Matrix Calculus

easy topic

- The derivative of f with respect to x is $\frac{\partial f}{\partial x}$.

$\frac{df}{dx}$

- Both x and f can be:
  - Scalar
  - Vector
  - Matrix

- This leads to 9 types of derivatives.

Calculus

- The derivative of f with respect to x is $\frac{\partial f}{\partial x}$.

$\frac{df}{dx}$

- Both x and f can be:
  - Scalar
  - Vector
  - Matrix

$\frac{\partial f}{\partial x}$

Why it is required?

loss function

MSE

$\underbrace{w}_{\text{vector}}^T x$ } scalar

- This leads to 9 types of derivatives.

$f(w) = (w^T x - y)^2$

$\frac{\partial f}{\partial w}$

scalar function

GO CLASSES

why we need matrix calculus?

> in machine learning, we will have loss functions

eg: $f(w) = \underbrace{(\overset{\text{scalar}}{y} - \underset{\text{vector}}{w^T}\underset{\text{vector}}{x})^2}$ } MSE

linear regression

$\frac{\partial f}{\partial w} = 0$

(you need to differenticate f w.r.t w)

scalar vector

Type $y \rightarrow$ $x \downarrow$	Scalar	Vector	Matrix
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{Y}}{\partial x}$
Vector	$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\frac{\partial y}{\partial \mathbf{X}}$		

Notation

- Matrix: $\mathbf{A}, \mathbf{X}, \mathbf{Y}$
  - bold capital letter
- Vector: $\mathbf{a}, \mathbf{x}, \mathbf{y}$ (**column**)
  - boldface lowercase letter
- Scalar: a, x, y
  - lowercase itatilc typeface

Good news: most of the rules you know and love from single variable calculus generalize well.

Scalar-by-Scalar

$$\frac{dy}{dx}$$

Scalar

Scalar

we already know this

e.g.

$$\frac{dx^2}{dx} = 2x$$

Scalar-by-Vector

- The derivative of y by $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is written as:

$y = x_1^2 + x_2^2$

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\frac{\partial y}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$$

Calculus

Scalar

$$\frac{\partial \mathrm{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \left[\frac{\partial y}{\partial x_1} \quad \frac{\partial y}{\partial x_2} \quad \cdots \quad \frac{\partial y}{\partial x_n}\right]$$

Numerator vs Denominator Layout

- Two main conventions used in vector/matrix calculus: numerator and denominator layouts.
- Numerator layout makes the dimension of the derivative be the numerator dimension by denominator dimension.
- For example, if y is a scalar and $x \in \mathbb{R}^N$ then

$$\frac{\partial y}{\partial x_n} \in \mathbb{R}^{1 \times N}$$

$\frac{\partial y}{\partial x}$ ← vector, ← scalar, $1 \times n$

Denominator layout

Denominator dim first

$\frac{\partial y}{\partial x} \in R^{N \times 1}$

Calculus

Numerator-Layout Notation	Denominator-Layout Notation
scalar $\frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1} \frac{\partial y}{\partial x_2} \cdots \frac{\partial y}{\partial x_n}\right]$ vector	scalar $\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$ vector
vector $\frac{\partial \mathbf{y}}{\partial \mathrm{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$ scalar	vector $\frac{\partial \mathbf{y}}{\partial \mathrm{x}} = \left[\frac{\partial y_1}{\partial x} \frac{\partial y_2}{\partial x} \cdots \frac{\partial y_m}{\partial x}\right],$ scalar

