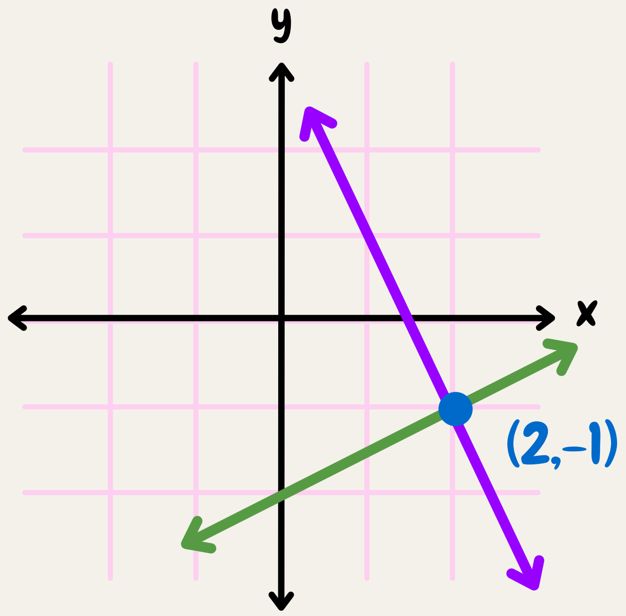


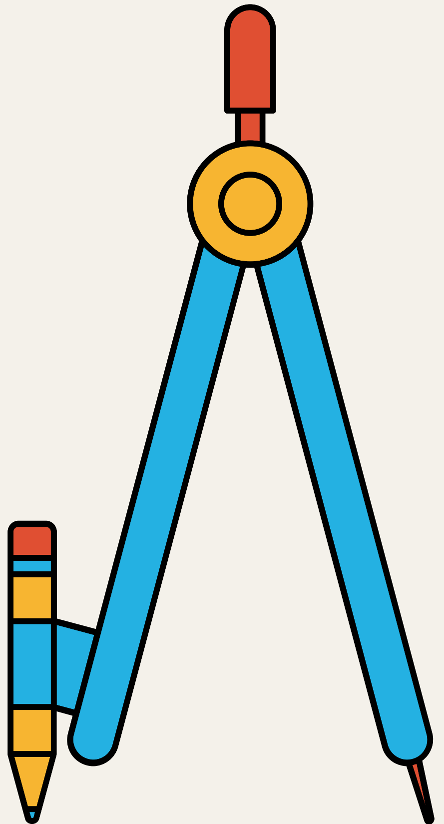
Linear Algebra Review

لاتنسونا من دعواتكم

الله يوفقنا ويوفقكم وبإذن الله نتقابل في الصيفية 🤲🏻

لو عندكم اي سؤال او اقتراح
هذي حسابتنا في X

Muath   Sara   Ice



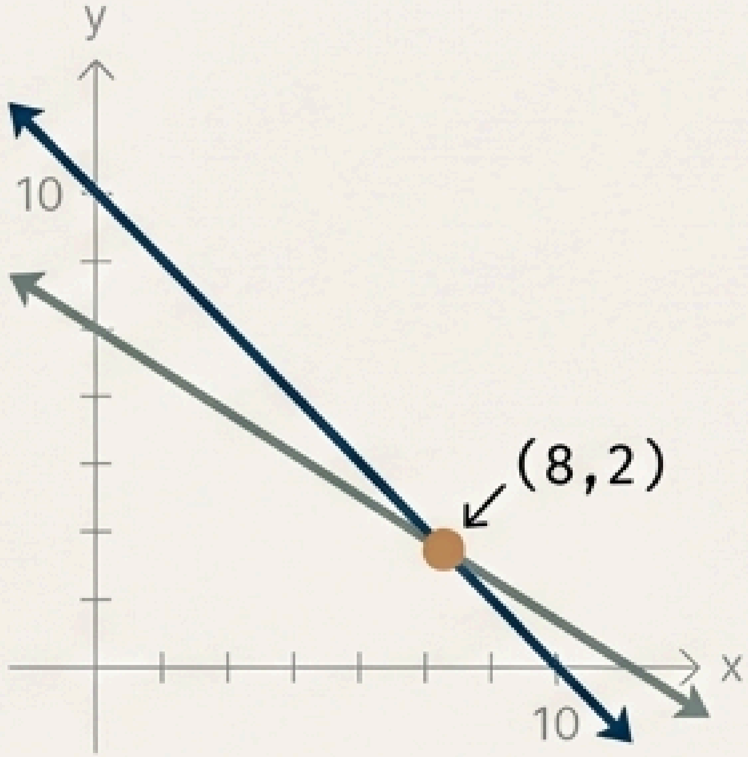
Every System of Equations Tells a Story

We can visualize systems of equations as the intersection of lines (in 2D) or planes (in 3D). The solution is the point where they meet. There are only three possible outcomes.

