Expanded and Detailed Study Guide: Mathematical Foundations for Machine Learning

This guide delves deeper into the essential mathematical concepts required for mastering machine learning. Each topic is explained thoroughly, with detailed explanations, real-world applications, and illustrative examples. This document will not only serve as a study guide but also as a comprehensive reference for revisiting key topics.

---

1. Systems of Linear Equations

Linear systems form the basis of solving real-world optimization problems in machine learning, such as regression analysis.

1.1 What is a Linear System?

A linear system consists of multiple linear equations involving the same set of variables:

$$Ax = b,$$

where:

- A: Coefficient matrix ($m \times n$).
- x: Column vector of unknowns ($n \times 1$).
- b: Column vector of constants ($m \times 1$).

The goal is to determine whether the system is **consistent** (has solutions) or **inconsistent** (no solutions).

1.2 Types of Solutions

1. **Unique Solution**:
   - Occurs when the determinant $\det(A) \neq 0$ (for square matrices).
   - Example: For $x + y = 3$ and $x - y = 1$, solving gives $x = 2, y = 1$.
2. **No Solution**:
   - Occurs when the system is inconsistent (e.g., parallel lines in 2D).
   - Example:
     $$x + y = 2, \quad x + y = 4.$$
     These lines are parallel and never intersect.
3. **Infinite Solutions**:
   - Occurs when the equations are dependent.
   - Example:
     $$x + y = 2, \quad 2x + 2y = 4.$$
     These lines overlap, leading to infinitely many solutions.

1.3 Solution Methods

Gaussian Elimination

This is a systematic method to solve linear systems by converting the coefficient matrix into row-echelon form.

Example: Solve the system:

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

Steps:

1. Write the augmented matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 2 & -1 & 1 & 3 \\ 1 & -1 & -1 & -2 \end{bmatrix}.$$

2. Use row operations to eliminate x from rows 2 and 3:

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & -3 & -1 & -9 \\ 0 & -2 & -2 & -8 \end{bmatrix}.$$

3. Eliminate y from row 3:

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & -3 & -1 & -9 \\ 0 & 0 & -4/3 & -10/3 \end{bmatrix}.$$

4. Solve using back-substitution:

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

LU Decomposition

Factorize $A = LU$, where:

- L: Lower triangular matrix.
- U: Upper triangular matrix.

Solve $Ly = b$, then $Ux = y$.

Example: For:

$$A = \begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix},$$

LU decomposition gives:

$$L = \begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 4 & 3 \\ 0 & -1.5 \end{bmatrix}.$$

---

2. Vector Spaces

Vector spaces provide the framework for representing data points, features, and models in machine learning.

2.1 Definition

A vector space is a set of vectors that satisfy closure properties under addition and scalar multiplication. These vectors can represent data points, parameters, or gradients in machine learning.

2.2 Key Concepts

1. **Basis**:
   - A minimal set of vectors that span the vector space.
   - Example: $\{[1, 0], [0, 1]\}$ is the basis for $\mathbb{R}^2$.
2. **Dimension**:
   - Number of vectors in the basis.
   - Example: $\mathbb{R}^3$ has dimension 3.
3. **Linear Independence**:
   - Vectors $\{v_1, v_2, \dots, v_k\}$ are independent if:

$$c_1 v_1 + c_2 v_2 + \dots + c_k v_k = 0 \implies c_1 = c_2 = \dots = c_k = 0.$$

2.3 Practical Applications

1. **Feature Representation**:
   - Data points as vectors in high-dimensional spaces.
   - Example: Image pixels represented as vectors.
2. **Dimensionality Reduction (PCA)**:
   - Projects data onto a lower-dimensional subspace.

---

3. Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are essential for understanding linear transformations in machine learning.

3.1 Definitions

If:

$$A\vec{v} = \lambda\vec{v},$$

then:

- λ: Eigenvalue.
- $\vec{v}$: Eigenvector.

3.2 Applications

1. **Principal Component Analysis (PCA)**:
   - Eigenvectors of the covariance matrix define principal components.
2. **Spectral Clustering**:
   - Uses eigenvectors of graph Laplacians to partition data.

3.3 Practical Example

Find eigenvalues and eigenvectors for:

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

Steps:

1. Compute the characteristic equation:

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = 0.$$

   Solve for λ: $\lambda_1 = 3$, $\lambda_2 = 1$.

2. Solve for eigenvectors:

$$\lambda_1 : \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \lambda_2 : \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

---

4. Matrix Factorizations

Matrix factorizations decompose matrices into simpler forms for efficient computation.

4.1 Key Types

1. **LU Decomposition**:
   - $A = LU$, used for solving linear systems efficiently.
2. **QR Decomposition**:
   - $A = QR$, where Q is orthogonal, and R is upper triangular.
3. **Singular Value Decomposition (SVD)**:
   - $A = U\Sigma V^T$, used in PCA and recommendation systems.

4.2 Applications

1. **Dimensionality Reduction**:
   - PCA uses SVD to reduce the feature space.
2. **Collaborative Filtering**:
   - Matrix factorization for recommendation systems.