2 Mathematical Backbone: Vector Essentials for Machine Learning

2.1 Vectors and Scalars: Understanding the Basics

2.1.1 What is a Vector? Magnitude Meets Direction

A **vector** is a quantity that has both **magnitude** (size) and **direction**. Geometrically, a vector can be visualized as a **directed line segment**, with the length representing its magnitude and the arrow indicating its direction. In contrast, a **scalar** is a single numerical value that only has magnitude and no direction.

As illustrated in Figure 1, the length of a vector represents its magnitude, while the arrow shows its direction. This distinction between vectors and scalars is fundamental in linear algebra and machine learning, as many algorithms rely on operations with vectors rather than just scalar numbers.

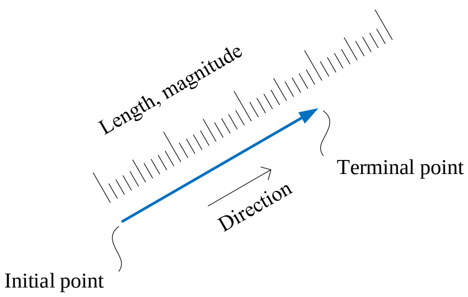

Figure 1. A vector has both magnitude and direction

2.1.2 Row and Column Vectors: Two Faces of the Same Coin

A **row vector** arranges its components in a single row. For example, a row vector with n components can be written as:

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \tag{1}$$

Here, the i-th component of the vector is a_i, with the index iii ranging from 1 to n. The **dimension** of a vector refers to the number of components it contains. Therefore, the vector above is an n-dimensional vector.

A row vector can also be viewed as a special case of a **matrix**. In general, a matrix is a rectangular array of numbers with multiple rows and columns, where each element is uniquely identified by a row index and a column index. A row vector can be considered a matrix with one row and multiple columns.

A **column vector** arranges its components in a single column. An n-dimensional column vector $\boldsymbol{a}$ can be written as:

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{2}$$

Column vectors can be seen as matrices with multiple rows and a single column. Unless otherwise specified, this book assumes all vectors are column vectors. In this book, column vectors are denoted using lowercase, bold, and italic letters, such as $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{x}$, and $\boldsymbol{y}$.

2.1.3 Transpose: Flipping Between Rows and Columns

The **transpose** operation swaps the rows and columns of a matrix. As shown in Figure 2, the transpose of a row vector produces a column vector, and vice versa.

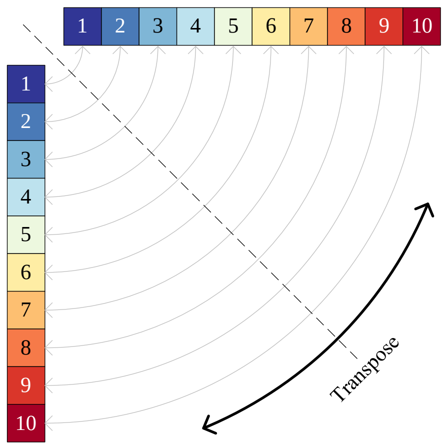

Figure 2. Transpose transforms row vectors into column vectors and column vectors into row vectors

In this book, the transpose of a matrix is denoted by a superscript $^{\mathrm{T}}$, for example

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{3}$$

Transposing a vector does not change the number of components or their relative order. It simply changes whether the vector is arranged in a row or a column.

2.2 One-Dimensional Thinking: The Real Number Line

2.2.1 Scalars on a Line: Origin, Direction, and Unit Length

As illustrated in Figure 3, the real number line, also called the real axis, is an infinitely extending straight line that provides a fundamental way to represent the position of real numbers, also known as scalars.

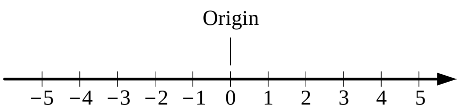

Figure 3. The real number line, showing the origin, positive and negative directions, and unit length.

2.2.2 From Numbers to Space: The First Step Toward Geometry

The three essential components of the number line are the origin, the positive direction, and the unit length.

