0.1 Vector Space

A vector space is an algebraic structure that consists of a set of vectors and two operations defined on these vectors: vector addition and scalar multiplication. The elements of the vector space are called vectors, which can be added together and multiplied by numbers, called scalars. The operations of vector addition and scalar multiplication must satisfy a set of axioms, which ensure that the resulting vectors remain within the vector space. It will be helpful to describe a vector space by first explaining the concept of a set and a group.

A set is a collection of mathematical objects such as numbers, lines or potentially other sets. They were first formalised by George Cantor in 1895 [1] as being either infinite or finite and containing distinct elements.

Definition 0.1.1 Let $\mathcal{V}$ be a set and $\bigotimes$ denote a binary operation between elements of this set such that $\bigotimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$. $G := (\mathcal{G}, \bigotimes)$ is called a group [2] if the following conditions are met and specific elements are present:

1. (Closure) The binary operation between any two elements of the set will result in an element which is also part of the set: $\bigotimes : \forall x, y \in \mathcal{G} : x \bigotimes y \in \mathcal{G}$.

2. (Associativity) The way in which elements of the set are combined within a larger expression does not affect the result: $\forall x, y, z \in \mathcal{G} : (x \bigotimes y) \bigotimes = x \bigotimes (y \bigotimes z)$.

3. (Neutral Element) A neutral element in a set is an element that, when combined with any other element in the set using the group operation,

results in the same element: $\exists e \in \mathcal{G} \forall x \in \mathcal{G} : x \bigotimes e = x$ and $e \bigotimes x = x$.

4. (Inverse Element) The inverse of an element in a set is an element that when combined with the original element using the group operation, results in the neutral element. It allows for the reversal of the group operation: $\forall x \in \mathcal{G} \exists y \in \mathcal{G} : x \bigotimes y = e$ and $y \bigotimes x = e$, where e is the neutral element. We can denote the inverse of an element x as x^{-1}.

Groups are very important in many areas of mathematics. Their rigorous and formal definition means that, if something is found to be a group, its properties can be better understood in the specific context in which it is relevant. An example of a group is the general linear group. This is the set of regular invertible matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with respect to matrix multiplication. Given a group G, if the order in which the group operation is performed does not matter i.e. $\forall x, y \in \mathcal{G} : x \bigotimes y = y \bigotimes x$, then $G = (\mathcal{G}, \bigotimes)$ is known as an *Abelian group* (commutative). An example of this would be $(\mathbb{Z}, +)$, the set of all integers under the addition operation.

A vector space [3] is a special type of group with some additional conditions.

Definition 0.1.2 (Vector Space) A real-valued *vector space* $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations:

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V} \quad \text{(Inner Operation)} \tag{1}$$

$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V} \quad \text{(Outer Operation)} \tag{2}$$

where:

1. $(\mathcal{V}, +)$ is an Abelian group
2. (Distributivity) The outer operation can be "Distributed" across elements either before or after the inner operation has occurred:
   (a) $\forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$
   (b) $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : (\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \mathbf{x}) = (\lambda\psi) \cdot \mathbf{x}$
4. Neutral element with respect to the outer operation: $\forall \mathbf{x} \in \mathcal{V} : 1 \cdot \mathbf{x} = \mathbf{x}$

We commonly think of vectors as being mathematical objects with both direction and magnitude. However, these are only geometric vectors. More generally, any set of objects which follows the definition of a vector space is known as a vector. For example, polynomials are also vectors. Two can be added together, resulting in another polynomial and they can be multiplied by a scalar $\lambda \in \mathbb{R}$ which again results in another polynomial.

0.2 Linear Independence

Linear independence [4] is a property of a set of vectors which describes whether there is any redundancy with respect to the linear combinations of these vectors.

Definition 0.2.1 (Linear Combination) Consider a vector space V and a finite number of vectors $\mathbf{x}_i, \ldots, \mathbf{x}_k \in V$. Then, every $\mathbf{v} \in V$ of the form

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i \in V \tag{3}$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a *linear combination* of the vectors $\mathbf{x}_1, \cdots, \mathbf{x}_k$.

Consider two vectors in $\mathbb{R}^2$, $\mathbf{e}_1 = (1,0)^\top$ and $\mathbf{e}_2 = (0,1)^\top$. It is common to see these written as $\mathbf{i}$ and $\mathbf{j}$ respectively. We can represent any vector in $\mathbb{R}^2$ as a linear combination of these two vectors.

