Math Foundations for Machine Learning

Lecture 1: The Definitive Guide

The Geometry and Algebra of Solving Equations

BITS Pilani
Prof. Saurabh

October 31, 2025

Contents

1 Welcome: Why Linear Algebra is the Language of Data

Welcome! Before we can teach computers how to learn, we first need a way to describe the world to them. Computers don't inherently understand images, sounds, or text the way humans do. They understand numbers.

Linear algebra is the language we use to translate the real world into numbers, and it provides the rules—the engine—for manipulating those numbers to discover hidden patterns.

1.1 Everything is a Vector

Imagine you are trying to describe a house to a computer. You must use numbers:

$$(Price, Bedrooms, Area, Age) = (300000, 4, 2200, 15)$$

This structured list of numbers is called a **vector**. It's a mathematical container for data. In machine learning (ML), this concept is everywhere:

- **Images:** A simple grayscale image is a grid of numbers (a **matrix**, which is just a collection of vectors) representing pixel brightness.
- **Text:** We translate words into vectors (called embeddings). If the vector for "King" is mathematically close to the vector for "Queen", the computer understands they are related concepts.
- **Recommendations:** Your profile on a streaming service is a long vector representing how much you liked every movie in their catalog.

1.2 The Engine of Machine Learning

If data is the fuel, linear algebra is the engine. It allows us to:

1. **Calculate Similarity:** How similar are two data points? This is calculated by measuring the "distance" or "angle" between their vectors.
2. **Transform Data:** Neural networks work by applying successive linear transformations (matrix multiplications) to input data.
3. **Optimize Models:** The very act of "training" an ML model (like linear regression) is often about solving a massive system of linear equations to find the parameters that minimize error.

In this guide, we will focus on this core operation: solving systems of linear equations, exploring both how to do it (the algebra) and what it means (the geometry).

2 The Core Problem: Systems of Linear Equations

2.1 A Motivating Example: The Factory

Let's start with a real-world scenario. A factory produces three products: Laptops (P_1), Phones (P_2), and Tablets (P_3). It requires resources: CPU Chips (R_1), Memory Units (R_2), and Assembly Hours (R_3).

The "recipe":

- **Laptop (P_1):** 1 chip, 2 memory units, 4 hours.
- **Phone (P_2):** 1 chip, 1 memory unit, 2 hours.

- **Tablet (P_3):** 1 chip, 2 memory units, 3 hours.

The factory has a daily supply: 15 chips, 25 memory units, and 40 hours.

The Question: How many laptops (x_1), phones (x_2), and tablets (x_3) should we build to use up *all* the resources exactly?

We translate this into mathematics by focusing on the constraints (the resources):

- **Chips:** $1x_1 + 1x_2 + 1x_3 = 15$
- **Memory:** $2x_1 + 1x_2 + 2x_3 = 25$
- **Hours:** $4x_1 + 2x_2 + 3x_3 = 40$

This is a **system of linear equations**. We are looking for a combination (x_1, x_2, x_3) that satisfies all three constraints simultaneously.

2.2 The Geometry of Solutions: What Does It Look Like?

Before we try to solve the factory problem (which is in 3D), let's visualize a system in 2D (2 variables, 2 equations).

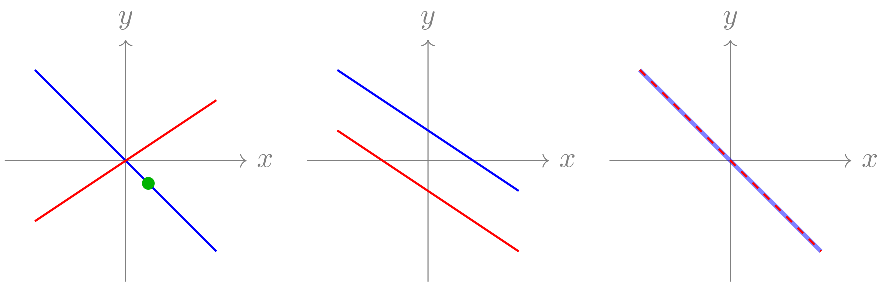
$$\begin{aligned} 2x + y &= 5 \\ x - y &= 1 \end{aligned}$$

Each of these equations defines a straight line in the 2D plane.