Unique Solution

$$a + b = 10$$
$$a + 2b = 12$$

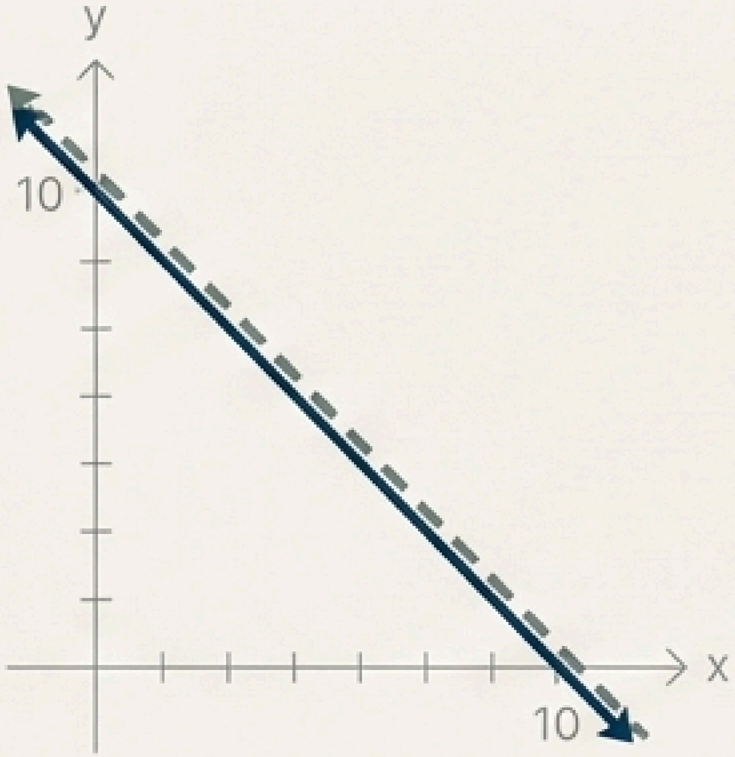
The equations intersect at exactly one point.



Infinite Solutions

$$a + b = 10$$
$$2a + 2b = 20$$

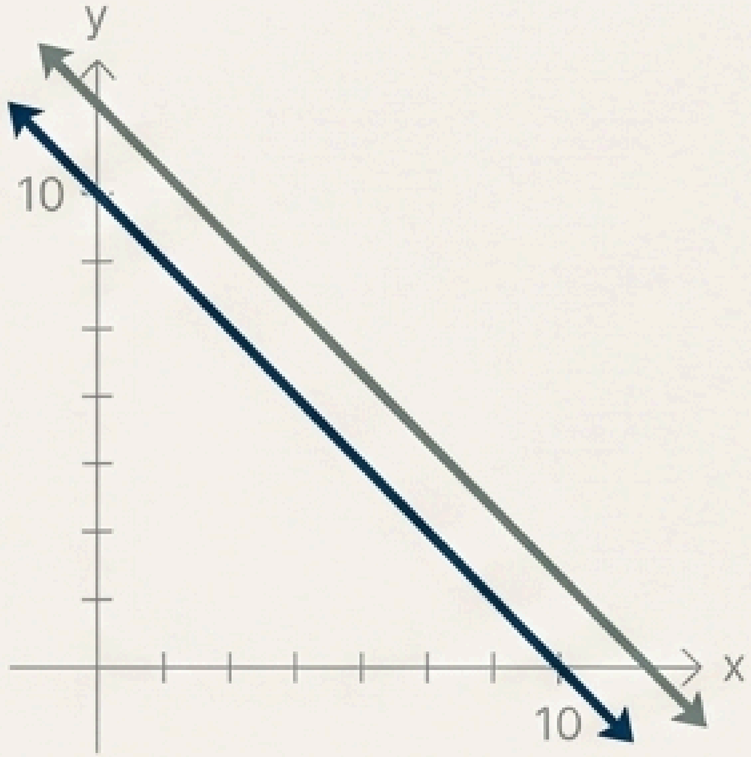
The equations represent the same line.



No Solution

$$a + b = 10$$
$$2a + 2b = 24$$

The lines are parallel and never intersect.



Matrices Are a Compact Way to Write Systems

Instead of writing full equations, we can represent the coefficients and solutions in a structure called a matrix. This is the foundation for manipulating data at scale.

System of Equations

$$\begin{aligned} 5a + 1b &= 17 \\ 4a - 3b &= 6 \end{aligned}$$

Augmented Matrix

$$\left[\begin{array}{cc|c} 5 & 1 & 17 \\ 4 & -3 & 6 \end{array}\right]$$

The Determinant: A Single Number to Test for a Unique Solution

The determinant is a scalar value calculated from a square matrix. Its value tells us if the system has a unique solution (non-singular) or not (singular). This is directly related to whether the rows of the matrix are linearly independent—meaning, no equation is redundant.

The Rule

$\det(A) \neq 0$ → **Non-singular (Unique Solution)**

$\det(A) = 0$ → **Singular (No or Infinite Solutions)**

Formula for 2x2 matrix

For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is

$$\det(A) = ad - bc$$

Examples

Non-singular

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

$$\det(A) = (1)(2) - (1)(1) = \mathbf{1}$$

Singular

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

$$\det(A) = (1)(2) - (1)(2) = \mathbf{0}$$

Vectors Represent Data Points in Space

A vector is an object with both magnitude (length) and direction. In ML, we use vectors to represent features of a data point. The length of a vector is calculated using its norm.

