Course Name: Mathematics for AI
Week-2: Matrix operations, matrix multiplication and transformations

Matrices: Matrices play a central role in linear algebra. They can be used to compactly represent systems of linear equations, but they also represent linear functions.

Definition: Let $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{N}$. A real-valued $(\boldsymbol{m}, \boldsymbol{n})$-matrix $\boldsymbol{A}$ is an $\boldsymbol{m} \times \boldsymbol{n}$ tuple of elements $\boldsymbol{a_{ij}} \in R$, where $\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{m}$ and $\boldsymbol{j} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{n}$, arranged according to a rectangular scheme consisting of $\boldsymbol{m}$ rows and $\boldsymbol{n}$ columns.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

By convention (1, n)-matrices are called rows and (m,1)-matrices are called row columns. These special matrices are also called row/column vectors.

Definition (Matrix Addition): Let $A = (a_{ij})$and $B = (b_{ij})$be two real-valued matrices of the same order $m \times n$. The sum of $A$and B, denoted by $A + B$, is the matrix $C = (c_{ij})$of order $m \times n$defined by $c_{ij} = a_{ij} + b_{ij}$,for all $i = 1,2, \ldots, m$ and $j = 1,2, \ldots, n$.

That is,

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

Note: Matrix addition is defined only when both matrices have the same dimensions.

A. Practise Problems

1. Let $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 5 & 1 \\ 0 & -2 \end{bmatrix}$. Find $A + B$.
2. If $A = [7 \quad -3 \quad 2], B = [-4 \quad 6 \quad 5]$, find $A + B$.
3. Let $A = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}, B = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$. Compute $A + B$.
4. If $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, B = \begin{bmatrix} -1 & 0 & 2 \\ 3 & -5 & 1 \end{bmatrix}$, find $A + B$.

5. Given $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, B = \begin{bmatrix} 2a & -b \\ c & 3d \end{bmatrix}$, express $A + B$.

6. If $A = \begin{bmatrix} x & y \\ z & w \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, find $A + B$ when $x = 2, y = -1, z = 0, w = 3$.

Definition (Matrix Multiplication):

Let $A = (a_{ij})$be a matrix of order $m \times n$ and $B = (b_{jk})$be a matrix of order $n \times p$.
The product of $A$and B, denoted by AB, is the matrix $C = (c_{ik})$of order $m \times p$, where each entry is defined by

$$c_{ik} = \sum_{j=1}^{n} a_{ij}\ b_{jk}, \text{for } i = 1,2, \dots, m \text{ and } k = 1,2, \dots, p.$$

That is, the element in the i-th row and k-th column of $AB$is obtained by multiplying corresponding entries of the i-th row of $A$and the k-th column of B, and then adding the products.

Important Notes:

1. Matrix multiplication is defined only when the number of columns of Aequals the number of rows of B.
2. In general, matrix multiplication is not commutative, i.e., $AB \neq BA$.
3. The order of the product AB is $m \times p$.

Practice Problems: Matrix Multiplication

1. Compute AB, where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 0 \\ 1 & 5 \end{bmatrix}$.

2. Find the product AB, where $A = \begin{bmatrix} 3 & -1 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}$.

3. If $A = \begin{bmatrix} 1 & 0 \\ -2 & 3 \end{bmatrix}$, compute $A^2 = A \cdot A$.

4. Determine whether the following products are defined. If yes, state the order of the result.

(a) $(2 \times 3)(3 \times 4)$ (b) $(3 \times 2)(3 \times 2)$ (c) $(4 \times 1)(1 \times 5)$

5. Let A be a 3×2 matrix and B be a 2×4 matrix. What is the order of AB?

6. Show by example that matrix multiplication is not commutative, i.e., $AB \neq BA$.

7. If $AB = 0$, does it necessarily imply that $A = 0$or $B = 0$? Explain with a counterexample.

8. If I is the identity matrix, show that $AI = IA = A$.

Geometric Interpretation of Matrix Multiplication

Matrix multiplication can be understood geometrically as a linear transformation of space.

1. Matrices as Transformations

An $n \times n$ matrix represents a linear transformation of $\mathbb{R}^n$:

- It maps vectors to vectors,
- It preserves straight lines and the origin,
- It can stretch, shrink, rotate, reflect, or shear space.

If A is a matrix and x is a vector, then Ax is the image of x after the transformation defined by A.

2. Column Interpretation:

For a matrix $A = \begin{bmatrix} | & | \\ \mathrm{a}_1 & \mathrm{a}_2 \\ | & | \end{bmatrix}$, the columns $\mathrm{a}_1, \mathrm{a}_2$ are the images of the standard basis vectors: $A\mathrm{e}_1 = \mathrm{a}_1, A\mathrm{e}_2 = \mathrm{a}_2$. Thus, the matrix tells us where the coordinate axes go.

