product	$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$	Matrix product
Tensor in	Tensor out	$\Delta y = \frac{\partial y}{\partial x}\Delta x$	Generalized Matrix product	$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$	Generalized Matrix product

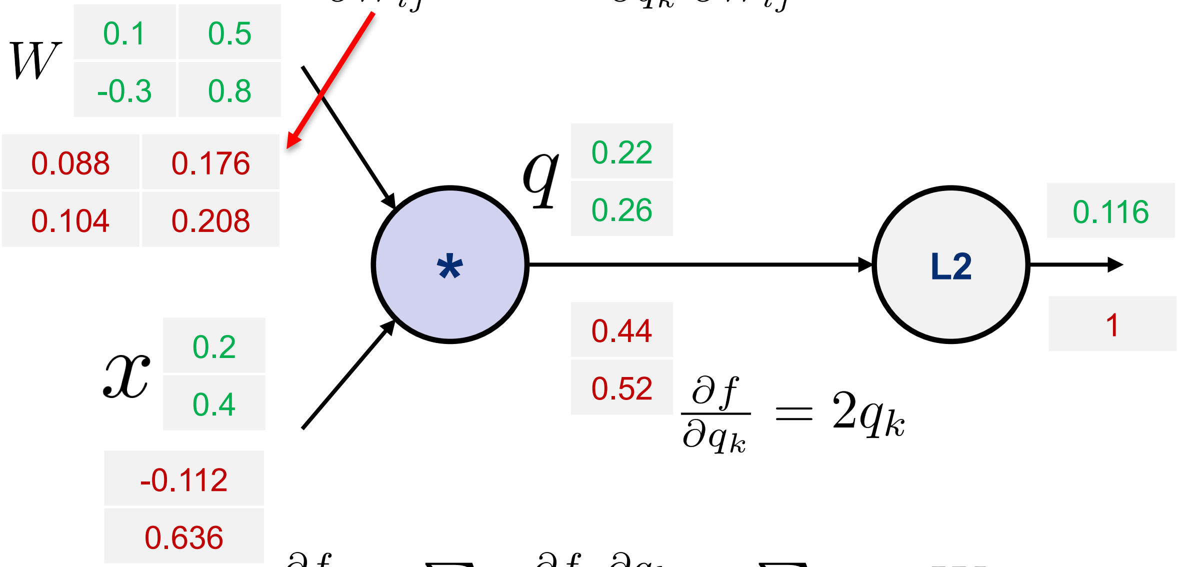
$$\left(\frac{\partial y}{\partial x}\Delta x\right)_j = \sum_i \left(\frac{\partial y}{\partial x}\right)_{i,j} (\Delta x)_i = \left(\frac{\partial y}{\partial x}\right)_{j,:} \cdot \Delta x$$

$$\left(\frac{\partial z}{\partial x}\right)_{i,j} = \sum_k \left(\frac{\partial z}{\partial y}\right)_{i,k} \left(\frac{\partial y}{\partial x}\right)_{k,j} = \left(\frac{\partial z}{\partial y}\right)_{i,:} \cdot \left(\frac{\partial y}{\partial x}\right)_{:,j}$$

Recall: A Vectorized Example

$$f(W, x) : \mathbb{R}^{n \times n} \times \mathbb{R}^n \mapsto \mathbb{R} \qquad f(W, x) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

$$\frac{\partial f}{\partial W_{ij}} = \sum_k \frac{\partial f}{\partial q_k} \frac{\partial q_k}{\partial W_{ij}} = 2 q_i x_j$$



