

Math for ML Notes

Ahmed Yasser

September 2025

Chapter 2

2.6.2 Rank

Definition (Rank)

The number of **linearly independent** columns of a matrix $A \in \mathbb{R}^{m \times n}$ equals the number of **linearly independent** rows of A and is called the **rank** of A , and is denoted by $\text{rk}(A)$

Remark (Matrix Rank Properties)

- **Rank equality:** For any matrix A , the column rank equals the row rank: $\text{rk}(A) = \text{rk}(A^\top)$
- **Column space (image/range):** The columns of $A \in \mathbb{R}^{m \times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(A)$. A basis for U can be found using Gaussian elimination to identify pivot columns.
- **Row space:** The rows of $A \in \mathbb{R}^{m \times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(A)$. A basis for W can be found by applying Gaussian elimination to A^\top .
- **Invertibility condition:** For square matrices $A \in \mathbb{R}^{n \times n}$, A is regular (invertible) if and only if $\text{rk}(A) = n$.
- **Linear system solvability:** For $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, the system $Ax = b$ can be solved if and only if $\text{rk}(A) = \text{rk}(A|b)$, where $A|b$ is the augmented matrix.
- **Null space (kernel):** For $A \in \mathbb{R}^{m \times n}$, the solution space of $Ax = 0$ has dimension $n - \text{rk}(A)$. This subspace is called the kernel or null space.
- **Full rank:** A matrix $A \in \mathbb{R}^{m \times n}$ has full rank when $\text{rk}(A) = \min(m, n)$, meaning its rank equals the maximum possible rank for its dimensions.
- **Rank deficient:** A matrix that does not have full rank is called rank deficient.

2.7 Linear Mappings

Definition (Linear Mapping)

For vector spaces V and W , a mapping $\Phi : V \rightarrow W$ is called a **linear mapping (or linear transformation/vector space homomorphism)** if

$$\Phi(x + y) = \Phi(x) + \Phi(y)$$

$$\Phi(\lambda x) = \lambda \Phi(x)$$

$\forall x, y \in V$ and $\lambda \in \mathbb{R}$

Remark

Consider a mapping $\Phi : \mathcal{V} \rightarrow \mathcal{W}$, where \mathcal{V} and \mathcal{W} can be arbitrary sets. Then Φ is called:

- **Injective (one-to-one)** if $\forall x, y \in \mathcal{V} : \Phi(x) = \Phi(y) \implies x = y$ i.e. there is no two different elements in \mathcal{V} that map to the same element in \mathcal{W} .
- **Surjective (onto)** if $\Phi(\mathcal{V}) = \mathcal{W}$ i.e. every element in \mathcal{W} can be reached from \mathcal{V} using Φ .
- **Bijective** if Φ is both **injective** and **surjective**.

Remark

A bijective mapping $\Phi : \mathcal{V} \rightarrow \mathcal{W}$ is reversible: there exists a mapping $\Psi : \mathcal{W} \rightarrow \mathcal{V}$ such that $\Psi \circ \Phi(x) = x$ and $\Phi \circ \Psi(y) = y$. This mapping Ψ is the **inverse** of Φ , denoted Φ^{-1} .

Remark Special cases of linear mappings between vector spaces

- **Isomorphism:** $\Phi : V \rightarrow W$ linear and bijective (maps between different spaces, reversible)
- **Endomorphism:** $\Phi : V \rightarrow V$ linear (maps a space to itself)
- **Automorphism:** $\Phi : V \rightarrow V$ linear and bijective (maps a space to itself, reversible)
- **Identity mapping:** $\text{id}_V : V \rightarrow V, x \mapsto x$ (leaves every vector unchanged)

Remark

Finite dimensional vector spaces V and W are **isomorphic** if and only if $\dim(V) = \dim(W)$

Remark

Consider vector spaces V, W, X . Then:

- If $\Phi : V \rightarrow W$ and $\Psi : W \rightarrow X$ are **linear** then $\Psi \circ \Phi : V \rightarrow X$ is **linear**.
- If $\Phi : V \rightarrow W$ is an **isomorphism** then $\Phi^{-1} : W \rightarrow V$ is an **isomorphism**.
- If $\Phi : V \rightarrow W$ and $\Psi : V \rightarrow W$ are **linear** then $\Phi + \Psi : V \rightarrow W$ and $\lambda\Phi : V \rightarrow W, \lambda \in \mathbb{R}$ are **linear**.

