Comprehensive and In-Depth Study Notes: Mathematical Foundations for Machine Learning

These study notes provide a **detailed explanation of concepts** required to excel in Mathematical Foundations for Machine Learning, making complex topics easy to understand and apply. Each topic is explained with theoretical concepts, practical examples, and real-world applications.

---

1. Systems of Linear Equations

What is a System of Linear Equations?

A system of linear equations is a collection of equations involving the same set of variables. It is often represented in matrix form:

$$Ax = b,$$

where:

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

Solution Types

1. **Unique Solution**:
   - Occurs when A has full rank (all rows are linearly independent).
   - Example: $2x + y = 5,\ x - y = 1$.
2. **No Solution**:

- Occurs when the system is inconsistent (e.g., parallel lines in 2D).
- Example: $x + y = 2, \ x + y = 4$.

3. **Infinite Solutions**:
   - Occurs when there are free variables (e.g., dependent equations).
   - Example: $x + y = 2, \ 2x + 2y = 4$.

Solution Methods

1. **Gaussian Elimination**:
   - Convert the system into row-echelon form using elementary row operations.
   - Solve using back-substitution.
2. **LU Decomposition**:
   - Factorize A as LU, where L is lower triangular, and U is upper triangular.
   - Solve $Ly = b$, then $Ux = y$.
3. **Matrix Inversion** (for square matrices):
   - If A is invertible:

$$x = A^{-1}b.$$

Geometrical Interpretation

1. **2D Case**:

- Each equation represents a line.
- Solution: The intersection of lines.

2. **3D Case**:
   - Each equation represents a plane.
   - Solution: The intersection of planes.

Practical Example

Solve:

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

Steps:

1. Represent as an augmented matrix:
$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 2 & -1 & 1 & 3 \\ 1 & -1 & -1 & -2 \end{bmatrix}.$$
2. Use Gaussian elimination:
   - Eliminate x from rows 2 and 3.
   - Reduce to row-echelon form.
3. Solve using back-substitution:
$$z = 1, \, y = 2, \, x = 3.$$

---

2. Vector Spaces

What is a Vector Space?

A vector space is a collection of vectors where:

- **Addition** and **scalar multiplication** are defined.
- Vectors satisfy closure properties (e.g., adding two vectors yields another vector in the space).

Key Concepts

1. **Linear Independence**:
    - Vectors $\{v_1, v_2, \dots, v_k\}$ are linearly independent if:

$$c_1v_1 + c_2v_2 + \ldots + c_kv_k = 0 \implies c_1 = c_2 = \ldots = c_k = 0.$$

2. **Basis**:
    - A set of linearly independent vectors that span the vector space.
    - Example: The standard basis for R^3 is $\{[1, 0, 0], [0, 1, 0], [0, 0, 1]\}$.
3. **Dimension**:
    - Number of vectors in the basis.

Applications

1. **Feature Representation**:

- Data points are represented as vectors in $\mathbb{R}^n$.

2. **Principal Component Analysis (PCA)**:
   - Projects data onto a lower-dimensional vector space.

---

3. Eigenvalues and Eigenvectors

Definitions

- **Eigenvalue** (λ): Scalar such that:

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

  where $\vec{v}$ is a non-zero eigenvector.

Key Properties

1. Eigenvalues are roots of the characteristic equation:

$$\det(A - \lambda I) = 0.$$

2. Eigenvectors point in directions invariant under A.

Applications

1. **PCA**:
   - Eigenvectors of the covariance matrix define principal components.
2. **Spectral Clustering**:
   - Uses eigenvectors of graph Laplacians.

Practical Example

Find eigenvalues and eigenvectors for:

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

1. Compute:

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

   Eigenvalues: $\lambda_1 = 3, \lambda_2 = 1.$
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 Decompositions

Key Types

1. **LU Decomposition**:
   - Factorize A into L (lower triangular) and U (upper triangular).
2. **QR Decomposition**:
   - $A = QR$, where Q is orthogonal, and R is upper triangular.
3. **SVD**:
   - $A = U\Sigma V^T$, where:
     - U, V: Orthogonal matrices.
     - Σ: Diagonal matrix of singular values.

Applications

1. **Dimensionality Reduction**:
   - PCA uses SVD to reduce feature space.
2. **Recommendation Systems**:
   - Matrix factorization for collaborative filtering.

---

5. Optimization Techniques

Optimization is at the heart of training machine learning models.

Key Methods

1. **Gradient Descent**:
   - Iteratively updates parameters:
     $$w \leftarrow w - \eta \nabla J(w),$$
     where η is the learning rate.
2. **Stochastic Gradient Descent (SGD)**:
   - Updates parameters for a single data point.
3. **Newton's Method**:
   - Uses second-order derivatives (Hessian) for faster convergence.

Applications

1. **Training Neural Networks**:
   - Backpropagation computes gradients.
2. **Logistic Regression**:
   - Gradient descent minimizes the cross-entropy loss.

---

6. Dimensionality Reduction

Principal Component Analysis (PCA)

Projects data onto top k eigenvectors of the covariance matrix:

$$C = \frac{1}{n} X^T X.$$

Applications

1. **Visualization**:
    - Reduce high-dimensional data to 2D/3D.
2. **Noise Reduction**:
    - Remove low-variance components.

Final Thoughts

These notes offer a deep dive into the core mathematical concepts that underpin machine learning.