The Intuition

When we "solve a system," we are looking for a point (x, y) that lies on *both* lines at the same time. We are looking for the intersection.

When you have two lines in a 2D plane, there are only three ways they can interact:



1. One Unique Solution (Lines intersect at one point) **2. No Solution** (Parallel lines never touch) **3. Infinite Solutions** (Lines are the same)

Figure 1: The Three Possibilities for Solutions in a 2x2 System.

Key Concept

A fundamental truth: Every system of linear equations, whether it has 2 variables or 2000, will have either **Zero**, **One**, or **Infinitely Many** solutions. There are no other possibilities (e.g., you can never have exactly two solutions).

What about 3D? In our factory problem (3 variables), each equation defines a flat **plane** in 3D space.

- **One Solution:** Three planes intersect at a single point (like the corner of a room where two walls meet the floor).
- **No Solution:** The planes do not share a common intersection (e.g., parallel planes, or forming a long triangular tunnel).
- **Infinite Solutions:** The planes intersect to form a line (like the spine of a book), or they are all the same plane.

2.3 The Augmented Matrix: A Shorthand

To solve large systems efficiently, we drop the variables and focus only on the numbers. We organize the system into a grid called an **augmented matrix**.

Our system:

$$1x_1 + 1x_2 + 1x_3 = 15$$

$$2x_1 + 1x_2 + 2x_3 = 25$$

$$4x_1 + 2x_2 + 3x_3 = 40$$

Becomes:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 15 \\ 2 & 1 & 2 & 25 \\ 4 & 2 & 3 & 40 \end{array}\right]$$

The left side contains the **coefficients** (Matrix A). The right side contains the **constants** (Vector $\mathbf{b}$). We denote this $[A|\mathbf{b}]$.

3 The Strategy: Elimination and Tidying Up

Our strategy is to transform this complex system into a much simpler, equivalent system where the answer is obvious.

3.1 The Goal: The "Ideal" System

What is the simplest possible system? Consider this:

$$1x_1 + 0x_2 + 0x_3 = 5$$

$$0x_1 + 1x_2 + 0x_3 = 0$$

$$0x_1 + 0x_2 + 1x_3 = 10$$

This is ideal because we can just read the answer: $x_1 = 5, x_2 = 0, x_3 = 10$.

In matrix form, the ideal system looks like this:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 10 \end{array}\right]$$

This form, where the coefficient matrix is the **Identity Matrix** (1s on the diagonal, 0s elsewhere), is our ultimate goal.

3.2 The Method: Connecting Algebra and Matrices

Let's look at a simple 2x2 system to see how we transform it algebraically, and how that connects to the matrix operations.

Algebra	**Matrix**
$x + y = 3$ $2x + 3y = 7$	$\left[\begin{array}{cc\|c} 1 & 1 & 3 \\ 2 & 3 & 7 \end{array}\right]$
Goal: Eliminate x from the second equation (the '2' in R2C1).	
We subtract 2 times the first equation from the second. *Operation:* $(Eq2) \to (Eq2) - 2(Eq1)$ $(2x + 3y) - 2(x + y) = 7 - 2(3)$	*We subtract 2 times the first row from the second.* *Operation:* $R_2 \to R_2 - 2R_1$ $[2\ 3 \mid 7] - 2[1\ 1 \mid 3] = [0\ 1 \mid 1]$
New System: $x + y = 3$ $y = 1$	*New Matrix:* $\left[\begin{array}{cc\|c} 1 & 1 & 3 \\ 0 & 1 & 1 \end{array}\right]$

The system is now triangular. We solve this using **back-substitution**. Since $y = 1$, we plug this into Eq 1: $x + (1) = 3 \implies x = 2$. The solution is $(2, 1)$.

