03 Matrix Addition , subtraction & Trace

Monower hossen

1. Matrix Addition

Definition

Two matrices can be **added** if they have the **same dimensions**.

$$A + B = C \text{ where } A, B \in \mathbb{R}^{m\times n}$$

Rule

Add **corresponding elements**:

$$C_{ij} = A_{ij} + B_{ij}$$

Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 1 + 5 & 6 + 2 \\ 3 + 7 & 4 + 8 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

ML & DS Usage

- Combining feature matrices
- Accumulating gradients
- Updating model parameters:

$$W_{\text{new}} = W - \eta \nabla W$$

2. Matrix Subtraction

Definition

Two matrices can be **subtracted** if they have the **same dimensions**.

$$A - B = C$$

Rule

Subtract **corresponding elements**:

$$C_{ij} = A_{ij} - B_{ij}$$

Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 1\text{-}5 & 2\text{-}6 \\ 3\text{-}7 & 4\text{-}8 \end{bmatrix}$$

$$A - B = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$$

ML & DS Usage

- Error calculation:

$$\text{Error} = Y_{\text{true}} - Y_{\text{pred}}$$

- Residuals in regression
- Gradient descent updates

3. What is Scalar Broadcasting?

Scalar broadcasting means applying a **single scalar value** (number) to **every element** of a vector or matrix **without explicitly copying the scalar**.
Mathematically:

$$A \in \mathbb{R}^{m \times n}, c \in \mathbb{R}$$

$$A + c \;\; \Rightarrow \;\; A_{ij} + c \forall i, j$$

3.1. Simple Example

Matrix + Scalar

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$A + 5 = \begin{bmatrix} 1+5 & 2+5 \\ 3+5 & 4+5 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix}$$

The scalar **5** is *broadcast* to every element.

3.2. Why Broadcasting Exists

Without broadcasting, we would need to create a matrix like:

$$\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$$

Broadcasting:

- Saves **memory**
- Improves **speed**
- Simplifies **code**

3.3. Broadcasting Rules (Core Intuition)

A scalar:

- Has **no dimensions**
- Can be broadcast to **any shape**

So operations like:

- $A + c$
- $A - c$
- $A \times c$
- A / c

are always valid.

3.4. Scalar Broadcasting in Machine Learning

1. Feature Scaling

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

- μ(mean) → broadcast across rows
- σ(std) → broadcast across rows

2. Bias Term in Neural Networks

$$Z = WX + b$$

- $b$is a **scalar or vector**
- Broadcast across all samples

3. Loss Functions

$$\text{MSE} = \frac{1}{n} \sum (y - \hat{y})^2$$

- $\frac{1}{n}$is broadcast across the entire sum

4. Trace of a Matrix

Definition

The **trace** of a square matrix is the **sum of its diagonal elements**.

$$\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}$$

Requirement

- Matrix must be **square** $(n \times n)$

Example

$$tr\left(\begin{bmatrix} 5 & 3 & 5 \\ 4 & -1 & 2 \\ -3 & 8 & 7 \end{bmatrix}\right) = 5 - 1 + 7 = 11.$$

4.1. Why Trace Matters in ML & AI

Key Applications

- **Loss functions** (e.g., Frobenius norm):

$$\| A \|_F^2 = \operatorname{tr}(A^T A)$$

- **Covariance matrices**
- **Regularization terms**
- **Gaussian distributions**
- **Matrix derivatives**

4.2. Important Trace Properties (Interview-Ready)

Linearity

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$

Scalar Multiplication

$$\mathrm{tr}(cA) = c \cdot \mathrm{tr}(A)$$

Cyclic Property

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$$

(only for valid matrix multiplication)

Order matters for matrices, but trace allows cyclic permutation.

4.3. Dense vs Sparse Context (Quick Insight)

- **Addition/Subtraction**: Efficient for both dense & sparse matrices
- **Trace**: Extremely fast, only diagonal entries matter
- Used heavily in **large-scale optimization problems**

4.4. Quick Summary

Concept	Key Idea	ML Relevance
Addition	Element-wise	Parameter updates
Subtraction	Element-wise	Error computation
Trace	Sum of diagonal	Loss, regularization