Definition 0.2.2 (Linear (In)dependence) Let us consider a vector space V with $k \in \mathbb{R}$ and $\mathbf{x}_1, \cdots, \mathbf{x}_k \in V$. If there is a non-trivial combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\mathbf{x}_1, \cdots, \mathbf{x}_k$ are *linearly dependent*. If only the trivial solution exists, i.e., $\lambda_1 = \cdots = \lambda_k = 0$ the vectors $\mathbf{x}_1, \cdots, \mathbf{x}_k$ are *linearly independent*.

Lets consider the following set $\mathcal{V}$ of vectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ where (respectively):

$$\mathcal{V} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ -4 \end{pmatrix}, \begin{pmatrix} -4 \\ -1 \end{pmatrix} \right\} \tag{4}$$

And let $\mathbf{y} \in \mathbb{R}^2$ be any linear combination that can be made from these vectors. Writing this out explicitly with $\alpha, \beta, \omega \in \mathbb{R}$:

$$\begin{aligned}
\mathbf{y} &= \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} -2 \\ -4 \end{pmatrix} + \omega \begin{pmatrix} -4 \\ -1 \end{pmatrix} && (5) \\
&= \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 2\beta \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \omega \begin{pmatrix} -4 \\ -1 \end{pmatrix} && (6) \\
&= (\alpha - 2\beta) \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \omega \begin{pmatrix} -4 \\ -1 \end{pmatrix} && (7) \\
& && (8)
\end{aligned}$$

What we have shown is that if a vector $\mathbf{y}$ can be represented as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ then we can represent $\mathbf{y}$ as a linear combination of $\mathbf{v}_2$ and $\mathbf{v}_3$. This is because $\mathbf{v}_1$ is a scaled version of $\mathbf{v}_2$ and vice versa. More formally:

$$\begin{pmatrix} -2 \\ -4 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \mathbf{0} \tag{9}$$

which following from definition 0.2.2 demonstrates that we have a linearly dependent set of vectors.

0.3 Generating Set and Span

A generating set and the span [5] of a set of vectors are concerned with the vector space produced by the linear combination of all the vectors in the set.

Definition 0.3.1 (Generating Set and Span) Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq V$. If every vector $\mathbf{v} \in V$ can be expressed as a linear combination of vectors of $\mathbf{x}_1, \ldots, \mathbf{x}_k$, $\mathcal{A}$ is called the generating set of V. The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans the vector space V, we write $V = span[\mathcal{A}]$.

Consider the vector space $V = (\mathbb{R}^2, +, \cdot)$ under matrix addition and scalar multiplication. Previously we have said that every vector $v \in V$ can be represented as a linear combination of $\mathbf{e}_1 = (1, 0)^\top$ and $\mathbf{e}_2 = (0, 1)^\top$. By definition 0.3.1, the set of vectors $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ is a generating set of $\mathbf{V}$.

On the other hand, the set of two-dimensional vectors $\mathbb{R}^2$ contains all linear combinations of the set B meaning that the $span[\mathcal{B}] = \mathbb{R}^2$. In this sense, a generating set and the span of a set are two sides of the same coin.

0.4 Basis

A basis is concerned with the minimum number of vectors that would be needed to span a particular vector space.

Definition 0.4.1 (Basis) Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and $\mathcal{A} \subseteq \mathcal{V}$. A generating set $\mathcal{A}$ of V is called *minimal* if there exists no smaller set $\tilde{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{V}$ that spans V. Every linearly independent generating set of V is minimal and is called a *basis* or V.

Consider the vector space $V = (\mathbb{R}^3, +, \cdot)$. A potential generating set for V is the set $\mathcal{A}$ where:

$$\mathcal{A} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \tag{10}$$

This is because every vector in $\mathbb{R}^3$ can be represented as linear combinations of the vectors in $\mathcal{A}$. Since this set of vectors is linearly independent, it forms a *minimal* generating set. We've described previously that the *span* of a set of vectors can be thought of as the collection of all potential linear combinations of the set. $\mathcal{A}$ forms a *minimal* generating set because no vectors may be removed without eliminating some linear combinations / reducing the overall span of the set. This is what we mean by a basis for the vector space V. As well as this, the number of dimensions of the vector space spanned by the basis is equal to the number of vectors in the basis for the vector space.

0.5 Row Rank and Column Rank

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, the row rank (denoted as $rk_{row}(\boldsymbol{A})$) and the column rank (denoted as $rk_{col}(\boldsymbol{A})$) describe the number of linearly independent row vectors and column vectors respectively.