The Intuition

Matrix row operations are just a structured, organized way to perform algebraic elimination. We are "tidying up" the matrix to reveal the solution.

3.3 The Tools: Elementary Row Operations (EROs)

There are three "legal" moves we can perform, called **Elementary Row Operations (EROs)**.

Key Concept

Crucially, EROs do **not** change the underlying solution set of the system. They only change its appearance.

1. **Swapping ($R_i \leftrightarrow R_j$):** Changing the order of the equations.
2. **Scaling ($kR_i \to R_i$):** Multiplying an entire equation by a *non-zero* constant. (e.g., $x + y = 3$ is the same as $2x + 2y = 6$).
3. **Replacement ($R_i \to R_i + kR_j$):** Adding a multiple of one equation to another. (The workhorse of elimination!)

3.4 The First Goal: The Staircase (Row Echelon Form - REF)

We use EROs systematically to create zeros below the main diagonal, organizing the matrix into a "staircase" shape. This process is called **Gaussian Elimination**, and the resulting shape is **Row Echelon Form (REF)**.

A matrix is in REF when:

1. The first non-zero number in any row (called the **pivot**) is strictly to the right of the pivot in the row above it.
2. All entries below a pivot are zero.

3. Any rows consisting entirely of zeros are at the bottom.

Visualizing REF: Let $\blacksquare$ represent a pivot (non-zero), and $*$ represent any number.

$$\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{Notice the staircase pattern.}$$

Why REF? It's the perfect setup for back-substitution.

3.5 The Ultimate Goal: The Perfect Tidy (Reduced Row Echelon Form - RREF)

We can take the "tidying" process further. Let's also create zeros *above* the pivots and make the pivots themselves equal to 1.

This ultimate form is **Reduced Row Echelon Form (RREF)**. The process of going all the way to RREF is called **Gauss-Jordan Elimination**.

RREF rules (stricter than REF):

1. It must be in REF.
2. Every pivot must be exactly 1.
3. Every pivot must be the *only* non-zero entry in its column (zeros below AND above).

Visualizing RREF:

$$\begin{bmatrix} 1 & 0 & * & 0 \\ 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{array}{l} \text{Pivots are 1.} \\ \text{Zeros above and below pivots.} \end{array}$$

Why RREF? It is the simplest possible form. You can read the answers directly without back-substitution.

Key Concept

Every matrix has one, and only one, RREF. It is a unique representation of the system.

4 Interpreting the RREF: The Three Cases Revisited

The RREF reveals exactly which of the three geometric possibilities we are dealing with.

4.1 Pivot Variables vs. Free Variables

We classify variables based on the columns in the RREF.

- **Pivot Variables (Basic Variables):** Variables whose columns contain a pivot (a leading '1').
- **Free Variables:** Variables whose columns do *not* contain a pivot.

Let's examine an RREF:

$$\left[\begin{array}{ccc|c} \mathbf{1} & 0 & 5 & 4 \\ 0 & \mathbf{1} & 2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Here, x_1 and x_2 (Columns 1 and 2) are pivot variables. x_3 (Column 3) is a free variable.

The Intuition

A "free" variable is literally free to be any value. The system doesn't provide enough constraints to solve for it uniquely. The pivot variables then *depend* on the value chosen for the free variable.

4.2 Analyzing the RREF

4.2.1 Case 1: No Solution (The Contradiction)

How to spot it: The RREF contains a row where all coefficients are zero, but the constant is non-zero.

$$\text{Example: } \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 0 & \mathbf{1} \end{array}\right]$$

The Meaning: The last row translates to the equation $0x_1 + 0x_2 = 1$, or $0 = 1$. This is impossible. The system is **inconsistent**. (Geometric parallel: Parallel lines).

