Week 2 Recitation

Krushna Shah
CS 190I Deep Learning

UC SANTA BARBARA

UC SANTA BARBARA
Computer Science

Outline

- **Recap of Linear Algebra Concepts**
- **Basics of Vector Calculus**
- **Homework 1 Question 4**

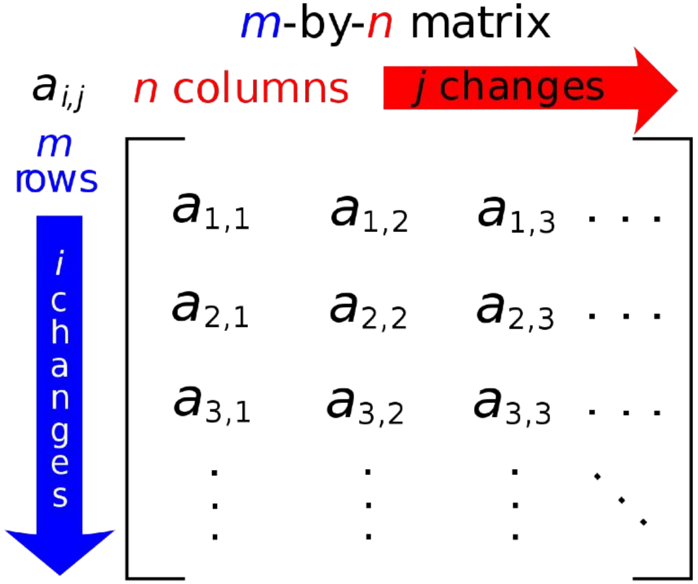
Scalars and Vectors

- *Scalars*: A scalar is just a single number. Example: 5, 10, 15
- *Vectors*: A vector is an ordered array of numbers. We can identify each individual number by its index in that ordering. Example:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrices and Tensors

- *Matrices*: A matrix is a rectangular array of numbers, and we can identify each number using its *row and column indices*.



- *Tensors*: A tensor is like a high-dimensional matrix that can be indexed similarly. For example, the element at (i, j, k) coordinate of a 3D tensor is denoted by $\mathbf{A}_{i,j,k}$.

Matrix and Vector Operations

- *Matrix addition:* Matrices can be added as long as their shapes match

$$C = A + B, \text{ where } C_{i,j} = A_{i,j} + B_{i,j}$$

- *Scalar multiplication and addition:* Scalars can be multiplied and added to each element of a matrix

$$D = a \cdot B + c, \text{ where } D_{i,j} = a \cdot B_{i,j} + c$$

- *Broadcasting:* Vectors can be added to matrices (shapes must match)

$$C = A + B, \text{ where } C_{i,j} = A_{i,j} + B_j$$

The vector gets added to every row in .

Matrix and Vector operations

- *Dot product* of two vectors $x^T y = y^T x = \Sigma_k x_k y_k$
- *Product of two matrices* A is defined as

$$C_{i,j} = \Sigma_k A_{i,k} B_{k,j}$$

$C_{i,j}$ is the *dot product* of the i th row of A and jth column of B .
Number of columns in A must match number of rows in B

- *Distributive law:* $A(B + C) = AB + AC$
- *Associativity:* $A(BC) = (AB)C$
- *Commutativity* does not always hold: $AB \neq BA$

Matrix and vector Operations

- *Transpose* of a matrix is obtained by “flipping” along the diagonal.

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow A^\top = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

- *Transpose of a product:*

$$(AB)^T = B^T A^T$$

Some Special Matrices

- *A square matrix* has the same number of rows and columns. *The identity matrix* is a square matrix with 1s along the diagonal:

$$I_1 = [1],\ I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \ldots,\ I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

- *The inverse of a square matrix* is a matrix A^{-1} that satisfies:

$$AA^{-1} = A^{-1}A = I$$

Norms

- *Norm* is a function that intuitively measures the size of a vector.

- L^1 norm : $\|x\|_1 = \Sigma_i |x_i|$

- L^2 norm : $\|x\|_2 = \sqrt{\Sigma_i |x_i|^2}$
- L^∞ norm : $\|x\|_\infty = \max_i |x_i|$. Also known as the max norm.