The origin, denoted as 0, serves as the central reference point. Numbers to the right of the origin are positive, while numbers to the left are negative. The positive direction is indicated by an arrow, and the unit length defines the distance corresponding to a step of 1.

The number line extends infinitely in both directions, approaching negative infinity ($-\infty$) on the left and positive infinity ($+\infty$) on the right.

Mathematically, the real number line is also referred to as a one-dimensional Euclidean space, denoted by $\mathbb{R}$, which is the set of all real numbers. This simple yet powerful concept forms the basis for more advanced ideas in mathematics and machine learning, where scalar quantities are often represented as points on this line.

2.3 Mapping Space: Cartesian Coordinate Systems

2.3.1 Two-Dimensional Planes: Plotting Vectors on a Grid

The Cartesian coordinate system, also known as the rectangular coordinate system, is a framework for uniquely specifying the location of points in space using numerical values.

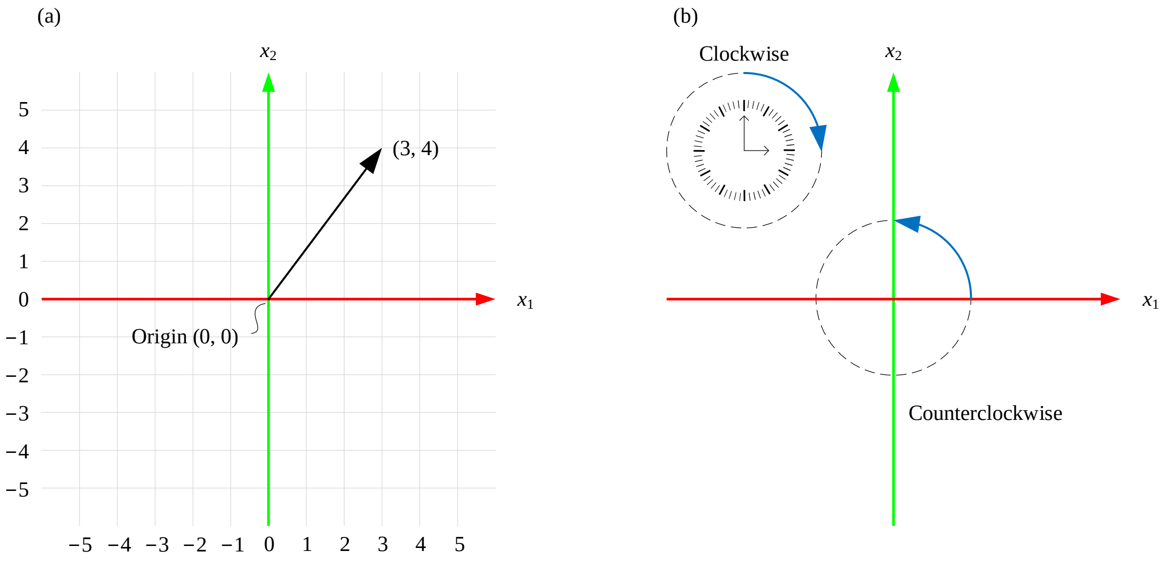

Figure 4. Two-dimensional Cartesian coordinate system with perpendicular axes, origin, and positive directions.

As shown in Figure 4 (a), when two number lines intersect perpendicularly at the origin, they form a two-dimensional Cartesian plane, also called a two-dimensional Euclidean space, denoted by $\mathbb{R}^2$.

A column vector $\boldsymbol{x} = [x_1, x_2]$ in this plane can be interpreted as a directed line segment starting at the origin (0, 0) and ending at the point (x_1, x_2).

The zero vector, ***0***, naturally lies at the origin. Unless stated otherwise, vectors are assumed to originate from the origin, i.e., ***0***.

In this coordinate system, angles are measured in a counterclockwise direction. For example, as illustrated in Figure 4 (b), rotating from the positive x_1-axis to the positive x_2-axis through 90 degrees defines the positive angular direction. In general, when measuring angles less than 180 degrees, the positive x_1-axis leads the positive x_2-axis in a counterclockwise sense.