4.2.2 Case 2: One Unique Solution (The Perfect Fit)

How to spot it: The system is consistent (no $0 = 1$ rows) AND there are **no free variables**.

$$\text{Example: } \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

The Meaning: Every variable is a pivot variable. Every variable is locked to a specific value. (Geometric parallel: Intersection at a point).

4.2.3 Case 3: Infinitely Many Solutions (The Under-constrained System)

How to spot it: The system is consistent AND there is **at least one free variable**.

$$\text{Example: } \left[\begin{array}{ccc|c} 1 & 0 & 5 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The Meaning: x_3 is free. We must describe the solution **parametrically**.

1. Assign a parameter to the free variable: Let $x_3 = t$ (where t can be any real number, $t \in \mathbb{R}$).
2. Express the pivot variables in terms of t.

- Row 2: $x_2 + 2x_3 = 3 \implies x_2 = 3 - 2t$
- Row 1: $x_1 + 5x_3 = 4 \implies x_1 = 4 - 5t$

The solution set is $(x_1, x_2, x_3) = (4 - 5t, 3 - 2t, t)$.

4.3 The Geometry of Infinite Solutions: Parametric Vector Form

When we have infinite solutions, we can rewrite the parametric solution in a way that reveals the underlying geometry. This is the **Parametric Vector Form**. We separate the constants from the parameters (t):

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4-5t \\ 3-2t \\ t \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} + \begin{pmatrix} -5t \\ -2t \\ 1t \end{pmatrix}$$

We factor out the t:

$$\mathbf{x} = \underbrace{\begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}}_{\text{A Specific Point}} + t \underbrace{\begin{pmatrix} -5 \\ -2 \\ 1 \end{pmatrix}}_{\text{Direction Vector}}$$

The Intuition

This is the equation of a line in 3D space! The solution set passes through the point $(4, 3, 0)$ and moves in the direction defined by the vector $\langle -5, -2, 1 \rangle$. If there is one free variable (t), the solution set is a line. If there are two free variables (e.g., s and t), the solution set is a plane.

5 Worked Examples: The Process in Action

Let's walk through the full Gauss-Jordan process, emphasizing the strategy behind each move.

5.1 Example 1: A Unique Solution (The Factory Problem)

Let's solve the motivating example from Section 2.1.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 15 \\ 2 & 1 & 2 & 25 \\ 4 & 2 & 3 & 40 \end{array}\right]$$

Elimination (to RREF):

Strategy: Use the pivot in R1C1 to eliminate the entries below it.

$$\xrightarrow[\substack{R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 4R_1}]{} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 15 \\ 0 & -1 & 0 & -5 \\ 0 & -2 & -1 & -20 \end{array}\right]$$

Strategy: Make the second pivot (R2C2) a 1.

$$\xrightarrow{R_2 \to -R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 15 \\ 0 & 1 & 0 & 5 \\ 0 & -2 & -1 & -20 \end{array}\right]$$

Strategy: Use the pivot in R2C2 to eliminate the entry below it.

$$\xrightarrow{R_3 \to R_3 + 2R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 15 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & -1 & -10 \end{array}\right]$$

(We have reached REF).

Strategy: Make the third pivot (R3C3) a 1.

$$\xrightarrow{R_3 \to -R_3} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 15 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 10 \end{array}\right]$$

Strategy: Work upwards (Gauss-Jordan step). Eliminate entries above the pivots.

$$\xrightarrow{R_1 \to R_1 - R_3} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 5 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 10 \end{array}\right] \xrightarrow{R_1 \to R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 10 \end{array}\right]$$

Interpretation: Consistent. No free variables. The unique solution is $x_1 = 0, x_2 = 5, x_3 = 10$. The factory should make 0 laptops, 5 phones, and 10 tablets.

5.2 Example 2: No Solution

Solve:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 3 \\ 2 & 3 & 2 & 6 \end{array}\right]$$

Elimination (to REF): *Strategy: Eliminate below the first pivot.*

$$\xrightarrow[\substack{R_2 \to R_2 - R_1 \\ R_3 \to R_3 - 2R_1}]{} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 2 \end{array}\right]$$

Strategy: Eliminate below the second pivot.

$$\xrightarrow{R_3 \to R_3 - R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Interpretation: We can stop here. The last row represents $0 = 1$. This is a contradiction. **No solution**.

