CIS3333: Mathematics of Machine Learning Fall 2025

Lecture: Linear Algebra Review II

Date: October 6th, 2025 *Author: Surbhi Goel*

1 Recap and Preview

Last time we formalized the notion of vector spaces: sets with addition and scalar multiplication that satisfy nice properties (Abelian group structure, distributivity, etc.). We saw that vector spaces can be familiar ($\mathbb{R}^n$) or exotic (polynomials, positive reals with unusual operations).

Today we'll develop tools to understand the *structure* of vector spaces: when can we represent every vector using a smaller set of "building blocks"? This leads to the concepts of linear independence, basis, and rank—which will directly answer our motivating question from last time: when does the system $X\theta = Y$ have a solution?

2 Linear Independence, Basis and Rank

2.1 Linear Combinations and Independence

- A linear combination is a combination of scaled vectors:

$$v = \sum_i \lambda_i x_i$$

- If $x_i \in \mathbb{R}^d$ then we will typically abbreviate this as $\lambda^\top X$ where X is the matrix of elements stacked in each row
- If there exists λ such that $0 = \sum_i \lambda_i x_i$ with at least one $\lambda_i \neq 0$ then they are linearly dependent. If no such non-zero solution exists, they are linearly independent.
- Properties of linear independence:
  - k vectors are either linearly dependent or independent
  - If a vector is 0 or if the same vector is repeated, they are dependent

2.2 Generating Sets, Span, and Basis

- Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and let $\mathcal{A} = \{x_1, \ldots, x_k\} \subset \mathcal{V}$. If every vector $v \in \mathcal{V}$ can be expressed as a linear combination of $\mathcal{A}$, then this is a *generating set* of $\mathcal{V}$.
- The set of all linear combinations of vectors in $\mathcal{A}$ is the *span* of $\mathcal{A}$.
- A generating set $\mathcal{A}$ is *minimal* if there does not exist a smaller $\bar{\mathcal{A}} \subsetneq \mathcal{A}$ that spans V.

- Every independent generating set of V is minimal and is called a *basis* of V.
- Let $\mathcal{B} \subseteq \mathcal{V}$, $\mathcal{B} \neq \emptyset$. The following statements are equivalent:
  - $\mathcal{B}$ is a basis
  - $\mathcal{B}$ is a minimal generating set
  - $\mathcal{B}$ is a maximally linearly independent set of vectors in V, i.e. adding any vector will make the set linearly dependent
  - Every vector $x \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique:
$$x = \sum_i \lambda_i b_i = \sum_i \psi_i b_i \Rightarrow \lambda_i = \psi_i$$
- **Example: Standard basis.** $\mathcal{B} = \{e_1, \ldots, e_k\}$ where e_i is zero everywhere except for the ith position which is 1.
- **Example: Alternative basis for $\mathbb{R}^3$.**
$$\mathcal{B} = \left\{ b_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, b_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, b_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$
  is also a basis for $\mathbb{R}^3$ (check: these are linearly independent and span $\mathbb{R}^3$).
- Every vector space has a basis $\mathcal{B}$. There can be multiple bases, i.e. they are not unique. However, they all have the same number of basis vectors.
- The *dimension* of V is the number of basis vectors of V, denoted as $\dim(V)$.
- A finite dimensional vector space is one where $\dim(V) < \infty$.
- **Example: Polynomial basis.** Recall from last time that $\mathcal{P}_n = \{p(x) = a_0 + a_1 x + \cdots + a_n x^n : a_i \in \mathbb{R}\}$ is a vector space. The set $\mathcal{B} = \{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathcal{P}_n$:
  - Linearly independent: if $a_0 \cdot 1 + a_1 \cdot x + \cdots + a_n \cdot x^n = 0$ for all x, then $a_0 = a_1 = \cdots = a_n = 0$
  - Generating set: every polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ is a linear combination of $\mathcal{B}$
  - Therefore $\dim(\mathcal{P}_n) = n + 1$

  This shows that a vector space of functions can have a finite basis!
- **Example: A countably infinite basis.** Consider the vector space $\mathcal{V}$ of all infinite sequences $v = (v_1, v_2, \ldots)$ where only a *finite* number of entries v_i are non-zero. This space has a *countably infinite*[1] basis $\mathcal{B} = \{e_1, e_2, e_3, \ldots\}$, where e_i is the sequence with a 1 in position i and 0s elsewhere. Any vector $v \in \mathcal{V}$ can be written as a *finite* sum $v = \sum_{i=1}^{k} v_i e_i$ for some k.

[1] A set is "countably infinite" if its elements can be put into one-to-one correspondence with the natural numbers (i.e., you can list them). This is the simplest "size" of infinity and serves as a stepping stone to more complex function spaces (like Hilbert spaces) which can have *uncountably infinite* dimensions, meaning their basis vectors cannot be listed out. Uncountably infinite spaces require tools for handling infinite sums (convergence).

2.3 Rank

- The number of linearly independent columns of a matrix A equals the number of linearly independent rows and is called the rank of A, denoted as $\text{rk}(A)$.
- Properties:
  - $\text{rk}(A) = \text{rk}(A^\top)$
  - The columns of A span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(A)$. This subspace is called the image or range of A.
  - Similarly, the rows of A span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(A)$
  - For square matrices $A \in \mathbb{R}^{n \times n}$, A is invertible (regular) if and only if $\text{rk}(A) = n$.
  - For all A, b the linear system $Ax = b$ can be solved if and only if $\text{rk}(A) = \text{rk}(A|b)$.
  - For $A \in \mathbb{R}^{m \times n}$, the subspace of solutions x such that $Ax = 0$ has dimension $n - \text{rk}(A)$. This subspace is called the kernel, or the null space.
  - A matrix has full rank if $\text{rk}(A) = \min(m, n)$. Otherwise, it is rank deficient.

