Tensor Derivatives

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu

Maryam Ramezani maryam.ramezani@sharif.edu

Introduction

Linearization

- Board 😄

Motivation

- Machine Learning training requires one to evaluate how one vector changes with respect to another
- How output changes with respect to parameters
- How do we find minimum of a scalar function?
- How do we find minimum of two variables?

Vector-Valued Function

- Derivative of a vector-valued function $f: \mathbb{R} \to \mathbb{R}^n$ with respect to scalar $x \in \mathbb{R}$:

$$\frac{\partial f(x)}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} \\ \frac{\partial f_2(x)}{\partial x} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x} \end{bmatrix}$$

$$f(x) \approx f(x_0) + m(x - x_0) \qquad m = \begin{bmatrix} f_1'(x_0) \\ .. \\ f_n'(x_0) \end{bmatrix}$$

Example

$$f(x) = \begin{bmatrix} \sin(x) \\ cos(x) \end{bmatrix}$$

Matrix-Valued Function

- Derivative of a matrix-valued function $f: \mathbb{R} \to \mathbb{R}^{m \times n}$ with respect to scalar $x \in \mathbb{R}$:

$$\frac{\partial f(x)}{\partial x} \triangleq \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x} & \frac{\partial f_{12}(x)}{\partial x} & \cdots & \frac{\partial f_{1n}(x)}{\partial x} \\ \frac{\partial f_{21}(x)}{\partial x} & \frac{\partial f_{22}(x)}{\partial x} & \cdots & \frac{\partial f_{2n}(x)}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{m1}(x)}{\partial x} & \frac{\partial f_{m2}(x)}{\partial x} & \cdots & \frac{\partial f_{mn}(x)}{\partial x} \end{bmatrix}$$

Example

- Rotation Matrix

Real-Valued Multivariant Function

- Derivative of a real-valued function $f: \mathbb{R}^n \to \mathbb{R}$ with respect to vector $\mathbf{x} \in \mathbb{R}^n$:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} \\ \frac{\partial f_2(x)}{\partial x} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x} \end{bmatrix}$$

$$f(x) - f(x_0) = m^T(x - x_0) \qquad m = \begin{bmatrix} f_1'(x_0) \\ .. \\ f_n'(x_0) \end{bmatrix}$$

- Gradient

Directional Derivative

- https://www.geogebra.org/m/bxhwxr2x

Product Rule

- $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$

Example

- $f, g: \mathbb{R} \to \mathbb{R}^n$ $\quad h(x) = f(x)^T g(x) \quad h'(x) = ?$
- $f: \mathbb{R}^n \to \mathbb{R} \quad g: \mathbb{R} \to \mathbb{R} \quad h(x) = f(x)g(x) \quad h'(x) = ?$
- H: $\mathbb{R} \to \mathbb{R}^{m\times n}$, F: $\mathbb{R} \to \mathbb{R}^{m\times p}$, G: $\mathbb{R} \to \mathbb{R}^{p\times n}$ $\quad H(x) = F(x)G(x)$

Chain Rule

- Board 😄

Motivation

- Derivative of a scalar function $f: \mathbb{R}^N \to \mathbb{R}$ with respect to vector $\mathbf{x} \in \mathbb{R}^N$:

  - $$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_N} \end{bmatrix}$$

- Derivative of a vector function $f: \mathbb{R}^N \to \mathbb{R}^M$ with respect to vector $\boldsymbol{x} \in \mathbb{R}^N$:

  - $$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_N} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\boldsymbol{x})}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_M(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_M(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_M(\boldsymbol{x})}{\partial x_N} \end{bmatrix}$$

Definitions

Definition

- Derivative of a scalar function $f: \mathbb{R}^{M\times N} \to \mathbb{R}$ with respect to matrix $\boldsymbol{X} \in \mathbb{R}^{M\times N}$:

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} \triangleq \begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial X_{1,1}} & \frac{\partial f(\boldsymbol{X})}{\partial X_{2,1}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial X_{M,1}} \\ \frac{\partial f(\boldsymbol{X})}{\partial X_{1,2}} & \frac{\partial f(\boldsymbol{X})}{\partial X_{2,2}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial X_{M,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial X_{1,N}} & \frac{\partial f(\boldsymbol{X})}{\partial X_{2,N}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial X_{M,N}} \end{bmatrix}$$

- Using the above definitions, we can generalize the chain rule, Given $\boldsymbol{u} = \boldsymbol{h}(x)$ (i.e. $\boldsymbol{u}$ is a function of x) and $\boldsymbol{g}$ is a vector function of $\boldsymbol{u}$, the vector-by-vector chain rule states:

$$\frac{\partial \boldsymbol{g}(\boldsymbol{u})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{g}(\boldsymbol{u})}{\partial \boldsymbol{u}}$$

