A-Z Linear Algebra & Calculus for AI, Data Science and Machine Learning

Instructor:

Rashedul Alam Shakil

Founder of Study Mart

Founder of aiQuest Intelligence

Employed in a Big Tech Company in Germany

Master in Data Science at FAU Erlangen

A-Z Linear Algebra & Calculus for AI, Data Science and Machine Learning

Instructor:
Rashedul Alam Shakil
Founder of Study Mart
Founder of aiQuest Intelligence
Master in Data Science at FAU Erlangen
Automation Programmer at Siemens Energy, Germany

Exam: Test your linear algebra skills!

Importance of Linear Algebra for AI/ML/DS

Watch Tutorial

Importance of Linear Algebra for AI/ML/DS

Linear algebra provides the foundation for many key mathematical equations used in Machine Learning (ML), Artificial Intelligence (AI), and Data Science (DS).

- **Core to Machine Learning Algorithms**
- Data Representation as Matrices & Vectors
- Matrix Operations Speed Up Computation
- Optimization & Gradient Descent
- Deep Learning & Neural Networks Depend on Matrices
- AI Applications Using Linear Algebra
- Dimensionality Reduction & Feature Engineering

Why is Linear Algebra Important for AI/ML?

- **Linear Regression:** Uses matrix equations to find the best-fit line.
- **Logistic Regression:** Uses dot product and sigmoid functions.
- **Support Vector Machines (SVMs):** Use vector spaces for classification.
- **Neural Networks:** Each layer applies matrix multiplications.
- **PCA (Dimensionality Reduction):** Uses eigenvalues & eigenvectors.

Importance of Linear Algebra for AI/ML/DS

- Core to Machine Learning Algorithms
- **Data Representation as Matrices & Vectors**
- **Matrix Operations Speed Up Computation**
- **Optimization & Gradient Descent**
- Deep Learning & Neural Networks Depend on Matrices
- AI Applications Using Linear Algebra
- Dimensionality Reduction & Feature Engineering

Why is Linear Algebra Important for AI/ML?

Marketing Spend	Administration	Transport
11452361	1368978	4717841
1625977	15137759	44389853
15344151	10114555	40793454
14437241	11867185	38319962
14210734	9139177	36616842
1318769	9981471	36286136
13461546	14719887	12771682
13029813	14553006	32387668
12054252	14871895	31161329

.xlsx or .csv file

$$\begin{bmatrix} 11452361 & 1368978 & 4717841 \\ 1625977 & 15137759 & 44389853 \\ 15344151 & 10114555 & 40793454 \\ 14437241 & 11867185 & 38319962 \\ 14210734 & 9139177 & 36616842 \\ 1318769 & 9981471 & 36286136 \\ 13461546 & 14719887 & 12771682 \\ 13029813 & 14553006 & 32387668 \\ 12054252 & 14871895 & 31161329 \end{bmatrix}$$

matrix of .csv file

Why is Linear Algebra Important for AI/ML?

- Matrix multiplication, transposition, and inverse make training models faster.
- Libraries like NumPy, TensorFlow, PyTorch optimize ML using vectorized operations.

- Linear regression equation:

$$y = X\beta + \epsilon$$

- Closed-form solution using the Normal Equation:

$$\beta = (X^T X)^{-1} X^T y$$

Why is Linear Algebra Important for AI/ML?

- Matrix multiplication, transposition, and inverse make training models faster.
- Libraries like NumPy, TensorFlow, PyTorch optimize ML using vectorized operations.
- Linear regression equation:

$$y = \beta_0 + \beta_1 X + \epsilon$$

$$\beta_1 = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

$$\beta_0 = \frac{\sum Y - \beta_1 \sum X}{N}$$

- Closed-form solution using the Normal Equation:

$$\beta = (X^T X)^{-1} X^T y$$

- Multivariate: For multivariate regression (multiple features), we need a more general form:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

$$\beta = (X^T X)^{-1} X^T y$$

β = Vector of coefficients $[\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]$ (Four Features)

Why is Linear Algebra Important for AI/ML?

- Multivariate:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

$$\beta = (X^T X)^{-1} X^T y$$

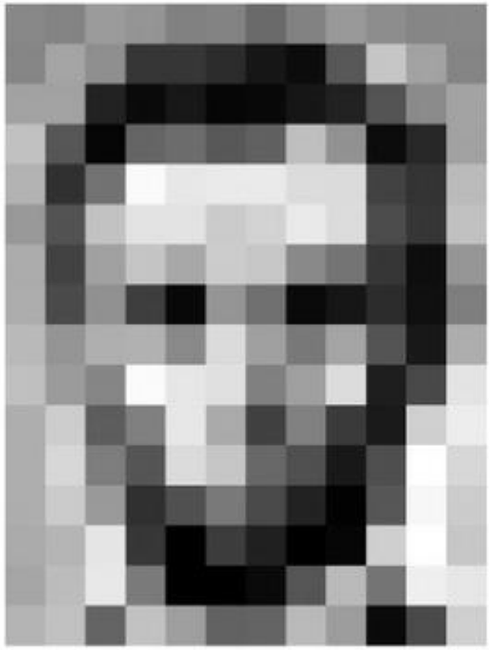
β = Vector of coefficients $[\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]$

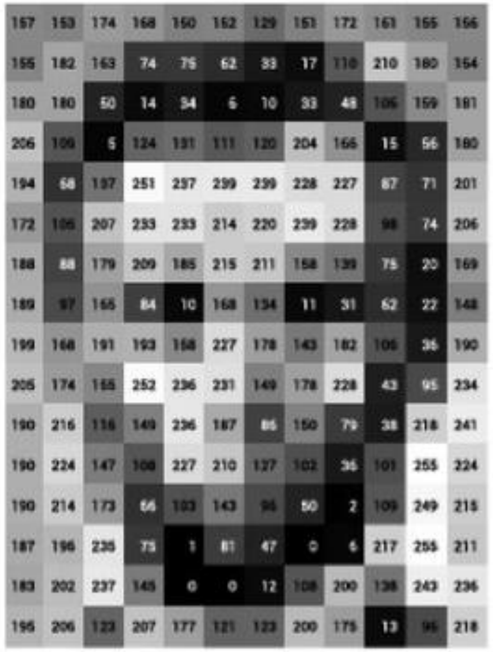
$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & X_{14} \\ 1 & X_{21} & X_{22} & X_{23} & X_{24} \\ 1 & X_{31} & X_{32} & X_{33} & X_{34} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Importance of Linear Algebra for AI/ML/DS

- Data Representation as Matrices & Vectors
- Core to Machine Learning Algorithms
- Matrix Operations Speed Up Computation
- Optimization & Gradient Descent
- **Deep Learning & Neural Networks Depend on Matrices**
- **AI Applications Using Linear Algebra**
- Dimensionality Reduction & Feature Engineering

Matrix Operations in Computer Vision





157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #1

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #2

0	1	1
0	1	0
1	-1	1

Kernel Channel #3

$$308 \quad + \quad -498 \quad + \quad 164 \quad + 1 = -25$$

Bias = 1

Output

-25				...
				...
				...
				...
...	...	...	...	...

Scaler, Vectors & Matrix

Watch Tutorial

Scaler, Vectors & Matrix

I will discuss the fundamental concepts in linear algebra, focusing on vectors, matrices, and their operations—essential for machine learning, deep learning, and data science.

- A scalar is a single number. Example: a=5

I will discuss the fundamental concepts in linear algebra, focusing on vectors, matrices, and their operations—essential for machine learning, deep learning, and data science.

- A scalar is a single number. Example: a=5
- A vector is just a **single-column** matrix (n×1). Used in **feature representation** in machine learning.

$$y = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix} \quad \text{(vector, size: } 3 \times 1\text{)}$$

Scaler, Vectors & Matrix

I will discuss the fundamental concepts in linear algebra, focusing on vectors, matrices, and their operations—essential for machine learning, deep learning, and data science.

- A scalar is a single number. Example: a=5
- A vector is just a **single-column** matrix (n×1). Used in **feature representation** in machine learning.

$$y = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix} \quad (\text{vector, size: } 3 \times 1)$$

Rows(n) x Columns(p)

Scenario: Predicting House Prices. Features of the house can be represented as a vector.

$$\mathbf{x} = \begin{bmatrix} \text{Size (sq ft)} \\ \text{Bedrooms} \\ \text{Age (years)} \end{bmatrix}$$

Scaler, Vectors & Matrix

I will discuss the fundamental concepts in linear algebra, focusing on vectors, matrices, and their operations—essential for machine learning, deep learning, and data science.

- A scalar is a single number. Example: a=5
- A vector is just a **single-column** matrix (n×1). Used in **feature representation** in machine learning.
- A matrix is a **grid of numbers** with multiple rows and columns (n×p). Used to **store datasets, perform transformations, and represent systems of equations**.

$$y = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix} \quad (\text{vector, size: } 3 \times 1)$$

Rows(n) x Columns(p)

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \quad (\text{matrix, size: } 2 \times 3)$$

Scaler, Vectors & Matrix

(1) **Feature Matrix (X)**

$$X = \begin{bmatrix} 2000 & 3 & 10 \\ 1500 & 2 & 5 \\ 2500 & 4 & 20 \end{bmatrix}_{3\times 3}$$

Shape: $(3, 3)$ → 3 rows (houses) × 3 columns (features)

Type: Matrix

Assume a Tabular Dataset

Size(sq. ft.)	Bedroom	Age	Price
2000	3	10	30000
1500	2	5	15000
2500	4	20	?

(2) **Target Vector (Y)**

$$Y = \begin{bmatrix} 30000 \\ 15000 \\ ? \end{bmatrix}_{3\times 1}$$

Shape: $(3, 1)$ → 3 rows × 1 column

Type: Column Vector

Notation & Terminologies in Linear Algebra

Notation & Terminologies

- $\mathbb{R}$ → Set of real numbers
- $\mathbb{R}^n$ → n-dimensional real space (vector space)
- $\mathbb{R}^{m \times n}$ → Set of all $m \times n$ matrices
- $A \in \mathbb{R}^{m \times n}$ → Matrix A with dimensions $m \times n$
- $\mathbf{v} \in \mathbb{R}^n$ → A vector in n-dimensional space
- A^T → Transpose of matrix A
- A^{-1} → Inverse of matrix A
- $\det(A)$ → Determinant of matrix A

Notation & Terminologies

- A^{-1} → Inverse of matrix A
- $\det(A)$ → Determinant of matrix A
- $\mathrm{tr}(A)$ → Trace of matrix A
- $||\mathbf{v}||$ → Norm (magnitude) of vector $\mathbf{v}$
- $\langle \mathbf{u}, \mathbf{v} \rangle$ → Inner product (dot product) of vectors $\mathbf{u}, \mathbf{v}$
- $A\mathbf{x} = \mathbf{b}$ → Linear system of equations
- λ → Eigenvalue of matrix
- $\mathbf{v}$ → Eigenvector
- $A\mathbf{v} = \lambda\mathbf{v}$ → Eigenvalue equation
- $U\Sigma V^T$ → Singular Value Decomposition (SVD)
- $||A||_F$ → Frobenius norm of matrix A

Notation & Terminologies

- Σ → Covariance matrix
- $\mathbb{E}[X]$ → Expectation (mean) of random variable X
- $\mathrm{Var}(X)$ → Variance of X
- $\mathrm{Cov}(X, Y)$ → Covariance between X and Y
- $KL(P||Q)$ → Kullback-Leibler divergence
- $D_{KL}(P||Q)$ → KL divergence between distributions P and Q
- $\sum$ → Summation (used in linear algebra and statistics)
- $\prod$ → Product notation
- Σ^{-1} → **Precision matrix** (inverse of the covariance matrix)
- $(A^T A)^{-1}$ → Inverse of a Gram matrix (used in least squares regression)

Sets, Subsets, Vector Space

Watch Tutorial

Sets & Subsets

- **Set** (S) → A collection of elements $S = \{a, b, c\}$
- **Subset** ($A \subseteq B$) → All elements of A are in B
- **Universal Set** (U) → The complete set under consideration
- **Empty Set** ($\emptyset$) → A set with no elements
- **Union** ($A \cup B$) → All elements from both A and B
- **Intersection** ($A \cap B$) → Common elements between A and B

1. **Set S → A collection of elements,** $S = \{a, b, c, d\}$
2. **Subset $A \subseteq B$ → All elements of A are in B,** $A = \{1, 2\}, \quad B = \{1, 2, 3, 4\}$
3. **Universal Set U → The complete set under consideration,** $U = \{1, 2, 3, 4, 5, 6, \ldots\}$
4. **Empty Set $\emptyset$ → A set with no elements,** $\emptyset = \{\}$
5. **Union $A \cup B$ → All elements from both A and B,**

$$A = \{1, 2, 3\}, \quad B = \{3, 4, 5\} \quad \text{So, } A \cup B = \{1, 2, 3, 4, 5\}$$

6. **Intersection $A \cap B$ → Common elements between A and B**

$$A = \{1, 2, 3\}, \quad B = \{3, 4, 5\}$$

$$A \cap B = \{3\}$$

Vector Space

A **vector space** (linear space) is a set of vectors that can be added and scaled while remaining in the same space. It follows **vector addition** and **scalar multiplication** rules.