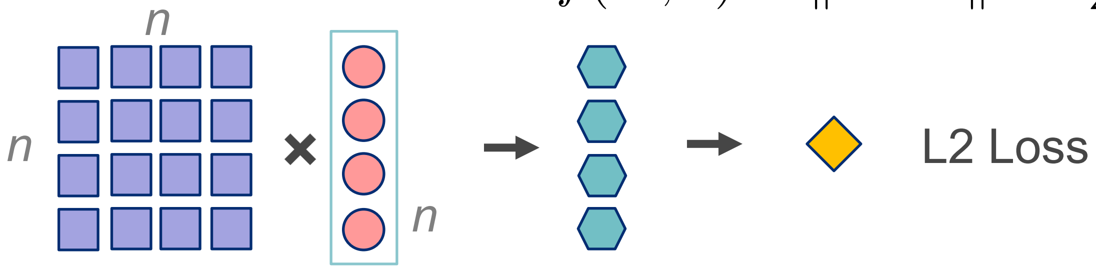
$$\frac{\partial f}{\partial q_k} = 2 q_k$$

$$\frac{\partial f}{\partial x_i} = \sum_k \frac{\partial f}{\partial q_k} \frac{\partial q_k}{\partial x_i} = \sum_k 2 q_k W_{ki}$$

A More Complicated Case

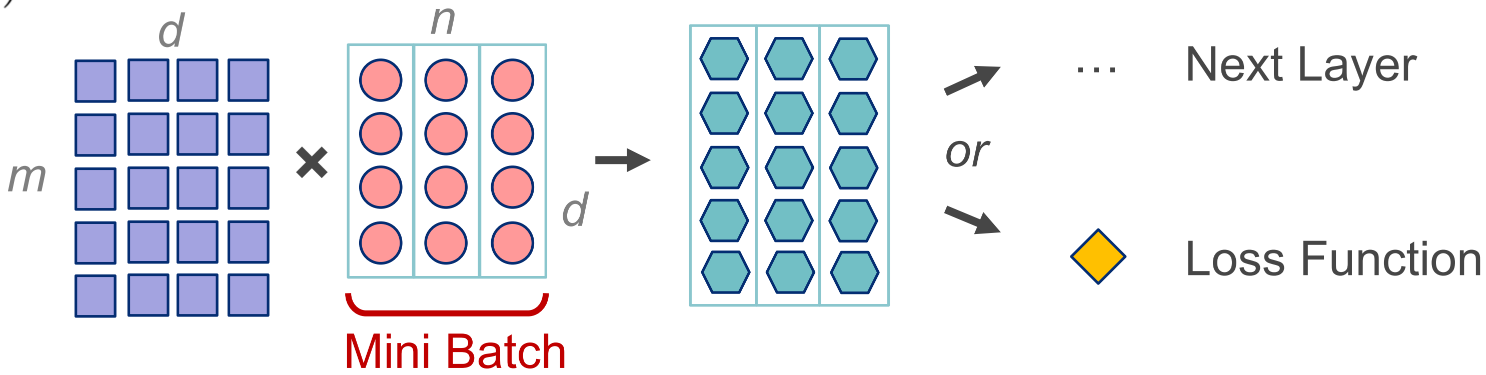
Vector:

$$f(W, x) : \mathbb{R}^{n\times n} \times \mathbb{R}^n \mapsto \mathbb{R} \qquad f(W, x) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



Matrix:

$$f(W, x) : \mathbb{R}^{m\times d} \times \mathbb{R}^{d\times n} \mapsto \mathbb{R}^{m\times n} \mapsto \mathbb{R}$$



An Matrix Example

$$f(W, x) : \boxed{\mathbb{R}^{m \times d} \times \mathbb{R}^{d \times n}} \mapsto \mathbb{R}^{m \times n} \mapsto \mathbb{R}$$

Tensor Tensor Scalar

We consider the case: m = 2, d = 2, n = 3

$$W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \qquad X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

$$Y = WX$$

Tensor in Tensor out

A Matrix Example

$$f(W, x) : \mathbb{R}^{m \times d} \times \mathbb{R}^{d \times n} \mapsto \mathbb{R}^{m \times n} \mapsto \boxed{\mathbb{R}}$$

We consider the case: m = 2, d = 2, n = 3

$$W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \qquad X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

$$\boxed{Y} = \boxed{W}\boxed{X}$$ Tensor in Tensor out

$$\boxed{L} = l(\boxed{Y})$$ Tensor in Scalar out

□ Tensor

□ scalar

A Matrix Example

$$f(W, x) : \mathbb{R}^{m\times d} \times \mathbb{R}^{d\times n} \mapsto \boxed{\mathbb{R}^{m\times n}} \mapsto \mathbb{R}$$

