UMassAmherst | Manning College of Information & Computer Sciences

Vector, Matrix and Tensor Derivatives


$$\frac{\partial h_i}{\partial W_{jk}}$$
Wait its all scalar derivatives?
Always has been.

Overview

- Matrix derivatives as scalar derivatives
- Writing everything in terms of scalars
- Chain rule
- Kronocker delta function
- Examples
- Easy implementation (einstein notation + np.einsum)

Scalar derivatives

$$L = f(W)$$

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial W_{11}} & \cdots & \frac{\partial L}{\partial W_{1m}} \\ \vdots & & \vdots \\ \frac{\partial L}{\partial W_{n1}} & \cdots & \frac{\partial L}{\partial W_{nm}} \end{bmatrix} = \left[\frac{\partial L}{\partial W_{ij}} \right]$$

Scalar derivatives

$$h = \sigma(Wx + b)$$

$$\frac{\partial h}{\partial W} \implies \frac{\partial h_i}{\partial W_{jk}}$$

dhdW.ndim = 3

Rewriting matrix equations in scalars

$$\frac{\partial y_i}{\partial W_{jk}}$$

$$y = Wx + b$$

$$y_i = (Wx)_i + b_i = \sum_j W_{ij} x_j + b_i$$

Chain rule

$f(x_1, \ldots, x_n) = f(x) \quad x \in \mathbb{R}^n$

$\frac{\partial f}{\partial W_{ij}}$

$$f(x_1(W), \ldots, x_n(W))$$

$$\frac{\partial f}{\partial W_{ij}} = \frac{\partial f}{\partial x_1}\frac{x_1}{W_{ij}} + \cdots + \frac{\partial f}{\partial x_n}\frac{x_1}{W_{ij}} = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\frac{x_k}{W_{ij}}$$

$\boxed{F(x, y)},\ x, y \in \mathbb{R}^n \Longrightarrow f(x_1, \ldots, x_n, y_1, \ldots, y_n)$

Remove nested equations

$$f(x) = ||Ax||^2$$

$$f(y) = ||y||^2, \quad y = Ax$$

$$\frac{\partial f}{\partial x_i} = \sum_k \frac{\partial f}{\partial y_k} \frac{\partial y_k}{\partial x_i}$$

Chain rule example

$$f(x) = ||Ax||^2$$

$$f(y) = ||y||^2 \longrightarrow \boxed{f = \sum_k y_k^2}$$

$$y = Ax \longrightarrow \boxed{y_i = \sum_j A_{ij} x_j}$$

$i_1, i_2, \ldots, i_K$

$$\frac{\partial f}{\partial x_\ell} = \boxed{\sum_a \boxed{\frac{\partial f}{\partial y_a}} \boxed{\frac{\partial y_a}{\partial x_\ell}}}$$

$$\frac{\partial f}{\partial y_a} = \frac{\partial}{\partial y_a} \sum_k y_k^2 = \sum_k \frac{\partial y_k^2}{\partial y_a} = \sum_k 2 y_k \underbrace{\frac{\partial y_k}{\partial y_a}}_{\delta_{ka}} = \boxed{2 y_a}$$

$$\frac{\partial y_a}{\partial x_\ell} = \frac{\partial}{\partial x_\ell}\left[\sum_j A_{aj} x_j\right] = \sum_j A_{aj} \frac{\partial x_j}{\partial x_\ell} = \sum_j A_{aj} \delta_{j\ell} = \boxed{A_{a\ell}}$$

$$\frac{\partial f}{\partial x_\ell} = \sum_a 2 y_a A_{a\ell} = 2 \sum_a A_{a\ell} y_a = 2 A^T y$$

Caution

Don’t repeat indices!

$$y = Wx + b$$

$$\frac{\partial y_i}{\partial W_{jk}} = \frac{\partial}{\partial W_{jk}} (\sum_l W_{il} x_l + b_i)$$

Kronecker delta function

Just a fancy way to write the identity matrix

$$\boxed{\frac{\partial y_a}{\partial y_k}} = \begin{cases} 1 & \text{when } a = k \\ 0 & \text{otherwise} \end{cases}$$

$$= \delta_{ak}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial W_{ij}}{\partial W_{ab}} = \delta_{ia} \delta_{jb}$$

$$I = \left[\delta_{ij} \right] = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \vdots & \ddots & & \\ & & \cdots & \delta_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Kronecker delta function