Representation

$\mathbf{u} = (4,\ 3)$

Magnitude (L2-Norm)

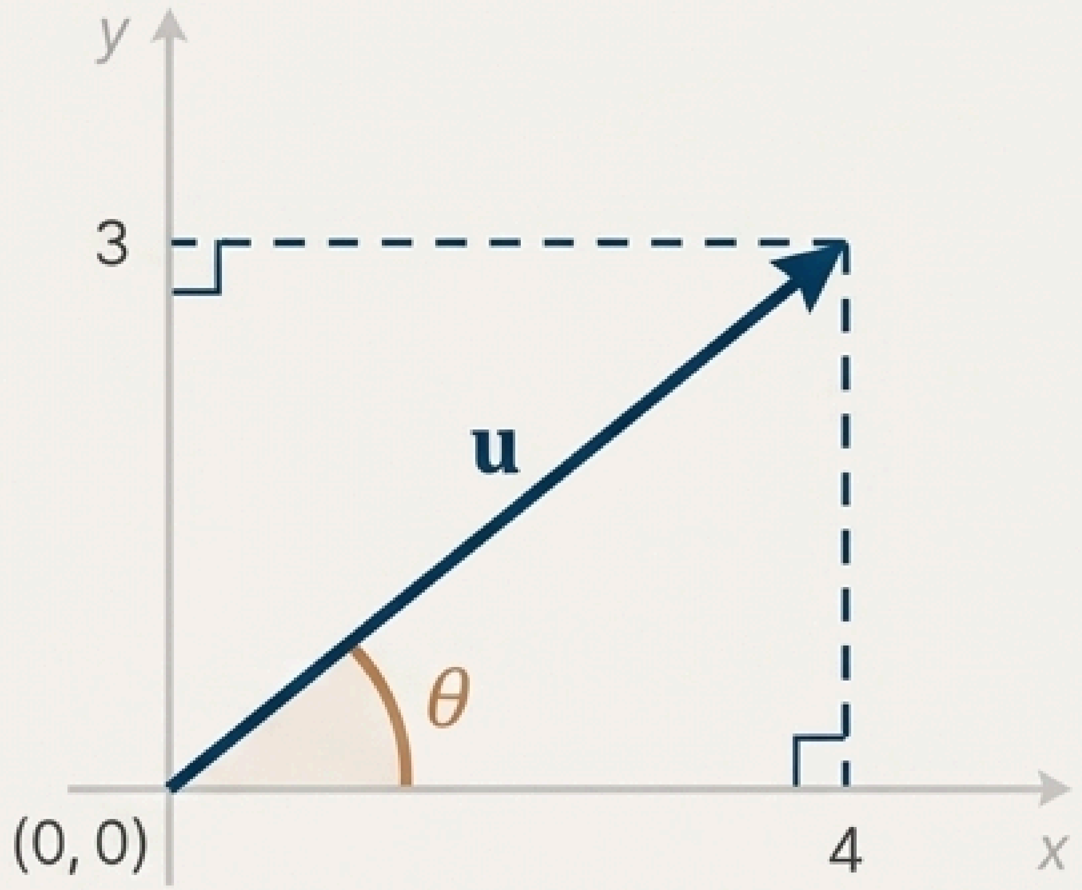
The straight-line distance from the origin.

$||\mathbf{u}||_2 = \sqrt{4^2 + 3^2} = \sqrt{25} = 5$

Direction

The angle it makes with an axis.