2.3.2 From Plane to Space: Introducing Three Dimensions

Extending this idea to three dimensions, we can introduce a third axis x_3 perpendicular to both x_1 and x_2, with its origin aligned at (0, 0, 0). This forms a three-dimensional Cartesian coordinate system, or $\mathbb{R}^3$.

In this space, a point is represented by an ordered triple (x_1, x_2, x_3), which in this book is the preferred notation.

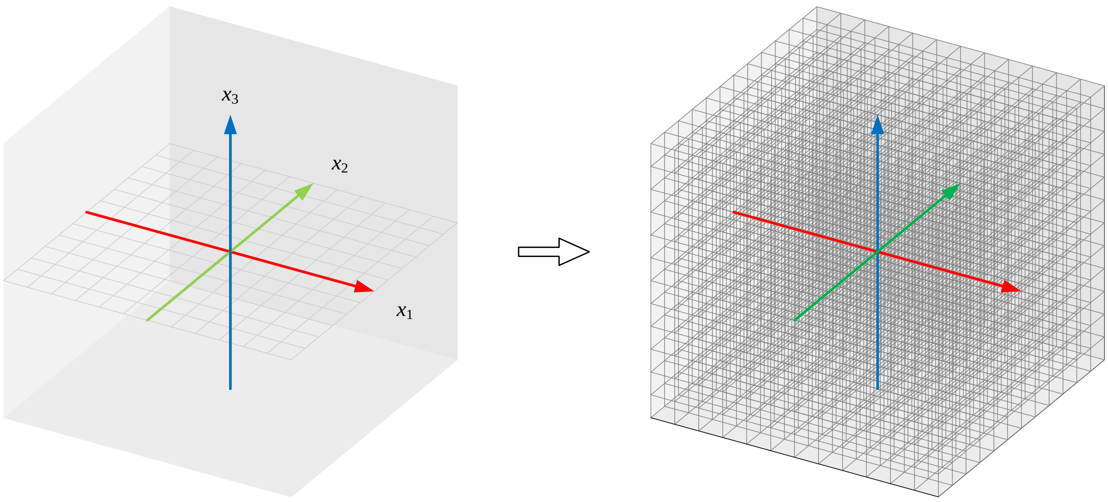

Figure 5. Three-dimensional Cartesian coordinate system represented as a grid, showing the origin, axes, and a sample vector.

2.3.3 Visualizing Vectors: Magnitude and Direction in Space

A three-dimensional vector $\boldsymbol{x} = [x_1, x_2, x_3]^T$ is a directed line segment originating from the origin $\boldsymbol{0} = [0, 0, 0]^T$, and terminating at ($x_1$, x_2, x_3), carrying both direction and magnitude information. Such vectors are commonly used to represent positions, forces, or other physical quantities in three-dimensional space, as illustrated in Figure 5.

The length (or magnitude, also called the L2 norm or Euclidean norm) of a two-dimensional vector $\boldsymbol{x} = [x_1, x_2]^T$ is calculated as

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2} \tag{4}$$

Similarly, the length of a three-dimensional vector $\boldsymbol{x} = [x_1, x_2, x_3]^T$ is

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2} \tag{5}$$

These lengths capture the vector's size, while the direction from the origin encodes orientation, making vectors essential for representing both geometric and physical properties in two- and three-dimensional spaces.

2.4 Playing with Vectors: Basic Operations

2.4.1 Adding and Subtracting Vectors: Parallelogram and Triangle Laws

Vector operations are fundamental tools in both mathematics and machine learning, as they allow us to combine, scale, and manipulate quantities that have both magnitude and direction.

