

Linear algebra is the study of vectors. At the most general level, vectors are **ordered finite lists of numbers**.

Vectors are the most fundamental mathematical object in machine learning

Representation: \vec{v}

Types of vectors

(1) geometric vectors; (2) polynomials; (3) elements of \mathbb{R}^n

Geometric vectors

Geometric vectors are oriented segments.

Polynomials

A polynomial is an expression like $f(x) = x^2 + y + 1$. This is, an expression adding multiple "terms" (nomials). Polynomials are vectors because they meet the definition of a vector: they can be added together to get another polynomial, and they can be multiplied together to get another polynomial.

Function addition is valid and multiplying by a scalar is valid.

Elements of \mathbb{R}^n space

Elements of \mathbb{R}^n are sets of real numbers. This type of representation is arguably the most important for applied machine learning. It is how data is commonly represented in computers to build machine learning models. For instance in \mathbb{R}^3 takes the shape of:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3$$

Indicating that it contains three dimensions.

Addition is valid and multiplying by a scalar is valid

Zero vector, unit vector, and sparse vector

Zero vector:

$$\vec{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Unit vectors:

$$\vec{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \vec{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \vec{x}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Sparse vectors, are vectors with most of their elements equal to zero. We denote the number of nonzero elements of a vector \vec{x} as $nnz(x)$. The sparse possible vector is the zero vector.

Vector dimensions and coordinate system

Vectors dimensions map into **coordinate systems** or **perpendicular axes**. Coordinate systems have an origin at $(0, 0, 0)$, hence when we define a vector:

$$\vec{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \in \mathbb{R}^3$$

We are saying: starting from the origin, move 3 units in the 1st perpendicular axis, 2 units in the 2nd perpendicular axis, and 1 unit in the 3rd perpendicular axis.

Basic vector operations

Vector-vector addition

Vector-vector addition is an element-wise operation, only defined for vectors of the same size (i.e., number of elements)

$$\vec{x} + \vec{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

Fundamental properties:

1. Commutativity: $x + y = y + x$
2. Associativity: $(x + y) + z = x + (y + z)$
3. Adding the zero vector has no effect
4. Subtracting a vector from itself returns the zero vector

Vector-scalar multiplication

It is defined as:

$$\alpha \vec{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

Vecto-scalar multiplication satisfies a series of important properties:

1. Associativity: $(\alpha\beta)\vec{x} = \alpha(\beta\vec{x})$
2. Left-distributive property: $(\alpha + \beta)\vec{x} = \alpha\vec{x} + \beta\vec{x}$
3. Right-distributive property: $\vec{x}(\alpha + \beta) = \vec{x}\alpha + \vec{x}\beta$
4. Right-distributive property for vector addition: $\alpha(\vec{x} + \vec{y}) = \alpha\vec{x} + \alpha\vec{y}$

Linear combination of vectors

Only two operations with vectors in linear algebra are legal: **addition** and **mutiplication by numbers**.

When we combine these, we get a **linear combination**.

$$\alpha \vec{x} + \beta \vec{y} = \alpha \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \beta \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{pmatrix}$$

Vector-vector multiplication: dot product

Vector-vector multiplication is commonly known as a **dot product** or **inner product**. The dot product of x and y is defined as:

$$\vec{x} \cdot \vec{y} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 \times y_1 + x_2 \times y_2$$

The **T** denotes the transpose of the vector.

Vector space, span, and subspace

Vector space

A vector space, also known as linear space, is a collection of objects that follow the rules defined for vectors in \mathbb{R}^n . A vector space is the set of proper vectors and all possible linear combinations of the vector set. In addition, vector addition and multiplication must follow these eight rules:

1. Commutativity: $x + y = y + x$
2. Associativity: $x + (y + z) = (x + y) + z$
3. Unique zero vector such that: $x + 0 = x \forall x$
4. $\forall x$ there is a unique vector x such that $x + -x = 0$
5. Identity element of scalar multiplication: $1 \times x = x$
6. Distributivity of scalar multiplication w.r.t vector addition: $x(y + z) = xz + xy$
7. $x(yz) = (xy)z$
8. $(y + z)x = yx + zx$

Vector span

If our vectors \vec{x} and \vec{y} point into different directions in the 2-dimensions in the 2-dimensional space, we get that the $\text{span}(x, y)$ is equal to the **entire 2-dimensional plane**

If you combine vectors which point in the same direction, you just can span a line. **Multicollinearity** is closely related to this issue: when two variables are "colinear", they are pointing in the same direction, hence they provide redundant information, so can drop one without information loss.

With three vectors pointing into different directions, we can span the entire 3-dimensional space or a **hyper-plane**.

Vector subspaces

A vector subspace (or linear subspace) is a vector space that lies within a larger vector space. These are also known as linear subspaces. Consider a subspace S . For a vector to be a valid subspace it has to meet three conditions:

1. Contains the zero vector, $0 \in S$
2. Closure under multiplication, $\forall \alpha \in \mathbb{R} \rightarrow \alpha \times s_i \in S$
3. Closure under addition, $\forall s_i \in S \rightarrow s_1 + s_2 \in S$

Intuitively, you can think of closure as being unable to "jump out" from space into another. A pair of vectors lying flat in the 2-dimensional space, can't by either addition or multiplication, "jump out" into the 3-dimensional space.