For Example:

The set of all 2D vectors $\mathbf{v} = (x, y)$ forms the **vector space** $\mathbb{R}^2$, since it follows vector addition and scalar multiplication properties.

Vector Space

A **vector space** (linear space) is a set of vectors that can be added and scaled while remaining in the same space. It follows **vector addition** and **scalar multiplication** rules.

Properties:

1️⃣ Closure Under Addition
2️⃣ Closure Under Scalar Multiplication
3️⃣ Commutativity of Vector Addition
4️⃣ Associativity of Vector Addition
5️⃣ Existence of a Zero Vector
6️⃣ Existence of Additive Inverses
7️⃣ Associativity of Scalar Multiplication
8️⃣ Distributive Property (Scalar over Vector Addition)
9️⃣ Distributive Property (Vector over Scalar Addition)
🔟 Multiplicative Identity

✅ If a Set Satisfies All 10 Properties, It is a Vector Space.

Subspaces & Linear Independence

A **subspace** is any subset of a vector space that is closed under **addition and scalar multiplication**.

Example:

- The set of all vectors $(x, 0)$ where x is any real number forms a **subspace** of $\mathbb{R}^2$, since it is a **line** passing through the origin. $v(x,y) = \begin{bmatrix} x \\ y \end{bmatrix}$ $v(0,y) = \begin{bmatrix} 0 \\ y \end{bmatrix}$

Subspaces & Linear Independence

A **subspace** is any subset of a vector space that is closed under **addition and scalar multiplication**.

Example:

- The set of all vectors $(x, 0)$ where x is any real number forms a **subspace** of $\mathbb{R}^2$, since it is a **line** passing through the origin.

$$v(x, y) = \begin{bmatrix} x \\ y \end{bmatrix} \qquad v(0, y) = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

A set of vectors is **linearly independent** if none of them can be written as a linear combination of the others.

Example:

Vectors $\mathbf{v_1} = (1, 0)$ and $\mathbf{v_2} = (0, 1)$ in $\mathbb{R}^2$ are **independent**. However, $(2, 4)$ and $(1, 2)$ are **dependent** because $(2, 4) = 2(1, 2)$.

What is Norm?

Norms

A norm is a function that assigns a non-negative length or magnitude to a vector. The distance between two vectors can be computed using a norm, specifically the L2 norm (Euclidean norm) or other norm-based distance metrics. Mathematically, the norm of a vector v is written as:

$$||v||$$

- Lp Norm (General Form)
- L1 Norm (Manhattan Norm)
- L2 Norm (Euclidean Norm)
- L∞ Norm (Chebyshev Distance)

Matrix Addition & Subtraction Operations in Linear Algebra

Watch Tutorial

Addition and Subtraction

1. **Addition:** If we have two matrices A and B of the same size $m \times n$, their sum C is obtained by **adding** corresponding elements: $C=A+B$.
2. **Subtraction:** Similarly, matrix subtraction is performed element-wise: $C=A-B$.

Addition:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$B = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 1+7 & 2+8 & 3+9 \\ 4+10 & 5+11 & 6+12 \end{bmatrix} = \begin{bmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}$$

Subtraction:

$$A = \begin{bmatrix} 10 & 9 & 8 \\ 7 & 6 & 5 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 10-1 & 9-2 & 8-3 \\ 7-4 & 6-5 & 5-6 \end{bmatrix} = \begin{bmatrix} 9 & 7 & 5 \\ 3 & 1 & -1 \end{bmatrix}$$

Can You Add or Subtract Matrices of Different Sizes?

❌ **No**, they must be the same size!

For example:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix}$$

Cannot be added or subtracted because their dimensions **do not match**.

Why is Scalar Addition Not Defined?

- In matrix operations, addition is only defined between two matrices of the **same dimensions**.
- A scalar is a single number that does not have the structure to align with each matrix element for direct addition.

Alternative Interpretation: Scalar Broadcasting

$$A + s = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + s = \begin{bmatrix} a_{11} + s & a_{12} + s \\ a_{21} + s & a_{22} + s \end{bmatrix}$$

Scaler Addition:

$$A = \begin{bmatrix} 2 & 5 \\ 7 & 1 \end{bmatrix}$$

$$A + 3 = \begin{bmatrix} 2+3 & 5+3 \\ 7+3 & 1+3 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 10 & 4 \end{bmatrix}$$

Why is Scalar Addition Not Defined?

- In matrix operations, addition is only defined between two matrices of the **same dimensions**.
- A scalar is a single number that does not have the structure to align with each matrix element for direct addition.

Alternative to Scaler Broadcasting:

$$A = \begin{bmatrix} 2 & 5 \\ 7 & 1 \end{bmatrix}$$

Create a One-Matrix, $J = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$$s \times J = 3 \times \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$$

$$A + (s \times J) = \begin{bmatrix} 2 & 5 \\ 7 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 2+3 & 5+3 \\ 7+3 & 1+3 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 10 & 4 \end{bmatrix}$$

Addition and Subtraction

Property	Matrix Addition ($A + B$)	Matrix Subtraction ($A - B$)
Commutative	✅ $A + B = B + A$	❌ $A - B \neq B - A$
Associative	✅ $(A + B) + C = A + (B + C)$	✅ $A - (B - C) = (A - B) + C$
Additive Identity	✅ $A + O = A$	✅ $A - O = A$
Additive Inverse	✅ $A + (-A) = O$	✅ $A - A = O$

Trace of Matrix

Trace Calculation

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \longrightarrow \mathrm{tr}(A) = 1 + 5 + 9 = 15$$

The trace itself is a scalar (single number), not a matrix. However, if you create a matrix where all diagonal elements are the trace value, it forms a scalar multiple of the identity matrix:

$$T = \mathrm{tr}(A)I = \begin{bmatrix} 15 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 15 \end{bmatrix} = ?$$

Identity Matrix, $I_n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Let’s Solve:

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 2 & 1 \\ 6 & 4 & 3 \\ 7 & 9 & 8 \end{bmatrix}$$

(i). Find A+B.

(ii). Find A-B.

Comment below: **tr(A+B)** and **tr(A+B).**

Solutions

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 2 & 1 \\ 6 & 4 & 3 \\ 7 & 9 & 8 \end{bmatrix}$$

(i). Find A+B.

(ii). Find A-B.

Question: **tr(A+B)** and **tr(A+B).**

$$A + B = \begin{bmatrix} 7 & 6 & 7 \\ 7 & 7 & 8 \\ 14 & 17 & 17 \end{bmatrix} \qquad \mathrm{tr}(A+B) = 7 + 7 + 17 = 31$$

$$A - B = \begin{bmatrix} -3 & 2 & 5 \\ -5 & -1 & 2 \\ 0 & -1 & 1 \end{bmatrix} \qquad \mathrm{tr}(A-B) = (-3) + (-1) + 1 = -3$$

Matrix Multiplication in Linear Algebra

Watch Tutorial

Rules for Matrix Multiplication

Given two matrices A and B:

- A must have **the same number of columns** as the number of rows in B.
- If A is of shape $(m \times n)$ and B is of shape $(n \times p)$, then the product AB will have shape $(m \times p)$.

$$\boxed{A_{m\times n} \times B_{n\times p} = C_{m\times p}}$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

2×2 2×2

$$C_{ij} = \sum(\text{Row of } A \times \text{Column of } B)$$

$$C = \begin{bmatrix} (1\times5+2\times7) & (1\times6+2\times8) \\ (3\times5+4\times7) & (3\times6+4\times8) \end{bmatrix}$$

$$C = \begin{bmatrix} (5+14) & (6+16) \\ (15+28) & (18+32) \end{bmatrix}$$

$$C = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

2×2

Rules for Matrix Multiplication

Given two matrices A and B:

- A must have **the same number of columns** as the number of rows in B.
- If A is of shape $(m \times n)$ and B is of shape $(n \times p)$, then the product AB will have shape $(m \times p)$.

$$\boxed{A_{m\times n} \times B_{n\times p} = C_{m\times p}}$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$C_{ij} = \sum(\text{Row of } A \times \text{Column of } B)$$

$$C = \begin{bmatrix} (1 \times 5 + 2 \times 7) & (1 \times 6 + 2 \times 8) \\ (3 \times 5 + 4 \times 7) & (3 \times 6 + 4 \times 8) \end{bmatrix}$$

$$C = \begin{bmatrix} (5 + 14) & (6 + 16) \\ (15 + 28) & (18 + 32) \end{bmatrix}$$

$$C = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Rules for Matrix Multiplication

Given two matrices A and B:

- A must have **the same number of columns** as the number of rows in B.
- If A is of shape $(m \times n)$ and B is of shape $(n \times p)$, then the product AB will have shape $(m \times p)$.

$$\boxed{A_{m\times n} \times B_{n\times p} = C_{m\times p}}$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$C_{ij} = \sum (\text{Row of } A \times \text{Column of } B)$$

$$C = \begin{bmatrix} (1\times5+2\times7) & (1\times6+2\times8) \\ (3\times5+4\times7) & (3\times6+4\times8) \end{bmatrix}$$

$$C = \begin{bmatrix} (5+14) & (6+16) \\ (15+28) & (18+32) \end{bmatrix}$$

$$C = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Rules for Matrix Multiplication

Given two matrices A and B:

- A must have **the same number of columns** as the number of rows in B.
- If A is of shape $(m \times n)$ and B is of shape $(n \times p)$, then the product AB will have shape $(m \times p)$.

$$\boxed{A_{m\times n} \times B_{n\times p} = C_{m\times p}}$$

$$C_{ij} = \sum(\text{Row of } A \times \text{Column of } B)$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$C = \begin{bmatrix} (1\times 5 + 2\times 7) & (1\times 6 + 2\times 8) \\ (3\times 5 + 4\times 7) & (3\times 6 + 4\times 8) \end{bmatrix}$$

$$C = \begin{bmatrix} (5+14) & (6+16) \\ (15+28) & (18+32) \end{bmatrix}$$

$$C = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Rules for Matrix Multiplication

Given two matrices A and B:

- A must have **the same number of columns** as the number of rows in B.
- If A is of shape $(m \times n)$ and B is of shape $(n \times p)$, then the product AB will have shape $(m \times p)$.

$$\boxed{A_{m\times n} \times B_{n\times p} = C_{m\times p}}$$

$$C_{ij} = \sum(\text{Row of } A \times \text{Column of } B)$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$C = \begin{bmatrix} (1\times 5 + 2\times 7) & (1\times 6 + 2\times 8) \\ (3\times 5 + 4\times 7) & (3\times 6 + 4\times 8) \end{bmatrix}$$

$$C = \begin{bmatrix} (5+14) & (6+16) \\ (15+28) & (18+32) \end{bmatrix}$$

$$C = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Rules for Matrix Multiplication

Given two matrices A and B:

- A must have **the same number of columns** as the number of rows in B.
- If A is of shape $(m \times n)$ and B is of shape $(n \times p)$, then the product AB will have shape $(m \times p)$.

$$\boxed{A_{m\times n} \times B_{n\times p} = C_{m\times p}}$$

$$C_{ij} = \sum(\text{Row of } A \times \text{Column of } B)$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$C = \begin{bmatrix} (1\times 5 + 2\times 7) & (1\times 6 + 2\times 8) \\ (3\times 5 + 4\times 7) & (3\times 6 + 4\times 8) \end{bmatrix}$$

$$C = \begin{bmatrix} (5+14) & (6+16) \\ (15+28) & (18+32) \end{bmatrix}$$

$$\boxed{C = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}}$$

Examples of Valid and Invalid Multiplication

Matrix A (size)	Matrix B (size)	Valid?	Result Size
2×3	3×4	✅ Yes	2×4
3×2	2×3	✅ Yes	3×3
2×2	2×2	✅ Yes	2×2
3×2	3×3	❌ No	Not possible

Matrix A has 2 columns, while matrix B has 3 rows.
Cols of A=Rows of B (2 ≠ 3), multiplication is not possible.