Numerator-Layout Notation	Denominator-Layout Notation
Scalar $\dfrac{\partial y^{1}}{\partial \mathbf{x}_{n}} = \left[\dfrac{\partial y}{\partial x_1}\,\dfrac{\partial y}{\partial x_2}\cdots\dfrac{\partial y}{\partial x_n}\right]$ $1\times n$ vector	Transpose Scalar $\dfrac{\partial y^{1}}{\partial \mathbf{x}_{n}} = \begin{bmatrix}\dfrac{\partial y}{\partial x_1}\\ \dfrac{\partial y}{\partial x_2}\\ \vdots \\ \dfrac{\partial y}{\partial x_n}\end{bmatrix}$ $n\times 1$ vector
vector $\dfrac{\partial \mathbf{y}^{m}}{\partial \mathrm{x}_{1}} = \begin{bmatrix}\dfrac{\partial y_1}{\partial x}\\ \dfrac{\partial y_2}{\partial x}\\ \vdots \\ \dfrac{\partial y_m}{\partial x}\end{bmatrix}$ $m\times 1$ Scalar	Transpose vector $\dfrac{\partial \mathbf{y}^{m}}{\partial \mathrm{x}_{1}} = \left[\dfrac{\partial y_1}{\partial x}\,\dfrac{\partial y_2}{\partial x}\cdots\dfrac{\partial y_m}{\partial x}\right],$ $1\times m$ Scalar

This lecture follows numerator layout convention. There is an alternative denominator layout convention, where several results are transposed. Do not mix different layout conventions.

https://souryadey.github.io/teaching/material/Matrix_Calculus.pdf

Question:

Consider the expression $\mathbf{x}^T\mathbf{b}$ where $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$.

What is the correct differentiation of $\mathbf{x}^T\mathbf{b}$ with respect to $\mathbf{x}$?

Scalar

vector

https://www.sfu.ca/~haiyunc/notes/matrix_calculus.pdf

Question:

Consider the expression $\mathbf{x}^T\mathbf{b}$ where $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$.

What is the correct differentiation of $\mathbf{x}^T\mathbf{b}$ with respect to $\mathbf{x}$?

Scalar

vector

$$\frac{\partial x^T b}{\partial x} = \left[\frac{\partial x^T b}{\partial x_1} \quad \frac{\partial x^T b}{\partial x_2} \quad \cdots \quad \frac{\partial x^T b}{\partial x_n} \right]$$

$$x^T b = x_1 b_1 + x_2 b_2 + \cdots x_n b_n$$

$$\frac{\partial x^T b}{\partial x} = \left[b_1 \quad b_2 \quad b_3 \quad \cdots \quad b_n \right] = b^T.$$

Ajay Sreenivas to Everyone 6:45 PM

AS

[b1 b2 b3 ....bn]

Aditya to Everyone 6:46 PM

A

b^t

Solution:

Given:

$$\mathbf{x}^T\mathbf{b} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b_1x_1 + \cdots + b_nx_n$$

The differentiation is:

$$\begin{aligned} \frac{\partial \mathbf{x}^T\mathbf{b}}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial \mathbf{x}^T\mathbf{b}}{\partial x_1} & \cdots & \frac{\partial \mathbf{x}^T\mathbf{b}}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}(b_1x_1 + \cdots + b_nx_n) & \cdots & \frac{\partial}{\partial x_n}(b_1x_1 + \cdots + b_nx_n) \end{bmatrix} \\ &= \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} \\ &= \mathbf{b}^T \end{aligned}$$

Question:

Consider the expression $\mathbf{x}^T\mathbf{x}$ where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

What is the differentiation of $\mathbf{x}^T\mathbf{x}$ with respect to $\mathbf{x}$?

Question:

Consider the expression $\mathbf{x}^T\mathbf{x}$ where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

What is the differentiation of $\underbrace{\mathbf{x}^T\mathbf{x}}_{\text{Scalar}}$ with respect to $\mathbf{x}$?

$$x^T x = x_1^2 + x_2^2$$

$$\frac{\partial x^T x}{\partial x} = \begin{bmatrix} \frac{\partial x^T x}{\partial x_1} & \frac{\partial x^T x}{\partial x_2} \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}$$

$$= 2x^T$$

Solution:

Given $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2$

$$\frac{\partial(\mathbf{x}^T\mathbf{x})}{\partial \mathbf{x}} = \left[\frac{\partial(x_1^2 + x_2^2)}{\partial x_1} \; \frac{\partial(x_1^2 + x_2^2)}{\partial x_2}\right]$$

$$\frac{\partial(\mathbf{x}^T\mathbf{x})}{\partial \mathbf{x}} = [2x_1 \; 2x_2] = 2\,[x_1 \; x_2] = 2\mathbf{x}^T$$

