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 $\Phi : V \to W$ 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 $\Phi : \mathcal{V} \to \mathcal{W}$, 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 $\Phi : \mathcal{V} \to \mathcal{W}$ 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}.

- **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 $(e_i, ..., e_n)$ is the **standard basis** of $\mathbb{R}^n$.

Definition (Transformation Matrix)

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

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

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

Corollary

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

$$\hat{y} = A_\Phi \hat{x}.$$

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

$$B = (b_1, ..., b_n) \quad \text{and} \quad \tilde{B} = (\tilde{b}_1, ..., \tilde{b}_n)$$

of V, and

$$C = (c_1, ..., c_m) \quad \text{and} \quad \tilde{C} = (\tilde{c}_1, ..., \tilde{c}_m)$$

of W.

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

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

where:
- $S \in \mathbb{R}^{n \times n}$ is the transformation matrix of Id_V that maps coordinates with respect to $\tilde{B}$ onto coordinates with respect to B in V
- $T \in \mathbb{R}^{m \times m}$ 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 A and $\tilde{A} \in \mathbb{R}^{m \times n}$ are **equivalent** if there exists **regular** matrices $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$, such that $\tilde{A} = T^{-1} A S$

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

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

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

Remark

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

Remark

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