

Mathematics for Machine Learning

Anirudh Gupta

November 20, 2020

Chapter 1

Vector Spaces

A vector space is the general mathematical structure we need to be able to talk about linear mappings. To introduce a vector space we need a lot of mathematical concepts. So we start with a group.

Definition 1. *A set G of elements with an operation $+: G \times G \rightarrow G$ is called a group if the following properties hold:*

G1 Associativity: $\forall a, b, c \in G : (a + b) + c = a + (b + c)$

G2 Identity element: $\exists e \in G : \forall g \in G : e + g = g + e = g$

G3 Inverse element: $\forall a \in G, \exists b \in G : a + b = b + a = e$

A group is called a commutative group (Abelian group) if we have additionally that $\forall a, b \in G : a + b = b + a$

Examples

- $(\mathbb{R}^n, +)$: This can be thought of as an n -dimensional vector with addition as the associated operation. The addition of three vectors can be done in any order and is thus associative. The identity element in this case is the zero vector. The inverse element is the negative of each element. Thus this combination forms a group.
- (\mathbb{R}^+, \cdot) : This is the set of positive real numbers with multiplication as the associated operation. This also forms a group.
- (\mathbb{R}^-, \cdot) : This is set of negative real numbers with associated operation as multiplication. It does not form a group, since multiplication of two negative real numbers gives us a positive real number and that is out of the set considered. We can also say that the set of negative real numbers is not closed with respect to multiplication.

Definition 2. *A set F with two operation $(+, \cdot) : F \times F \rightarrow F$ is called a field if the following properties hold:*

F1 $(F, +)$ is a commutative group with identity element 0.

F2 $(F \setminus \{0\}, \cdot)$ is a commutative group with identity element 1.

F3 Distributivity: $\forall a, b, c \in F : a \cdot (b + c) = a \cdot b + a \cdot c$

The two most common fields are the real numbers $(\mathbb{R}, +, \cdot)$ and complex numbers $(\mathbb{C}, +, \cdot)$ with defined addition and multiplication.

Definition 3. Let F be a field with identity elements 0 and 1. A vector space over the field F is a set V with a mapping: $(+): V \times V \rightarrow V$ (vector addition) and a mapping $(\cdot): F \times V \rightarrow V$ (scalar multiplication) such that:

V1 $(V, +)$ is a commutative group.

V2 Multiplicative identity: $\forall v \in V : 1 \cdot v = v$

V3 Distributive property: $\forall a, b \in F$ and $\forall u, v \in V$

$$a \cdot (u + v) = a \cdot u + a \cdot v$$

$$(a + b)u = a \cdot u + b \cdot u$$

Elements of V are called vectors and elements of F are called scalars. Depending on whether the field is real or complex we call the space as real vector space or complex vector space.

Examples

- \mathbb{R}^n with standard operations of adding vectors and multiplying vectors with a scalar.
- Function spaces: These are spaces which consists of functions. And essentially the whole field of functional analysis exploits the fact if you group them in a vector space, there are many properties that you can find about functions without even explicitly looking at what type of functions you are talking about.

$\mathbb{R}^\chi : \{f : \chi \rightarrow \mathbb{R}\}$ the space of all real valued functions on a set χ .
Now we define the two operations:

$$+ : \mathbb{R}^\chi \times \mathbb{R}^\chi \rightarrow \mathbb{R}^\chi, (f + g)(x) := f(x) + g(x)$$

$$\cdot : \mathbb{R} \times \mathbb{R}^\chi \rightarrow \mathbb{R}^\chi, (\lambda \cdot f)(x) := \lambda \cdot (f(x))$$

Then $(\mathbb{R}^\chi, +, \cdot)$ is a vector space.

Chapter 2

Basis and Dimension

Definition 4. Let V be a vector space and $U \subset V$ a non-empty set. We call U a subspace of V if it is closed under linear combinations.

$\forall \lambda, \mu \in F$ and $\forall u, v \in U : \lambda u + \mu v \in U$, where F is the field where scalars come from.

So essentially subspace is a vector space on its own. And it is closed means that we can add two elements and multiply by scalars and still remain in that space.

Examples

- $C(\chi)$ is a subspace of \mathbb{R}^χ where $C(\chi)$ is the set of continuous functions on domain χ and \mathbb{R}^χ is any real valued function. Sum of two continuous functions is continuous and scalar multiplication of a continuous function is a continuous function as well.
- The set S of symmetric matrices of size $n \times m$ is a subspace of $\mathbb{R}^{n \times m}$. Since sum of two symmetric matrices is a symmetric matrix and the same goes for scalar multiplication.

Definition 5. Let V be a vector space over F and $u_1, \dots, u_n \in V$, $\lambda_1, \dots, \lambda_n \in F$. Then $\sum_{i=1}^n \lambda_i u_i$ is called a linear combination. The set of all linear combinations of u_1, \dots, u_n is called the span (linear hull) of u_1, \dots, u_n .
 $\text{span}(u_1, \dots, u_n) := \{\sum_{i=1}^n \lambda_i u_i \mid \lambda_i \in F\}$

The set $U := \{u_1, \dots, u_n\}$ is the generator of $\text{span}(u_1, \dots, u_n)$

Definition 6. A set of vectors v_1, \dots, v_n is called linearly independent if the following holds:

$$\sum_{i=1}^n \lambda_i v_i = 0 \implies \lambda_1 = \dots = \lambda_n = 0$$

Examples

- The vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \in \mathbb{R}^3$ are linearly independent.

