Tensor Derivatives

CE282: Linear Algebra

Computer Engineering Department

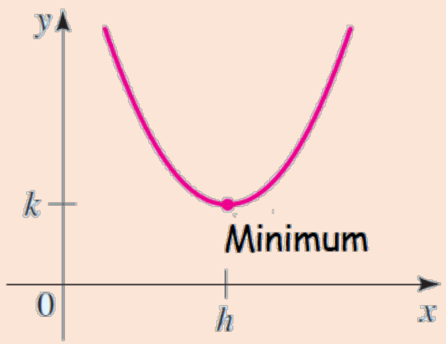
Sharif University of Technology

Hamid R. Rabiee

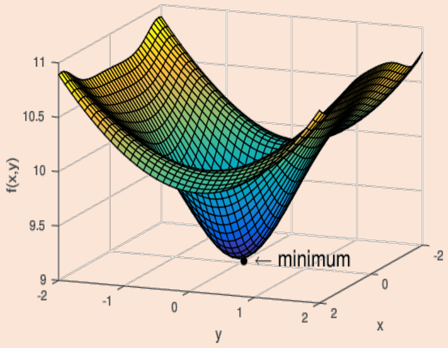
Maryam Ramezani

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?



Good Resources

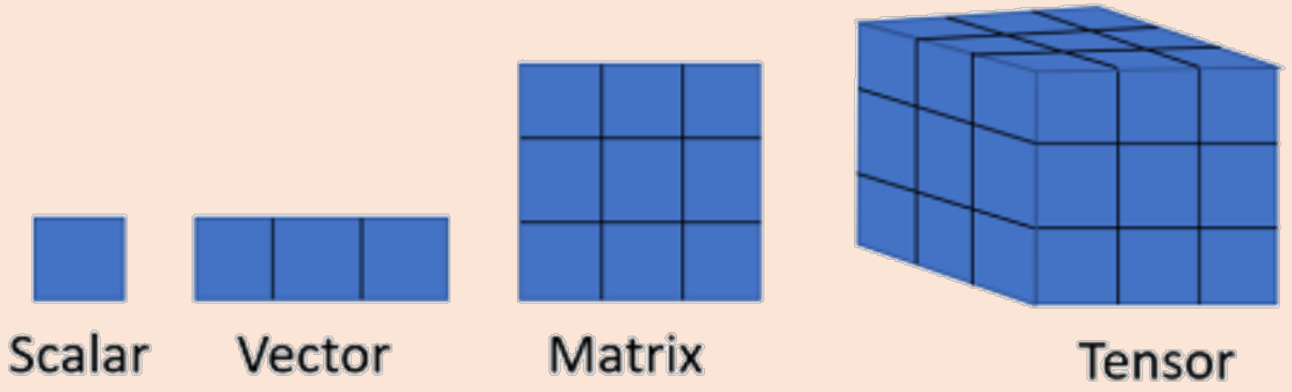
Resources

- Https://en.Wikipedia.org/wiki/matrix_calculus
- Https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf
- Https://www.kamperh.com/notes/kamper_matrixcalculus13.pdf

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

Definition

- 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 $\boldsymbol{x} \in \mathbb{R}^N$:

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

Definition

- Derivative of a scalar function $f\colon \mathbb{R}^{M\times N} \to \mathbb{R}$ 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:

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

Resource

- https://www.youtube.com/watch?v=IgAr5kzza78&list=RDCMUCYa1WtI-vb_bx-anHdmpNfA&index=3

Example

Find the derivative of quadratic form

Yet another approach using the Frobenius product notation.

For a column vector $x \in \mathbb{R}^n$, and a matrix $A \in \mathbb{R}^{n \times n}$ we can write:

$$x^T Ax = Tr(x^T Ax) = x : Ax$$

Then we take the differential and derivative as

$$\begin{aligned} d(x : Ax) &= dx : Ax + x : Adx \\ &= Ax : dx + A^T x : dx \\ &= (Ax + A^T x) : dx \\ \frac{\partial(x^T Ax)}{\partial x} &= (Ax + A^T x) = (A + A^T)x \end{aligned}$$

Conclusion

Conclusion

- $\frac{\partial(\boldsymbol{u}(\boldsymbol{x})+\boldsymbol{v}(\boldsymbol{x}))}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}(\boldsymbol{x})}{\partial x} + \frac{\partial \boldsymbol{v}(\boldsymbol{x})}{\partial \boldsymbol{x}}$
- $\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}$
- $\frac{\partial(\boldsymbol{x}^T\boldsymbol{a})}{\partial \boldsymbol{x}} = \boldsymbol{a}$
- $\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = (\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}$
- $\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$ 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!

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

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

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

Notes

Note

$$\frac{\partial\left(y^T A X\right)}{\partial A} = yX^T$$

$$\frac{\partial\left(X^T A X\right)}{\partial A} = XX^T$$

Geometry:

https://www.youtube.com/watch?v=bohL918kXQk

https://www.youtube.com/watch?v=ahvmZX9WvVY
https://www.youtube.com/watch?v=mhTvwrNz1Qk

Derivative of a vector with respect to a matrix

https://www.youtube.com/watch?v=iMPhPwmcixM

Derivative of a matrix with respect to a matrix

https://www.youtube.com/watch?v=DIXLbZZzc_Q

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.

Derivative of matrix inverse with respect to a scalar

Resources

- https://www.youtube.com/watch?v=OZRsegSgqy0

Resources

- https://vedadian.com/matrices-and-diffrentiation-2/
- https://www.youtube.com/watch?v=6Vub-tiPhlI

Derivative of a trace with respect to a matrix

- https://www.youtube.com/watch?v=9fc-kdSRE7Y