Just a fancy way to write the identity matrix

$$\sum_k y_k \delta_{ak} = y_1 \cancel{\delta_{a1}}^{0} + \cdots + \underbrace{\boxed{y_a \delta_{aa}}}_{= y_a} + \cdots + y_n \cancel{\delta_{an}}^{0}$$

$$= y_a$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_i \sum_j y_i y_j \delta_{ij} \rightarrow \sum_i y_i \sum_j y_j \delta_{ij}$$

$$= \sum_i y_i y_i$$

Overview

1. Modularize your equations
2. Chain rule
3. Rewrite everything in terms of scalars

Example 1

$$\sum_l \delta_{il}\delta_{lj} = \delta_{ij}$$

$$h(W) = \sigma(Wx + b)$$

$$\frac{\partial h}{\partial W} = \frac{\partial h_i}{\partial W_{jk}}$$

$h(y) = \sigma(y) \longrightarrow h_i = \sigma(y_i)$

$y = Wx + b \longrightarrow y_i = \sum_j W_{ij} x_j + b_i$

$$\frac{\partial h_i}{\partial W_{jk}} = \sum_l \boxed{\frac{\partial h_i}{\partial y_l}} \boxed{\frac{\partial y_l}{\partial W_{jk}}}$$

$$= \sum_l \frac{\partial \sigma}{\partial y_i} \delta_{il} \delta_{lj} x_k$$

$$= \boxed{\frac{\partial \sigma}{\partial y_i} \delta_{ij} x_k}$$

$$\frac{\partial h_i}{\partial y_l} = \frac{\partial \sigma(y_i)}{\partial y_l} = \frac{\partial \sigma}{\partial y_i}\frac{\partial y_i}{\partial y_l} = \boxed{\frac{\partial \sigma}{\partial y_i}\delta_{il}}$$

$$\frac{\partial y_l}{\partial W_{jk}} = \frac{\partial}{\partial W_{jk}}\left(\sum_a W_{la} x_a + b_l\right)$$

$$= \sum_a \frac{\partial W_{la}}{\partial W_{jk}} x_a = \sum_a \delta_{lj}\delta_{ak} x_a = \boxed{\delta_{lj} x_k}$$

$Y(p(\theta), q(\theta))$

Example 2

$$\frac{\partial y}{\partial \theta} \rightarrow \frac{\partial Y_{ij}}{\partial \theta}$$

$$y(\theta) = (\underbrace{Ax(\theta)}_{p})(\underbrace{Bx(\theta)}_{q})^T$$

$$Y_{ij} = p_i q_j$$

$$p_i = \sum_u A_{iu} x_u$$

$$q_j = \sum_v B_{jv} x_v$$

$$\frac{\partial Y_{ij}}{\partial \theta} = \sum_k \boxed{\frac{\partial Y_{ij}}{\partial p_k}} \boxed{\frac{\partial p_k}{\partial \theta}} + \boxed{\frac{\partial Y_{ij}}{\partial q_k}} \frac{\partial q_k}{\partial \theta}$$

$$\frac{\partial Y_{ij}}{\partial p_k} = \frac{\partial p_i q_j}{\partial p_k} = \frac{\partial p_i}{\partial p_k} q_j = \delta_{ik} q_j \longrightarrow \frac{\partial Y_{ij}}{\partial q_k} = \delta_{jk} p_i$$

$$\frac{\partial p_k}{\partial \theta} = \sum_u A_{ku} \frac{\partial x_u}{\partial \theta}$$

$$\frac{\partial q_k}{\partial \theta} = \sum_v B_{kv} \frac{\partial x_v}{\partial \theta}$$

$$= \sum_k \delta_{ik} q_j \sum_u A_{ku} \frac{\partial x_u}{\partial \theta} + \delta_{jk} p_i \sum_v B_{kv} \frac{\partial x_v}{\partial \theta}$$

$$= q_j \sum_u A_{iu} \frac{\partial x_u}{\partial \theta} + p_i \sum_v B_{jv} \frac{\partial x_v}{\partial \theta}$$

Example 3

$$L = tr(W)$$

$$tr(W) = \sum_i W_{ii}$$

$$\frac{\partial L}{\partial W_{jk}} = \frac{\partial \sum_i W_{ii}}{\partial W_{jk}} = \sum_i \frac{\partial W_{ii}}{\partial W_{jk}} = \sum_i \delta_{ij}\,\delta_{ik} = \delta_{jk}$$

Implementation with einsum

- Einstein notation -> remove the sum symbol

$$\boxed{y = ABx}$$

$$y_i = \sum_j A_{ij} \sum_k B_{jk} x_k \implies A_{ij} B_{jk} x_k = y_i$$

$$A_{ij} B_{jk} x_k = y_i$$

```
y = np.einsum("ij,jk,k->i", A, B, x)
```

Contraction indices

einops

einops.rearrange("(h 2) (w 2) c → h w (c 4)", x)

Einsum example 2

$$L = \sum_i \sum_j \boxed{tr(X^{(i)^T} X^{(j)})}$$

$X^{(i)} \in \mathbb{R}^{n \times m}$

$i = 1, \ldots, K$

$$L = \sum_i \sum_j \sum_k (X^{(i)^T} X^{(j)})_{kk} = \sum_i \sum_j \sum_k \sum_l X^{(i)}_{lk} X^{(j)}_{lk}$$

L = np.einsum("ilk,jlk->", X, X)

Depends on X.shape