- The functions $\sin(x)$ and $\cos(x) \in \mathfrak{R}$ are linearly independent.
- Any set of $d + 1$ vectors in \mathfrak{R}^d is linearly dependent.

Definition 7. A subset B of a vector space V is called a (Hammel) basis if:

- $\text{span}(B) = V$
- B is linearly independent.

This means that any vector in V can be written as a linear combination of vectors in B but basis vectors themselves cannot be written in terms of each other since they are independent.

Examples

- The canonical basis of \mathfrak{R}^3 is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.
- Another basis of \mathfrak{R}^3 is given by \mathfrak{R}^3 is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Proposition 1. If $U = \{u_1, \dots, u_n\}$ spans a vector space V , then the set U can be reduced to a basis of V .

Proof. If U is already linearly independent then we are done. If U is linearly dependent, $\exists u \in U$ that is a linear combination of other vectors in U . We repeat this step until we reach a point where all vectors in U are linearly independent. \square

Definition 8. A vector space is called finite dimensional if it has a finite basis.

Proposition 2. Let $U = \{u_1, \dots, u_n\} \subset V$ be a set of linearly independent vectors and let V be a finite dimensional vector space. Then U can be extended to a basis of V .

Proof. Let w_1, \dots, w_m be a basis of V . Consider a set $\{u_1, \dots, u_n, w_1, \dots, w_m\}$. Remove vectors from the end until remaining vectors are linearly independent. The remaining set $\text{spans}(V)$, is linearly independent by construction and contains U . \square

Corollary 1. Let V be a finite dimensional vector space. Then any two basis of V have the same length.

Definition 9. The length of the basis of a finite dimensional vector space is called its dimension.

We have defined what a basis and we also know what a subspace is. Another notion which brings these two things together is known as the *sum* and *direct sum* of subspaces.

Definition 10. Assume that we have two subspaces U_1 and U_2 of a vector space V . The sum of the two spaces is defined as :

$$U_1 + U_2 := \{u_1 + u_2 | u_1 \in U_1, u_2 \in U_2\}$$

The sum which has just been seen is known as the direct sum. If each element in the sum can be written in exactly one way $(U_1 \oplus U_2)$.

Proposition 3. *Suppose V is finite-dimensional, and $U \subset V$ is a subspace, then there exists a subspace $W \subset V$ such that $U \oplus W = V$.*

Proof. Let the set $\{u_1, \dots, u_k\}$ be a basis of U . Extend it to a basis of V , say the resulting set is $\{u_1, \dots, u_k, v_1, \dots, v_m\}$. Define $W = \text{span}\{v_1, \dots, v_m\}$ \square

Chapter 3

Linear Maps, Kernals and Range

Definition 11. Let U and V be vector spaces over the same field F . A mapping $T : U \rightarrow V$ is linear if $\forall u_1, u_2 \in U, \forall \lambda \in F$

$$\begin{aligned}T(u_1 + u_2) &= T(u_1) + T(u_2) \\T(\lambda u_1) &= \lambda T(u_1)\end{aligned}$$

The set of all linear mappings from $U \rightarrow V$ is denoted by $\mathcal{L}(U, V)$. If $U = V$, then linear mappings are denoted by $\mathcal{L}(U)$, which is a set of linear mappings from a set onto itself.

Examples

- $T : \mathcal{C}[a, b] \rightarrow \mathbb{R}, f \rightarrow \int_a^b f(x)dx$. Integration is linear operator since it does not matter whether you add two functions and then integrate or vice versa. Also, we can pull out scalar multiples out of the integrals.
- $D : \mathcal{C}^\infty[a, b] \rightarrow \mathcal{C}^\infty[a, b], f \rightarrow f'$. The derivative of the sum of functions is the same as taking the sum and then derivatives.

Definition 12. $T \in \mathcal{L}(U, V)$. Then the kernal of T (null space) is defined as :

$$\ker(T) := \text{null}(T) := \{u \in U | Tu = 0\}$$

Proposition 4. $\ker(T)$ is a subspace of U .

T is injective iff $\ker(T) = \{0\}$. Means if we take two input points in the input space then they are always mapped to different points in the output space. So 0 is always an element in the kernal since a linear map always maps 0 to 0. Thus the kernal is never empty. The smallest kernal of any vector space is just the 0 vector.

Definition 13. The range of T (image of T) is defined as:

$$\text{range}(T) := \text{Image}(T) := \{Tu | u \in U\}$$

Proposition 5. The range is always a subspace of V

T is surjective iff $\text{range}(T) = V$

Definition 14. $V' \subset V$ where V' is any set. The pre-image of V' is defined as:

$$T^{-1}(V') := \{u \in U \mid Tu \in V'\}$$

Proposition 6. If $V' \subset V$ is a subspace of V , then $T^{-1}(V')$ is a subspace of U .