3. Meaning of the Product AB

Let A and B be square matrices. The product AB represents composition of transformations: $AB\mathrm{x} = A(B\mathrm{x})$.

Geometric meaning:

1. First apply the transformation B to vector x,
2. Then apply the transformation A to the result.

2D Example

Consider $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

- B: rotation by $90°$counterclockwise,
- A: stretches the x-direction by a factor of 2.
- AB: rotate first, then stretch,
- BA: stretch first, then rotate.

The final images are different, showing $AB \neq BA$.

Area and Volume Interpretation

- The absolute value of the determinant of a matrix gives the area (in 2D) or volume (in 3D) scaling factor.
- For products: $\det(AB) = \det(A)\det(B)$, meaning the total scaling equals the product of individual scaling.

Practise Problems:

1. A map uses a linear transformation represented by $A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ to enlarge regions. If the original region has area 5 square units, find the area after transformation.

2. A square of area 10 square units undergo the transformation $A = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$. Find the area of the transformed square and comment on the role of k.

3. A triangular region in the plane is transformed by $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. If the original area is 8 square units:
   (a) find the area after transformation, (b) explain the sign of the determinant.

5. Rotation Matrix
   A rotation in the plane is represented by $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Explain why the area of any region remains unchanged after rotation.

Neural-Network–Style Geometric Interpretation (Linear Layers)

In neural networks, each linear layer is a matrix multiplication, and geometry provides the intuition for how data is transformed layer by layer.

1. Linear Layer as a Geometric Transformation

A typical linear layer is written as $\mathrm{y} = W\mathrm{x} + \mathrm{b}$,

where

- $\mathrm{x} \in \mathbb{R}^n$= input vector,
- $W \in \mathbb{R}^{m\times n}$= weight matrix,
- $\mathrm{b} \in \mathbb{R}^m$= bias vector.

Geometric meaning:

- $W\mathrm{x} \rightarrow$ linear transformation (stretching, rotating, shearing, projecting),
- $+\mathrm{b} \rightarrow$ translation of space.

So, a linear layer reshapes the space and then shifts it.

2. Column View (Very Important in NN)

Let $W = \begin{bmatrix} | & | & & | \\ \mathrm{w}_1 & \mathrm{w}_2 & \cdots & \mathrm{w}_n \\ | & | & & | \end{bmatrix}$. Then $W\mathrm{x} = x_1\mathrm{w}_1 + x_2\mathrm{w}_2 + \cdots + x_n\mathrm{w}_n$.

Interpretation:

- Each column w_j is a direction in output space,

- Input coordinates x_j decide how much we move along each direction.

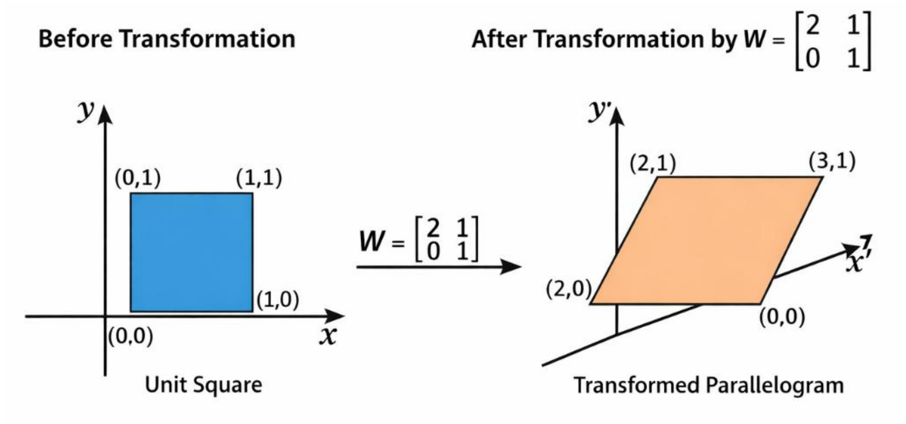
Conclusion: A linear layer re-expresses the input vector in a new coordinate system.