We consider the case: m = 2, d = 2, n = 3

$$W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \qquad X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

$$\begin{aligned} Y &= WX \\ &= \begin{pmatrix} w_{11}x_{11} + w_{12}x_{21} & w_{11}x_{12} + w_{12}x_{22} & w_{11}x_{13} + w_{12}x_{23} \\ w_{21}x_{11} + w_{22}x_{21} & w_{21}x_{12} + w_{22}x_{22} & w_{21}x_{13} + w_{22}x_{23} \end{pmatrix} \\ &= \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{pmatrix} \end{aligned}$$

An Matrix Example

By the chain rule, we know that

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y}\frac{\partial Y}{\partial X}$$

A Matrix Example

By the chain rule, we know that

$$\frac{\partial L}{\partial X} = \boxed{\frac{\partial L}{\partial Y}} \frac{\partial Y}{\partial X}$$

$$\frac{\partial L}{\partial Y} = \begin{pmatrix} \frac{\partial L}{\partial y_{11}} & \frac{\partial L}{\partial y_{12}} & \frac{\partial L}{\partial y_{13}} \\ \frac{\partial L}{\partial y_{21}} & \frac{\partial L}{\partial y_{22}} & \frac{\partial L}{\partial y_{23}} \end{pmatrix}$$

A Matrix Example

By the chain rule, we know that

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y}\boxed{\frac{\partial Y}{\partial X}} \quad \Rightarrow \quad \begin{pmatrix} \frac{\partial Y}{\partial x_{11}} & \frac{\partial Y}{\partial x_{12}} & \frac{\partial Y}{\partial x_{13}} \\ \frac{\partial Y}{\partial x_{21}} & \frac{\partial Y}{\partial x_{22}} & \frac{\partial Y}{\partial x_{23}} \end{pmatrix}$$

A Matrix Example

By the chain rule, we know that

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y}\boxed{\frac{\partial Y}{\partial X}} \Rightarrow \begin{pmatrix} \frac{\partial Y}{\partial x_{11}} & \frac{\partial Y}{\partial x_{12}} & \frac{\partial Y}{\partial x_{13}} \\ \frac{\partial Y}{\partial x_{21}} & \frac{\partial Y}{\partial x_{22}} & \frac{\partial Y}{\partial x_{23}} \end{pmatrix}$$

$$Y = WX$$

$$= \begin{pmatrix} w_{11}x_{11} + w_{12}x_{21} & w_{11}x_{12} + w_{12}x_{22} & w_{11}x_{13} + w_{12}x_{23} \\ w_{21}x_{11} + w_{22}x_{21} & w_{21}x_{12} + w_{22}x_{22} & w_{21}x_{13} + w_{22}x_{23} \end{pmatrix}$$

$$\frac{\partial Y}{\partial x_{11}} = \begin{pmatrix} w_{11} & 0 & 0 \\ w_{21} & 0 & 0 \end{pmatrix}$$

A Matrix Example

By the chain rule, we know that

$$\boxed{\frac{\partial L}{\partial X}} = \frac{\partial L}{\partial Y}\frac{\partial Y}{\partial X}$$

$$\frac{\partial L}{\partial Y} = \begin{pmatrix} \frac{\partial L}{\partial y_{11}} & \frac{\partial L}{\partial y_{12}} & \frac{\partial L}{\partial y_{13}} \\ \frac{\partial L}{\partial y_{21}} & \frac{\partial L}{\partial y_{22}} & \frac{\partial L}{\partial y_{23}} \end{pmatrix} \qquad \frac{\partial Y}{\partial x_{11}} = \begin{pmatrix} w_{11} & 0 & 0 \\ w_{21} & 0 & 0 \end{pmatrix}$$