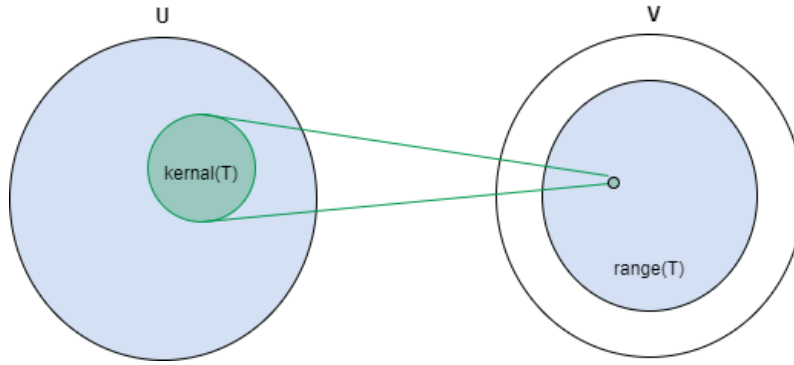


Figure 3.1: Kernel and range of a $T : U \rightarrow V$

We can see in the figure above that a part of U which is mapped to 0 is known as the kernel of the map T . The range lives in V and the kernel lives in U .

Theorem 1. Let V be finite dimensional W be any vector space and $T \in \mathcal{L}(V, W)$. Let u_1, \dots, u_n be a basis of $\ker(T) \subset V$. Let w_1, \dots, w_m be a basis of the $\text{range}(T) \subset W$. Then $u_1, \dots, u_n, T^{-1}(w_1), \dots, T^{-1}(w_m)$ forms a basis of V . In particular, $\dim(V) = \dim(\ker(T)) + \dim(\text{range}(T))$

Proof. Denote $T^{-1}(w_1) := z_1, \dots, T^{-1}(w_m) := z_m$. If $\{u_1, \dots, u_n, z_1, \dots, z_m\}$ for a basis of V , then they must span the whole space V and be linearly independent. Let's do each step by step.

Step 1: Prove that $V \subset \text{span}\{u_1, \dots, u_n, z_1, \dots, z_m\}$

Let $v \in V$, and $Tv \in \text{range}(T)$. Since we know the basis of $\text{range}(T)$,

$$\begin{aligned} \exists \lambda_1, \dots, \lambda_m : Tv &= \lambda_1 w_1 + \dots + \lambda_m w_m \\ &= \lambda_1 T(z_1) + \dots + \lambda_m T(z_m) \\ &= T(\lambda_1 z_1) + \dots + T(\lambda_m z_m) \end{aligned}$$

$\implies Tv - T(\lambda_1 z_1 + \dots + \lambda_m z_m) = T(v - (\lambda_1 z_1 + \dots + \lambda_m z_m)) = 0$
 $\implies (v - (\lambda_1 z_1 + \dots + \lambda_m z_m)) \in \ker(T)$ since this vector is mapped to 0 by the linear mapping. Since we also know the basis of the kernel any vector in the kernel can be written as:

$$\begin{aligned} (v - (\lambda_1 z_1 + \dots + \lambda_m z_m)) &= \mu_1 u_1 + \dots + \mu_n u_n \\ v &= \lambda_1 z_1 + \dots + \lambda_m z_m + \mu_1 u_1 + \dots + \mu_n u_n \end{aligned}$$

Step 2 : Prove that $u_1, \dots, u_n, z_1, \dots, z_m$ are linearly independent.

Assume that $\mu_1 u_1 + \dots + \mu_n u_n + \lambda_1 z_1 + \dots + \lambda_m z_m = 0$

If these vectors are linearly independent then all the coefficients will simultaneously go to zero.

$$\begin{aligned}\lambda_1 w_1 + \dots + \lambda_m w_m &= \lambda_1 T(z_1) + \dots + \lambda_m T(z_m) \\ &= \lambda_1 T(z_1) + \dots + \lambda_m T(z_m) + \mu_1 T(u_1) + \dots + \mu_n T(u_n) \\ &= T(\lambda_1 z_1 + \dots + \lambda_m z_m + \mu_1 u_1 + \dots + \mu_n u_n) = T(0) = 0\end{aligned}$$

We can add $\mu_1 T(u_1) + \dots + \mu_n T(u_n)$ to the expression since it goes to zero by definition of a kernel. Everything inside the operator is zero due to the assumption at the beginning of step 2. Since $\lambda_1 w_1 + \dots + \lambda_m w_m = 0$ and w_1, \dots, w_m is a basis $\implies \lambda_1 = \dots = \lambda_m = 0$

Now $\mu_1 u_1 + \dots + \mu_n u_n = 0$ since we have already proved the other half of the assumption to be zero. Since u_1, \dots, u_n is the basis of the kernel $\implies \mu_1 = \dots = \mu_n = 0$. \square

Proposition 7. Let $T \in \mathcal{L}(V, W)$ and V, W be finite dimensional, then the following statements are equivalent.

- T is injective
- T is surjective
- T is bijective

This proposition does not hold in infinite dimensional spaces.

Chapter 4

Matrices