Linear dependence and independence

A set of vectors is linearly dependent if at least one vector can be obtained as a linear combination of other vectors in the set. There is a more rigorous definition of linear dependence. Consider a set of vectors x_1, \dots, x_k and scalars $\beta \in \mathbb{R}$. If there is a way to get $0 = \sum_{i=1}^k \beta_i x_i$ with at least one $\beta \neq 0$, we have a linear combination of the vectors in the set, with weights that are **not all zero**, we have a linearly dependent set.

A set of vectors is **linearly independent** if none vector can be obtained as a linear combination of other vectors in the set. Again, consider a set of vectors x_1, \dots, x_k and scalars $\beta \in \mathbb{R}$. If the only way to get $0 = \sum_{i=1}^k \beta_i x_i$ requires all $\beta_i = 0, i = 0, \dots, k$, then we have linearly independent vectors.

Vector null space

Intuitively, the null space of a set of vectors are all linear combinations that "map" into the zero vector.

Vector norms

Norms "map" vectors to non-negative values. In this sense are functions that assign length $\|\vec{x}\| \in \mathbb{R}^n$ to a vector \vec{x} . To be valid, a norm has to satisfy these properties:

1. Absolutely homogeneous: $\forall \alpha \in \mathbb{R}, \|\alpha \vec{x}\| = |\alpha| \|\vec{x}\|$.
2. Triangle inequality: $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$

3. Positive definite: $\|\vec{x}\| \geq 0$ and $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$

Euclidean norm

It is defined as:

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^n \vec{x}_i^2} = \sqrt{\vec{x}^T \vec{x}}$$

Hence, in two dimensions the L_2 norm is:

$$\|\vec{x}\|_2 \in \mathbb{R}^2 = \sqrt{x_1^2 + x_2^2}$$

Which is equivalent to the formula for the hypotenuse a triangle with sides x_1^2 and x_2^2 .

The same pattern follows for higher dimensions of \mathbb{R}^n .

Manhattan norm

It is defined as:

$$\|\vec{x}\|_1 := \sum_{i=1}^n |\vec{x}_i|$$

Where $|\vec{x}_i|$ is the absolute value. The L_1 norm is preferred when discriminating between elements that are exactly zero and elements that are small but not zero.

Max norm

The max norm or infinity norm is simply the absolute value of the largest element in the vector. It is defined as:

$$\|\vec{x}\|_\infty := \max_i |\vec{x}_i|$$

Vector inner product, length, and distance.

Inner products are a more general concept than **dot products**, with a series of additional properties. In other words, every dot product is an inner product, but not every inner product is a dot product.

The notation for the inner product is usually a pair of angle brackets as $\langle \dots \rangle$. For instance, the scalar inner product is defined as:

$$\langle \vec{x}, \vec{y} \rangle := \vec{x} \cdot \vec{y}$$

In \mathbb{R}^n the inner product is a dot product defined as:

$$\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle := \vec{x} \cdot \vec{y} = \sum_{i=1}^n x_i y_i$$

Length is a concept from geometry. We say that geometric vectors have length and that vectors in \mathbb{R}^n have norm. In practice, many machine learning textbooks use these concepts interchangeably. For instance, we can compute the length of a directed segment (i.e. geometrical vector) \vec{x} by taking the square root of the inner product with itself as:

$$\|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{y}} = \sqrt{x^2 + y^2}$$

Distance is a relational concept. It refers to the length (or norm) of the difference between two vectors. Hence, we use norms and lengths to measure the distance between vectors. Consider the vectors \vec{x} and \vec{y} , we define the distance $d(\vec{x}, \vec{y})$ as:

$$d(\vec{x}, \vec{y}) := \|\vec{x} - \vec{y}\| = \sqrt{\langle \vec{x} - \vec{y}, \vec{x} - \vec{y} \rangle}$$

When the inner product $\langle x - y, x - y \rangle$ is the dot product, the distance equals to the Euclidean distance.

In machine learning, unless made explicit, we can safely assume that an inner product refers to the dot product.

As with the inner product, usually, we can safely assume that distance stands for the Euclidean distance or L_2 norm unless otherwise noted.

Vector angles and orthogonality

In the same manner, inner products are used to define **angles** and **orthogonality**.

In machine learning, the **angle** between a pair of vectors is used as a measure of vector similarity. Consider a pair of non-zero vectors \vec{x} and $\vec{y} \in \mathbb{R}^n$. The **Cauchy-Schwarz inequality** states that:

$$\| \langle x, y \rangle \| \leq \|x\| \|y\|$$

The only case where both sides of the expression are equal is when the vectors are colinear.

The definition of the angle between vectors can be thought as a generalization of the law of cosines in trigonometry, which defines for a triangle with sides a , b , and c , and an angle θ are related as:

$$c^2 = a^2 + b^2 - 2ab \cos \theta$$

We can replace this expression with vectors lengths as:

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2(\|x\| \|y\|) \cos \theta$$

With a bit of algebraic manipulation, we can clear the previous equation to:

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$

And there we have a definition for (\cos) angle θ . Further, from the Cauchy-Schwarz inequality we know that $\cos \theta$ must be:

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \|y\|} \leq 1$$

Orthogonality is often used interchangeably with "independence" although they are mathematically different concepts. Orthogonality can be seen as a generalization of perpendicularity to vectors in any number of dimensions.

We say that a pair of vectors x and y are orthogonal if their inner product is zero, $\langle x, y \rangle = 0$. The notation for a pair of orthogonal vectors is $x \perp y$. In the 2-dimensional plane, this equals to a pair of vectors forming a 90° angle.

Systems of linear equations

A system of linear equations involve multiple equations that have to be solved simultaneously.