Calculus

- Derivative of Sums

$$y = u + v$$
$$\frac{\partial y}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}$$

- Product Rule

$$\frac{d(uv)}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}$$

- Chain Rule

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

A Brief Note about Numerator and Denominator Layout

	Numerator Layout	**Denominator Layout**
$\frac{\partial y}{\partial \mathbf{x}}$	1-D row vector	1-D column vector
$\frac{\partial \mathbf{y}}{\partial x}$	1-D column vector	1-D row vector
$\frac{\partial \mathbf{a}^T \mathbf{z}}{\partial \mathbf{z}}$	$\mathbf{a}^\mathbf{T}$	$\mathbf{a}$
$\frac{\partial \mathbf{M}\mathbf{z}}{\partial \mathbf{z}}$	$\mathbf{M}$	$\mathbf{M}^\mathbf{T}$

Numerator Layout

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad \frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

d**y**/dx is a column vector

Numerator Layout

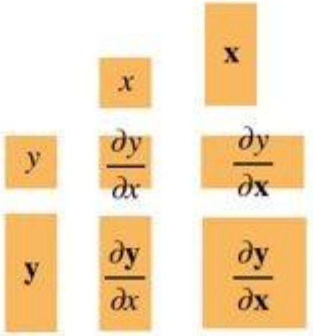
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_n}\right]$$

dy/d**x** is a row vector

Numerator Layout

$$\mathbf{x} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \in \mathbb{R}^{m \times n}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$



$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} \\ \frac{\partial y_2}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial y_m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}, \frac{\partial y_1}{\partial x_2}, \ldots, \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1}, \frac{\partial y_2}{\partial x_2}, \ldots, \frac{\partial y_2}{\partial x_n} \\ \vdots \\ \frac{\partial y_m}{\partial x_1}, \frac{\partial y_m}{\partial x_2}, \ldots, \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

Differentiation Rules for Matrices and Vectors:

The following rules for vector and matrix differentiation are good to remember. Note here that a and z are vectors and M is a matrix.

1. $\frac{\partial \mathbf{a}^T \mathbf{z}}{\partial \mathbf{z}} = \mathbf{a}$
2. $\frac{\partial M \mathbf{z}}{\partial \mathbf{z}} = M^T$
3. $\frac{\partial \mathbf{z}^T M \mathbf{z}}{\partial \mathbf{z}} = (M + M^T)\mathbf{z}$

Example 1

$$y = x_1^2 + 2x_2^2$$

$$\frac{\partial y}{\partial \mathbf{x}}$$

Example 1

$$y = x_1^2 + 2x_2^2$$

$$\frac{\partial y}{\partial \mathbf{x}} = [2x_1, 4x_2]$$

$$y = x_1^2 + 2x_2^2$$

$$\frac{dy}{dx} = \left[\frac{dy}{dx_1}, \frac{dy}{dx_2} \right]$$

$$= \left[\frac{d(x_1^2 + 2x_2^2)}{dx_1}, \frac{d(x_1^2 + 2x_2^2)}{dx_2} \right]$$