2.7.1 Matrix Representation of Linear Mappings

Remark (Notation)

- $B = \{b_1, \dots, b_n\}$ is an **unordered** basis
- $B = (b_1, \dots, b_n)$ is an **ordered** basis
- $B = [b_1 \ \dots \ b_n]$ is a **matrix** whose columns are the vectors b_1, \dots, b_n

Definition (Coordinates)

Consider a vector space V and an **ordered** basis $B = (b_1, \dots, b_n)$ of V . For any vector $x \in V$ we obtain a **unique** representation (linear combination)

$$x = \alpha_1 b_1 + \dots + \alpha_n b_n$$

of x with respect to B . Then $\alpha_1, \dots, \alpha_n$ are the *coordinates* of x with respect to B , and the vector

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$$

is the *coordinate vector/coordinate representation* of x with respect to the ordered basis B .

Remark

A basis effectively defines a coordinate system and any basis of the vector space defines a valid coordinate system. The coordinates of a vector may be different between different basis.

Remark

For an n -dimensional vector space V and an ordered basis B of V , the mapping $\Phi : \mathbb{R}^n \rightarrow V, \Phi(e_i) = b_i, i = 1, \dots, n$, is **linear** and **bijective** (since V and \mathbb{R}^n are of the same dimension), where (e_1, \dots, e_n) is the **standard basis** of \mathbb{R}^n .

Definition (Transformation Matrix)

Consider vector spaces V and W with corresponding **ordered** bases $B = (b_1, \dots, b_n)$ and $C = (c_1, \dots, c_m)$. Also consider a **linear mapping** $\Phi : V \rightarrow W$. For $j \in \{1, \dots, n\}$

$$\Phi(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \alpha_{2j}\mathbf{c}_2 + \dots + \alpha_{mj}\mathbf{c}_j = \sum_{i=1}^m \alpha_{ij}\mathbf{c}_i$$

is the unique representation (linear combination) of $\Phi(\mathbf{b}_j)$ with respect to the C . Then we call the $m \times n$ matrix \mathbf{A}_Φ , whose elements are given by

$$A_{\Phi(i,j)} = \alpha_{ij},$$

The *transformation matrix* of Φ with respect to the ordered bases B of V and C of W .

Remark

From the definition of the transformation matrix we can see that the coordinates of $\Phi(\mathbf{b}_j)$ with respect to the ordered basis C of W are the j -th column of \mathbf{A}_Φ

Corollary

Consider *finite dimensional* vector spaces V, W with ordered basis B, C and a linear mapping $\Phi : V \rightarrow W$ with transformation matrix \mathbf{A}_Φ . If $\hat{\mathbf{x}}$ is the **coordinate vector** of $\mathbf{x} \in V$ with respect to B and $\hat{\mathbf{y}}$ is the **coordinate vector** of $\mathbf{y} = \Phi(\mathbf{x}) \in W$ with respect to C , then

$$\hat{\mathbf{y}} = \mathbf{A}_\Phi \hat{\mathbf{x}}.$$

This means that the transformation matrix can be used to map coordinates with respect to an ordered basis in V to coordinates with respect to an ordered basis in W .

2.7.2 Basis Change

Theorem (Basis Change)

Let $\Phi : V \rightarrow W$ be a linear mapping between vector spaces with ordered bases

$$B = (\mathbf{b}_1, \dots, \mathbf{b}_n) \quad \text{and} \quad \tilde{B} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n)$$

of V , and

$$C = (\mathbf{c}_1, \dots, \mathbf{c}_m) \quad \text{and} \quad \tilde{C} = (\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_m)$$

of W .

