

What is the Rank of a Matrix?

The **rank** of a matrix is the maximum number of **linearly independent** rows or columns in the matrix. It gives an indication of the number of dimensions in the space spanned by the rows or columns of the matrix. In simpler terms, the rank tells us how many linearly independent vectors (either rows or columns) are in the matrix.

- If a matrix has full rank, it means that all of its rows (or columns) are linearly independent.
- If a matrix has deficient rank, it means some rows (or columns) are linearly dependent on others, making the matrix singular and not invertible.

For an $n \times m$ matrix A :

- The **rank** of A is at most $\min(n, m)$.
- A **square matrix** A (i.e., where $n = m$) is invertible if and only if its rank is equal to n (i.e., it has full rank).

Example of Matrix Rank

Consider the following 3×3 matrix A :

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

Here, the second row is a multiple of the first row (i.e., $\text{Row}_2 = 2 \times \text{Row}_1$), which means the second row is linearly dependent on the first row. Therefore, the rank of this matrix is 2 (since only two rows are linearly independent).

How to Invert a Matrix

To **invert a matrix** means finding another matrix A^{-1} such that:

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

Where I is the identity matrix (a matrix with 1's on the diagonal and 0's elsewhere). For a matrix to be invertible:

1. The matrix must be square (same number of rows and columns).
2. The matrix must have **full rank** (its rows/columns are linearly independent).
3. Determinant must be non-zero.

Steps to Invert a Matrix (2x2 Matrix Example)

Let's take a 2×2 matrix A :

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

The inverse of A , if it exists, is given by the formula:

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Where:

- $ad - bc$ is the **determinant** of the matrix A .
- The matrix is invertible if and only if the determinant $ad - bc \neq 0$.

Example:

Consider the matrix:

$$A = \begin{pmatrix} 4 & 7 \\ 2 & 6 \end{pmatrix}$$

1. Determinant:

$$\det(A) = (4 \times 6) - (7 \times 2) = 24 - 14 = 10$$

Since the determinant is non-zero ($10 \neq 0$), the matrix is invertible.

2. Inverse:

Using the formula for a 2×2 matrix:

$$A^{-1} = \frac{1}{10} \begin{pmatrix} 6 & -7 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{pmatrix}$$

Thus, the inverse of A is:

$$A^{-1} = \begin{pmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{pmatrix}$$

How to Invert a Matrix (General Case)

For larger matrices (i.e., $n \times n$ where $n > 2$), the inversion process is more complicated. Common methods include:

1. **Gaussian elimination:** Reducing the matrix to its reduced row echelon form and solving.
2. **LU decomposition:** Factorizing the matrix as a product of a lower triangular matrix L and an upper triangular matrix U .
3. **Adjugate and determinant method:** Using the matrix of minors, cofactors, and the determinant.
4. **Numerical methods:** For very large matrices, methods like Singular Value Decomposition (SVD) or QR decomposition are used.

Inverting $X^\top X$ in Linear Regression

In linear regression, we often deal with matrices that are not square, so we use the **normal equation** to compute the coefficient vector β :

$$\beta = (X^\top X)^{-1} X^\top Y$$

Here, $X^\top X$ is guaranteed to be square and invertible if X has full column rank (i.e., the columns of X are linearly independent). If $X^\top X$ is singular (i.e., not invertible), we need to regularize the problem (e.g., using Ridge regression).

Example of Rank and Inversion in Linear Regression

Suppose we have the following design matrix X for two data points and two features:

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

- The matrix X is not square, but $X^\top X$ is:

$$X^\top X = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}$$

- Now, compute the inverse of $X^\top X$:

$$(X^\top X)^{-1} = \frac{1}{(3)(14) - (6)(6)} \begin{pmatrix} 14 & -6 \\ -6 & 3 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 14 & -6 \\ -6 & 3 \end{pmatrix}$$

Thus, $(X^\top X)^{-1}$ can be computed, and it is invertible because the determinant of $X^\top X$ is non-zero.

In linear regression, you're trying to solve for the vector of coefficients β in the equation:

$$X\beta = Y$$

Where:

- \mathbf{X} is the design matrix (containing the input data, or features).
- $\boldsymbol{\beta}$ is the vector of coefficients (the parameters you're trying to estimate).
- \mathbf{Y} is the output vector (the target variable, or dependent variable).

Why can't we directly solve $\mathbf{X}\boldsymbol{\beta} = \mathbf{Y}$?

In general, the matrix \mathbf{X} is not square; it has more rows (data points) than columns (features).

Therefore, you can't directly compute the inverse of \mathbf{X} because:

1. \mathbf{X} might not be invertible (it could be rank-deficient, meaning some columns are linearly dependent).
2. Even if \mathbf{X} is square, the number of rows and columns may not match, so \mathbf{X} may not have an inverse.

The role of $\mathbf{X}^\top \mathbf{X}$ and why we use it ?

The standard solution to the linear regression problem comes from minimizing the **least squares error**, which means you're minimizing the sum of squared differences between the predicted values ($\mathbf{X}\boldsymbol{\beta}$) and the actual values \mathbf{Y} .

The **least squares solution** can be written as:

$$\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

Least Squares Problem:

The goal is to minimize the residual sum of squares (RSS), which is:

$$\text{RSS} = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

To minimize this, you take the derivative with respect to $\boldsymbol{\beta}$ and set it equal to zero:

$$\frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{Y}^\top \mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}) = 0$$

This simplifies to:

$$\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y}$$

This is called the **normal equation**.