$$= [2x_1, 4x_2]$$

Example 2

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

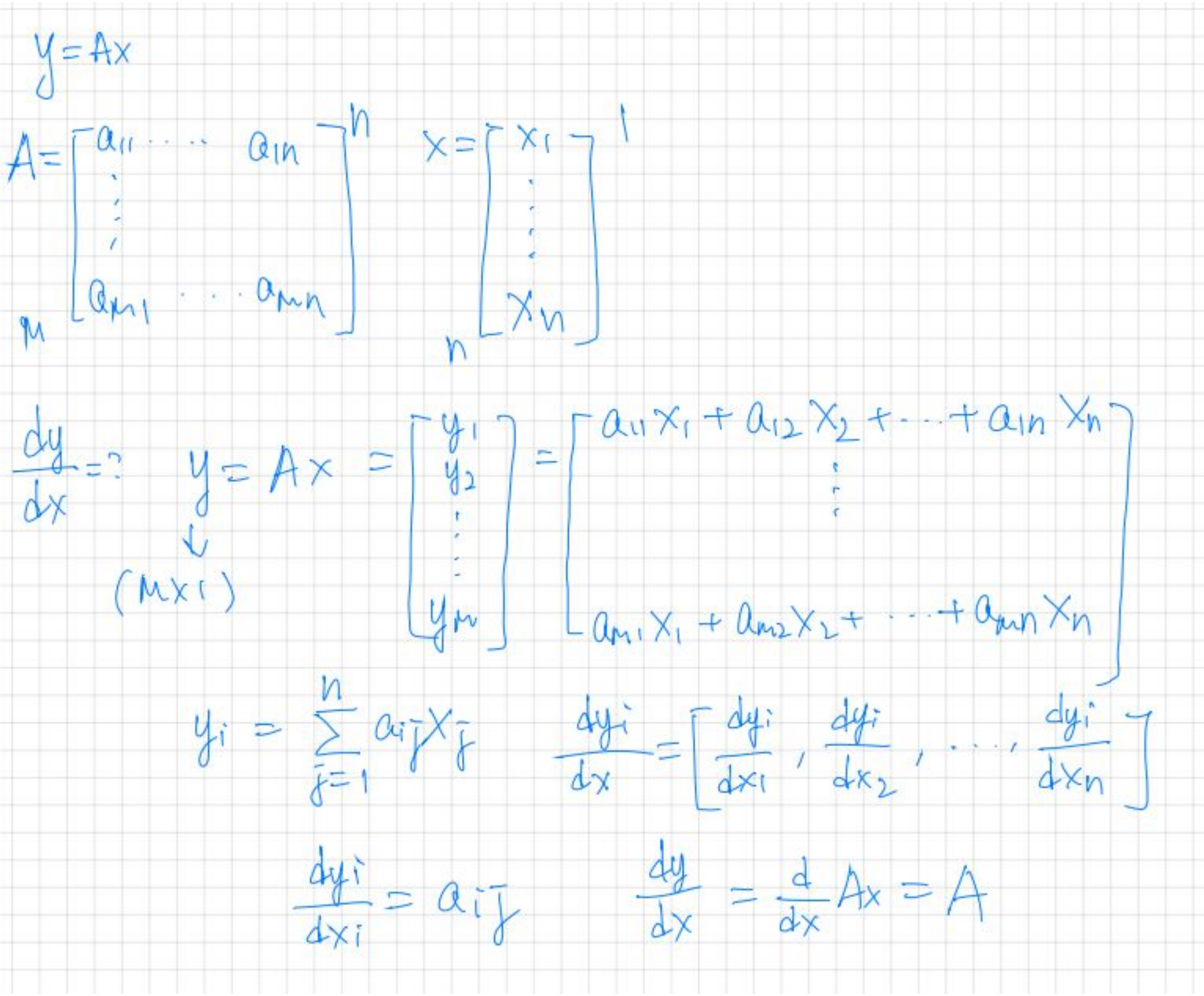
Example 2

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$y_i = \sum_{k=1}^{n} a_{ik} x_k$$

$$\frac{\partial y_i}{\partial x_j} = a_{ij}$$

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



Example 3

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

Example 3

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

$$y_i = \sum_{k=1}^{n} x_k a_{ki}$$

$$\frac{\partial y_i}{\partial x_j} = a_{ji}$$

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

$$y = x^T A \qquad \frac{dy}{dx} = ?$$

$$x^T = [x_1 \cdots x_m] \quad \downarrow (1 \times m)$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & & a_{mn} \end{bmatrix} \quad \downarrow (m \times n)$$

$$y = x^T A = [a_{11}x_1 + \cdots + a_{m1}x_m, \cdots, a_{1n}x_1 + \cdots + a_{mn}x_m]$$

$$\downarrow (1 \times n)$$

$$y_i = \sum_{j=1}^{m} a_{ji} x_j \qquad \frac{dy_i}{dx_j} = a_{ji}$$

$$\frac{dy}{dx} = \frac{d}{dx} x^T A = A^T$$

Example 4

$$y = x^T A x$$

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

Example 4

$$y = x^T A x$$

$$b = Ax$$

$$y = x^T b$$

$$\frac{\partial y}{\partial x} = b^T \frac{\partial x}{\partial x} + x^T \frac{\partial b}{\partial x}$$