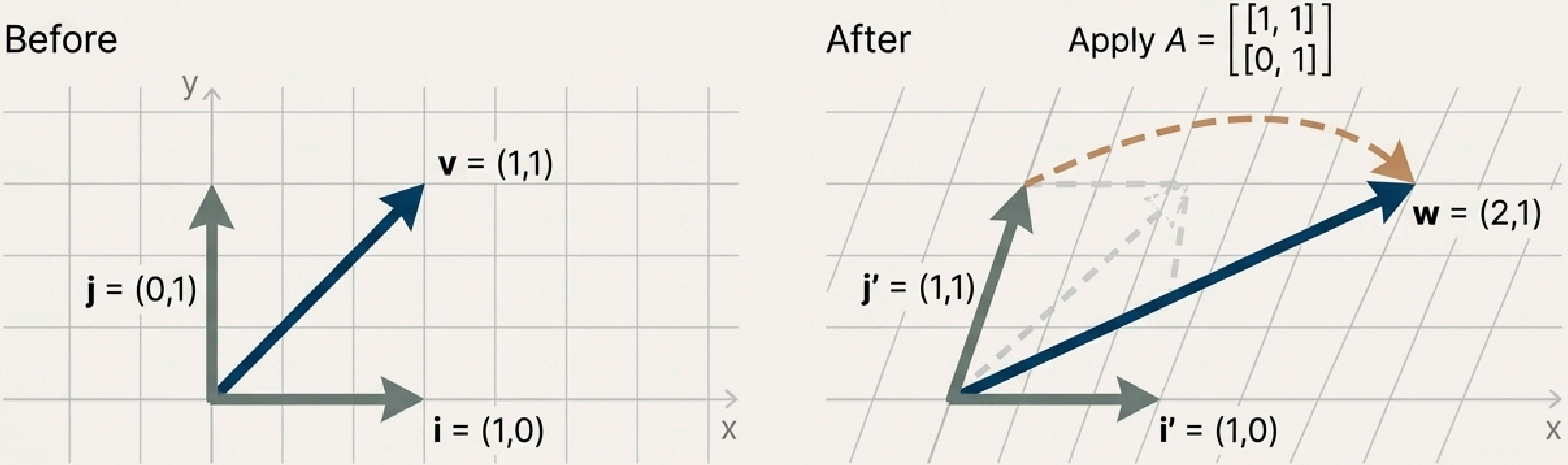
$\theta = \arctan(3/4) \approx 36.87°$



Why it matters for ML: A user profile with 10 features (age, location, etc.) can be represented as a single vector in a 10-dimensional space.

Matrices Act as Transformations on Vectors

Multiplying a matrix by a vector (Av = **w**) produces a new vector. This can be thought of as a transformation that moves, stretches, rotates, or shears the original vector.



$$\begin{bmatrix} [1,\ 1] \\ [0,\ 1] \end{bmatrix} * \begin{bmatrix} [1 \\ \ 1 \end{bmatrix} = [(1*1 + 1*1),\ (0*1 + 1*1))] = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The Matrix Inverse Undoes a Transformation

The inverse of a matrix **A**, written as A^{-1}, is a matrix that reverses the transformation of **A**. A matrix can only be inverted if it is non-singular ($\det(A) \neq 0$).

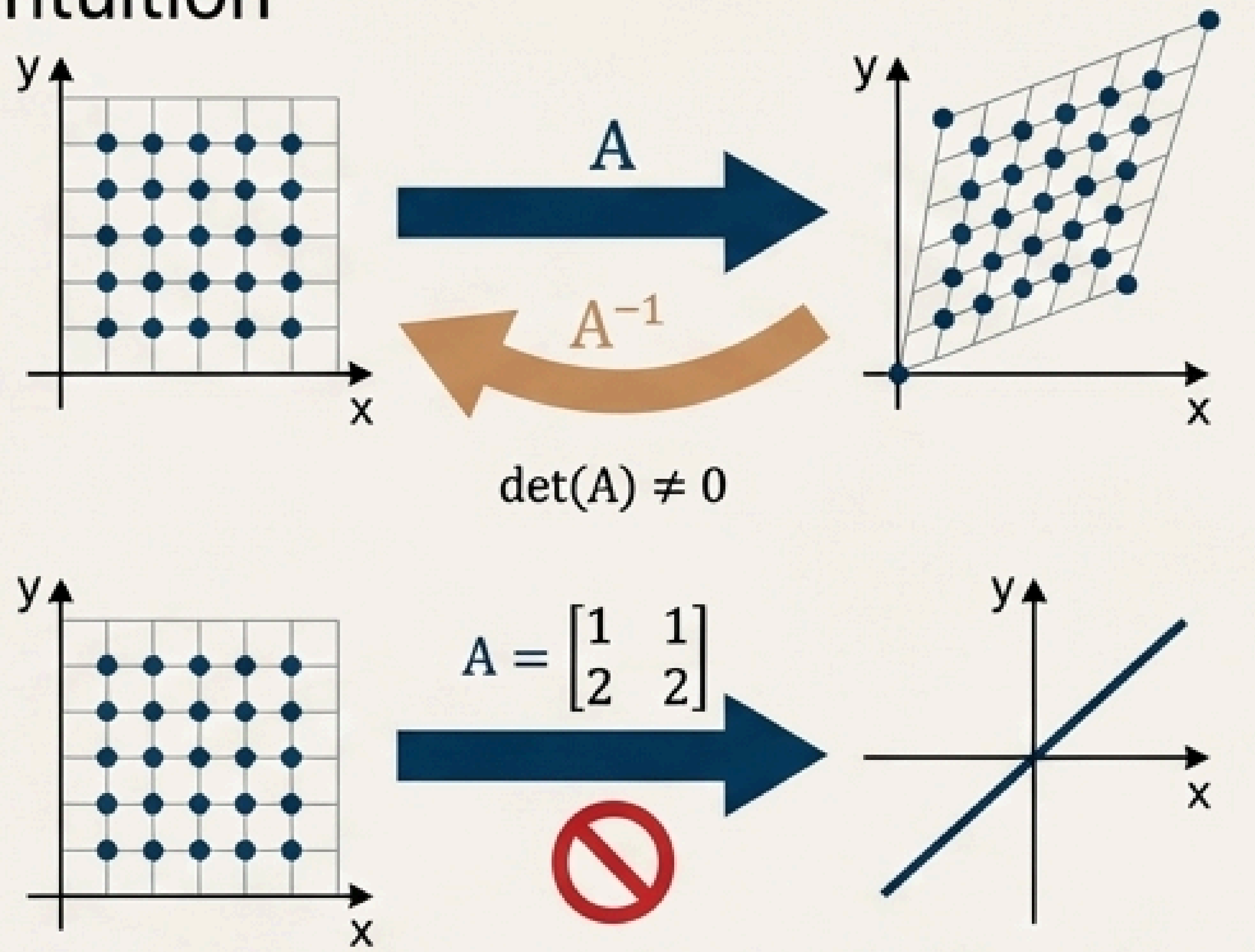
$$A * A^{-1} = I$$

(The Identity Matrix)

The Rule

A matrix is invertible if and only i its determinant is non-zero.