Key Rule: Cols of A = Rows of B for multiplication to be valid.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad (\text{Size: } 3 \times 2)$$

Not possible

$$B = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{bmatrix} \quad (\text{Size: } 3 \times 3)$$

Properties of Matrix Multiplication

Property	Formula	Holds Always?
Associative	$(AB)C = A(BC)$	✅ Yes
Distributive	$A(B+C) = AB + AC$	✅ Yes
Identity Matrix	$AI = IA = A$	✅ Yes
Zero Matrix	$AO = OA = O$	✅ Yes
Commutative	$AB = BA$	❌ No (Not always)

Identity Matrix, $I_n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Let's Solve:

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 2 & 1 \\ 6 & 4 & 3 \\ 7 & 9 & 8 \end{bmatrix}$$

(i). Find $A*B$.

(ii). Find $A*B$.

Comment below: **tr(A*B)** and **tr(A*B).**

Solution:

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 2 & 1 \\ 6 & 4 & 3 \\ 7 & 9 & 8 \end{bmatrix}$$

(i). Find A*B.

(ii). Find B*A.

Comment below: **tr(A*B)** and **tr(B*A).**

$$A \times B = \begin{bmatrix} 76 & 74 & 62 \\ 58 & 59 & 50 \\ 146 & 127 & 103 \end{bmatrix} \qquad B \times A = \begin{bmatrix} 19 & 34 & 49 \\ 37 & 60 & 83 \\ 79 & 119 & 159 \end{bmatrix}$$

$$\boxed{\operatorname{tr}(A \times B) = 238} \qquad \boxed{\operatorname{tr}(B \times A) = 238}$$

Diagonal Matrix | Identity Matrix | Transpose Matrix | Orthogonal Matrix

Watch Tutorial

Diagonal Matrix

A diagonal matrix has **nonzero elements only on its main diagonal**. In diagonal matrix D, it is always equal to its transpose.

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

Properties:

- Multiplication with a vector scales each component.
- **Transpose is itself**: $D^T = D$.
- **Easy to compute the inverse** (if diagonal elements are nonzero):

$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & 0 \\ 0 & \frac{1}{d_2} & 0 \\ 0 & 0 & \frac{1}{d_3} \end{bmatrix}$$

Identity Matrix

Identity Matrix

An identity matrix is a square matrix with 1s on the main diagonal and 0s everywhere else.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Properties:

- Multiplication Identity: $AI = IA = A$ for any matrix A.
- Only square matrices have an identity matrix.
- Inverse of itself: $I^{-1} = I$.

Transpose Matrix

Transpose of a Matrix

If a matrix A has a shape of (m,n), then its transpose A will have a shape of (n,m)

- Rows become columns R → C
- Columns become rows C → R

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

2×3

Then its **transpose** A^T is:

$$A^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

3×2

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Then its **transpose** A^T is:

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Transpose of a Matrix

Property	Mathematical Expression	Explanation
Double Transpose	$(A^T)^T = A$	Taking the transpose twice returns the original matrix.
Transpose of a Sum	$(A + B)^T = A^T + B^T$	The transpose of a sum is the sum of transposes.
Transpose of a Product	$(AB)^T = B^T A^T$	The transpose of a product reverses the order of multiplication.
Scalar Multiplication	$(cA)^T = cA^T$	The scalar factor c remains unchanged when transposing.
Symmetric Matrix Condition	$A = A^T$	A matrix is symmetric if it equals its transpose.
Skew-Symmetric Matrix	$A^T = -A$	A matrix is skew-symmetric if its transpose is its negative.
Orthogonal Matrix	$A^T A = I$	A matrix is orthogonal if its transpose equals its inverse.

Orthogonal Matrix

An **orthogonal matrix** is a **square matrix** whose columns and rows are orthonormal vectors.

Mathematically, a matrix Q is **orthogonal** if it satisfies:

$$\boxed{Q^T Q = Q Q^T = I}$$

where:

- Q^T is the **transpose** of Q.
- I is the **identity matrix**.

This means that the inverse of an orthogonal matrix is equal to its transpose: $\boxed{Q^{-1} = Q^T}$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2\times 2}$$

Orthogonal Matrix

An **orthogonal matrix** is a **square matrix** whose columns and rows are orthonormal vectors.

Mathematically, a matrix Q is **orthogonal** if it satisfies:

$$Q^TQ = QQ^T = I$$

where:

- Q^T is the **transpose** of Q.
- I is the **identity matrix**.

This means that the inverse of an orthogonal matrix is equal to its transpose: $Q^{-1} = Q^T$

$$Q^TQ = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}$$

$\rightarrow$

$$a^2 + c^2 = 1$$

$$ab + cd = 0$$

$$b^2 + d^2 = 1$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Orthogonal Matrix

An **orthogonal matrix** is a **square matrix** whose columns and rows are orthonormal vectors.

Mathematically, a matrix Q is **orthogonal** if it satisfies:

$$Q^T Q = QQ^T = I$$

where:

- Q^T is the **transpose** of Q.
- I is the **identity matrix**.

This means that the inverse of an orthogonal matrix is equal to its transpose: $Q^{-1} = Q^T$

$$Q = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix} \qquad Q^T Q = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \times \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}$$

$$= \begin{bmatrix} (2 \times 2 + 5 \times 5) & (2 \times 3 + 5 \times 7) \\ (3 \times 2 + 7 \times 5) & (3 \times 3 + 7 \times 7) \end{bmatrix} = \begin{bmatrix} 29 & 41 \\ 41 & 58 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

An **orthogonal matrix** is a **square matrix** whose columns and rows are orthonormal vectors.

Mathematically, a matrix Q is **orthogonal** if it satisfies:

$$Q^TQ = QQ^T = I$$

where:

- Q^T is the **transpose** of Q.
- I is the **identity matrix**.

This means that the inverse of an orthogonal matrix is equal to its transpose: $Q^{-1} = Q^T$

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Let's See,

$$Q^TQ = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{1}{2} + \frac{1}{2}\right) & \left(-\frac{1}{2} + \frac{1}{2}\right) \\ \left(\frac{1}{2} - \frac{1}{2}\right) & \left(\frac{1}{2} + \frac{1}{2}\right) \end{bmatrix} = \boxed{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} = I$$

Since $Q^TQ = I$, the matrix Q is **orthogonal**.

Let's Solve:

$$A = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Comment below on whether the mentioned two matrices are orthogonal or not orthogonal.

Solution:

$$A = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Comment below on whether the mentioned two matrices are orthogonal or not orthogonal.

- Matrix A is orthogonal because $A^T A = I$.
- Matrix B is not orthogonal because $B^T B \neq I$.

Singular & Non-Singular Matrix

Watch Tutorial

Singular Matrix

A singular matrix is a square matrix whose determinant is zero: $\det(A) = 0$

Properties:

- **No Inverse Exists:** A^{-1} does not exist.
- **Dependent Rows/Columns:** One row/column is a **linear combination** of others.
- **Causes Issues in Computations:** Cannot be used in solving equations via matrix inversion.

Non-Singular Matrix

A singular matrix is a square matrix whose determinant is non-zero: $\det(A) \neq 0$

Properties:

- **Inverse Exists**: A^{-1} can be computed.
- **Independent Rows/Columns**: Each row and column is linearly independent.
- **Can Solve Linear Equations**: The system $AX = B$ has a unique solution.

Determinant of Matrix

Determinant of Matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \longrightarrow \det(A) = ad - bc$$

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

$$\det(A) = (2 \times 4) - (3 \times 1) = 8 - 3 = 5$$

Why determinant is important?

- Whether the matrix is **invertible** ($\det(A) \neq 0$).
- Volume scaling in **linear transformations**.
- Whether a system of linear equations has a **unique solution**.

Determinant of 3x3 Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \longrightarrow \det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Where each 2×2 determinant is called a minor.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \longrightarrow \det(A) = 1 \times \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \times \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \times \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

Continue…

Determinant of 3x3 Matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \longrightarrow \det(A) = 1 \times \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \times \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \times \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

$$(5 \times 9 - 6 \times 8) = (45 - 48) = -3$$

$$(4 \times 9 - 6 \times 7) = (36 - 42) = -6$$

$$(4 \times 8 - 5 \times 7) = (32 - 35) = -3$$

$$\det(A) = (1 \times -3) - (2 \times -6) + (3 \times -3)$$

$$= -3 + 12 - 9 = 0$$

Since $\det(A) = 0$, the matrix is **singular** (not invertible).

Let's Solve:

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 3 & 4 \\ 5 & -2 & 1 \end{bmatrix}$$

Comment below on whether the two mentioned matrices are **invertible** **or** **not invertible**.

Solution:

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 3 & 4 \\ 5 & -2 & 1 \end{bmatrix}$$

Comment below on whether the two mentioned matrices are **invertible or not invertible**.

The determinant of matrix A is **0**, meaning A is **not invertible**.

The determinant of matrix B is **3**, which is nonzero, meaning B is **invertible**.

Inverse of a Matrix

Watch Tutorial

Inverse Matrix

The inverse of a matrix is a matrix that, when **multiplied by the original matrix**, results in the **identity matrix**.

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

When Does a Matrix Have an Inverse?

A **square matrix** A (of size $n \times n$) has an inverse if and only if:

1. $\det(A) \neq 0$ (non-zero determinant).
2. **All rows/columns are linearly independent**.

If $\det(A) = 0$, the matrix is **singular** and does **not have an inverse**.

Inverse Matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{\det(A) = ad - bc} A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If $\det(A) = 0$, the inverse **does not exist**.

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \longrightarrow \det(A) = (2 \times 4) - (3 \times 1) = 8 - 3 = 5$$

Since $\det(A) \neq 0$, A is **invertible**.

$$A^{-1} = \frac{1}{5} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{bmatrix}$$ **(Final Result)**

$$\boxed{\textbf{Let's Prove:} \quad AA^{-1} = A^{-1}A = I}$$

Multiply A and A^{-1}: $\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \times \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix}$

$$= \begin{bmatrix} (2 \times 4/5 + 3 \times -1/5) & (2 \times -3/5 + 3 \times 2/5) \\ (1 \times 4/5 + 4 \times -1/5) & (1 \times -3/5 + 4 \times 2/5) \end{bmatrix}$$

$$= \begin{bmatrix} (8/5 - 3/5) & (-6/5 + 6/5) \\ (4/5 - 4/5) & (-3/5 + 8/5) \end{bmatrix}$$

$$= \begin{bmatrix} 5/5 & 0 \\ 0 & 5/5 \end{bmatrix} \quad = \boxed{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} = I$$

(Proved)

Let's Solve:

$$A = \begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

Calculate the **trace** of A^{-1} and B^{-1}.

Solution:

$$A = \begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

Calculate the **trace** of A^{-1} and B^{-1}.

$$A^{-1} = \begin{bmatrix} 2 & -2.5 \\ -1 & 1.5 \end{bmatrix} \qquad \mathrm{Tr}(A^{-1}) = 2 + 1.5 = 3.5$$

$$\det(B) = (1 \times 6) - (2 \times 3) = 6 - 6 = 0$$

Since $\det(B) = 0$, matrix B is **not invertible**.