Derivative of vector w.r.t. scalar

If x is a scalar,

$$\frac{\partial \mathbf{y}}{\partial x} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}_{m \times 1}$$

GO CLASSES

Question :

$$y = \begin{bmatrix} 2x \\ 5x^2 \\ -x \end{bmatrix}$$

$x \leftarrow$ scalar

$$\frac{\partial y}{\partial x} = ?$$

Question :

$$y = \begin{bmatrix} 2x \\ 5x^2 \\ -x \end{bmatrix}$$

$x \leftarrow$ scalar

$$\frac{\partial y}{\partial x} = \begin{bmatrix} 2 \\ 10x \\ -1 \end{bmatrix}$$

Derivative of vector w.r.t. vector

where $\mathbf{x}$ and $\mathbf{y}$ are vectors

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

Derivative of vector w.r.t. vector

where $\mathbf{x}$ and $\mathbf{y}$ are vectors

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}_{n \times m}$$

n

m

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

Question:

vector

Consider the expression $\mathbf{Ax}$ where $\mathbf{A}$ is an $m \times n$ matrix that is not a function of $\mathbf{x}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

What is the differentiation of $\mathbf{Ax}$ with respect to $\mathbf{x}$?

$$\frac{\partial Ax}{\partial x}$$

← vector

← vector

Solution:

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{Ax} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

$$\frac{\partial Ax}{\partial x} = \begin{bmatrix} \frac{\partial (a_{11}x_1 + a_{12}x_2)}{\partial x_1} & \frac{\partial (a_{11}x_1 + a_{12}x_2)}{\partial x_2} \\ \frac{\partial (a_{21}x_1 + a_{22}x_2)}{\partial x_1} & \frac{\partial (a_{21}x_1 + a_{22}x_2)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$= A$$

Solution:

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{Ax} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

$$\frac{\partial(\mathbf{Ax})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial(a_{11}x_1+a_{12}x_2)}{\partial x_1} & \frac{\partial(a_{11}x_1+a_{12}x_2)}{\partial x_2} \\ \frac{\partial(a_{21}x_1+a_{22}x_2)}{\partial x_1} & \frac{\partial(a_{21}x_1+a_{22}x_2)}{\partial x_2} \end{bmatrix}$$

$$\frac{\partial(\mathbf{Ax})}{\partial \mathbf{x}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}$$

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathbf{Ax} = \begin{bmatrix} \sum_{i=1}^{n} a_{1i}x_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}x_i \end{bmatrix}$$

$$\frac{\partial(\mathbf{Ax})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial(\sum_{i=1}^{n} a_{1i}x_i)}{\partial x_1} & \cdots & \frac{\partial(\sum_{i=1}^{n} a_{1i}x_i)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial(\sum_{i=1}^{n} a_{mi}x_i)}{\partial x_1} & \cdots & \frac{\partial(\sum_{i=1}^{n} a_{mi}x_i)}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial(\mathbf{Ax})}{\partial \mathbf{x}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \mathbf{A}$$

Question:

Consider the expression $\mathbf{A}^T\mathbf{x}$ where $\mathbf{A}$ is an $n \times m$ matrix that is not a function of $\mathbf{x}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

What is the differentiation of $\mathbf{A}^T\mathbf{x}$ with respect to $\mathbf{x}$?

Question:

Consider the expression $\mathbf{A}^T\mathbf{x}$ where $\mathbf{A}$ is an $n \times m$ matrix that is not a function of $\mathbf{x}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

What is the differentiation of $\mathbf{A}^T\mathbf{x}$ with respect to $\mathbf{x}$?

Answer: A^T

Question:

Consider the expression $\mathbf{x}^T\mathbf{A}$ where $\mathbf{A}$ is an $n \times m$ matrix that is not a function of $\mathbf{x}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

What is the differentiation of $\mathbf{x}^T\mathbf{A}$ with respect to $\mathbf{x}$?