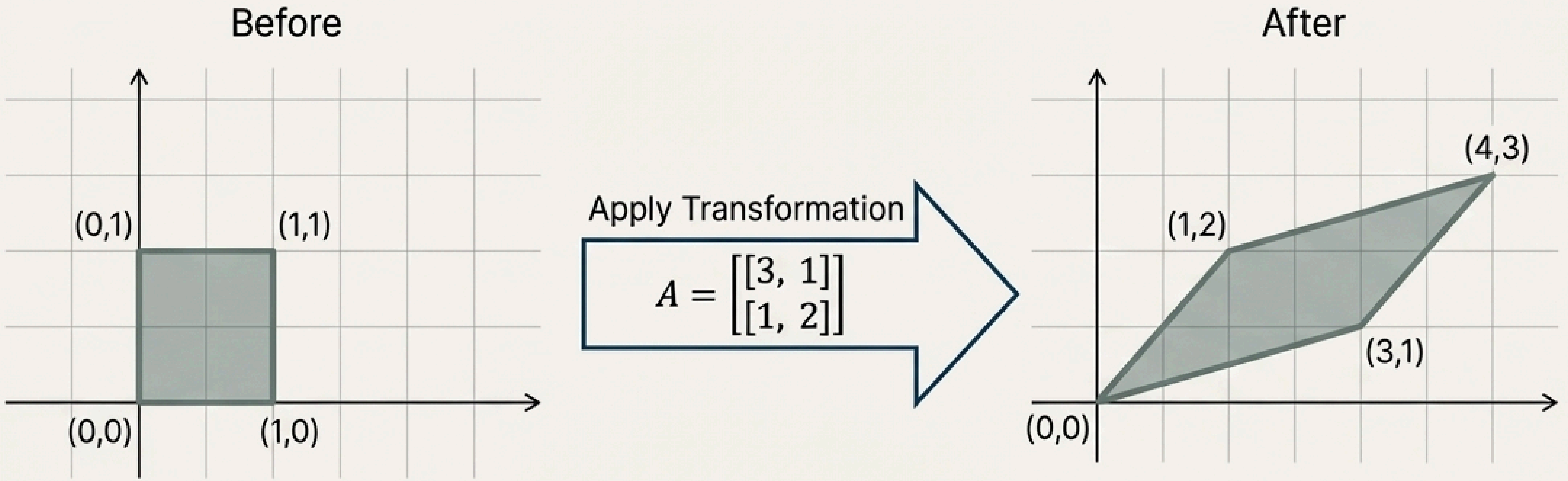
The Intuition



Information is lost. You cannot "un-squash" a line back into a plane. This is why a singular matrix has no inverse. $\det(A) = 0$.

The Determinant Measures How Much a Transformation Scales Space

The absolute value of the determinant, $|\det(A)|$, represents the scaling factor for area (in 2D) or volume (in 3D). It tells us how much space expands or contracts under the transformation A.



Area = 1

$\det(A) = (3)(2) - (1)(1) = 5$

The new area is $1 \times |\det(A)| = 5$.

Why it matters for ML: Large determinants in feature transformations can indicate that features are being amplified, which might affect model stability.

The Soul of a Matrix: Eigenvectors and Eigenvalues

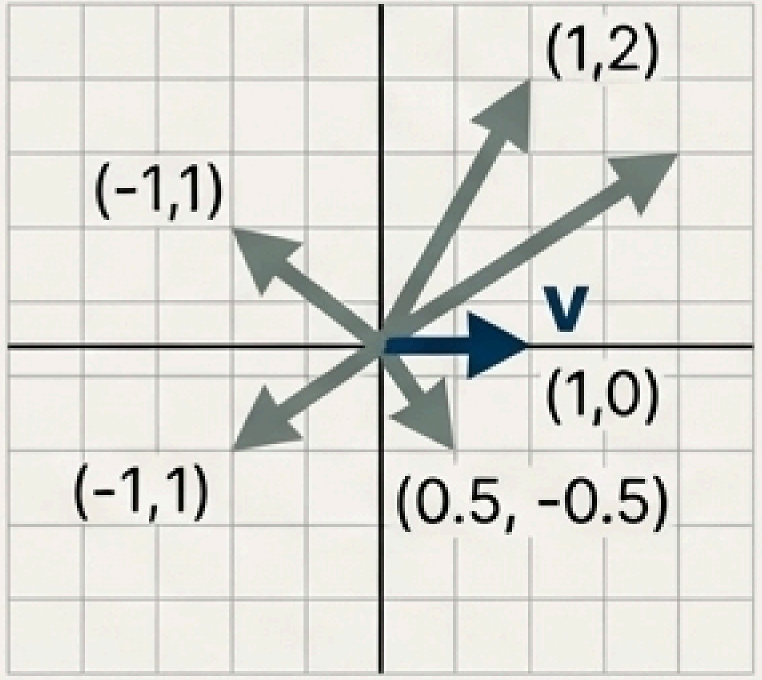
For any given transformation A, there are special vectors that do not change their direction. They only scale in length. These are the **eigenvectors**. The factor by which they scale is the **eigenvalue (λ)**.