Sparse/Dense Matrix

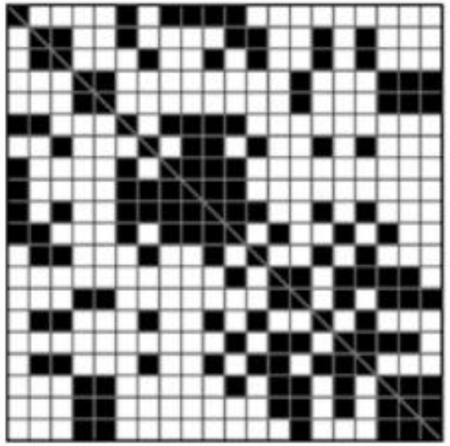
Watch Tutorial

Sparse Matrix

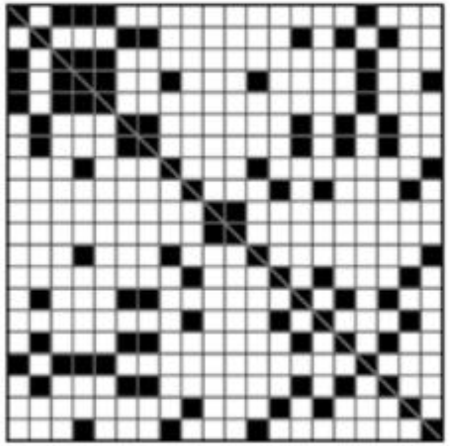
A sparse matrix is a matrix that has **mostly zero elements** and only a few nonzero values.

Characteristics of a Sparse Matrix:

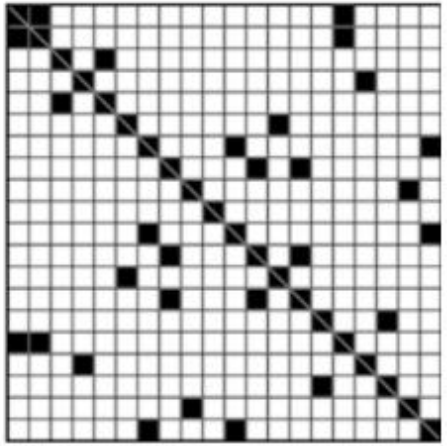
- Most elements are zero (typically, **more than 50% are zero**).
- Efficient storage techniques are used to save memory.
- Specialized operations (compressed storage formats, optimized computations).



20x20



20x20



20x20

Sparse Matrix in Machine Learning

Sparse data is commonly found in several domains of machine learning where datasets contain a high number of dimensions with many missing or zero values.

- Natural Language Processing (NLP)
- Recommender Systems
- Computer Vision
- Anomaly Detection
- Genomics & Bioinformatics
- Graph-Based Data (Social Networks, Knowledge Graphs)
- High-Dimensional Feature Spaces

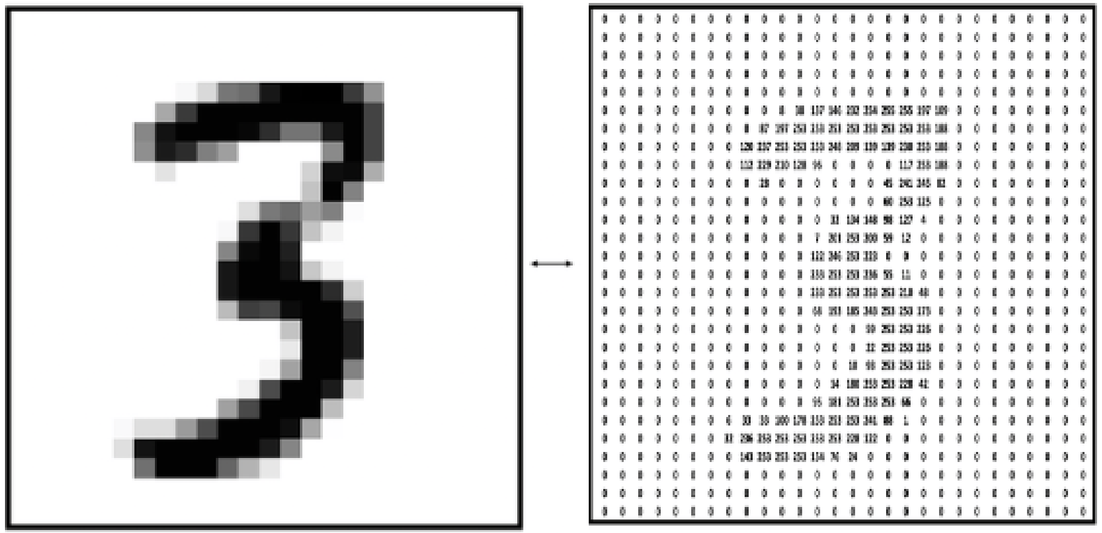
Sparse Matrix in Machine Learning (NLP)

	text
0	Eddard Stark is a king in the north.
1	A king but one king : kings are everywhere.
2	Hodor was different : he was not a king .
3	But the North could not change without him.

sparse

	king	was	the	not	But	him	one	north	kings	is	in	he	Eddard	everywhere	different	could	change	but	are	Stark	North	Hodor	without
0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0
1	2	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	0
2	1	2	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0
3	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1

Sparse Matrix in Computer Vision



Advantages of Sparse Data in ML/AI

- Efficient Storage & Computation
- Reduced Overfitting in High-dimensional Space
- Easier Interpretability
- Faster Training for Certain Algorithms
- Better Scalability

Dense Matrix

Dense Matrix

A dense matrix is a matrix where most of the elements are nonzero. It contrasts with a sparse matrix.

Characteristics of a Dense Matrix:

- Few or **no zero** elements
- Standard storage format (2D array/matrix)
- Computationally **expensive** for large sizes
- Used in general numerical computations

$$A = \begin{bmatrix} 4 & 1 & 3 \\ 2 & 8 & 6 \\ 7 & 5 & 9 \end{bmatrix}$$

Dense Vs. Sparse Matrix

Dense Matrix

1	2	31	2	9	7	34	22	11	5
11	92	4	3	2	2	3	3	2	1
3	9	13	8	21	17	4	2	1	4
8	32	1	2	34	18	7	78	10	7
9	22	3	9	8	71	12	22	17	3
13	21	21	9	2	47	1	81	21	9
21	12	53	12	91	24	81	8	91	2
61	8	33	82	19	87	16	3	1	55
54	4	78	24	18	11	4	2	99	5
13	22	32	42	9	15	9	22	1	21

Sparse Matrix

1	.	3	.	9	.	3	.	.	.
11	.	4	.	.	.	.	.	2	1
.	.	1	.	.	.	4	.	1	.
8	.	.	.	3	1	.	.	.	.
.	.	.	9	.	.	1	.	17	.
13	21	.	9	2	47	1	81	21	9
.	.	.	.	.	.	.	.	.	.
.	.	.	.	19	8	16	.	.	55
54	4	.	.	.	11	.	.	.	.
.	.	2	.	.	.	.	22	.	21

Rank of Matrix

Watch Tutorial

The rank of a matrix is the maximum number of linearly independent rows or columns in the matrix. It represents the dimension of the column space or row space of the matrix. **Let's see the column rank method:**

For Example: $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$

Step 1: Identify the Column Vectors

$$C_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \quad C_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

We check if these columns are **linearly dependent** by expressing one column as a combination of others.

- Observe that $C_2 = 2C_1$ **and** $C_3 = 3C_1$.
- Since all columns are multiples of C_1, there is **only one linearly independent column.**

Thus, the **rank of** A **is 1**.

Properties of Rank

- A square matrix A of size $n \times n$ is **full rank** if its rank is n (i.e., all rows/columns are linearly independent).
- If a matrix has dependent rows or columns, its rank is **less than** the total number of rows or columns.
- If A is a square matrix and $\det(A) \neq 0$, then A has full rank n.
- If $\det(A) = 0$, the matrix is singular and has **rank < n**.

Properties of Rank

- If $\det(A) = 0$, the matrix has **dependent rows or columns**, meaning its rank is **less than n** (not necessarily 0). Det(A) = 0 does NOT mean rank is 0

- A matrix has **rank = 0 only if** all its elements are zero.

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

  - Here, all rows and columns are **dependent** (zero vectors), so **rank = 0**.

Let’s Solve:

Since row rank = column rank, You only need to check either row rank or column rank, not both.

$$A = \begin{bmatrix} 2 & 4 & 1 \\ 6 & 12 & 3 \\ 1 & -1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Comment below the rank of A and B Matrix.

Solution:

Comment below the rank of A and B Matrix.

$$A = \begin{bmatrix} 2 & 4 & 1 \\ 6 & 12 & 3 \\ 1 & -1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

- Matrix A has **2 linearly independent rows/columns.**
- Matrix B also has **2 linearly independent rows/columns.**
- Both matrices are **rank-deficient**, meaning they do not have full rank.

Alternative to Matrix Division

Watch Tutorial

Alternative to Matrix Division: Multiplying by the Inverse

If A is a **square, invertible matrix**, we define matrix "division" as:

$$A^{-1} \times A = I$$

So instead of $A \div B$, we write:

$$X = A^{-1}B \quad \left| \quad \begin{array}{l} ax = b \quad \Rightarrow \quad x = \dfrac{b}{a} = a^{-1}b \ \text{(Scaler)} \\ AX = B \end{array} \right.$$

where A^{-1} is the **inverse** of matrix A.

The Inverse Matrix Acts Like an Alternative to Division

Determinant of a Matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

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

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(Formula)

Inverse of a Matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

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

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(Formula)

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix},$$

$$B = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\det(A) = (2 \times 4) - (3 \times 1) = 8 - 3 = 5$$

$$A^{-1} = \frac{1}{5} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{bmatrix}$$

(Inverse is Done)

Alternative to Matrix Division

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

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

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(Formula)

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix},$$

$$B = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\det(A) = (2 \times 4) - (3 \times 1) = 8 - 3 = 5$$

$$A^{-1} = \frac{1}{5} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{bmatrix}$$

(Inverse is Done)

Finally, $X = A^{-1}B$

$$X = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$X = \begin{bmatrix} \frac{4}{5} \times 5 + (-\frac{3}{5} \times 6) \\ -\frac{1}{5} \times 5 + \frac{2}{5} \times 6 \end{bmatrix}$$

$$X = \begin{bmatrix} \frac{20}{5} - \frac{18}{5} \\ -\frac{5}{5} + \frac{12}{5} \end{bmatrix}$$

$$X = \begin{bmatrix} 0.4 \\ 1.4 \end{bmatrix}$$

(Final Result)

Is Scalar Division Possible in a Matrix?

No, scalar division is not directly defined in matrix algebra. However, you can achieve a similar effect using scalar multiplication with the reciprocal.

Alternative to Scalar Division:

If you want to divide a matrix A by a scalar k, you can **multiply by its reciprocal**:

$$A \div k = \frac{1}{k} A$$

For Example: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$A \div k = \frac{1}{k} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \frac{a}{k} & \frac{b}{k} \\ \frac{c}{k} & \frac{d}{k} \end{bmatrix}$$

Let's Solve:

$$A = \begin{bmatrix} 3 & 5 \\ 1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 2 \\ 7 & 5 \end{bmatrix}$$

Comment below the Trace of $A^{-1}B$

Solution:

$$A = \begin{bmatrix} 3 & 5 \\ 1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 2 \\ 7 & 5 \end{bmatrix}$$

Comment below the Trace of $A^{-1}B$

$$A^{-1}B = \begin{bmatrix} -11/7 & -17/7 \\ 15/7 & 13/7 \end{bmatrix}$$

$$\text{Trace}(A^{-1}B) = \left(-\frac{11}{7}\right) + \left(\frac{13}{7}\right)$$

$$= \frac{-11 + 13}{7} = \frac{2}{7} \approx 0.2857$$

Symmetric & Asymmetric Matrix

Watch Tutorial

Symmetric Matrix

A symmetric matrix is a **square matrix** that is **equal to its transpose**.

$$A^T = A$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} \longrightarrow A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$$

Properties:

- Always square $(n \times n)$.
- Main diagonal remains unchanged.
- Eigenvalues are always real.
- Used in PCA, covariance matrices, and physics applications.

Asymmetric Matrix

An asymmetric matrix is a **square matrix** that is **not equal to its transpose**.

$$A^T \neq A$$

$$B = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 3 & 6 \\ 7 & 8 & 9 \end{bmatrix} \longrightarrow B^T = \begin{bmatrix} 1 & 2 & 7 \\ 4 & 3 & 8 \\ 5 & 6 & 9 \end{bmatrix}$$

Properties:

- Not necessarily square ($m \times n$ or $n \times n$).
- Rows and columns are different.
- Can have complex or real eigenvalues.
- Used in directed graphs, adjacency matrices, and transformations.