Solution:

$$\mathbf{x}^T\mathbf{A} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{A} = \begin{bmatrix} x_1a_{11} + x_2a_{21} \\ x_1a_{12} + x_2a_{22} \end{bmatrix}$$

$$\frac{\partial x^T A}{\partial x} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = A^T$$

$$\frac{\partial(\cdot)}{\partial x_1} \qquad \frac{\partial(\ )}{\partial x_2}$$

Question:

Given the vectors

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

and the equations

$$y_1 = x_1^2 - x_2$$
$$y_2 = x_3^2 + 3x_2$$

the partial derivative matrix $\partial \mathbf{y} / \partial \mathbf{x}$ is:

Question:

Given the vectors

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

and the equations

$$y_1 = x_1^2 - x_2$$
$$y_2 = x_3^2 + 3x_2$$

$\frac{\partial y}{\partial x_1}$

the partial derivative matrix $\partial \mathbf{y} / \partial \mathbf{x}$ is:

$$\frac{\partial y}{\partial x} = \begin{bmatrix} 2x_1 & -1 & 0 \\ 0 & 3 & 2x_3 \end{bmatrix}$$

$\frac{\partial y}{\partial x_3}$

Solution:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \end{bmatrix}$$

$$y_1 = x_1^2 - x_2$$
$$y_2 = x_3^2 + 3x_2$$

For y_1:

$$\frac{\partial y_1}{\partial x_1} = 2x_1, \quad \frac{\partial y_1}{\partial x_2} = -1, \quad \frac{\partial y_1}{\partial x_3} = 0$$

For y_2:

$$\frac{\partial y_2}{\partial x_1} = 0, \quad \frac{\partial y_2}{\partial x_2} = 3, \quad \frac{\partial y_2}{\partial x_3} = 2x_3$$

Partial derivative matrix $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$:

$$\begin{bmatrix} 2x_1 & -1 & 0 \\ 0 & 3 & 2x_3 \end{bmatrix}$$

Handy Derivatives

$\mathbf{y}$	$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{Ax}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}^T$
$\mathbf{x}^T\mathbf{Ax}$	$\mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T$

http://media.ee.ntu.edu.tw/personal/pcwu/tutorials/matrix_calculus.pdf

Handy Derivatives

$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{Ax}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}^T$
$\mathbf{x}^T\mathbf{Ax}$	$\mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T$

$$\frac{\partial\, x^T x}{\partial x} = \frac{\partial x^T}{\partial x} \cdot x^T + x^T \cdot \frac{\partial x}{\partial x}$$

$$= x^T + x^T$$

$$= 2x^T$$

$$\frac{\partial\, x^T A x}{\partial x} = \frac{\partial x^T}{\partial x} \underline{(Ax)^T} + x^T A \frac{\partial x}{\partial x}$$

$$= x^T A^T + x^T A$$

http://media.ee.ntu.edu.tw/personal/pcwu/tutorials/matrix_calculus.pdf

Handy Derivatives

$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{Ax}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}^T$
$\mathbf{x}^T\mathbf{Ax}$	$\mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T$

- Hint: Derive $\mathbf{x}$
  - If you have to differentiate $\mathbf{x}^T$, transpose the rest.
  - If you have two $\mathbf{x}$-terms, differentiate them separately in turn and then sum up the two derivatives.

MSQ Question:

Let the scalar α be defined by

$$\alpha = \mathbf{y}^\top \mathbf{A}\mathbf{x}$$

where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ is independent of $\mathbf{x}$ and $\mathbf{y}$. Then, which of the following are correct? (use numerator layout)

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

B. $\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^\top \mathbf{A}^\top$

C. $\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{y}$

D. $\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{A}\mathbf{x}$

https://atmos.washington.edu/~dennis/MatrixCalculus.pdf

MSQ Question:

Let the scalar α be defined by

$$\alpha = \mathbf{y}^\top \mathbf{A}\mathbf{x}$$

where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ is independent of $\mathbf{x}$ and $\mathbf{y}$.

