Mathematics for Machine Learning

Anirudh Gupta

November 26, 2020

Chapter 1

Vector Spaces

A vector space is the general mathemtical 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 $+$: $G \times G \to G$ 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

- $(\Re^n, +)$: 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.
- $(\Re^+, .)$: This is the set of positive real numbers with multiplication as the associated operation. This also forms a group.
- $(\Re^-, .)$: 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 $(+, \cdot) : F \times F \to F$ 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 ral numbers $(\Re, +, \cdot)$ 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: $(+) : V \times V \to V$ (vector addition) and a mapping $(\cdot) : F \times V \to V$ (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

- $\Re^n$ 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.

  $\Re^\chi$: $\{f : \chi \to \Re\}$ the space of all real valued functions on a set χ. No we define the two operations:

  $$+ : \Re^\chi \times \Re^\chi \to \Re^\chi \ , \ (f + g)(x) := f(x) + g(x)$$
  $$\cdot : \Re \times \Re^\chi \to \Re^\chi, (\lambda \cdot f)(x) := \lambda \cdot (f(x))$$

  Then $(\Re^\chi, +, \cdot)$ 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* $\forall u, v \in V : \lambda u + \mu v \in U$*, 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 $\Re^{\chi}$ where $C(\chi)$ is the set of continuous functions on domain χ and $\Re^{\chi}$ is any real valued function. Sum of two continuous functions is continuous and scalar multiplication of a continous function is a continous function as well.
- The set S of symmetric matrices of size $n \times m$ is a subspace of $\Re^{n \times m}$. 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 $u_1, ..., u_n \in V$, $\lambda_1, ...\lambda_n \in F$. Then $\sum_{i=1}^{n} \lambda_i u_i$ is called a linear combination. The set of all linear combinations of $u_1, ..., u_n$ is called the span (linear hull) of $u_1, ..., u_n$.*
$span(u_1, ..., u_n) := \{\sum_{i=1}^{n} \lambda_i u_i | \lambda_i \in F\}$

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

Definition 6. *A set of vectors $v_1, ...v_n$ is called linearly independent if the following holds:*
$\sum_{i=1}^{n} \lambda_i v_i = 0 \implies \lambda_1 = ... = \lambda_n = 0$

Examples

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

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

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

- *$span(B) = V$*
- *B is linearly independent.*

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

Examples

- The canonical basis of $\Re^3$ 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 $\Re^3$ is given by $\Re^3$ 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 $U = \{u_1, ...u_n\}$ 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 untill we reach a point where all vectors in U are linearly independent. □

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

Proposition 2. *Let $U = \{u_1, ..., u_n\} \subset V$ 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 $w_1, ..., w_m$ be a basis of V. Consider a set $\{u_1, ...u_n, w_1, ..., w_m\}$. Remove vectors from the end untill remaining vectors are linearly independent. The remaining set $spans(V)$, is linearly independent by construction and contains U. □

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

Chapter 3

Linear Maps, Kernals and Range

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

$$T(u_1 + u_2) = T(u_1) + T(u_2)$$
$$T(\lambda u_1) = \lambda T(u_1)$$

The set of all linear mappings from $U \to V$ 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

- $T : \mathcal{C}[a, b] \to \Re,\ f \to \int_a^b f(x)dx$. 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.
- $D : \mathcal{C}^\infty[a, b] \to \mathcal{C}^\infty[a, b],\ f \to f'$. The derivative of the sum of functions is the same as taking the sum and then derivatves.

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

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

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:*

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

Proposition 5. *The range is always a subspace of V*

T is surjective iff $range(T) = V$

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

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

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

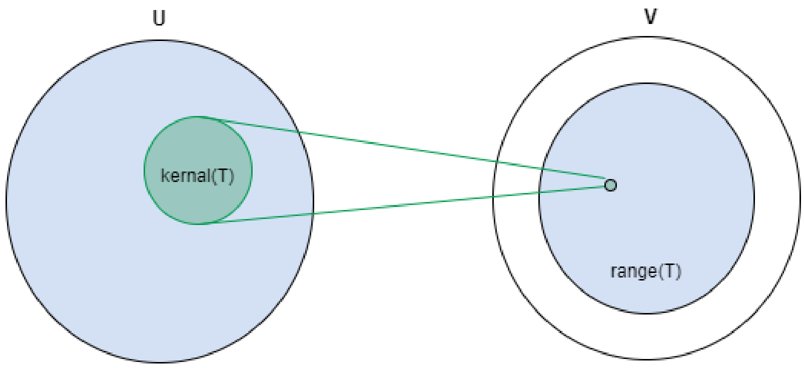

Figure 3.1: Kernal and range of a $T : U \to V$

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

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

Proof. Denote $T^{-1}(w_1) := z_1, ..., T^{-1}(w_m) := z_m$. If $\{u_1, ...u_n, z_1, ...z_m\}$ 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 $V \subset span\{u_1, ...u_n, z_1, ...z_m\}$
Let $v \in V$, and $Tv \in range(T)$. Since we know the basis of $range(T)$,

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

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

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

Step 2 : Prove that $u_1, ..., u_n, z_1, ..., z_m$ are linearly independent.
Assume that $\mu_1 u_1 + ...\mu_n u_n + \lambda_1 z_1 + ... + \lambda_m z_m = 0$
If these vectors are linearly independent then all the coefficients will simultaneously go to zero.