3. One Linear Layer (2D → 2D Example)

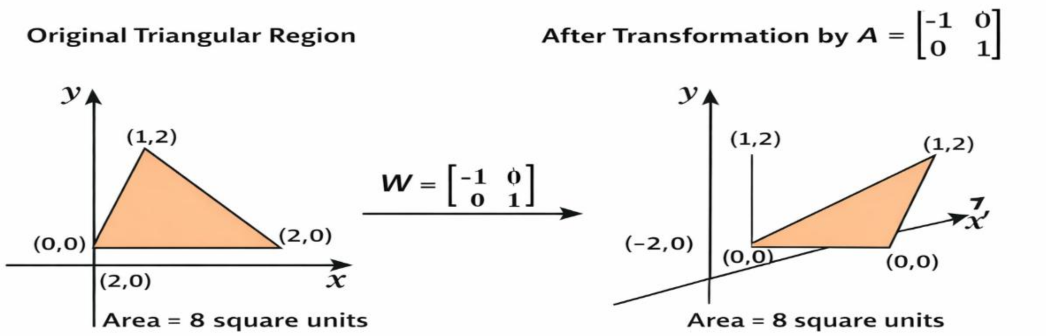
Suppose input points lie in the plane $\mathbb{R}^2$. Let $W = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$.

Geometric effect:

- The unit square becomes a parallelogram,



- Axes are skewed,



- Relative positions of points are preserved (linearity).

All input points are simultaneously transformed in the same way.

4. Multiple Linear Layers = Composition of Transformations

Consider two linear layers: $\mathrm{h} = W_1 \mathrm{x}, \mathrm{y} = W_2 \mathrm{h}$. Then $\mathrm{y} = (W_2 W_1)\mathrm{x}$.

Geometric meaning:

- First layer W_1: reshapes input space,
- Second layer W_2: reshapes the new space again,
- The product $W_2 W_1$is a single combined transformation.

Conclusion: This is exactly matrix multiplication as composition of maps.

5. Why Depth Without Nonlinearity Is Limited

If a network has only linear layers: $y = W_k \cdots W_2 W_1 x$, it is still just one linear transformation.

Geometrically:

- Multiple stretches/rotations collapse into one,
- No bending or warping of space.

Remark: Nonlinear activation functions (ReLU, tanh, sigmoid) are needed to:

- Fold space,
- Create curved decision boundaries,
- Separate complex data.

6. Geometric View of Classification

- Linear layer → maps data to a new space,
- Hyperplanes become decision boundaries,
- Training adjusts W so classes become linearly separable.

In 2D: A line separates classes

In higher dimensions: A hyperplane separates data clouds.

Case Study: Predicting Student Performance Using a Neural Network Linear Layer

Problem Context

Suppose we want to predict a **student's final performance score** based on three factors:

1. Hours studied per week (x_1)
2. Attendance percentage (x_2)
3. Previous exam score (x_3)

We use a **simple neural network with one linear layer** having **2 neurons**.

Solution: Step 1: Input Vector (Student Data) For a particular student:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

Step 2: Weight Matrix of the Linear Layer

Assume the network has learned the following weights:

$$W = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.1 & 0.6 & 0.4 \end{bmatrix}$$

- Neuron 1 focuses more on study hours

- Neuron 2 focuses more on attendance and previous score

Step 3: Bias Vector

$$b = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

Step 4: Linear Transformation (Matrix Multiplication) $Z = WX + b$

First compute WX:

$$WX = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.1 & 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

Neuron 1 Output: $(0.5)(6) + (0.2)(8) + (0.3)(7) = 3 + 1.6 + 2.1 = 6.7$

Neuron 2 Output: $(0.1)(6) + (0.6)(8) + (0.4)(7) = 0.6 + 4.8 + 2.8 = 8.2$

Now add bias: $Z = \begin{bmatrix} 6.7 \\ 8.2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 7.7 \\ 8.7 \end{bmatrix}$

Step 5: Interpretation of the Transformation

- Input space: 3-dimensional (study, attendance, score)
- Output space: 2-dimensional feature space
- Matrix multiplication projects and reshapes the input into new features

Remark: Each neuron computes a weighted combination of all inputs.

Step 6: Activation Function (Optional) Apply ReLU (**Rectified Linear Unit**) activation:

Mathematical Definition

For any real number z, $\text{ReLU}(z) = \max(0, z)$

That means:

- If $z > 0$, output is z
- If $z \leq 0$, output is 0

$$A = \max(0, Z)$$
$$A = \begin{bmatrix} 7.7 \\ 8.7 \end{bmatrix}$$

ReLU is applied **element by element** to the vector Z. That is, No change since values are positive.