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

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

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

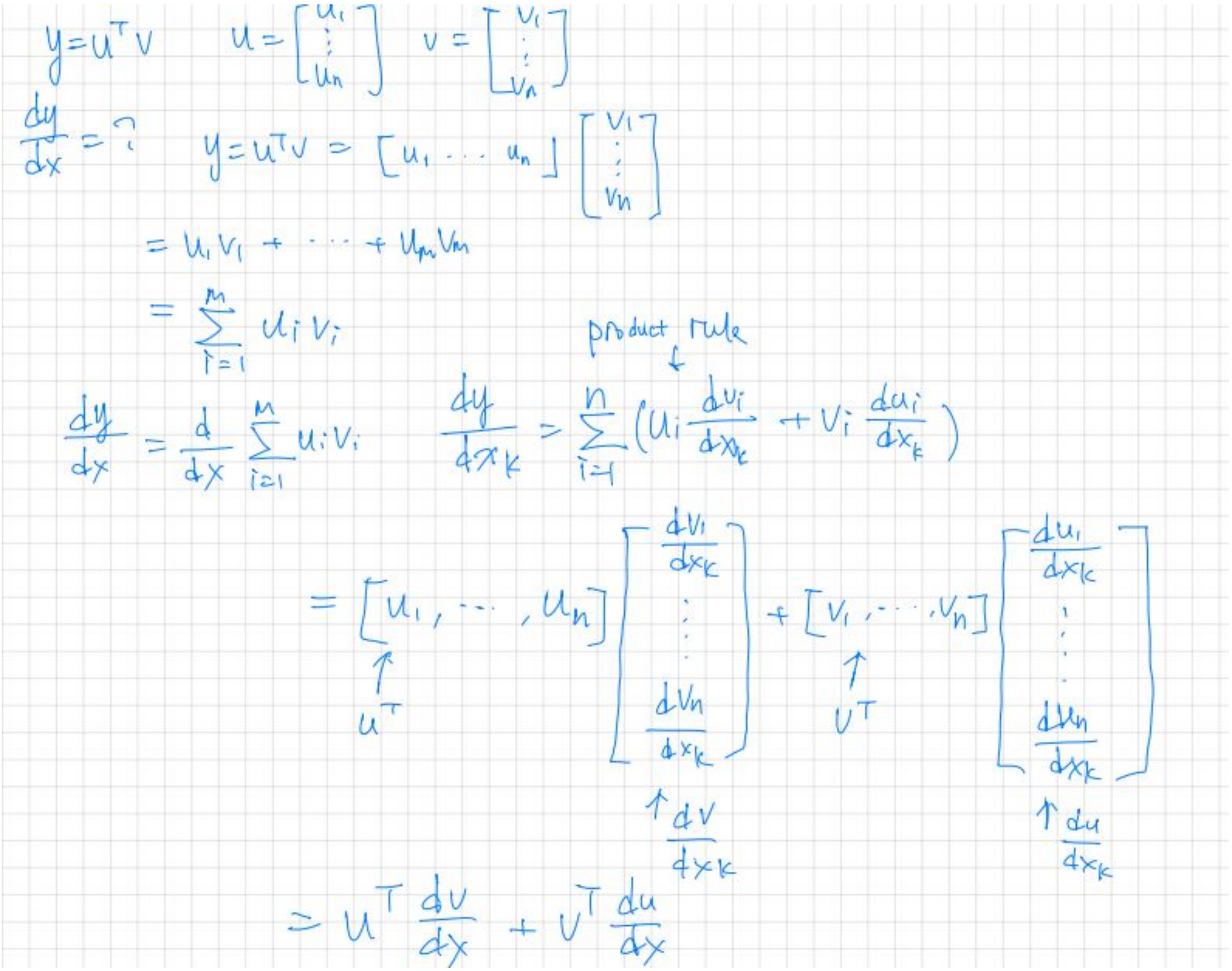
Example 5

$$y = \mathbf{u}^T \mathbf{v}$$

$$y = \sum_{i=1}^{n} u_i v_i$$

$$\frac{\partial y}{\partial x_j} = \sum_{i=1}^{n} \left(v_i \frac{\partial u_i}{\partial x_j} + u_i \frac{\partial v_i}{\partial x_j} \right)$$

$$\frac{\partial y}{\partial \mathbf{x}} = \mathbf{v}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}^T \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$$



Example 4

$$y = \mathbf{u}^T \mathbf{v}$$

$$y = \sum_{i=1}^{n} u_i v_i$$

$$\frac{\partial y}{\partial x_j} = \sum_{i=1}^{n} \left(v_i \frac{\partial u_i}{\partial x_j} + u_i \frac{\partial v_i}{\partial x_j} \right)$$

$$\frac{\partial y}{\partial \mathbf{x}} = \mathbf{v}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}^T \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$$

$$\frac{\partial y}{\partial \mathbf{x}} = \frac{\partial y}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial y}{\partial \mathbf{v}} \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$$

$$\frac{\partial y}{\partial \mathbf{x}} = \mathbf{v}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}^T \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$$

Homework 1 Problem 4

Suppose x is a 3-d vector.

$$f(x) = |e^{A\cdot x+b} - c|_2^2$$

where

$$A = \begin{bmatrix} 2 & -2 & 3 \\ 3 & 2.5 & -1 \end{bmatrix}, b = \begin{bmatrix} -2 \\ -3 \end{bmatrix}, c = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$$

$|\cdot|_2$ is 2-norm: $|x|_2 = \sqrt{x_1^2 + x_2^2 + \ldots}$.

What is the differential $\frac{\partial f}{\partial x}$?

Reference for Basics of PyTorch

https://towardsdatascience.com/understanding-pytorch-with-an-example-a-step-by-step-tutorial-81fc5f8c4e8e#3a3f

Any Questions?