Scalar & Vectors

- Board 😄

Vectors & Vectors

- Board 😄

Tensor

Definition

- Multi-dimensional array of numbers

```
w = torch.empty(3)
x = torch.empty(2, 3)
y = torch.empty(2, 3, 4)
z = torch.empty(2, 3, 2, 4)
```

Scalar	Vector	Matrix	Rank-3 Tensor
(rank 0)	(rank 1)	(rank 2)	(rank 3)

Tensors Addition

- Adding tensors with same size
- Adding scalar to tensor
- Adding tensors with different size: if broacastable

Tensors Addition

- Two tensors are "broadcastable" if the following rules hold:
  - Each tensor has at least one dimension.
  - When iterating over the dimension sizes, starting at the trailing dimension, the dimension sizes must either be equal, one of them is 1, or one of them does not exist.
- Example
  - T1: (5,7,3) T2:(5,7,3)
  - T1: (5,3,4,1) T2:(3,1,1)

Notes

- Matrix Product on tensors

$$(m \times n) \cdot (n \times k) = (m \times k)$$

product is defined

Conclusion

Conclusion

- $\frac{\partial(\boldsymbol{u}(x)+\boldsymbol{v}(x))}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}(x)}{\partial x} + \frac{\partial \boldsymbol{v}(x)}{\partial \boldsymbol{x}}$
- $\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}$
- $\frac{\partial(\boldsymbol{x}^T\boldsymbol{a})}{\partial \boldsymbol{x}} = \boldsymbol{a}$
- $\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{x}$
- $\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$ if $\boldsymbol{A}$ is symmetric
- $\frac{\partial|\boldsymbol{X}|}{\partial \boldsymbol{X}} = |\boldsymbol{X}|(\boldsymbol{X}^{-1})^T$
- $\frac{\partial(\ln(|\boldsymbol{X}|))}{\partial \boldsymbol{X}} = (\boldsymbol{X}^{-1})^T$

Hint!

$$A\vec{x} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_1x_1 + a_2x_2 \\ a_3x_1 + a_4x_2 \end{bmatrix}$$

$$\frac{dA\vec{x}}{dx} = \begin{bmatrix} \dfrac{\partial(a_1x_1 + a_2x_2)}{\partial x_1} & \dfrac{\partial(a_1x_1 + a_2x_2)}{\partial x_2} \\ \dfrac{\partial(a_3x_1 + a_4x_2)}{\partial x_1} & \dfrac{\partial(a_3x_1 + a_4x_2)}{\partial x_2} \end{bmatrix}$$

$$= \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = A$$

Derivative of a vector with respect to a matrix

Derivative of a matrix with respect to a matrix

Conclusion

Important

1. Derivative of a linear function:

$$\frac{\partial}{\partial \vec{x}} \vec{a}.\vec{x} = \frac{\partial}{\partial \vec{x}} \vec{a}^T \vec{x} = \frac{\partial}{\partial \vec{x}} \vec{x}^T \vec{a} = \vec{a}$$

(If you think back to calculus, this is just like $\frac{\mathrm{d}}{\mathrm{dx}} \mathrm{ax} = \mathrm{a}$).

2. Derivative of a quadratic function:

$$\frac{\partial}{\partial x} \vec{x}^T A \vec{x} = 2A\vec{x}$$

(Again, if you think back to calculus, this is just like $\frac{\mathrm{d}}{\mathrm{dx}} \mathrm{ax}^2 = 2\mathrm{ax}$).

If you ever need it, the more general rule (for non-symmetric A) is:

$$\frac{\partial}{\partial x} \vec{x}^T A \vec{x} = (A + A^T)\vec{x}$$

which of course is the same thing as $2A\vec{x}$ when A is symmetric.

Good Examples!!!

Proof the followings:

- $\frac{\partial(y^T A X)}{\partial A} = y X^T$
- $\frac{\partial(X^T A X)}{\partial A} = X X^T$
- $\frac{\partial(A(t))^{-1}}{\partial t} = -A(t)^{-1} \frac{\partial(A(t))}{\partial t} A(t)^{-1}$
- $\frac{\partial \det(A)}{\partial A} = \det(A)\, A^{-1}$
- $\frac{\partial \det(A(t))}{\partial t} = \det(A)\, trace(A^{-1} \frac{\partial(A(t))}{\partial t})$
- $\frac{\partial trace(BA^{-1})}{\partial A} = -A^{-1} B A^{-1}$

References

- Linear Algebra and Its Applications, David C. Lay
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares
- https://en.Wikipedia.org/wiki/matrix_calculus
- https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf
- https://www.kamperh.com/notes/kamper_matrixcalculus13.pdf