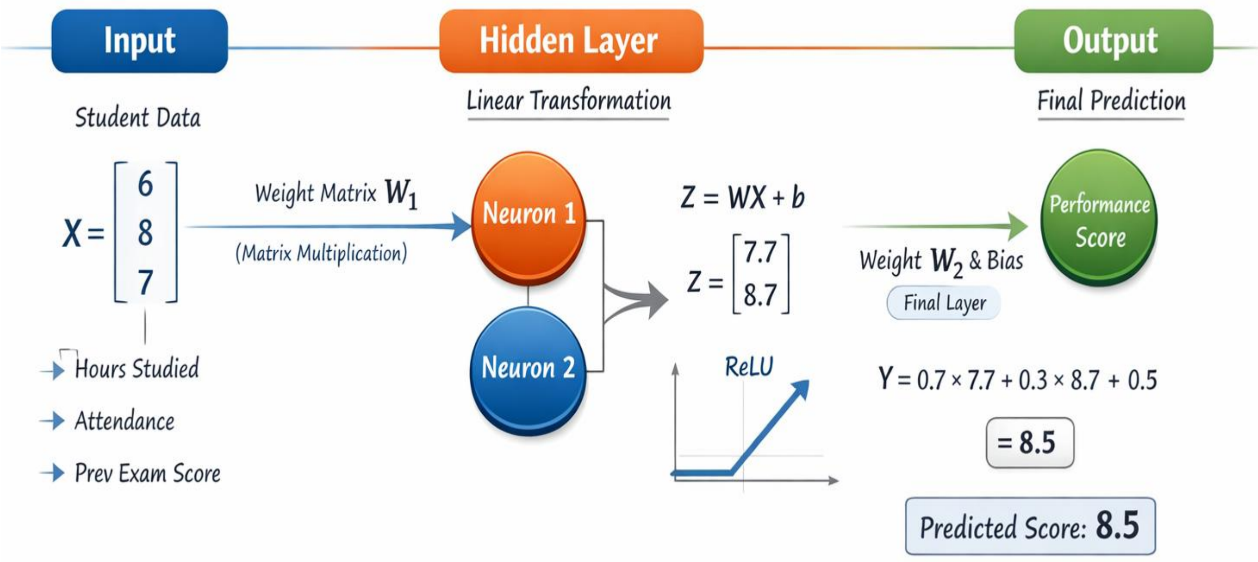
Step 7: Second Layer (Final Prediction)

Assume a final layer to produce **one performance score**.

Weight and Bias: $W_2 = [0.7 \quad 0.3], b_2 = 0.5$

$$Y = W_2 A + b_2$$
$$Y = (0.7)(7.7) + (0.3)(8.7) + 0.5$$
$$Y = 5.39 + 2.61 + 0.5 = 8.5$$

Final Output: Predicted Performance Score = 8.



Problem-2:

Suppose we want to predict a **system performance index** based on three input parameters:

- x_1: Input load level
- x_2: Resource utilization
- x_3: System response time

  We model this using a **multi-layer neural network** with:

- **Input layer**: 3 neurons
- **Hidden layer 1**: 2 neurons
- **Hidden layer 2**: 2 neurons
- **Output layer**: 1 neuron

Solution: Step-1 Input Vector For a particular system: $X = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$

Step 2: First Hidden Layer

Weight Matrix and Bias $W_1 = \begin{bmatrix} 0.4 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}, b_1 = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$ respectively.

Linear Transformation

$$Z_1 = W_1X + b_1$$
$$Z_1 = \begin{bmatrix} (0.4)(4) + (0.3)(6) + (0.2)(5) + 0.6 \\ (0.2)(4) + (0.5)(6) + (0.1)(5) + 0.4 \end{bmatrix} = \begin{bmatrix} 5.0 \\ 4.7 \end{bmatrix}$$

Activation (ReLU)

$$A_1 = \max(0, Z_1) = \begin{bmatrix} 5.0 \\ 4.7 \end{bmatrix}.$$

Step 3: Second Hidden Layer

Weight Matrix and Bias $W_2 = \begin{bmatrix} 0.6 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}, b_2 = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix}$

Linear Transformation $Z_2 = W_2A_1 + b_2$

$$Z_2 = \begin{bmatrix} (0.6)(5.0) + (0.2)(4.7) + 0.3 \\ (0.3)(5.0) + (0.7)(4.7) + 0.2 \end{bmatrix} = \begin{bmatrix} 4.64 \\ 5.99 \end{bmatrix}$$

Activation (ReLU)

$$A_2 = \begin{bmatrix} 4.64 \\ 5.99 \end{bmatrix}$$

Step 4: Output Layer

Weight and Bias $W_3 = [0.5 \quad 0.5], b_3 = 0.4$

Output Computation $Y = W_3A_2 + b_3$

$$Y = (0.5)(4.64) + (0.5)(5.99) + 0.4$$
$$Y = 2.32 + 2.995 + 0.4 = 5.715$$