5.3 Example 3: Infinitely Many Solutions

Solve:

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \\ 2 & 3 & 5 & 7 \end{array}\right]$$

Elimination (to RREF): *Strategy: Eliminate below the first pivot.*

$$\xrightarrow[\substack{R_2 \to R_2 - R_1 \\ R_3 \to R_3 - 2R_1}]{} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{array}\right]$$

Strategy: Eliminate below the second pivot.

$$\xrightarrow{R_3 \to R_3 - R_2} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Strategy: This is REF. Now eliminate above the second pivot.

$$\xrightarrow{R_1 \to R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Interpretation: Consistent. x_3 is a free variable. Let $x_3 = t$. $x_1 = 2 - t$ $x_2 = 1 - t$

Parametric Vector Form:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 - t \\ 1 - t \\ t \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

The solution is a line in 3D passing through $(2, 1, 0)$ in the direction $\langle -1, -1, 1 \rangle$.

6 A Different Perspective: The Inverse Matrix

For systems where the coefficient matrix A is square ($n \times n$, same number of equations as variables), we have another powerful theoretical tool.

6.1 The Concept of "Undoing"

In simple algebra, we solve $5x = 10$ by multiplying by the inverse of 5 (which is $1/5$): $x = 2$.

In matrix algebra, $A\mathbf{x} = \mathbf{b}$. If we can find an "Inverse Matrix" A^{-1} such that $A^{-1}A = I$ (where I is the Identity Matrix, the matrix equivalent of '1'), we can solve the system instantly:

$$A\mathbf{x} = \mathbf{b} \implies A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} \implies I\mathbf{x} = A^{-1}\mathbf{b} \implies \mathbf{x} = A^{-1}\mathbf{b}$$

6.2 When Does an Inverse Exist? (Invertibility)

Not all matrices have an inverse. A matrix is **invertible** (or non-singular) only if the system it represents has a unique solution.

Key Concept

A square matrix A ($n \times n$) is invertible if and only if its RREF is the identity matrix I. This means it must have n pivots (no free variables).

If a system has no solution or infinite solutions (like Examples 2 and 3 above), the matrix A is singular (not invertible).

6.3 A Note on Computation

Practical Note

While the inverse matrix A^{-1} is a crucial theoretical concept, it is rarely used to solve a single system $A\mathbf{x} = \mathbf{b}$ in practice (like in large-scale machine learning). Calculating A^{-1} is computationally expensive and can be numerically unstable (prone to rounding errors). The elimination methods (like Gaussian elimination or related techniques like LU decomposition) are generally faster and more reliable.

7 Summary Checklist

We have explored the foundations of solving linear systems, bridging the gap between geometry and algebra. Here is your checklist for solving any system using **Gauss-Jordan Elimination**:

1. **Setup:** Convert the system into an Augmented Matrix $[A|\mathbf{b}]$.
2. **Forward Phase (Gaussian Elimination to REF):** Use EROs to create the staircase (zeros below the pivots), working left to right.
3. **Consistency Check:** Analyze the REF. If you see a row like '[0 0 0 | 1]', stop. **NO SOLUTION**.
4. **Backward Phase (Gauss-Jordan step to RREF):** If consistent, ensure pivots are 1, and create zeros above the pivots, working right to left.
5. **Interpretation:** Analyze the RREF by identifying Pivot Variables and Free Variables.
   - If there are **no free variables**: **ONE UNIQUE SOLUTION**.
   - If there is **at least one free variable**: **INFINITELY MANY SOLUTIONS**. Use parameters (t, s, etc.) and express the solution in Parametric Vector Form.

Mastering this process is the first, crucial step in understanding the mathematical machinery behind machine learning.