Importance of Symmetric Matrix

- **Principal Component Analysis (PCA) & Machine Learning:**
  - Covariance matrices (used in PCA) are symmetric, ensuring real eigenvalues that help in dimensionality reduction.
  - Kernel matrices in support vector machines (SVM) are symmetric, aiding in efficient classification.
- **Symmetric matrices require storing only half the elements, reducing memory usage.**
- **Graph Theory & Networks**
- **Applications in Physics & Engineering**

Let's Solve

$$A = \begin{bmatrix} 4 & 2 & -1 \\ 2 & 3 & 5 \\ -1 & 5 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 7 & 1 & 3 \\ 1 & 5 & 2 \\ 3 & 2 & 8 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Comment below whether a matrix is symmetric or asymmetric.

Let's Solve

Comment below whether a matrix is symmetric or asymmetric.

$$A = \begin{bmatrix} 4 & 2 & -1 \\ 2 & 3 & 5 \\ -1 & 5 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 7 & 1 & 3 \\ 1 & 5 & 2 \\ 3 & 2 & 8 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

- Matrix A: Symmetric ✅
- Matrix B: Symmetric ✅
- Matrix C: Asymmetric ❌

Covariance Vs. Correlation

Watch Tutorial

Covariance

Covariance is a statistical measure that quantifies the **relationship** between two random variables. It indicates the direction of their relationship—whether they tend to increase or decrease together. Mathematically, the covariance between two variables X and Y is given by:

$$\Sigma = \mathrm{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

where:

- X_i, Y_i are individual data points,
- $\bar{X}$ and $\bar{Y}$ are the mean (average) of X and Y,
- n is the number of data points.

Interpretation of Covariance

1. **Positive Covariance** ($\text{Cov}(X, Y) > 0$)
    - If X **increases,** Y **also increases** (positive correlation).
    - Example: Height and weight; taller people tend to weigh more.
2. **Negative Covariance** ($\text{Cov}(X, Y) < 0$)
    - If X **increases,** Y **decreases** (negative correlation).
    - Example: Speed of a car and time to reach a destination; as speed increases, time decreases.
3. **Zero Covariance** ($\text{Cov}(X, Y) = 0$)
    - No linear relationship between X and Y.
    - Example: Shoe size and IQ—no connection.

Correlation: Correlation Standardizes Covariance

By dividing covariance by the standard deviations of X and Y, $\sigma_X \sigma_Y$ correlation removes unit dependence.

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

The **correlation coefficient** (ρ) always ranges between **-1 and 1**, making interpretation easier:

- $\rho = 1$ → Perfect positive relationship.
- $\rho = -1$ → Perfect negative relationship.
- $\rho = 0$ → No linear relationship.

Correlation: Correlation Standardizes Covariance

By dividing covariance by the standard deviations of X and Y, $\sigma_X \sigma_Y$ correlation removes unit dependence.

$$X = [10, 20, 30, 40, 50] \qquad Y = [5, 10, 15, 20, 25]$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

Covariance $\text{Cov}(X, Y) = 100$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$= \frac{100}{(14.14 \times 7.07)}$$

$$= \frac{100}{100} = 1$$

Correlation $\rho = 1$, meaning a perfect positive relationship.

$$\sigma_X = \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2} = 14.14$$

$$\sigma_Y = \sqrt{\frac{1}{n} \sum (Y_i - \bar{Y})^2} = 7.07$$

Covariance Matrix

Covariance Matrix

A **covariance matrix** is a **square matrix** that contains the covariance values of multiple variables. It is widely used in statistics, machine learning, and finance to measure relationships between features.

For p **variables** $(X_1, X_2, \ldots, X_p)$, the **covariance matrix** is:

$$C = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \ldots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \mathrm{Cov}(X_p, X_2) & \ldots & \mathrm{Var}(X_p) \end{bmatrix}$$

We also can write

$$\boxed{\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}}$$

$$\Sigma = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(Y, X) & \mathrm{Var}(Y) \end{bmatrix}$$

Where:

- $\mathrm{Var}(X_i)$ is the **variance** of X_i.
- $\mathrm{Cov}(X_i, X_j)$ is the **covariance** between X_i and X_j.
- **Diagonal Elements**: $\mathrm{Var}(X_i)$ (the variance of each variable).
- **Off-Diagonal Elements**: $\mathrm{Cov}(X_i, X_j)$ (the covariance between different variables).

Can We Calculate Covariance with X and X (Itself)?

Are Covariance and Variance the Same?

❌ No, covariance and variance are not the same!

However, **variance is a special case of covariance**, specifically when a variable is compared to **itself.**

The covariance between two variables X and Y is:

$$\mathrm{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

Are Covariance and Variance the Same?

❌ No, covariance and variance are not the same!

However, **variance is a special case of covariance**, specifically when a variable is compared to **itself.**

The covariance between two variables X and Y is:

$$\mathrm{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

What Happens When $X = Y$?

If we calculate $\mathrm{Cov}(X, X)$, we replace Y with X:

$$\mathrm{Cov}(X, X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})$$

$$\mathrm{Cov}(X, X) = \frac{1}{n} \sum (X_i - \bar{X})^2$$

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X)$$

Are Covariance and Variance the Same?

❌ No, covariance and variance are not the same!

However, **variance is a special case of covariance**, specifically when a variable is compared to **itself.**

The covariance between two variables X and Y is:

$$\mathrm{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

What Happens When $X = Y$?

If we calculate $\mathrm{Cov}(X,X)$, we replace Y with X:

$$\mathrm{Cov}(X,X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})$$

$$\mathrm{Cov}(X,X) = \frac{1}{n}\sum(X_i - \bar{X})^2$$

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X)$$

Properties:

- Covariance measures the relationship between two variables.
- Variance measures how much a single variable spreads out from its mean.
- Since a variable is always 100% correlated with itself, the covariance reduces to variance.

Continue…

Variance

Variance is a special case of covariance where a variable is compared with itself. Variance measures how spread out a set of numbers is around the mean. If variance is high, the variable is more spread out from its mean. If variance is low, the values are closer to the mean.

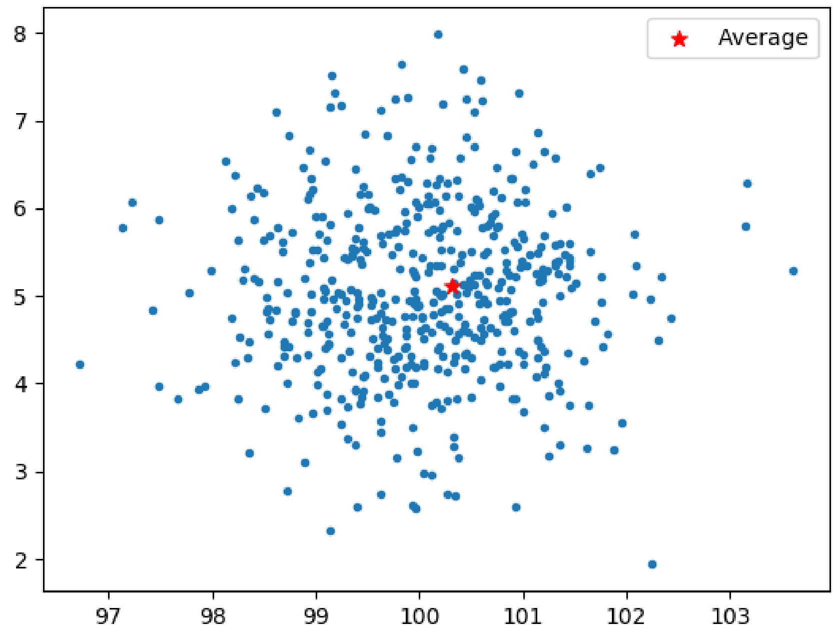
$$\Sigma = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

The formula for the variance of X is:

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

where:

- X_i are the data points of X,
- $\bar{X}$ is the mean of X,
- n is the number of data points.



Covariance Matrix (Part-02)

Watch Tutorial

Covariance Matrix

Let $X = (X_1, X_2, X_3, X_4)$ be a random vector distributed as: $X \sim \mathcal{N}(0, \Sigma)$

$$\Sigma = \frac{1}{8} \begin{array}{cc} \begin{array}{cccc} x_1 & \quad x_2 & \quad x_3 & \quad x_4 \end{array} & \\ \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Now, define a new random vector $Y = (Y_1, Y_2)'$ as a linear transformation of X:

$$Y_1 = X_1 + X_4, \quad Y_2 = X_2 - X_4$$

Covariance Matrix

Let $X = (X_1, X_2, X_3, X_4)$ be a random vector distributed as: $X \sim \mathcal{N}(0, \Sigma)$

$$\begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \Sigma = \frac{1}{8} & \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Now, define a new random vector $Y = (Y_1, Y_2)'$ as a linear transformation of X:

$$Y_1 = X_1 + X_4, \quad Y_2 = X_2 - X_4$$

Calculate the covariance matrix of Y. $\mathrm{Cov}(Y_1, Y_2) = \Sigma_Y = \begin{bmatrix} \Sigma_{Y_1Y_1} & \Sigma_{Y_1Y_2} \\ \Sigma_{Y_2Y_1} & \Sigma_{Y_2Y_2} \end{bmatrix}$

Covariance values $\Sigma_{Y_1Y_2}$ and $\Sigma_{Y_2Y_1}$ will always be the same because the covariance matrix is symmetric.

Covariance Matrix

$$\Sigma = \frac{1}{8} \begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ & \left[\begin{array}{cccc} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{array}\right] \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Set $Y = (Y_1, Y_2)'$ with

$$Y_1 = X_1 + X_4 \text{ and } Y_2 = X_2 - X_4.$$

Calculate the covariance matrix of Y.

Key Formula for Variance of a Sum/Difference of Two Random Variables:

$$\mathrm{Var}(A \pm B) = \mathrm{Var}(A) + \mathrm{Var}(B) \pm 2\mathrm{Cov}(A, B)$$

For $Y_1 = X_1 + X_4$:

$$\mathrm{Var}(Y_1) = \mathrm{Var}(X_1) + \mathrm{Var}(X_4) + 2\mathrm{Cov}(X_1, X_4)$$

For $Y_2 = X_2 - X_4$:

$$\mathrm{Var}(Y_2) = \mathrm{Var}(X_2) + \mathrm{Var}(X_4) - 2\mathrm{Cov}(X_2, X_4)$$

Covariance Matrix

$$\Sigma = \frac{1}{8} \begin{array}{cc} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} & \\ \left[\begin{array}{cccc} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{array}\right] & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Set $Y = (Y_1, Y_2)'$ with

$$Y_1 = X_1 + X_4 \text{ and } Y_2 = X_2 - X_4.$$

Calculate the covariance matrix of Y.

$$\operatorname{Var}(Y_1) = \operatorname{Var}(X_1) + \operatorname{Var}(X_4) + 2\operatorname{Cov}(X_1, X_4)$$

$$\operatorname{Var}(Y_1) = \frac{0.71}{8} + \frac{0.2}{8} + 2(0) = \frac{0.91}{8}$$

Covariance Matrix

$$\Sigma = \frac{1}{8}\begin{array}{cc} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} & \\ \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Set $Y = (Y_1, Y_2)'$ with

$$Y_1 = X_1 + X_4 \text{ and } Y_2 = X_2 - X_4.$$

Calculate the covariance matrix of Y.

$$\text{Var}(Y_1) = \text{Var}(X_1) + \text{Var}(X_4) + 2\text{Cov}(X_1, X_4)$$

$$\text{Var}(Y_1) = \frac{0.71}{8} + \frac{0.2}{8} + 2(0) = \frac{0.91}{8}$$

$$\text{Var}(Y_2) = \text{Var}(X_2) + \text{Var}(X_4) - 2\text{Cov}(X_2, X_4)$$

$$\text{Var}(Y_2) = \frac{0.46}{8} + \frac{0.2}{8} - 2(0) = \frac{0.66}{8}$$

Covariance Matrix

$$\Sigma = \frac{1}{8}\begin{array}{cc} \begin{array}{cccc} \quad x_1 & \quad x_2 & \quad x_3 & \quad x_4 \end{array} & \\ \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Set $Y = (Y_1, Y_2)'$ with

$$Y_1 = X_1 + X_4 \text{ and } Y_2 = X_2 - X_4.$$

Calculate the covariance matrix of Y.