A Matrix Example

By the chain rule, we know that

$$\boxed{\frac{\partial L}{\partial X}} = \frac{\partial L}{\partial Y}\frac{\partial Y}{\partial X}$$

$$\frac{\partial L}{\partial Y} = \begin{pmatrix} \frac{\partial L}{\partial y_{11}} & \frac{\partial L}{\partial y_{12}} & \frac{\partial L}{\partial y_{13}} \\ \frac{\partial L}{\partial y_{21}} & \frac{\partial L}{\partial y_{22}} & \frac{\partial L}{\partial y_{23}} \end{pmatrix} \qquad \frac{\partial Y}{\partial x_{11}} = \begin{pmatrix} w_{11} & 0 & 0 \\ w_{21} & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} \frac{\partial L}{\partial x_{11}} &= \boxed{\frac{\partial L}{\partial Y}\frac{\partial Y}{\partial x_{11}}} \longrightarrow \text{dot multiplication} \\ &= \sum_{k=1}^{m}\sum_{l=1}^{n}\frac{\partial L}{\partial y_{kl}}\frac{\partial y_{kl}}{\partial x_{11}} \\ &= \frac{\partial L}{\partial y_{11}}w_{11} + \frac{\partial L}{\partial y_{21}}w_{21} \end{aligned}$$

A Matrix Example

By the chain rule, we know that

$$\boxed{\frac{\partial L}{\partial X}} = \frac{\partial L}{\partial Y} \frac{\partial Y}{\partial X}$$

$$\frac{\partial L}{\partial x_{11}} = \frac{\partial L}{\partial y_{11}} w_{11} + \frac{\partial L}{\partial y_{21}} w_{21} \qquad \frac{\partial L}{\partial x_{12}} = \frac{\partial L}{\partial y_{12}} w_{11} + \frac{\partial L}{\partial y_{22}} w_{21} \qquad \frac{\partial L}{\partial x_{13}} = \frac{\partial L}{\partial y_{13}} w_{11} + \frac{\partial L}{\partial y_{23}} w_{21}$$

$$\frac{\partial L}{\partial x_{21}} = \frac{\partial L}{\partial y_{11}} w_{12} + \frac{\partial L}{\partial y_{21}} w_{22} \qquad \frac{\partial L}{\partial x_{22}} = \frac{\partial L}{\partial y_{12}} w_{12} + \frac{\partial L}{\partial y_{22}} w_{22} \qquad \frac{\partial L}{\partial x_{23}} = \frac{\partial L}{\partial y_{13}} w_{12} + \frac{\partial L}{\partial y_{23}} w_{22}$$

A Matrix Example

By the chain rule, we know that

$$\boxed{\frac{\partial L}{\partial X}} = \frac{\partial L}{\partial Y}\frac{\partial Y}{\partial X}$$

$$\begin{aligned}\frac{\partial L}{\partial X} &= \begin{pmatrix} \frac{\partial L}{\partial y_{11}} w_{11} + \frac{\partial L}{\partial y_{21}} w_{21} & \frac{\partial L}{\partial y_{12}} w_{11} + \frac{\partial L}{\partial y_{22}} w_{21} & \frac{\partial L}{\partial y_{13}} w_{11} + \frac{\partial L}{\partial y_{23}} w_{21} \\ \frac{\partial L}{\partial y_{11}} w_{12} + \frac{\partial L}{\partial y_{21}} w_{22} & \frac{\partial L}{\partial y_{12}} w_{12} + \frac{\partial L}{\partial y_{22}} w_{22} & \frac{\partial L}{\partial y_{13}} w_{12} + \frac{\partial L}{\partial y_{23}} w_{22} \end{pmatrix} \\ &= \boxed{\begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{pmatrix} \begin{pmatrix} \frac{\partial L}{\partial y_{11}} & \frac{\partial L}{\partial y_{12}} & \frac{\partial L}{\partial y_{13}} \\ \frac{\partial L}{\partial y_{21}} & \frac{\partial L}{\partial y_{22}} & \frac{\partial L}{\partial y_{23}} \end{pmatrix}} \\ &= \boxed{W^T \frac{\partial L}{\partial Y}}\end{aligned}$$

matrix multiplication

A Matrix Example

By the chain rule, we know that

$$Y = WX \qquad L = l(Y)$$

$$\frac{\partial L}{\partial X} = \boxed{W^T \frac{\partial L}{\partial Y}} \qquad \frac{\partial L}{\partial W} = \boxed{\frac{\partial L}{\partial Y} X^T}$$

Practice yourself!

Compute Graphs and Backpropagation

Questions?