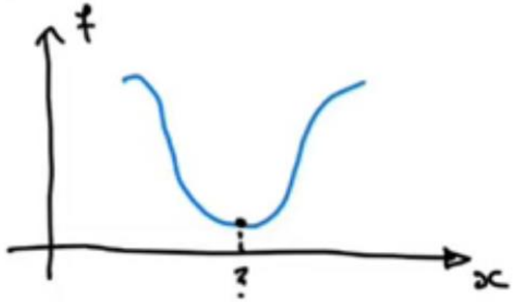
Vector, Matrix, Tensor Derivatives

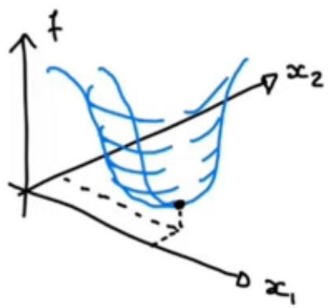
CE40282-1: Linear Algebra
Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

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?



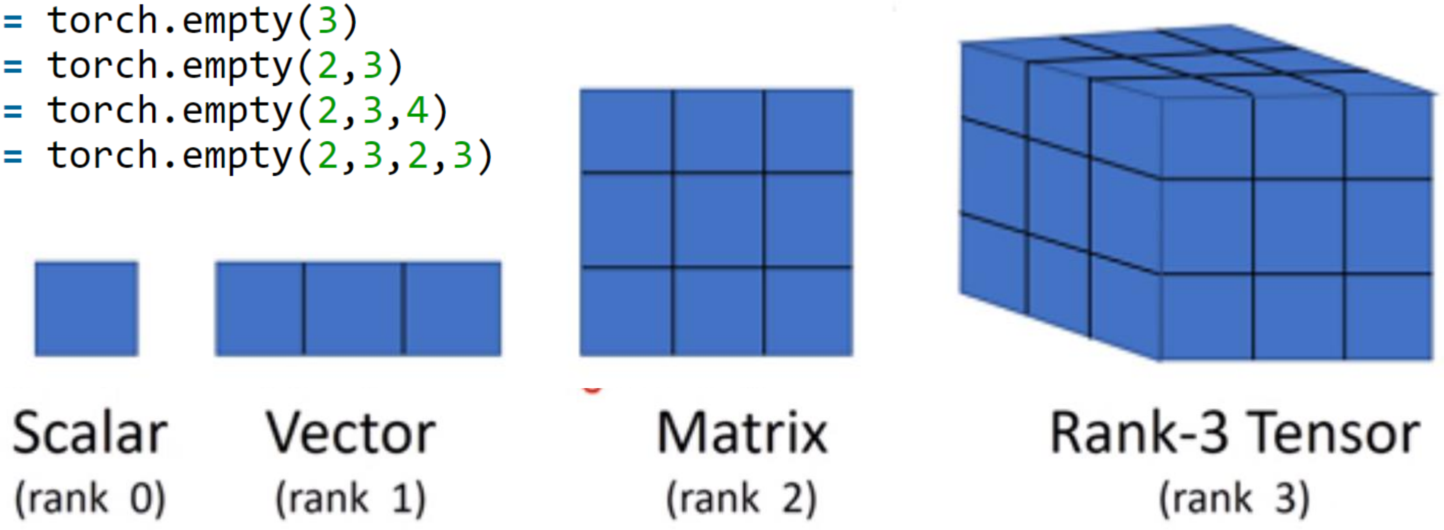
Good Resource

- http://en.wikipedia.org/wiki/Matrix_calculus
- https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf
- http://www.kamperh.com/notes/kamper_matrixcalculus13.pdf

Tensor

- Multi-dimensional array of numbers

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



Definitions

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

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

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

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

Definitions

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

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

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

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

Scalar and vectors

Vectors and vectors

Matrices and vectors

Another View

- ## Finding the Derivative

  To find $f'(x)$, we use a four-step process:

  Step 1. Find $f(x+h)$

  Step 2. Find $f(x+h)-f(x)$

  Step 3. Find $\dfrac{f(x+h)-f(x)}{h}$

  Step 4. Find $\lim\limits_{h\to 0} \dfrac{f(x+h)-f(x)}{h}$

  $h = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \to h_i$

  vector $\lim\limits_{h \to 0} h = \lim\limits_{h_i \to 0} h_i$

- Example: find the derivation of quadratic form

Conclusion

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

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}}=\mathbf{A}^{\top}$$

$$\frac{\partial \mathbf{x}^{\top}\mathbf{a}}{\partial \mathbf{x}}=\mathbf{a}$$

$$\frac{\partial \mathbf{x}^{\top}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}}=(\mathbf{A}+\mathbf{A}^{\top})\mathbf{x}$$

$$\frac{\partial \mathbf{x}^{\top}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}}=2\mathbf{A}\mathbf{x} \text{ if } \mathbf{A} \text{ is symmetric}$$

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}}=|\mathbf{X}|(\mathbf{X}^{-1})^{\top}$$

$$\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}}=(\mathbf{X}^{-1})^{\top}$$

Hint!

$$Ax = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} a_1x_1 + a_2x_2 \\ a_3x_1 + a_4x_2 \end{bmatrix}$$

$$\frac{dAx}{dx} \overset{\text{ژاکوبین}}{=} \begin{bmatrix} \frac{\partial(a_1x_1 + a_2x_2)}{\partial x_1} & \frac{\partial(a_3x_1 + a_4x_2)}{\partial x_1} \\ \frac{\partial(a_1x_1 + a_2x_2)}{\partial x_2} & \frac{\partial(a_3x_1 + a_4x_2)}{\partial x_2} \end{bmatrix}$$

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

Notes

- $$\frac{\partial (y^T A x)}{\partial A} = y x^T$$

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

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

Derivative of a vector with respect to a matrix

Derivative of a matrix with respect to a matrix

Conclusion

1. Derivative of a linear function:

$$\frac{\partial}{\partial \vec{x}} \vec{a} \cdot \vec{x} \quad = \quad \frac{\partial}{\partial \vec{x}} \vec{a}^\top \vec{x} \quad = \quad \frac{\partial}{\partial \vec{x}} \vec{x}^\top \vec{a} \quad = \quad \vec{a}$$

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

2. Derivative of a quadratic function: if A is symmetric, then

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

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

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

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

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

Derivative of matrix inverse with respect to a scalar

Derivative of a Determinant with respect to a Matrix