Then, which of the following are correct? (use numerator layout)

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

B. $\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^\top \mathbf{A}^\top$

C. $\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{y}$

D. $\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{A}\mathbf{x}$

$$\frac{\partial \alpha}{\partial x} = y^T A$$

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

https://atmos.washington.edu/~dennis/MatrixCalculus.pdf

Proof: *Define*

$$\mathbf{w}^T = \mathbf{y}^T \mathbf{A}$$

and note that

$$\alpha = \mathbf{w}^T \mathbf{x}$$

Hence, by Proposition 5 we have that

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{w}^T = \mathbf{y}^T \mathbf{A}$$

which is the first result. Since α *is a scalar, we can write*

$$\alpha = \alpha^T = \mathbf{x}^T \mathbf{A}^T \mathbf{y}$$

and applying Proposition 5 as before we obtain

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^T \mathbf{A}^T$$

q.e.d.

$$\underbrace{y^T A}_{w^T} x$$

$$x^T A^T y$$

Handy Matrix Differentiation Rules

- $y = Ax$ where A does not depend on x.

$$\frac{\partial y}{\partial x} = A$$

- $\alpha = y^T Ax$ where A does not depend on x or y.

$$\frac{\partial \alpha}{\partial x} = y^T A, \frac{\partial \alpha}{\partial y} = x^T A^T$$

- $y = A^T xB$ where A and B do not depend on x.

$$\frac{\partial y}{\partial x} = AB^T$$

http://www.yisongyue.com/courses/cs155/2017_winter/lectures/recitation2.pdf

Question:

Consider the function $f(\mathbf{w}) = \left(y - \mathbf{w}^T\mathbf{x}\right)^2$. $= (y - w^T x)^T (y - w^T x)$

Find $\frac{\partial f}{\partial \mathbf{w}}$, where $\mathbf{w}$ and $\mathbf{x}$ are vectors, and y is a scalar.

A) $\frac{\partial f}{\partial \mathbf{w}} = -2y\mathbf{x}^T$

B) $\frac{\partial f}{\partial \mathbf{w}} = 2y\mathbf{x}^T - 2\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}^T$

C) $\frac{\partial f}{\partial \mathbf{w}} = -2y\mathbf{x} + 2\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}$

D) $\frac{\partial f}{\partial \mathbf{w}} = -2y\mathbf{x}^T + 2\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}^T$

$(y - w^T x)^T (y - w^T x)$

$= (y^T - x^T w)(y - w^T x)$

$= y^2 - y^T w^T x - x^T w y$
$+ x^T w \cdot w^T x$

https://souryadey.github.io/teaching/material/Matrix_Calculus.pdf

Question:

Consider the function $f(\mathbf{w}) = \left(y - \mathbf{w}^T\mathbf{x}\right)^2$. $= (y - w^T x)^T (y - w^T x)$

Find $\frac{\partial f}{\partial \mathbf{w}}$, where $\mathbf{w}$ and $\mathbf{x}$ are vectors, and y is a scalar.

A) $\frac{\partial f}{\partial \mathbf{w}} = -2y\mathbf{x}^T$

B) $\frac{\partial f}{\partial \mathbf{w}} = 2y\mathbf{x}^T - 2\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}^T$

C) $\frac{\partial f}{\partial \mathbf{w}} = -2y\mathbf{x} + 2\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}$

D) $\frac{\partial f}{\partial \mathbf{w}} = -2y\mathbf{x}^T + 2\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}^T$

$$(y - w^T x)^T (y - w^T x)$$

$$2(y - w^T x) \cdot (-x^T)$$

$$2(y - w^T x) \cdot (-x^T) = -2y x^T + 2 (w^T x) x^T$$

Examples (Least Squares)

- Real valued function $f : \mathbb{R}^n \to \mathbb{R}$ where for $x \in \mathbb{R}^n$,

$$f(x) = ||Ax - b||_2^2 = (Ax - b)^T(Ax - b)$$

$$= x^TA^TAx - b^TAx - x^TA^Tb + b^Tb$$

- To find the x that minimizes f, we need to compute the gradient with respect to x.

$$\nabla_x f(x) = \nabla_x(x^TA^TAx - b^TAx - x^TA^Tb + b^Tb)$$

$$= \nabla_x(x^TA^TAx) - \nabla_x(b^TAx) - \nabla_x(x^TA^Tb) + \nabla_x(b^Tb)$$

$$= \mathbf{2A^TAx - 2A^Tb}$$

http://www.yisongyue.com/courses/cs155/2017_winter/lectures/recitation2.pdf

Next topic :

Positive Definite Matrices

PD

PSD

Semi