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

We can add $\mu_1 T(u_1) + ... + \mu_n T(u_n)$ to the expression since it goes to zero by definition of a kernal. Everything inside the operator is zero due to the assumption at the beginning of step 2. Since $\lambda_1 w_1 + ... + \lambda_m w_m = 0$ and $w_1, ..., w_m$ is a basis $\implies \lambda_1 = ... = \lambda_m = 0$
Now $\mu_1 u_1 + ... + \mu_n u_n = 0$ since we have already proved the other half of the assumption to be zero. Since $u_1, ..., u_n$ is the basis of the kernal $\implies \mu_1 = ...\mu_n = 0$. □

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 and Linear Maps

Each linear map on a finite dimensional map corresponds to a matrix and

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots n}}$$

m is the number of rows and n is the number of columns.

Now we want to see how matrices can be related to linear mappings. Consider $T \in \mathcal{L}(U, V)$ where V, W are finite dimensional vector spaces. Let $v_1, ...v_n$ be a basis for V and $w_1, ...w_m$ be a basis of W. The linear mapping is well defined as soon as we know how it acts on the basis vectors. Because then we can express it as linear combination and extend it any vecor in the space.

$$\begin{aligned} v &= \lambda_1 v_1 + ... + \lambda_n v_n, v \in V \\ T(v) &= T(\lambda_1 v_1 + ... + \lambda_n v_n) \\ &= \lambda_1 T(v_1) + ... + \lambda_n T(v_n) \end{aligned}$$

A matrix is a compact way of doing the above!

Each image vector vector $T(v_j)$ can be expressed in the basis $w_1, ..., w_m$. Here $a_{ij}...a_{mj} \in \Re$:

$$T(v_j) = a_{1j} w_1 + ... + a_{mj} w_m$$

We now stack these coefficients in a matrix.

$$\begin{pmatrix} a_{11} & \dots & a_{ij} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix}$$

This is the matrix of mapping T with respect to the basis $v_1, ..., v_n$ of V and $w_1, ...w_m$ of W. There are m rows, one for each basis vector of W. There are n columns, one for each basis vector of V.

Notation: Let $T : V \to W$ be a linear mapping and let $\mathcal{B}$ be a basis of V and $\mathcal{C}$ be a basis of W. We denote $H(T, \mathcal{B}, \mathcal{C})$ the matrix corresponding to T with basis $\mathcal{B}$ and $\mathcal{C}$

Convinient properties of matrices: Let V, W be vector spaces and consider the basis is fixed. Let $S, T \in \mathcal{L}(V, W)$.

- $H(S+T) = H(S) + H(T)$
- $H(\lambda S) = \lambda H(S)$
- For $v = \lambda_1 v_1 + ... + \lambda_n v_n$ we have that $T(v) = H(T)\begin{pmatrix}\lambda_1 \\ \vdots \\ \lambda_n\end{pmatrix}$, where $v_1, ...v_n$ are the basis of V.
- $H(S \circ T) = H(S) \cdot H(T)$

4.1 Invertiable maps and matrices

Definition 15. *Let $T \in \mathcal{L}(V, W)$ is called invertible if there exists another linear map $S \in \mathcal{L}(W, V)$:*

$$S \circ T = \mathcal{I}_v \text{ and } T \circ S = \mathcal{I}_w$$

The map S is called the inverse of T, denoted by T^{-1}.

Proposition 8. *Inverese maps are unique.*

Proposition 9. *A linear map is invertible if and only if it is injective and surjective.*

Proof. Invertible $\implies$ injective:
Suppose $T(u) = T(v)$. Then $u = T^{-1}(T(u)) = T^{-1}(T(v)) = v$. This means the mapping is injective.
Invertible $\implies$ surjective:
Let $w \in W$. Then $w = T(T^{-1}(w)) \implies w \in range(T)$. This means that mapping is surjective.
Injective and surjective $\implies$ invertible
Let $w \in W$, there exists unique $v \in V$ such that $T(v) = w$. Now define the mapping $S(w) = v$. Clearly we have $T \circ S = \mathcal{I}$. Let $v \in V$, then -

$$T((S \circ T)(v)) = (T \circ S)(Tv) = \mathcal{I} \circ Tv = Tv$$

$\implies (S \circ T)v = v \implies S \circ T = \mathcal{I}$ □

4.1.1 Inverse Matrix

Definition 16. *A square matrix $A \in F^{n \times m}$ is invertible if there exists a square matrix $B \in F^{n \times m}$ such that -*

$$A \cdot B = B \cdot A = \mathcal{I}$$

The matrix B is called the inverse matrix and is denoted by A^{-1}

Proposition 10. *The inverse matrix represents the inverse of the corresponding linear map, that is: $T : V \to V$*

$$M(T^{-1}) = (M(T))^{-1}$$

Essentially we are saying that the matrix of the inverse map is the same taking the matrix of the map and then doing the inverse.