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1 + X_4, X_2 - X_4)$$

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, -X_4) + \mathrm{Cov}(X_4, X_2) + \mathrm{Cov}(X_4, -X_4)$$

$$\mathrm{Cov}(Y_1, Y_2) = \frac{-0.43}{8} + 0 + 0 - \frac{0.2}{8} = \frac{-0.63}{8}$$

Thus, the final covariance matrix is: $\Sigma_Y = \frac{1}{8}\begin{bmatrix} 0.91 & -0.63 \\ -0.63 & 0.66 \end{bmatrix}$

Covariance Matrix

$$\begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \Sigma = \frac{1}{8} & \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \end{array}$$

Set $Y = (Y_1, Y_2)'$ with

$$Y_1 = X_1 + X_4 \text{ and } Y_2 = X_2 - X_4.$$

Calculate the covariance matrix of Y.

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1 + X_4, X_2 - X_4)$$

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, -X_4) + \mathrm{Cov}(X_4, X_2) + \mathrm{Cov}(X_4, -X_4)$$

$$\mathrm{Cov}(Y_1, Y_2) = \frac{-0.43}{8} + 0 + 0 - \frac{0.2}{8} = \frac{-0.63}{8}$$

Thus, the final covariance matrix is: $\Sigma_Y = \frac{1}{8}\begin{bmatrix} 0.91 & -0.63 \\ -0.63 & 0.66 \end{bmatrix}$

$$\boxed{\Sigma_Y = \begin{bmatrix} \Sigma_{Y_1Y_1} & \Sigma_{Y_1Y_2} \\ \Sigma_{Y_2Y_1} & \Sigma_{Y_2Y_2} \end{bmatrix}}$$

Basic Logarithmic Properties & Exponential Rules in Mathematics

Watch Tutorial

Logarithmic Properties

1. **Product Rule**

$$\log_b(xy) = \log_b x + \log_b y$$

Example:

$$\log_2(8 \times 4) = \log_2 8 + \log_2 4 = 3 + 2 = 5$$

1. **Product Rule**

$$\log_b(xy) = \log_b x + \log_b y$$

Example:

$\log_2(8 \times 4) = \log_2 8 + \log_2 4 = 3 + 2 = 5$

2. **Quotient Rule**

$$\log_b\left(\frac{x}{y}\right) = \log_b x - \log_b y$$

Example:

$\log_3\left(\frac{27}{3}\right) = \log_3 27 - \log_3 3 = 3 - 1 = 2$

Logarithmic Properties

3. **Power Rule**

$$\log_b(x^n) = n \log_b x$$

Example:

$$\log_5(25^3) = 3 \log_5 25 = 3 \times 2 = 6$$

3. **Power Rule**

$$\log_b(x^n) = n\log_b x$$

Example:

$$\log_5(25^3) = 3\log_5 25 = 3 \times 2 = 6$$

4. **Change of Base Formula**

$$\log_b x = \frac{\log_k x}{\log_k b}$$

(Commonly used with $k = 10$ or e for calculator computations.)

Example:

$$\log_2 10 = \frac{\log 10}{\log 2} = \frac{1}{0.301} \approx 3.32$$

5. **Logarithm of 1**

$$\log_b 1 = 0$$

Example:

$\log_7 1 = 0$ because $7^0 = 1$.

- $\log_2(1) = 0$
- $\log_{10}(1) = 0$
- $\ln(1) = \log_e(1) = 0$

5. **Logarithm of 1**

$$\log_b 1 = 0$$

- $\log_2(1) = 0$
- $\log_{10}(1) = 0$
- $\ln(1) = \log_e(1) = 0$

Example:

$\log_7 1 = 0$ because $7^0 = 1$.

6. **Logarithm of the Base**

$$\log_b b = 1$$

Example:

$\log_9 9 = 1$ because $9^1 = 9$.

This means that if the base and the value are the same, the logarithm is always 1.

Properties of e & Natural Logarithms ($\ln x$)

The constant **e≈2.718** is the base of the natural logarithm. The **natural logarithm** is denoted as:

$$\ln x = \log_e x$$

Properties of e:

1. **Definition of Natural Logarithm**

$$\ln e = 1 \quad \text{since} \quad e^1 = e$$

2. **Exponent and Log Relationship**

$$e^{\ln x} = x \quad \text{and} \quad \ln(e^x) = x$$

Example:
$e^{\ln 5} = 5$ and $\ln(e^4) = 4$

3. **Derivative of e^x**

$$\frac{d}{dx}e^x = e^x$$

4. **Derivative of $\ln x$**

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

5. **Integral of e^x and $\ln x$**

$$\int e^x dx = e^x + C$$

$$\int \ln x \, dx = x \ln x - x + C$$

All Exponent Rules with Examples

Exponential Rules

1. Product Rule (Exponent Addition Rule)

$$a^m \cdot a^n = a^{m+n}$$

Example: $2^3 \cdot 2^4 = 2^{3+4} = 2^7 = 128$

$$x^5 \cdot x^2 = x^{5+2} = x^7$$

Exponential Rules

1. Product Rule (Exponent Addition Rule)

$$a^m \cdot a^n = a^{m+n}$$

Example: $$2^3 \cdot 2^4 = 2^{3+4} = 2^7 = 128$$

$$x^5 \cdot x^2 = x^{5+2} = x^7$$

2. Quotient Rule (Exponent Subtraction Rule)

$$\frac{a^m}{a^n} = a^{m-n}$$

$$\frac{5^6}{5^2} = 5^{6-2} = 5^4 = 625$$

$$\frac{x^8}{x^3} = x^{8-3} = x^5$$

3. Power of a Power Rule

$$(a^m)^n = a^{m \cdot n}$$

Example:

$$(3^2)^4 = 3^{2\times 4} = 3^8 = 6561$$

$$(x^5)^3 = x^{5\times 3} = x^{15}$$

3. Power of a Power Rule

$$(a^m)^n = a^{m \cdot n}$$

Example:

$$(3^2)^4 = 3^{2\times4} = 3^8 = 6561$$

$$(x^5)^3 = x^{5\times3} = x^{15}$$

4. Power of a Product Rule

$$(ab)^m = a^m \cdot b^m$$

Example:

$$(2 \cdot 3)^4 = 2^4 \cdot 3^4 = 16 \cdot 81 = 1296$$

$$(xy)^3 = x^3 \cdot y^3$$

AI Quest Intelligence STUDY MART

5. Power of a Quotient Rule

$$\left(\frac{a}{b}\right)^m = \frac{a^m}{b^m}$$

Example:

$$\left(\frac{4}{2}\right)^3 = \frac{4^3}{2^3} = \frac{64}{8} = 8$$

$$\left(\frac{x}{y}\right)^5 = \frac{x^5}{y^5}$$

6. Zero Exponent Rule

$$a^0 = 1, \quad \text{for } a \neq 0$$

Example:

$$5^0 = 1$$

$$x^0 = 1$$

7. Negative Exponent Rule

$$a^{-m} = \frac{1}{a^m}$$

Example:

$$5^{-2} = \frac{1}{5^2} = \frac{1}{25}$$

$$x^{-3} = \frac{1}{x^3}$$

8. Fractional Exponent Rule

$$a^{\frac{m}{n}} = \sqrt[n]{a^m}$$

Example:

$$8^{\frac{1}{3}} = \sqrt[3]{8} = 2$$

$$16^{\frac{1}{4}} = \sqrt[4]{16} = 2$$

Positive Definite, Negative Definite, and Positive Semi-Definite Matrices

Watch Tutorial

What Are Eigenvalues?

Eigenvalues are **special numbers** associated with a square matrix that indicate how the matrix **scales** or **transforms** vectors.

Mathematically, for a square matrix A, an **eigenvalue** λ is a scalar that satisfies the equation:

$$Av = \lambda v$$

- A is an $n \times n$ matrix,
- v is a **nonzero vector** (called the **eigenvector**),
- λ is the **eigenvalue**.

Since λv represents **scalar multiplication** of the vector v, λ must be a **scalar** (a single real or complex number), not a vector or matrix.

Eigenvalues are found by solving the **characteristic equation**:

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

What Are Eigenvalues?

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

The eigenvalues λ are found by solving: $\det(A - \lambda I) = 0$

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

$$= (2-\lambda)^2 - 1 \;= 4 - 4\lambda + \lambda^2 - 1 \;= \lambda^2 - 4\lambda + 3$$

Set the determinant equal to zero: $\lambda^2 - 4\lambda + 3 = 0$

$\lambda^2 - 3\lambda - \lambda + 3 = 0$

$\lambda(\lambda - 3) - 1(\lambda - 3) = 0$

$$(\lambda - 3)(\lambda - 1) = 0$$

$\lambda = 3, 1 \rightarrow$ Positive

Positive Definite Matrix

A symmetric matrix A is **positive definite** if for all nonzero vectors x:

$$x^T A x > 0$$

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

Since all eigenvalues (3,1) are positive, A is positive definite.

- All **eigenvalues are positive** ($\lambda_i > 0$ for all i).
- This means the quadratic form $x^T A x$ is always positive for any nonzero x.

Negative Definite Matrix

A symmetric matrix A is **negative definite** if for all nonzero vectors x:

$$x^T A x < 0$$

$$A = \begin{bmatrix} -2 & 1 \\ 1 & -3 \end{bmatrix}$$

If all eigenvalues are negative, A is negative definite.

- All **eigenvalues are negative** ($\lambda_i < 0$ for all i).
- This means the quadratic form $x^T A x$ is always negative for any nonzero x.

Positive Semi-Definite Matrix

A symmetric matrix A is **positive semi-definite** if for all vectors x:

$$x^T A x \geq 0$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Since one eigenvalue is zero and the other is positive, A is positive semi-definite.

- All **eigenvalues are non-negative** ($\lambda_i \geq 0$ for all i).
- Some eigenvalues may be zero, but none are negative.
- The quadratic form $x^T A x$ is never negative but can be zero.

Indefinite Matrix & Its Importance

Watch Tutorial

Indefinite Matrix

An indefinite matrix is a square matrix with **positive** and **negative** eigenvalues. This means that when you take its quadratic form, it can produce positive and negative values depending on the input vector.

$$\exists\, x_1, x_2 \neq 0 \text{ such that } x_1^T A x_1 > 0 \text{ and } x_2^T A x_2 < 0$$

$$A = \begin{bmatrix} 2 & 4 \\ 4 & -3 \end{bmatrix}$$

$$\lambda = \frac{-1 \pm \sqrt{1+88}}{2} = \frac{-1 \pm \sqrt{89}}{2}$$

Indefinite Matrix

An indefinite matrix is a square matrix with **positive** and **negative** eigenvalues. This means that when you take its quadratic form, it can produce positive and negative values depending on the input vector.

$$\exists\, x_1, x_2 \neq 0 \text{ such that } x_1^T A x_1 > 0 \text{ and } x_2^T A x_2 < 0$$

$$A = \begin{bmatrix} 2 & 4 \\ 4 & -3 \end{bmatrix}$$

$$\lambda = \frac{-1 \pm \sqrt{1+88}}{2} = \frac{-1 \pm \sqrt{89}}{2}$$

Can an Indefinite Matrix Have Zero as an Eigenvalue?

No. An indefinite matrix must have both positive and negative eigenvalues, meaning at least one eigenvalue must be strictly positive and another strictly negative.

Indefinite Matrix

$$A = \begin{bmatrix} 2 & 4 \\ 4 & -3 \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & 4 \\ 4 & -3-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)(-3-\lambda) - (4)(4) = 0$$

$$-6 - 2\lambda + 3\lambda + \lambda^2 - 16 = 0$$

$$\lambda^2 + \lambda - 22 = 0$$

$$\lambda_1 \approx \frac{-1 + 9.43}{2} = \frac{8.43}{2} = 4.215$$

$$\lambda_2 \approx \frac{-1 - 9.43}{2} = \frac{-10.43}{2} = -5.215$$

Since one eigenvalue is **positive** and the other is **negative**, the matrix is **indefinite**.