The un-rotated vector

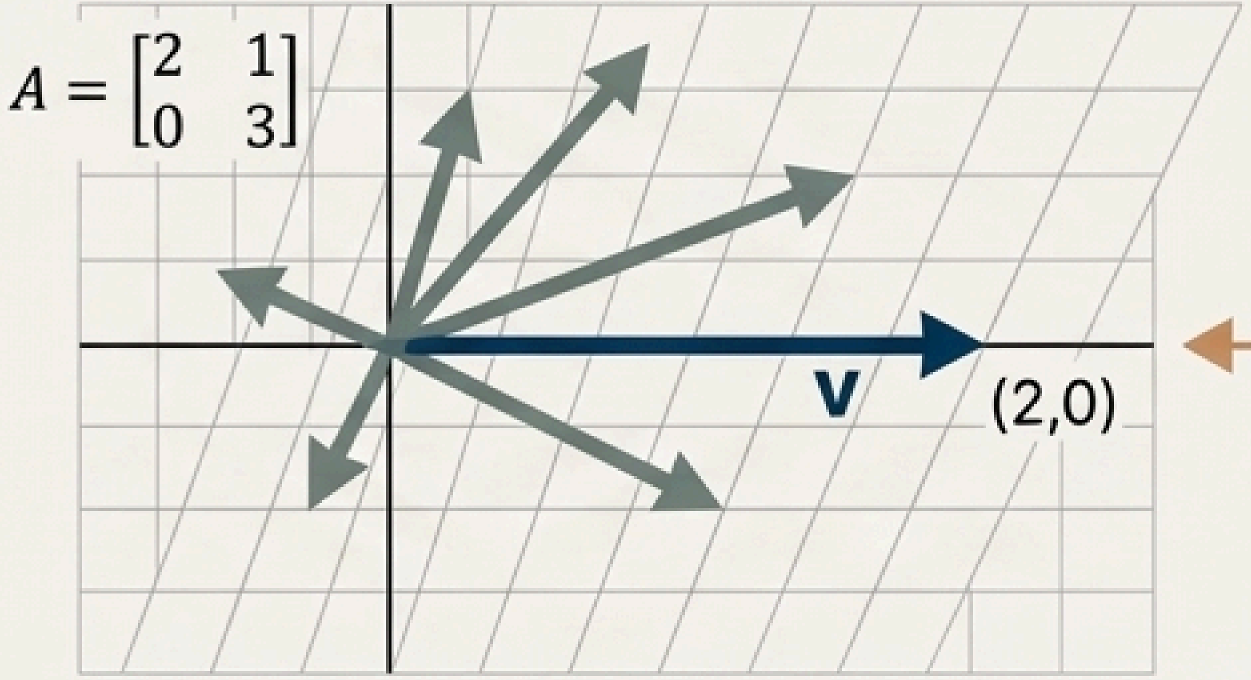
The transformation $\quad Av = \lambda v$

The scaling factor

Before



After



Eigenvector: (1,0). It only stretched. The new vector (2,0). The new vector (2,0) is `2 * (1,0). So, the **Eigenvalue** $\lambda = 2$.

How to Calculate Eigenvalues and Eigenvectors

To find the eigenvalues, we solve the characteristic equation $\det(A - \lambda I) = 0$. Once we have the eigenvalues, we plug them back into $(A - \lambda I)\mathbf{v} = 0$ to find the corresponding eigenvectors.

Example using $A = \left[\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}\right]$

Step 1: Find Eigenvalues (λ)

Set up the equation: $\det\left(\begin{bmatrix} 9-\lambda & 4 \\ 4 & 3-\lambda \end{bmatrix}\right) = 0$ ⟵ Intop the equation

Solve the polynomial: $(9-\lambda)(3-\lambda) - 16 = 0$

$\lambda^2 - 12\lambda + 11 = 0$

$(\lambda-11)(\lambda-1) = 0$

Result: $\lambda_1 = 11, \lambda_2 = 1$

Step 2: Find Eigenvector for $\lambda = 11$

Solve $(A - 11I)\mathbf{v} = 0$: $\left[\begin{bmatrix} -2 & 4 \\ 4 & -8 \end{bmatrix}\right] * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