For two vectors of the same dimension, **vector addition** is performed by adding their corresponding components. For example, given two n-dimensional column vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, their sum is a new n-dimensional vector obtained by adding each component of $\boldsymbol{a}$ to the corresponding component of $\boldsymbol{b}$:

$$\boldsymbol{a}+\boldsymbol{b}=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}+\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}=\begin{bmatrix} a_1+b_1 \\ a_2+b_2 \\ \vdots \\ a_n+b_n \end{bmatrix} \tag{6}$$

Geometrically, vector addition can be visualized using the **parallelogram law**, as shown in Figure 6. Each sub-figure illustrates different relative orientations of the two non-zero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

Figure 6(a) shows the case of unit vectors forming a unit square, Figure 6 (b) a square, Figure 6 (c) a rectangle, Figure 6 (d) a parallelogram with sides parallel to the horizontal axis, and Figure 6 (e) a parallelogram with sides parallel to the vertical axis.

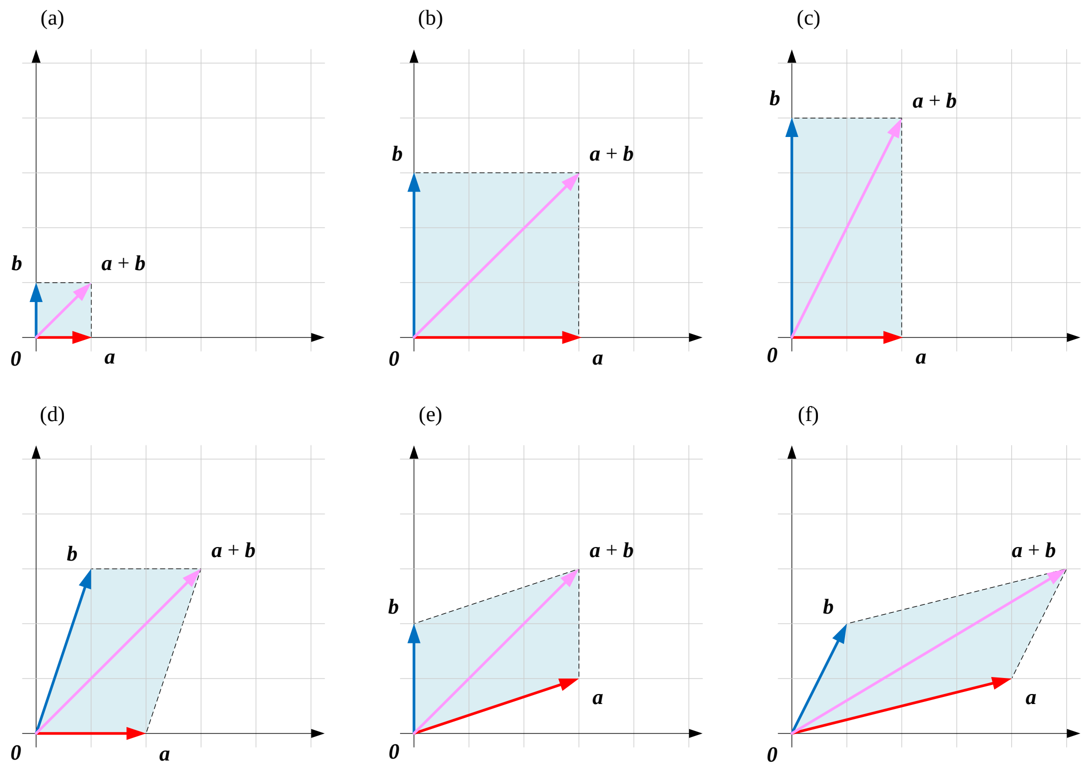

Figure 6. Examples of vector addition using the parallelogram law with different relative orientations. Figures can be generated by Ch02_01_Parallelogram_Law.ipynb.

Another geometric approach is the triangle law of vector addition, which constructs a triangle by placing the tail of the second vector at the head of the first vector. The directed line from the start of the first vector to the end of the second vector represents the sum $\boldsymbol{a} + \boldsymbol{b}$. The same sums computed using the parallelogram law in Figure 6 can also be obtained using the triangle law, as shown in Figure 7.