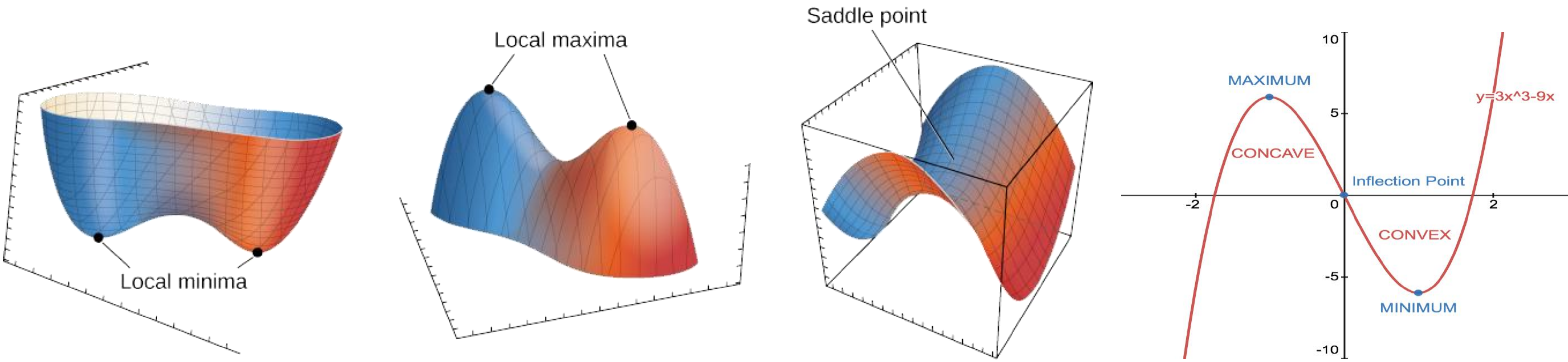
Indefinite Matrix

Key Properties of an Indefinite Matrix:

- Has both positive and negative eigenvalues.
- Neither positive definite nor negative definite.
- Occurs in **saddle-point** problems and optimization.

- Convex (Minimum) → Hessian is Positive Definite ($\lambda > 0$)
- Concave (Maximum) → Hessian is Negative Definite ($\lambda < 0$)
- Saddle Point → Hessian is Indefinite (some $\lambda > 0$, some $\lambda < 0$)

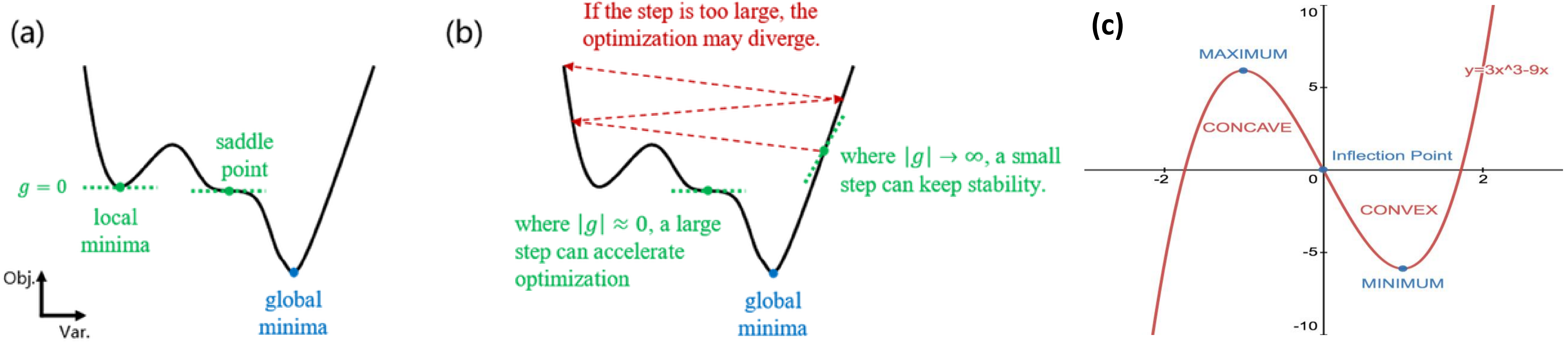
Convex and Non-Convex Function



(a) (b) (c) (d)

- Convex (Minimum) → Hessian is Positive Definite ($\lambda > 0$)
- Concave (Maximum) → Hessian is Negative Definite ($\lambda < 0$)
- Saddle Point → Hessian is Indefinite (some $\lambda > 0$, some $\lambda < 0$)

Convex and Non-Convex Function



Jensen's Inequality

Watch Tutorial

Jensen’s Inequality in the Deterministic Form

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y), \quad \forall x, y \text{ and } \lambda \in [0,1]$$

The function evaluated at a mixture of x and y is less than or equal to the mixture of the function values at x and y when the function is convex. For a convex function, the value at the average point is less than or equal to the average of the values.

Jensen’s Inequality in the Deterministic Form

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \text{ and } \lambda \in [0,1]$$

Example 1: Convex Function $f(x) = x^2$

Let's take $f(x) = x^2$, which is a convex function because its second derivative $f''(x) = 2$ is positive.

Step 1: Choose Two Points and λ

Let's pick two points: $x = 1$, $y = 3$, and let $\lambda = 0.5$.

Jensen's Inequality in the Deterministic Form

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \text{ and } \lambda \in [0,1]$$

Example 1: Convex Function $f(x) = x^2$

Let's take $f(x) = x^2$, which is a convex function because its second derivative $f''(x) = 2$ is positive.

Step 1: Choose Two Points and λ

Let's pick two points: $x = 1$, $y = 3$, and let $\lambda = 0.5$.

Step 2: Compute Both Sides of the Inequality

1. Compute the left-hand side:

$$f(\lambda x + (1-\lambda)y) = f(0.5(1) + 0.5(3)) = f(2) = 2^2 = 4$$

2. Compute the right-hand side:

$$\lambda f(x) + (1-\lambda)f(y) = 0.5(1^2) + 0.5(3^2) = 0.5(1) + 0.5(9) = 0.5 + 4.5 = 5$$

Since 4<5, So it is Convex Function

Jensen's Inequality in the Deterministic Form

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \text{ and } \lambda \in [0,1]$$

Example 2: Non-Convex Function $f(x) = -x^2$ (Concave Case)

Let's try the function $f(x) = -x^2$, which is **concave** since its second derivative is negative $f''(x) = -2$.

Using the same points $x = 1$, $y = 3$, and $\lambda = 0.5$:

1. Compute the left-hand side:
$$f(2) = -(2^2) = -4$$
2. Compute the right-hand side:
$$0.5(-1^2) + 0.5(-3^2) = 0.5(-1) + 0.5(-9) = -0.5 - 4.5 = -5$$

Since, $-4 \not\leq -5$

the inequality **fails**, proving that $f(x) = -x^2$ is **not convex**.

Derivative Basics (Calculus)

Basic Derivative Rule

Rule: $\frac{d}{dx}(x^n) = nx^{n-1}$ $\frac{d}{dx}(x^2) = 2x$ $\frac{d}{dx}(x^3) = 3x^2$

First: $f'(x) = \frac{d}{dx}f(x)$ **Example:** $\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(x^2) + \frac{d}{dx}(y^2) = 2x + 0 = 2x$

Basic Derivative Rule

Rule: $\frac{d}{dx}(x^n) = nx^{n-1}$ $\frac{d}{dx}(x^2) = 2x$ $\frac{d}{dx}(x^3) = 3x^2$

First: $f'(x) = \frac{d}{dx}f(x)$ **Example:** $\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(x^2) + \frac{d}{dx}(y^2) = 2x + 0 = 2x$

Second: $f''(x) = \frac{d}{dx}f'(x) = \frac{d^2}{dx^2}f(x)$

Mixed: $\frac{\partial^2 f}{\partial x \partial y}$

1. First, take the **partial derivative** of $f(x, y)$ with respect to x.
2. Then, differentiate the result **again** with respect to y.

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(x^2 + y^2) = 2x + 0 = 2x$$

$\frac{\partial}{\partial y}(2x)$ = 0 The derivative of a constant is 0.

Hessian Matrix for Optimization

Watch Tutorial

Hessian Matrix

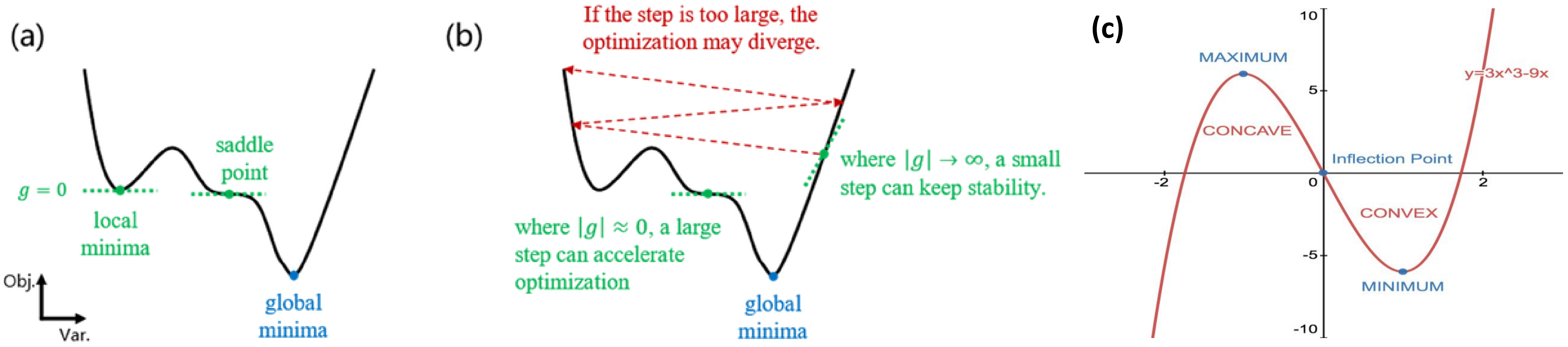
The Hessian matrix is a square, symmetric matrix of **second-order partial derivatives** of a scalar-valued function. It describes the local curvature of a function and is widely used in optimization, machine learning, and numerical analysis.

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} \qquad H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- In machine learning and other optimization problems, the Hessian **helps determine critical points**' nature (minima, maxima, or saddle points).
- If the Hessian is **positive definite**, you're at a **local minimum**.
- If **negative definite**, it's a **local maximum**.
- If **indefinite**, it's a **saddle point**.
- Newton's optimization method uses the Hessian to improve convergence when finding function minima or maxima.

The function is **strictly convex**, and **any local minimum is also a global minimum**

Convex and Non-Convex Function



- If a function $f(x)$ is **concave**, then $-f(x)$ is **convex**.
- Similarly, if $f(x)$ is **convex**, then $-f(x)$ is **concave**.

The function is **strictly convex**, and **any local minimum is also a global minimum**

Hessian Matrix

The Hessian matrix is a square, symmetric matrix of **second-order partial derivatives** of a scalar-valued function. It describes the local curvature of a function and is widely used in optimization, machine learning, and numerical analysis.

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Conditions:

You can just look at the sign of $f''(x)$ directly:

- Positive → local **minimum**
- Negative → local **maximum**
- Zero → **inconclusive**

(Higher order-> until you find the lowest non-zero derivative)

Let's Calculate for, $f(x) = x^2$ $\quad \nabla f(x) = \frac{d}{dx}(x^2) = 2x$

$$H(f) = \left[\frac{d^2 f}{dx^2}\right] \qquad \frac{d^2 f}{dx^2} = 2 \qquad H(f) = [2] \quad 1\times1$$

Since, 2>0, meaning the Hessian is positive definite. It is always a Local Minimum.

Function of $f(x, y) = x^2 + y^2$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$ **Final Matrix**

Let's see step by step:

1st Order derivative:

$$\frac{\partial f}{\partial x} = \frac{d}{dx}(x^2 + y^2) = 2x$$

$$\frac{\partial f}{\partial y} = \frac{d}{dy}(x^2 + y^2) = 2y$$

2nd Order:

$$\frac{\partial^2 f}{\partial x^2} = \frac{d}{dx}(2x) = 2 \qquad \frac{\partial^2 f}{\partial y^2} = \frac{d}{dy}(2y) = 2$$

Final Matrix $H(f) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

$$H(f) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Let's Calculate the Eigenvalue $\det(H - \lambda I) = 0$

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)(2-\lambda) = 0$$

$$\lambda_1 = 2, \quad \lambda_2 = 2$$

- Since **both eigenvalues are positive** ($\lambda_1, \lambda_2 > 0$), the Hessian matrix is **positive definite**.
- A **positive definite Hessian** confirms that $f(x, y)$ is **strictly convex**.