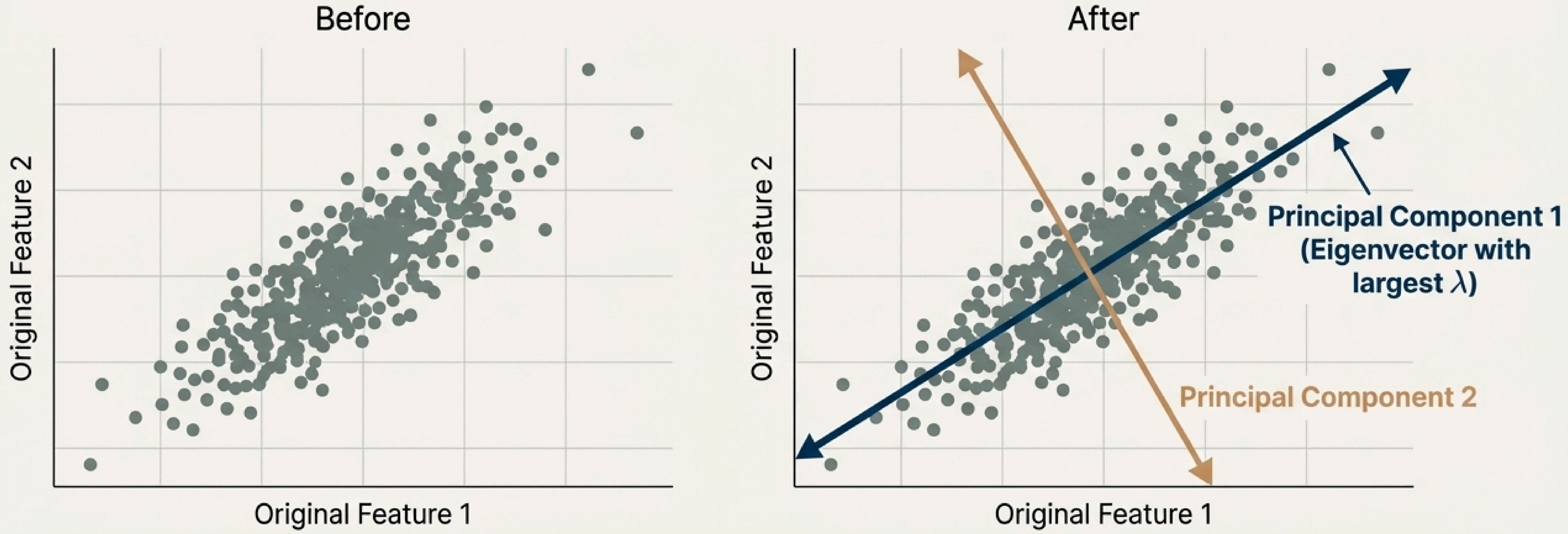
This gives the equation $-2x + 4y = 0$, which simplifies to $x = 2y$.

Result: A possible eigenvector is $\mathbf{v}_1 = (2, 1)$.

Putting It All Together: Principal Component Analysis (PCA)

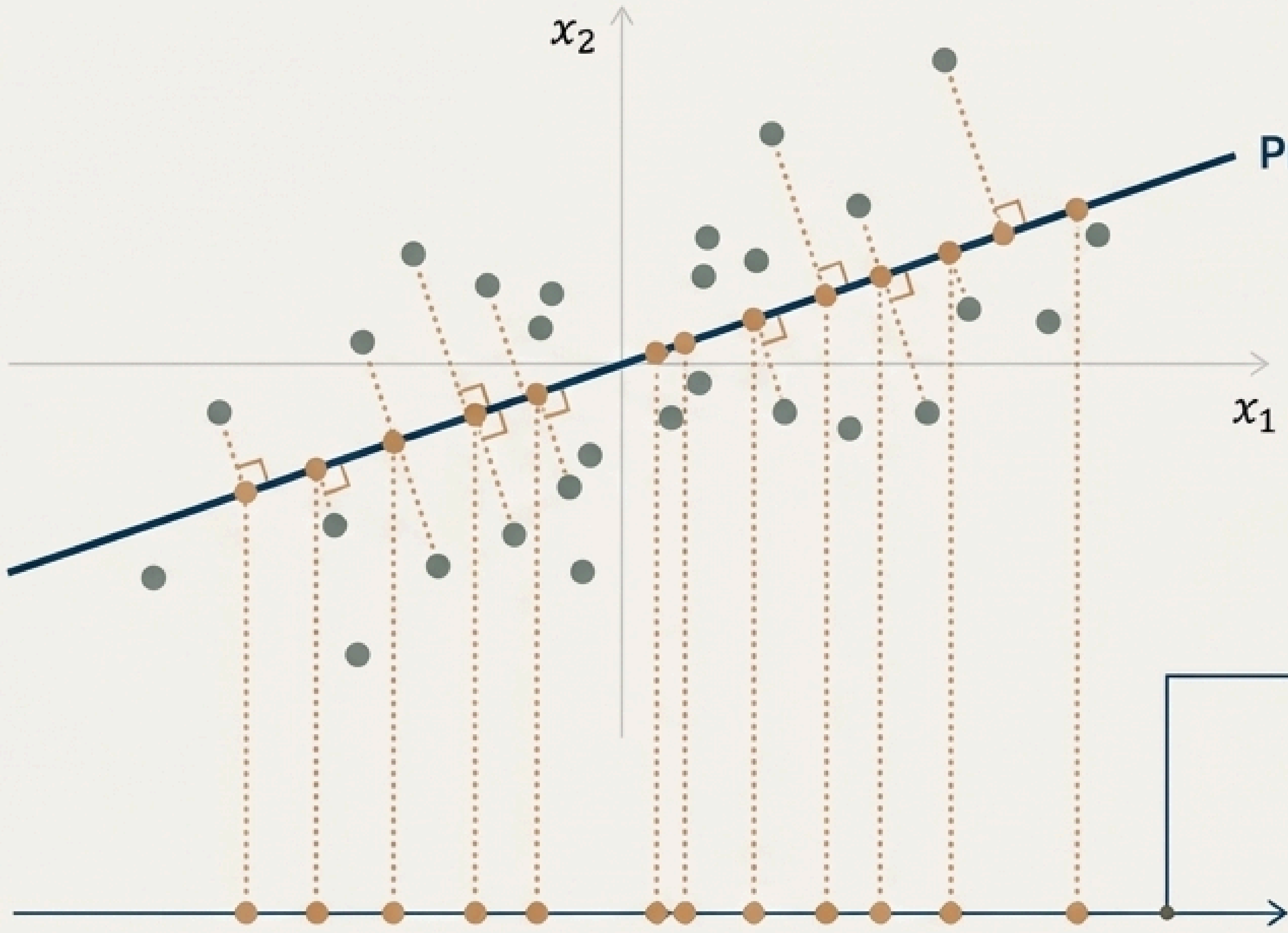
PCA is a technique to reduce the number of variables (dimensions) in a dataset while preserving as much information (variance) as possible. It finds a new, more optimal set of axes for the data.