If \mathbf{A}_Φ is the transformation matrix of Φ with respect to bases B and C , then the transformation matrix $\tilde{\mathbf{A}}_\Phi$ with respect to bases \tilde{B} and \tilde{C} is given by:

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1} \mathbf{A}_\Phi \mathbf{S}$$

where:

- $\mathbf{S} \in \mathbb{R}^{n \times n}$ is the transformation matrix of Id_V that maps coordinates with respect to \tilde{B} onto coordinates with respect to B in V
- $\mathbf{T} \in \mathbb{R}^{m \times m}$ is the transformation matrix of Id_W that maps coordinates with respect to \tilde{C} onto coordinates with respect to C in W

Definition (Equivalence)

Two matrices \mathbf{A} and $\tilde{\mathbf{A}} \in \mathbb{R}^{m \times n}$ are **equivalent** if there exists **regular** matrices $\mathbf{S} \in \mathbb{R}^{n \times n}$ and $\mathbf{T} \in \mathbb{R}^{m \times m}$, such that $\tilde{\mathbf{A}} = \mathbf{T}^{-1} \mathbf{A} \mathbf{S}$

in other words, two matrices \mathbf{A} and $\tilde{\mathbf{A}}$ are **equivalent** if they represent the **same linear transformation** $\Phi : V \rightarrow W$.

Definition (Similarity)

Two matrices A and $\tilde{A} \in \mathbb{R}^{n \times n}$ are **similar** if there exists a **regular** matrix $S \in \mathbb{R}^{n \times n}$, such that $\tilde{A} = S^{-1}AS$

in other words, two matrices A and \tilde{A} are **similar** if they represent the **same linear transformation** $\Phi : V \rightarrow V$.

Remark

Similar matrices are always **equivalent**. However, **equivalent** matrices are not necessary **similar**.

Remark

The composition of two linear transformations, $\Phi : V \rightarrow W$ and $\Psi : W \rightarrow X$, results in another linear transformation $\Psi \circ \Phi : V \rightarrow X$. The matrix representing this combined transformation, $A_{\Psi \circ \Phi}$, is the product of the individual transformation matrices: $A_{\Psi \circ \Phi} = A_{\Psi}A_{\Phi}$.

Definition (Kernel (Null Space))

For a linear transformation $\Phi : V \rightarrow W$, the **kernel** is the set of all vectors $v \in V$ that are mapped to the zero vector of W .

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{v \in V \mid \Phi(v) = \mathbf{0}_W\}$$

Definition (Image (Range))

For a linear transformation $\Phi : V \rightarrow W$, the **image** is the set of all vectors $w \in W$ that can be “reached” by the transformation from some vector in V .

$$\text{Im}(\Phi) := \Phi(V) = \{w \in W \mid \exists v \in V, \Phi(v) = w\}$$

Intuitively, the **kernel** is the set of vector $v \in V$ that Φ maps onto the zero vector $\mathbf{0}_W \in W$. The **image** is the set of all vectors $w \in W$ that can be *reached* by Φ from any vector in V .

Remark

For $\Phi : V \rightarrow W$, we call V the **domain** and W the **codomain**

The kernel is the set of vectors $v \in V$ that Φ maps onto the *neutral element* $\mathbf{0}_W \in W$

Remark

Consider a linear mapping $\Phi : V \rightarrow W$, where V and W are vector spaces:

- It always holds that $\Phi(\mathbf{0}_V) = \mathbf{0}_W$, therefore, $\mathbf{0}_V \in \ker(\Phi)$. In particular, the null spaces is never empty.
- $\text{Im}(\Phi) \subseteq W$ is a **subspace** of W , and $\ker(\Phi) \subseteq V$ is a **subspace** of V .

Theorem (Rank-Nullity Theorem or Fundamental Theorem of Linear Mappings)

For vector spaces V and W and linear mapping $\Phi : V \rightarrow W$ it holds that

$$\dim(\ker(\Phi)) + \dim(\text{Im}(\Phi)) = \dim(V)$$

Remark (Direct Consequences of the Rank-Nullity Theorem)

- If $\dim(\text{Im}(\Phi)) < \dim(V)$, then the $\ker(\Phi)$ is **non-trivial**, i.e., the kernel contains more than $\mathbf{0}_V$ and $\dim(\ker(\Phi)) \geq 1$
- If A_{Φ} is the transformation matrix of Φ with respect to an ordered basis and $\dim(\text{Im}(\Phi)) < \dim(V)$, then the system of linear equations $A_{\Phi}x = \mathbf{0}$ has **infinitely many solutions**.