Hessian Test for Local Min/Max & Convexity

Hessian Type	Eigenvalues λ	Local Min/Max Classification	Convexity Classification
Positive Definite	$\lambda_1 > 0, \lambda_2 > 0$	Local Minimum	Strictly Convex
Negative Definite	$\lambda_1 < 0, \lambda_2 < 0$	Local Maximum	Strictly Concave
Indefinite	$\lambda_1 > 0, \lambda_2 < 0$	Saddle Point (Neither Min nor Max)	Non-Convex
Positive Semi-Definite	$\lambda_1 \geq 0, \lambda_2 \geq 0$	Possibly Local Minimum (Need Higher-Order Test)	Convex (Possibly Flat in Some Directions)
Negative Semi-Definite	$\lambda_1 \leq 0, \lambda_2 \leq 0$	Possibly Local Maximum (Need Higher-Order Test)	Concave (Possibly Flat in Some Directions)

The function is **strictly convex**, and **any local minimum is also a global minimum**

System of Linear Equations

Watch Tutorial

A System of Linear Equations is a collection of two or more linear equations involving the same set of variables.

Key Concepts:

1. **Linear Equation:** An equation where the **highest power of the variable(s) is 1**.
   Example: $2x + 3y = 6$

2. **System:** When you **group multiple linear equations** together. So, A system of linear equations is a set of two or more linear equations that involve the same variables. Example of a system with two variables:
$$\begin{cases} 2x + 3y = 6 \\ x - y = 4 \end{cases}$$

Types of Solutions:

A system of linear equations can have:

1. One solution – the equations intersect at a single point (consistent and independent).

2. No solution – the equations are parallel and never intersect (inconsistent).

3. Infinitely many solutions – the equations represent the same line (consistent and dependent).

System of Linear Equations

A System of Linear Equations is a collection of two or more linear equations involving the same set of variables.

Key Concepts:

1. **Linear Equation:** An equation where the **highest power of the variable(s) is 1**.

   Example: $2x + 3y = 6$

2. **System:** When you **group multiple linear equations** together. So, A system of linear equations is a set of two or more linear equations that involve the same variables. Example of a system with two variables:

$$\begin{cases} 2x + 3y = 6 \\ x - y = 4 \end{cases}$$

How to Solve These Systems?

1. Substitution
2. Cramer's Rule
3. Gaussian Elimination

Systems of Linear Equations in Machine Learning

- **Linear Regression**: Solves $Xw = y$ to find best-fit line.
- **Neural Networks**: Each layer uses $y = Wx + b$ — a linear system.
- **PCA**: Solves homogeneous systems to find **eigenvectors** for dimensionality reduction.
- **Feature Selection**: Detect **linear dependence** among input variables.
- **SVM**: Finds optimal decision boundaries using linear constraints.
- **Gradient Descent**: Uses derivatives from linear systems to optimize models.
- **Matrix Rank**: Tells number of independent equations.
- **Eigenvalues/Eigenvectors**: Found by solving $(A - \lambda I)v = 0$.

What is a Homogeneous System?

A **homogeneous system of linear equations** is a system where **all constant terms are zero**. That means every equation is set equal to 0.

General Form:

$$\begin{cases} a_1x + b_1y + c_1z = 0 \\ a_2x + b_2y + c_2z = 0 \\ \vdots \end{cases}$$

Three Variables:

$$\begin{cases} x + y + z = 0 \\ 2x - y + z = 0 \\ x + 2y - 3z = 0 \end{cases}$$

How to Solve These Systems?

1. **Substitution**
2. Cramer's Rule
3. Gaussian Elimination

Let's solve with the substitution method: $\begin{cases} x + y = 7 \quad (1) \\ 3x - y = 5 \quad (2) \end{cases}$

- **Step 1: Solve Equation (1) for x** $\longrightarrow x = 7 - y$
- **Step 2: Substitute into Equation (2)** $\longrightarrow$

$$3(7-y) - y = 5$$
$$21 - 3y - y = 5$$
$$21 - 4y = 5$$

$$-4y = 5 - 21 = -16 \Rightarrow y = \frac{-16}{-4} = 4$$
$$x = 7 - y = 7 - 4 = 3$$

Final Answer: $x = 3, \quad y = 4$

Let’s Solve:

System of Equations:

$$\begin{cases} x + 2y - z = 4 & (1) \\ 2x - y + 3z = 1 & (2) \\ -x + 3y + 2z = 7 & (3) \end{cases}$$

✅ Goal: Find the x, y, and z values that satisfy **all three equations**.

Solution:

📘 **System of Equations:**

$$\begin{cases} x + 2y - z = 4 \quad (1) \\ 2x - y + 3z = 1 \quad (2) \\ -x + 3y + 2z = 7 \quad (3) \end{cases}$$

✅ Goal: Find the x, y, and z values that satisfy **all three equations**.

Answer: $\boxed{x = \frac{8}{15}, \quad y = \frac{31}{15}, \quad z = \frac{2}{3}}$

Go & Solve: https://matrixcalc.org/slu.html

Cramer’s Rule

Watch Tutorial

$$
\begin{cases}
x + y + z = 6 & (1) \\
2x + 3y + z = 14 & (2) \\
x + 2y + 3z = 14 & (3)
\end{cases}
$$

Let’s Solve using Cramer’s Rule:

Coefficient Matrix: $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix} \left| \begin{matrix} 6 \\ 14 \\ 14 \end{matrix} \right.$

Compute Determinant: $D = \det(A)$

$$
\begin{aligned}
D &= 1(3 \cdot 3 - 1 \cdot 2) - 1(2 \cdot 3 - 1 \cdot 1) + 1(2 \cdot 2 - 3 \cdot 1) \\
&= 1(9 - 2) - 1(6 - 1) + 1(4 - 3) = 7 - 5 + 1 = 3
\end{aligned}
$$

Important: *This works only when D≠0* (the system has a **unique solution**).

Cramer's Rule

- **Step 3: Compute D_x, D_y, D_z**

  Replace column 1 (x-column) with constants $[6, 14, 14]$:

$$D_x = \begin{vmatrix} 6 & 1 & 1 \\ 14 & 3 & 1 \\ 14 & 2 & 3 \end{vmatrix} = 6(3\cdot 3 - 1\cdot 2) - 1(14\cdot 3 - 1\cdot 14) + 1(14\cdot 2 - 3\cdot 14)$$

$$= 6(9-2) - (42-14) + (28-42) = 6(7) - 28 - 14 = 42 - 28 - 14 = 0$$

  Replace column 2 (y-column):

$$D_y = \begin{vmatrix} 1 & 6 & 1 \\ 2 & 14 & 1 \\ 1 & 14 & 3 \end{vmatrix} = 1(14\cdot 3 - 1\cdot 14) - 6(2\cdot 3 - 1\cdot 1) + 1(2\cdot 14 - 14\cdot 1)$$

$$= 1(42-14) - 6(6-1) + 1(28-14) = 28 - 30 + 14 = 12$$

- **Step 3: Compute D_x, D_y, D_z**

Replace column 3 (z-column):

$$D_z = \begin{vmatrix} 1 & 1 & 6 \\ 2 & 3 & 14 \\ 1 & 2 & 14 \end{vmatrix} = 1(3 \cdot 14 - 14 \cdot 2) - 1(2 \cdot 14 - 14 \cdot 1) + 6(2 \cdot 2 - 3 \cdot 1)$$
$$= 1(42 - 28) - (28 - 14) + 6(4 - 3) = 14 - 14 + 6 = 6$$

$$x = \frac{D_x}{D} = 0, \quad y = \frac{12}{3} = 4, \quad z = \frac{6}{3} = 2$$

✅ **Final Answer:** $x = 0, \quad y = 4, \quad z = 2$

If $D = 0$:

Case 1: Infinitely Many Solutions

This happens when:

- $D = 0$, and
- All of D_x, D_y, D_z are also **0**

This means the equations are **dependent** (e.g., one is a combination of others), and the system has **infinitely many solutions** (like a plane of solutions).

🔴 If $D = 0$:

✅ Case 1: Infinitely Many Solutions

This happens when:

- $D = 0$, and
- All of D_x, D_y, D_z are also **0**

🟢 This means the equations are **dependent** (e.g., one is a combination of others), and the system has **infinitely many solutions** (like a plane of solutions).

❌ Case 2: No Solution (Inconsistent)

This happens when:

- $D = 0$, and
- At least **one** of D_x, D_y, D_z is **not zero**

🔴 This means the system is **inconsistent** — the equations contradict each other (e.g., parallel planes that never intersect).

Let's Solve:

📘 **System of Equations:** Cramer's Rule

$$\begin{cases} x - y + z = 3 \\ 2x + y - 2z = 0 \\ 3x + 2y + z = 8 \end{cases}$$

✅ Goal: Find the x, y, and z values that satisfy **all three equations**.

Solution:

System of Equations:

$$\begin{cases} x - y + z = 3 \\ 2x + y - 2z = 0 \\ 3x + 2y + z = 8 \end{cases}$$

Goal: Find the x, y, and z values that satisfy **all three equations**.

Go & Solve: https://matrixcalc.org/slu.html#solve-using-Gaussian-elimination

Gaussian Elimination

Watch Tutorial

Gaussian Elimination

$$\begin{cases} x+y+z=6 \\ 2x+3y+z=14 \\ x+2y+3z=14 \end{cases}$$

Let's Solve using Gaussian Elimination:

Write as augmented matrix: $\begin{bmatrix} 1 & 1 & 1 & | & 6 \\ 2 & 3 & 1 & | & 14 \\ 1 & 2 & 3 & | & 14 \end{bmatrix}$

- Row2 = Row2 - 2×Row1: $[2\ 3\ 1 \mid 14] - 2 \times [1\ 1\ 1 \mid 6] = [0\ 1\ -1 \mid 2]$
- Row3 = Row3 - Row1: $[1\ 2\ 3 \mid 14] - [1\ 1\ 1 \mid 6] = [0\ 1\ 2 \mid 8]$

Now matrix: $\begin{bmatrix} 1 & 1 & 1 & | & 6 \\ 0 & 1 & -1 & | & 2 \\ 0 & 1 & 2 & | & 8 \end{bmatrix}$

Now matrix: $\begin{bmatrix} 1 & 1 & 1 & | & 6 \\ 0 & 1 & -1 & | & 2 \\ 0 & 1 & 2 & | & 8 \end{bmatrix}$

- Row3 = Row3 - Row2: $[0\ 1\ 2\ |\ 8] - [0\ 1\ \ -1\ |\ 2] = [0\ 0\ 3\ |\ 6]$

Targeted matrix: $\begin{bmatrix} 1 & 1 & 1 & | & 6 \\ 0 & 1 & -1 & | & 2 \\ 0 & 0 & 3 & | & 6 \end{bmatrix}$

- Row 3: $3z = 6 \Rightarrow z = 2$
- Row 2: $y - z = 2 \Rightarrow y = 2 + 2 = 4$
- Row 1: $x + y + z = 6 \Rightarrow x = 6 - 4 - 2 = 0$

✅ **Final Answer:** $x = 0, \quad y = 4, \quad z = 2$

📘 **System of Equations:** Gaussian Elimination

$$\begin{cases} 2x + y - z = 5 \\ -3x + 4y + 2z = 7 \\ x - 2y + 3z = -1 \end{cases}$$

✅ Goal: Find the x, y, and z values that satisfy **all three equations**.

Go & Solve: https://matrixcalc.org/slu.html#solve-using-Gaussian-elimination

Best Method for Machine Learning (ML)

Method	Best For	Manual Use	Scales to Large Systems	Matrix-Friendly	Used in ML Internally
Substitution	Small (2-variable) problems	✅ Easy	❌ Becomes complex	❌ No	❌ Not practical
Cramer's Rule	Small square systems (2×2, 3×3)	✅ Clear	❌ Determinants expensive	❌ Not scalable	❌ Rare in ML
Gaussian Elimination	3+ variable systems, structured solving	✅ Systematic	✅ Works well	✅ Yes	✅ Core of many ML methods

A-Z Linear Algebra & Calculus for AI, Data Science and Machine Learning

Instructor:

Rashedul Alam Shakil

Founder of Study Mart

Founder of aiQuest Intelligence

Master in Data Science at FAU Erlangen

Automation Programmer at Siemens Energy, Germany

Conditional Independence

Best Method for Machine Learning (ML)

Exam: Test Your Knowledge!

Thank you