1. Introduction to Linear Algebra

Linear algebra is the mathematical framework that deals with vectors, matrices, and linear transformations. It is crucial for machine learning as it enables us to:

1. Represent data compactly using vectors and matrices.
2. Transform and manipulate data efficiently.
3. Solve equations systematically.

Key Concepts

- **Scalars, Vectors, and Matrices:**
  - Scalars: Single numerical values (e.g., $a = 5$).
  - Vectors: One-dimensional arrays (e.g., $v = [3, 4, 5]^T$).
  - Matrices: Two-dimensional arrays (e.g., $A = [[1, 2], [3, 4]]$).

Practical Example

For a machine learning dataset:

- Each data point is a row vector $[x_1, x_2, ..., x_n]$.
- The entire dataset is represented as a matrix A where each row is a data point.

---

2. Vector Operations

Addition and Scalar Multiplication

Vectors can be added or scaled (multiplied by scalars). This is used in many machine learning models.

- **Addition**: Add corresponding elements.

$$v_1 = [1, 2, 3], \quad v_2 = [4, 5, 6] \quad \Rightarrow \quad v_1 + v_2 = [5, 7, 9]$$

- **Scalar Multiplication**: Multiply each element by a scalar.

$$c = 2, \quad v = [1, 2, 3] \quad \Rightarrow \quad c \cdot v = [2, 4, 6]$$

Practical Example

In a linear regression model, weights w are vectors, and predictions involve:

$$y = w^T x + b$$

---

3. Dot Product and Norms

Dot Product

The dot product measures the similarity between two vectors. It is calculated as:

$$a \cdot b = \sum_{i=1}^{n} a_i b_i$$

Example:

$$a = [1, 2], \quad b = [3, 4] \quad \Rightarrow \quad a \cdot b = (1 \cdot 3) + (2 \cdot 4) = 11$$

Norms

The norm of a vector measures its magnitude (length). For a vector v:

$$||v|| = \sqrt{\sum_{i=1}^{n} v_i^2}$$

Example:

$$v = [3, 4] \quad \Rightarrow \quad ||v|| = \sqrt{3^2 + 4^2} = 5$$

---

4. Linear Transformations

Linear transformations map one vector space to another while preserving vector operations. These are represented as matrix-vector multiplications.

Practical Example

In image processing:

- A rotation matrix can rotate an image by θ degrees.
- A scaling matrix can enlarge or shrink the image.

Example: Scaling a vector $v = [2, 3]$ by a factor of 2:

$$S = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad v' = S \cdot v = [4, 6]$$

---

5. Systems of Linear Equations

Recap:

A system of linear equations can be written as $Ax = b$. Depending on A and b, the system can have:

1. **No solution** (inconsistent system).

2. **A unique solution** (independent equations).
3. **Infinitely many solutions** (dependent equations).

Advanced Example

Solve:

$$x + y + z = 6, \quad x - y + z = 4, \quad 2x + 3y + 4z = 18$$

Step 1: Write as an Augmented Matrix

$$\begin{bmatrix} 1 & 1 & 1 & | & 6 \\ 1 & -1 & 1 & | & 4 \\ 2 & 3 & 4 & | & 18 \end{bmatrix}$$

Step 2: Apply Gaussian Elimination

1. Eliminate x from rows 2 and 3 using row operations.
2. Back-substitute to find z, y, x.

Solution:

$$x = 2, y = 1, z = 3$$

---

6. Matrix Inversion

Advanced Notes

The inverse A^{-1} exists if $\det(A) \neq 0$. For a 2×2 matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Example

Find A^{-1} for:

$$A = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}$$

Solution:

$$\det(A) = (4)(6) - (7)(2) = 10$$

$$A^{-1} = \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$$

---

7. Eigenvalues and Eigenvectors

Detailed Explanation

An eigenvalue λ and eigenvector v satisfy:

$$Av = \lambda v$$

Steps to Compute Eigenvalues:

1. Solve $\det(A - \lambda I) = 0$.
2. Solve for λ.

Steps to Compute Eigenvectors:

1. Solve $(A - \lambda I)v = 0$.
2. Find v.

Example

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

1. Compute λ: Solve $\det(A - \lambda I) = 0$.
2. Compute v: Substitute λ into $(A - \lambda I)v = 0$.

---

8. Singular Value Decomposition (SVD)

Expanded Example

For $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$:

1. Compute $A^T A$.
2. Solve $\det(A^T A - \lambda I) = 0$ to find singular values.
3. Find U, V, Σ.

---

9. Applications in Machine Learning

Dimensionality Reduction

PCA uses eigenvalues to project high-dimensional data into a lower-dimensional space.

Optimization

Linear regression solves:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Data Compression

SVD reduces the size of matrices while retaining key features.