Comprehensive Study Guide: Mathematical Foundations for Machine Learning

This guide has been developed to provide a thorough understanding of mathematical concepts that form the foundation of machine learning. These concepts are essential for designing, analyzing, and implementing algorithms effectively.

Section 1: Vectors and Vector Spaces

Vectors represent quantities that have both magnitude and direction, making them fundamental for representing data and features in machine learning.

Core Concepts

1. What is a Vector?

A vector is an ordered set of numbers, often written as:

$$\vec{v} = [v_1, v_2, ..., v_n],$$

where $v_1, v_2, ..., v_n$ are the components of the vector.

2. Vector Operations

1. **Addition**: Combine corresponding components of two vectors:

$$\vec{u} + \vec{v} = [u_1 + v_1, u_2 + v_2, ..., u_n + v_n].$$

   Example:

$$\vec{u} = [1, 2], \quad \vec{v} = [3, 4] \quad \implies \quad \vec{u} + \vec{v} = [4, 6].$$

2. **Scalar Multiplication**: Multiply each component by a scalar:

$$c\vec{v} = [cv_1, cv_2, ..., cv_n].$$

Example:

$$2\vec{v} = 2[3, 4] = [6, 8].$$

3. **Dot Product**: Measures the similarity of two vectors:

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i.$$

Example:

$$\vec{u} = [1, 2], \vec{v} = [3, 4] \quad \Longrightarrow \quad \vec{u} \cdot \vec{v} = (1)(3) + (2)(4) = 11.$$

4. **Cross Product (3D only)**: Produces a vector orthogonal to both inputs:

$$\vec{u} \times \vec{v} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

3. Linear Independence

Vectors are linearly independent if no vector in the set can be expressed as a combination of others:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_k\vec{v}_k = \vec{0}.$$

If the only solution is $c_1 = c_2 = \ldots = c_k = 0$, the vectors are independent.

Example: For $\vec{v}_1 = [1, 0]$ and $\vec{v}_2 = [0, 1]$, they are independent as no scalar combination produces the other.

4. Basis and Dimension

- **Basis**: A set of vectors that spans the vector space (e.g., $[1, 0]$ and $[0, 1]$ in R^2).
- **Dimension**: Number of vectors in the basis.

Applications in Machine Learning

1. **Feature Representation**: Each feature in a dataset is a vector component. For example, an image might be represented as a vector of pixel intensities.
2. **Word Embeddings**: Words are mapped into a vector space where similar words have closer representations.
3. **Data Projection**: Reduce high-dimensional data to 2D or 3D for visualization (e.g., using PCA).

Section 2: Matrices and Linear Transformations

Matrices are essential for transforming data and representing linear systems in machine learning.

Core Concepts

1. What is a Matrix?

A matrix is a two-dimensional array of numbers:

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

2. Key Operations

1. **Addition**:

$$A + B = [a_{ij} + b_{ij}].$$

2. **Scalar Multiplication**:

$$cA = [ca_{ij}].$$

3. **Matrix Multiplication**:

$$C = AB \quad \text{where} \quad c_{ij} = \sum_k a_{ik}b_{kj}.$$

3. Identity Matrix

Diagonal entries are 1; all others are 0. It acts as the multiplicative identity for matrices:

$$A \cdot I = A.$$

4. Transpose

Interchange rows and columns:

$$A^T = [a_{ij}]^T = [a_{ji}].$$

Applications in Machine Learning

1. **Neural Networks**: Weights of a neural network are stored in matrices and applied to input data.
2. **Dimensionality Reduction**: Techniques like PCA involve matrix decomposition.
3. **Linear Regression**: Solving $Ax = b$ to find optimal weights.

Section 3: Systems of Linear Equations

These systems model relationships between variables.

Core Concepts

1. Representation

$$Ax = b,$$

where A is the coefficient matrix, x is the unknown vector, and b is the constant vector.

2. Solution Types

1. **Unique Solution**: Occurs if A is invertible.
2. **Infinite Solutions**: Free variables result in multiple solutions.
3. **No Solution**: Contradictory constraints lead to inconsistency.

Solution Methods

1. Gaussian Elimination

Steps:

1. Convert to Row-Echelon Form (REF).
2. Use back-substitution to solve for variables.

Example: Solve:

$$x + y + z = 6, \quad 2x + 3y + z = 14, \quad 3x + y + 2z = 14.$$

Steps:

1. Write as augmented matrix.
2. Row reduce to:

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

3. Back-substitution gives:

$$x = 2, y = 4, z = 0.$$

2. Gauss-Jordan Elimination

Further reduces to Reduced Row-Echelon Form (RREF).

Applications in Machine Learning

1. **Optimization Problems**: Many ML algorithms solve $Ax = b$ to find model parameters.
2. **Graph Analysis**: Representing graphs as adjacency matrices.