Proposition 0.5.1 For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$:

$$rk_{col}(\boldsymbol{A}) = dim(span[\boldsymbol{A}_{.1}, \ldots, \boldsymbol{A}_{.m}]) \tag{11}$$

Example: Consider a matrix $A = [\boldsymbol{A}_{.1} | \boldsymbol{A}_{.2}]$ where $\boldsymbol{A}_{.1}$ and $\boldsymbol{A}_{.2}$ are linearly dependent. The *span* of $\mathcal{B} = \{\boldsymbol{A}_{.1}, \boldsymbol{A}_{.2}\}$ is all the linear combinations that can be created from the two vectors. Let $s \in span[B]$:

$$s = \lambda_1 \boldsymbol{A}_{.1} + \lambda_2 \boldsymbol{A}_{.2}, \quad \forall \lambda_1, \lambda_2 \in \mathbb{R} \tag{12}$$

Because the column vectors are linearly dependent, by definition 0.2.2, $\exists \lambda_3, \lambda_4$ such that:

$$\lambda_3 \boldsymbol{A}_{.1} + \lambda_4 \boldsymbol{A}_{.2} = \boldsymbol{0} \tag{13}$$

$$\boldsymbol{A}_{.1} = \frac{\lambda_4}{\lambda_3} \boldsymbol{A}_{.2} \tag{14}$$

Meaning that every linear combination can be written as:

$$
\begin{aligned}
s &= \lambda_1 \boldsymbol{A}_{.1} + \lambda_2 \boldsymbol{A}_{.2} && (15)\\
&= (-\lambda_1 \frac{\lambda_4}{\lambda_3} + \lambda_4)\boldsymbol{A}_{.2} && (16)
\end{aligned}
$$

This means that $\mathcal{C} = \{\boldsymbol{A}_{.2}\}$ forms a basis. Because the number of elements in this set is 1, the dimension of $span[C]$ is 1. Because both column vectors of $\boldsymbol{A}$ are linearly independent, the $rk_{col}(\boldsymbol{A}) = 1$ and so proposition 0.5.1 is shown to be true for the case of a matrix with two column vectors. Similar methods can show that this is true for larger matrices. Furthermore, the same reasoning can be used to conclude that $rk_{row}(\boldsymbol{A}) = dim(span[\boldsymbol{A}_{1.}, \ldots, \boldsymbol{A}_{n.}])$.

Proposition 0.5.2 Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, the following statement is always true:

$$
rk_{row}(\boldsymbol{A}) = rk_{col}(\boldsymbol{A}) \tag{17}
$$

where $rk_{row}(\boldsymbol{A})$ and $rk_{col}(\boldsymbol{A})$ denote the row rank and the column rank respectively.

Proof: Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a matrix and let $u = rk_{col}(\boldsymbol{A})$. There exists a basis $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_u\}$ of $m \times 1$ column vectors that spans the same space spanned by the columns of $\mathbf{A}$. Let $\boldsymbol{B} = [\mathbf{c}_1|\mathbf{c}_2|\ldots|\mathbf{c}_{u-1}\mathbf{c}_u]$ be a matrix where $[\mathbf{v}_1|\mathbf{v}_2]$ denotes the concatenation between two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. By definition 0.2.2, each column of $\boldsymbol{A}$ can be expressed as a linear combination of vectors in $\boldsymbol{B}$. If we collect the coefficients of these linear combinations into a matrix $\boldsymbol{D}$, we can represent $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{D}$ where $\boldsymbol{D} \in \mathbb{R}^{m \times n}$. Equally, we can see that the rows of $\boldsymbol{BD}$ can be expressed as a linear combination of $\boldsymbol{D}$ with coefficients taken from B. $\therefore$, the span of the rows of $\boldsymbol{A}$ is no greater than the span of the rows of $\boldsymbol{D}$ because linear combinations of the rows of $\boldsymbol{A}$ can be written as linear combinations of the rows in $\boldsymbol{D}$. $\boldsymbol{D}$ has u rows. If they are linearly independent, the dimension of their span is u. Otherwise, it has dimension $< u$. $\therefore$ the row rank is $\leq u$ and so $rk_{row}(\boldsymbol{A}) \leq rk_{col}(\boldsymbol{A})$. Similar reasoning will show that the $rk_{col}(\boldsymbol{A}) \leq rk_{row}(\boldsymbol{A})$ and so $rk_{row}(\boldsymbol{A}) = rk_{col}(\boldsymbol{A})$.

Bringing it all together, we get the rank of a matrix:

Definition 0.5.1 (Rank) The number of linearly independent columns of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows and is called the *rank* of $\mathbf{A}$ and is denoted by $rk(\mathbf{A})$.

Bibliography

[1] Georg Cantor. “Beiträge zur Begründung der transfiniten Mengenlehre”. In: *Mathematische Annalen* 46.4 (1895), pp. 481–512.

[2] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020, p. 36.

[3] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020, p. 37.

[4] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020, p. 40.

[5] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020, p. 40.