Math for ML Notes

Ahmed Yasser

August 2025

Chapter 2

2.6.2 Rank

Definition (Rank)

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

Remark (Matrix Rank Properties)

- **Rank equality**: For any matrix $\boldsymbol{A}$, the column rank equals the row rank: $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^\top)$
- **Column space (image/range)**: The columns of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(\boldsymbol{A})$. A basis for U can be found using Gaussian elimination to identify pivot columns.
- **Row space**: The rows of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(\boldsymbol{A})$. A basis for W can be found by applying Gaussian elimination to $\boldsymbol{A}^\top$.
- **Invertibility condition**: For square matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{A}$ is regular (invertible) if and only if $\text{rk}(\boldsymbol{A}) = n$.
- **Linear system solvability**: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be solved if and only if $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}|\boldsymbol{b})$, where $\boldsymbol{A}|\boldsymbol{b}$ is the augmented matrix.
- **Null space (kernel)**: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the solution space of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has dimension $n - \text{rk}(\boldsymbol{A})$. This subspace is called the kernel or null space.
- **Full rank**: A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank when $\text{rk}(\boldsymbol{A}) = \min(m, n)$, 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 $\Phi : V \to W$ is called a **linear mapping (or linear transformation/vector space homomorphism)** if

$$\Phi(\boldsymbol{x} + \boldsymbol{y}) = \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})$$
$$\Phi(\lambda \boldsymbol{x}) = \lambda \Phi(\boldsymbol{x})$$

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

Remark

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

- **Injective (one-to-one)** if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y}$ 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 $\Phi : \mathcal{V} \to \mathcal{W}$ is reversible: there exists a mapping $\Psi : \mathcal{W} \to \mathcal{V}$ 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**: $\Phi : V \to W$ linear and bijective (maps between different spaces, reversible)
- **Endomorphism**: $\Phi : V \to V$ linear (maps a space to itself)
- **Automorphism**: $\Phi : V \to V$ linear and bijective (maps a space to itself, reversible)
- **Identity mapping**: $\mathrm{id}_V : V \to V$, $x \mapsto x$ (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 $\Phi : V \to W$ and $\Psi : W \to X$ are **linear** then $\Psi \circ \Phi : V \to X$ is **linear**.
- If $\Phi : V \to W$ is an **isomorphism** then $\Phi^{-1} : W \to V$ is an **isomorphism**.
- If $\Phi : V \to W$ and $\Psi : V \to W$ are **linear** then $\Phi + \Psi : V \to W$ and $\lambda\Phi : V \to W, \lambda \in \mathbb{R}$ are **linear**.

2.7.1 Matrix Representation of Linear Mappings

Remark (Notaion)

- $B = \{b_1, ..., b_n\}$ is an **unordered** basis
- $B = (b_1, ..., b_n)$ is an **ordered** basis
- $\boldsymbol{B} = \begin{bmatrix} b_1 & ... & b_n \end{bmatrix}$ is a **matrix** whose columns are the vectors $b_1, ..., b_n$

Definition (Coordinates)

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

$$\boldsymbol{x} = \alpha_1 b_1 + ... + \alpha_n b_n$$

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

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

is the *coordinate vector*/*coordinate representaion* 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 $\Phi : \mathbb{R}^n \to V, \Phi(e_i) = b_i,\ i = 1, ..., n$, is **linear** and **bijective** (since V and $\mathbb{R}^n$ are of the same dimension), where $(\boldsymbol{e_i}, ..., \boldsymbol{e_n})$ is the **standard basis** of $\mathbb{R}^n$.

Definition (Transformation Matrix)

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

$$\Phi(\boldsymbol{b_j}) = \alpha_{1j}\boldsymbol{c_1} + \alpha_{2j}\boldsymbol{c_2} + \ldots + \alpha_{mj}\boldsymbol{c_j} = \sum_{i=1}^{m} \alpha_{ij}\boldsymbol{c_i}$$

is the unique representaion (linear combination) of $\Phi(\boldsymbol{b_j})$ with respect to the C. Then we call the $m \times n$ matrix $\boldsymbol{A}_\Phi$, 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 $\Phi(\boldsymbol{b_j})$ with respct to the ordered basis C of W are the j-th column of $\boldsymbol{A}_\Phi$

Corollary

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

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

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