> The eigenvectors of the data's covariance matrix are the Principal Components. The first principal component (eigenvector with the largest eigenvalue) is the direction of maximum variance in the data.



PCA Reduces Dimensions by Projecting Data

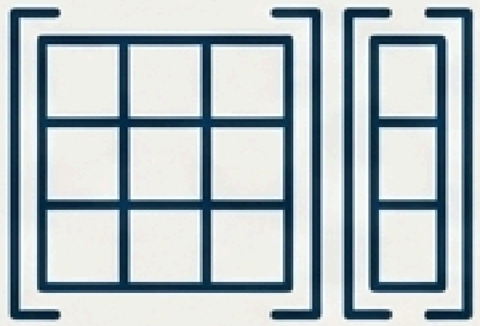
To reduce our data from 2D to 1D, we project the data points onto the first principal component. This single new feature captures the most important information that was previously spread across two features.



We've transformed a 2D dataset (x_1, x_2) into a 1D dataset (z_1) **while retaining the maximum possible variance.**

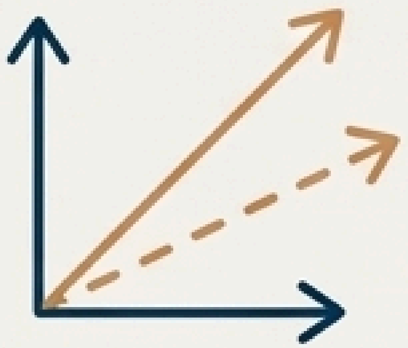
Linear Algebra is the Bedrock of Machine Learning

Understanding these building blocks provides the intuition behind how many machine learning algorithms work.



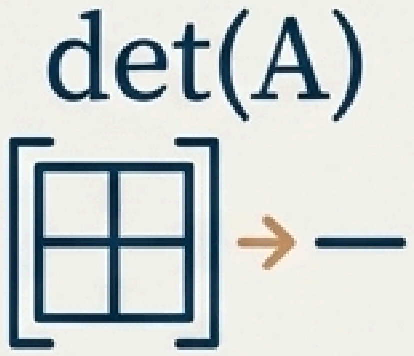
Vectors & Matrices

Representing data, features, and the weights of neural network layers.



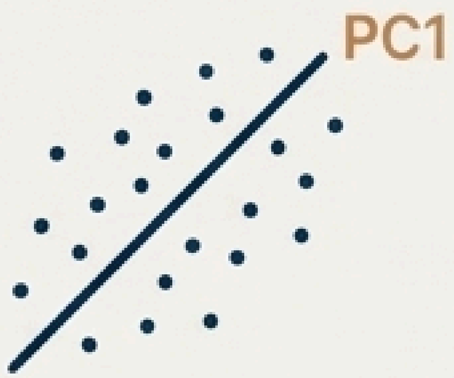
Matrix Multiplication & Transformations

Processing data through model layers, applying rotations and scaling in computer graphics.



Determinants & Inverses

Understanding model stability, solving for parameters in linear regression.



Eigenvectors & PCA

Dimensionality reduction, feature extraction, and identifying the most important patterns in data.