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 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 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 $\boldsymbol{Ax} = \boldsymbol{b}$ 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 $\boldsymbol{Ax} = \boldsymbol{0}$ 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

Theorem (Basis Change)

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

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

of V, and

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

of W.

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

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

where:

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

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

Definition (Similarity)

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: $\boldsymbol{A}_{\Psi\circ\Phi} = \boldsymbol{A}_{\Psi}\boldsymbol{A}_{\Phi}$.

Definition (Kernel (Null Space))

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

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

Definition (Image (Range))

For a linear transformation $\Phi : V \to W$, the **image** is the set of all vectors $\boldsymbol{w} \in W$ that can be "reached" by the transformation from some vector in V.

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

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

Remark

For $\Phi : V \to W$, 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 $\Phi : V \to W$, 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.
- $\mathrm{Im}(\Phi) \subseteq W$ 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 $\Phi : V \to W$ it holds that

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

Remark (Direct Consequences of the Rank-Nullity Theorem)

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

2.8.1 Affine Subspaces

Definition (Affine Subspace)

let V be a vector space, $\boldsymbol{x_0} \in V$ and $U \subseteq V$ a subspace. Then the subset

$$\begin{aligned} L = \boldsymbol{x_0} + U &:= \{\boldsymbol{x_0} + \boldsymbol{u} : u \in U\} \\ &= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x_0} + \boldsymbol{u}\} \subseteq V \end{aligned}$$

is called **affine subspace** or **linear manifold** of V. U is called **direction** or **direction space**, and $\boldsymbol{x_0}$ is called **support point**.

Remark

The definition of an affine subspace *excludes* $\mathbf{0}$ if $\boldsymbol{x_0} \notin U$. Therefore, an affine subspace is not a **linear** subspace (vector subspace) of V for $\boldsymbol{x}_0 \notin U$.

Remark

Consider two affine subspaces $L = \boldsymbol{x_0} + U$ and $\tilde{L} = \widetilde{\boldsymbol{x_0}} + \tilde{U}$ of a vector space V. Then, $L \subseteq \tilde{L}$ if and only if $U \subseteq \tilde{U}$ and $\boldsymbol{x_0} \in \tilde{L}$

Remark

Affine subspaces are often described by parameters: Consider a k-dimensional affine space $L = \boldsymbol{x_0} + U$ of V. if $(\boldsymbol{b_1}, ..., \boldsymbol{b_k})$ is an ordered basis of U, then every element $\boldsymbol{x} \in L$ can be uniquely written as

$$\boldsymbol{x} = \boldsymbol{x_0} + \lambda_1 \boldsymbol{b_1} + ... + \lambda_k \boldsymbol{b_k}$$

where $\lambda_1, ..., \lambda_k \in \mathbb{R}$. Ths representaion is called the **parametric equation** of L with **directional vectors** $\boldsymbol{b_1}, ..., \boldsymbol{b_k}$ and **parameter** $\lambda_1, ..., \lambda_k$.

Remark

In $\mathbb{R}^n$, the $(n-1)$-dimensional affine subspaces are called **hyperplanes**, and the corresponding parametric equation is:

$$\boldsymbol{y} = \boldsymbol{x}_0 + \sum_{i=1}^{n-1} \lambda_i \boldsymbol{b}_i$$

where $\boldsymbol{b}_1, ..., \boldsymbol{b}_{n-1}$ form a basis of an $(n-1)$-dimensional subspace U of $\mathbb{R}^n$.

This means that a hyperplane is defined by:

- A **support point** $\boldsymbol{x}_0$
- $(n-1)$ linearly independent vectors $\boldsymbol{b}_1, ..., \boldsymbol{b}_{n-1}$ that span the direction space

Examples:

- In $\mathbb{R}^2$, a line is a hyperplane
- In $\mathbb{R}^3$, a plane is a hyperplane

Remark (Inhomogeneous systems of linear equations and affine subspaces)

For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the solution of the system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is either the **empty set** or an **affine subspace** of $\mathbb{R}^n$ of dimension $n - \text{rk}(\boldsymbol{A})$.

In particular, the solution of the linear equation

$$\lambda_1 \boldsymbol{b}_1 + ... + \lambda_n \boldsymbol{b}_n = \boldsymbol{x}$$

where $(\lambda_1, ..., \lambda_n) \neq (0, ..., 0)$, is a hyperplane in $\mathbb{R}^n$.

In $\mathbb{R}^n$, every k-dimensional affine subspace is the solution of an inhomogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\mathrm{rk}(\boldsymbol{A}) = n - k$.

Note: Recall that for homogeneous equation systems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ the solution was a vector subspace, which we can also think of as a special affine space with support point $\boldsymbol{x}_0 = \boldsymbol{0}$.