Comprehensive Study Guide with Problem Solutions: Mathematical Foundations for Machine Learning

This guide includes detailed explanations for the concepts, paired with **practical examples** and **solutions** to the problems from the uploaded handout, "Lecture 0 - Some Motivating Problems." By mastering these topics and exercises, you'll develop a strong mathematical foundation for tackling machine learning challenges.

---

1. Linear Systems and Their Solutions

Why Does a Linear System $Ax = b$ Have Only No, One, or Infinite Solutions?

A linear system involves:

- A: A matrix of size $m \times n$.
- b: A vector of size $m \times 1$.
- x: An unknown vector of size $n \times 1$.

The solution count depends on the rank of A and the augmented matrix $[A|b]$:

1. **No Solution**:
    - Occurs if b lies outside the column space of A (system is inconsistent).
2. **One Solution**:
    - Occurs if A has full column rank, making the system consistent and independent.
3. **Infinite Solutions**:
    - Occurs if A does not have full column rank but the system is consistent (free variables exist).

Why Not Two Solutions?

- If $Ax_1 = b$ and $Ax_2 = b$, then subtracting the equations gives:

$$A(x_1 - x_2) = 0 \implies x_1 - x_2 \in \text{null}(A).$$

Hence, if $x_1 \neq x_2$, infinitely many solutions exist along the null space.

---

2. Solving a 10 × 5 Linear System $Ax = b$ in Two Halves

Is Splitting the System Viable?

Given:

- A: A 10×5 matrix.
- b: A 10×1 vector.
- Split into two smaller systems:
  - $A_1 x = b_1$ (first 5 rows).
  - $A_2 x = b_2$ (last 5 rows).

Challenges:

1. Solving $A_1 x = b_1$ and $A_2 x = b_2$ separately might yield **different solutions** because the constraints from the missing rows in A_2 affect the solution in A_1, and vice versa.
2. Linear systems require all equations to be solved together for consistency.

Conclusion: Splitting $Ax = b$ may fail to preserve the original solution space.

---

3. Counting Operations for REF to RREF

Gaussian Elimination Complexity

For an $n \times n$ matrix:

1. **From REF to RREF**:
   - Pivot adjustments for each row: $O(n^2)$.
2. **Total Operations**:
   - Additions/Subtractions: $O(n^3)$.
   - Multiplications/Divisions: $O(n^3)$.

Example: For a 5×5 matrix, transitioning from REF to RREF involves approximately $O(125)$ operations.

---

4. Consistency in Combined Systems

Given Systems:

1. $Ax_1 = b_1$.
2. $Ax_2 = b_2$.
3. Combined: $Ax = b_1 + b_2$.

Analysis:

1. Linearity of A:

$$A(x_1 + x_2) = b_1 + b_2.$$

If x_1, x_2 are solutions, $x_1 + x_2$ is also a solution.

2. **Conclusion**:
   - $Ax = b_1 + b_2$ is consistent if x_1 and x_2 are valid solutions.

---

5. Non-Zero Matrices A, B Such That $AB = 0$ and $BA \neq 0$

Example:

1. Let:

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

2. Compute:

$$AB = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0.$$

3. Check BA:

$$BA \neq 0.$$

---

6. Proving Positive Definiteness

Problem: Prove $S = A^T A$ is positive definite.

Proof:

1. For any non-zero vector x:

$$x^T S x = x^T (A^T A) x = (Ax)^T (Ax) = ||Ax||^2.$$

   Since $||Ax||^2 \geq 0$ and $Ax \neq 0$, $x^T S x > 0$.

2. **Diagonal Entries**:
   - Diagonal entries of S are sums of squares of columns of A, hence non-negative.

---

7. Determinant Using Recursion

Recursive Algorithm:

1. Base Case:
   - $\det(A) = a_{11}$ for 1×1 matrix.
2. Recursive Step:
   - Expand along the first row:

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}),$$

     where A_{1j} is the minor matrix.

Complexity:

1. Recursive algorithm: $O(n!)$.
2. Gaussian elimination: $O(n^3)$.