Connection to $X\theta = Y$. Recall our motivating question: when does $X\theta = Y$ have a solution? The answer: if and only if $\text{rk}(X) = \text{rk}(X|Y)$, i.e., when Y lies in the column space of X. If $X \in \mathbb{R}^{N \times D}$ has full column rank ($\text{rk}(X) = D$), then the solution is unique. If $\text{rk}(X) < D$, there are infinitely many solutions (the null space has dimension $D - \text{rk}(X)$).

3 Linear Mappings

Bases provide structure for vector spaces. We now consider mappings between vector spaces that *preserve* this structure—just as group homomorphisms preserve group structure.

3.1 Definition and Properties

- Let V, W be two vector spaces. A linear mapping $\Phi : V \to W$ satisfies

$$\forall x, y \in V, \forall \lambda, \psi \in \mathbb{R} : \Phi(\lambda x + \psi y) = \lambda \Phi(x) + \psi \Phi(y)$$

  These are sometimes also called vector space homomorphisms or linear transformations.
- In the vector space $\mathbb{R}^n$, we can represent linear mappings as matrices: if $\Phi(x) = Ax$, then Φ is linear.
- A mapping $\Phi : \mathcal{V} \to W$ on arbitrary sets $\mathcal{V}, \mathcal{W}$ is called:
  - Injective if $\forall x, y : \Phi(x) = \Phi(y) \Rightarrow x = y$ (different vectors map to different outputs)
  - Surjective if $\Phi(\mathcal{V}) = \mathcal{W}$ (all elements can be reached)
  - Bijective if it is both injective and surjective (operation can be undone)
- $\Phi : V \to W$ is an isomorphism if it is both linear and bijective

- $\Phi : V \to V$ is an endomorphism if it is linear
- $\Phi : V \to V$ is an automorphism if it is both linear and bijective
- $\mathrm{id}_V : V \to V$ is the identity mapping, which is an automorphism.
- Theorem: Finite dimensional vector spaces V, W are isomorphic if and only if $\dim(V) = \dim(W)$ (Axler 2015).
- Intuitively, this means that vector spaces with the same dimension are "equivalent"—you can transform from one to the other without losing information. For example, we can treat $\mathbb{R}^{m \times n}$ matrices as equivalent to $\mathbb{R}^{mn}$ vectors.
- More properties:
  - Let V, W, X be vector spaces. If $\Phi : V \to W$ and $\Psi : W \to X$ are linear mappings, then $\Psi \circ \Phi : V \to X$ is also linear.
  - If $\Phi : V \to W$ is an isomorphism, then $\Phi^{-1} : W \to V$ is an isomorphism.
  - If $\Phi : V \to W, \Psi : V \to W$ are linear, then $\Phi + \Psi$ and $\lambda\Phi$ are also linear.

3.2 Transformation Matrices

The key insight: any n dimensional vector space is isomorphic to $\mathbb{R}^n$. This means we can represent abstract vectors as concrete tuples of numbers once we choose a basis.

- Let $B = (b_1, \ldots, b_n)$ be an ordered basis of V. For any $x \in V$, we can uniquely write $x = \sum_i \alpha_i b_i$. We call $\alpha_1, \ldots, \alpha_n$ the coordinates of x with respect to B, and the vector $\hat{x} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ is the coordinate vector (or coordinate representation) of x with respect to B.
- Think of a basis as defining a coordinate system. The coordinate vector $\hat{x} \in \mathbb{R}^n$ is the representation of x in that coordinate system.
- **Example: Different coordinate vectors for the same vector.** Consider the vector $v = (2, 3, 5) \in \mathbb{R}^3$ and the alternative basis $\mathcal{B}$ from above.
  - With respect to the standard basis:

    $$v = 2e_1 + 3e_2 + 5e_3$$

    So the coordinate vector is $\hat{v} = (2, 3, 5)$.
  - With respect to the basis $\mathcal{B} = \{b_1, b_2, b_3\}$:

    $$v = \alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3 = \alpha_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_2 + \alpha_3 \\ \alpha_3 \end{bmatrix}$$

    Setting this equal to $(2, 3, 5)$: we get $\alpha_3 = 5$, $\alpha_2 + 5 = 3$ so $\alpha_2 = -2$, and $\alpha_1 - 2 + 5 = 2$ so $\alpha_1 = -1$. So the coordinate vector is $\hat{v} = (-1, -2, 5)$.

Same geometric vector, different "addresses"! This is why the choice of basis matters—it changes how we represent vectors numerically.

- Transformation matrix: Let V, W be vector spaces with bases B, C, and consider a linear mapping $\Phi : V \to W$. For $j \in \{1, \ldots, n\}$ let

$$\Phi(b_j) = \sum_i \alpha_{ij} c_i$$

be the unique representation of $\Phi(b_j)$ with respect to C. Then, if A is the matrix given by $A_{ij} = \alpha_{ij}$, we call A the transformation matrix of Φ with respect to bases B and C.

- Key point: *Any linear mapping between finite dimensional spaces can be represented as a matrix.* Just pick bases for the domain and codomain, and compute the coefficients!

- 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 $\hat{y} = A_\Phi \hat{x}$ where A_Φ is the transformation matrix of Φ.

Why different bases matter. The same linear transformation can have very different matrix representations depending on the choice of basis. Some choices make computations much easier—for example, a diagonal matrix is trivial to invert or raise to a power. Next time, we'll see how to systematically change between bases to find these "nice" representations.