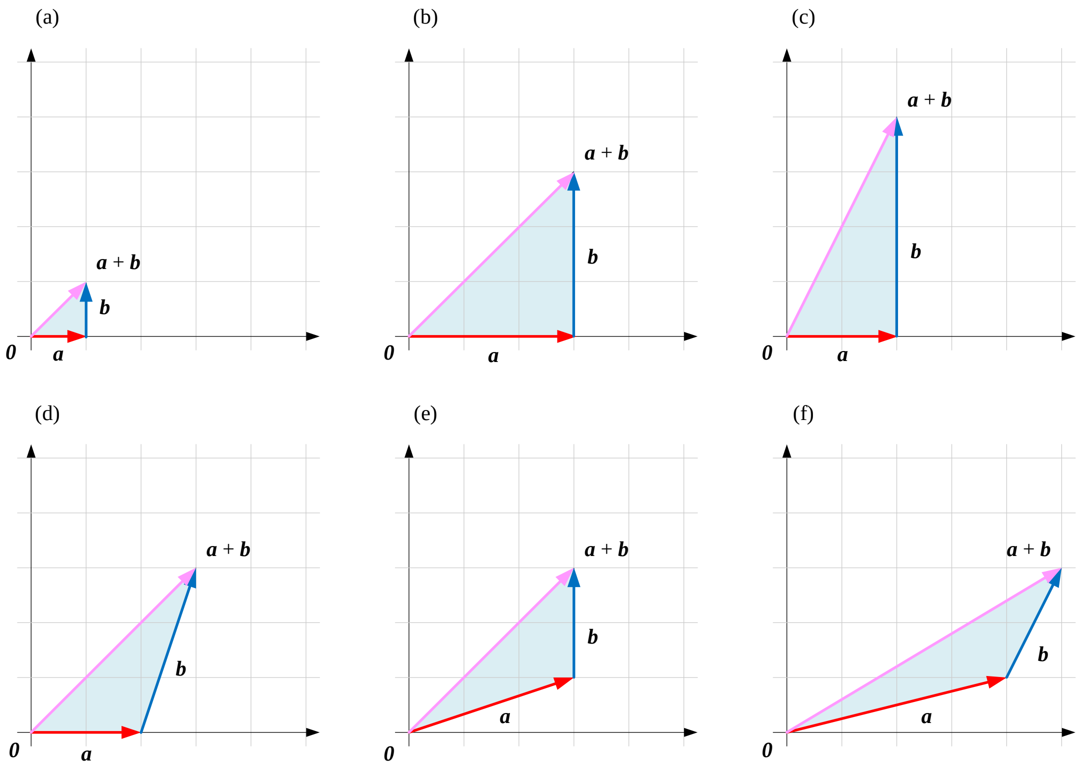

Figure 7. Examples of vector addition using the triangle law for the same vectors. Figures can be generated by Ch02_02_Triangle_Law.ipynb.

Vector subtraction is defined similarly but involves subtracting corresponding components of two vectors. For two n-dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, their difference is

$$\boldsymbol{a} - \boldsymbol{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \vdots \\ a_n - b_n \end{bmatrix} \tag{7}$$

2.4.2 Scaling Vectors: The Power of Scalar Multiplication

Finally, **scalar multiplication** scales a vector by a single number, or scalar. Each component of the vector is multiplied by the same scalar. For a non-zero n-dimensional vector $\boldsymbol{a}$ and a scalar k, scalar multiplication is defined as

$$k\boldsymbol{a} = k\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} ka_1 \\ ka_2 \\ \vdots \\ ka_n \end{bmatrix} \tag{8}$$

This operation changes the vector's magnitude while preserving its direction if $k > 0$, or reversing it if $k < 0$.

Together, these operations form the foundation for manipulating vectors in geometry, physics, and machine learning, where vectors often represent positions, features, gradients, or forces.

2.5 Measuring Relationships: Inner Products and Projections

2.5.1 Dot Product: Linking Vectors Through Angles

The **inner product** (or dot product) of two vectors is a fundamental operation in both mathematics and machine learning. For two column vectors of the same dimension, $\boldsymbol{a}$ and $\boldsymbol{b}$, the inner product is computed by multiplying their corresponding components and summing the results, producing a scalar:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \langle \boldsymbol{a}, \boldsymbol{b} \rangle = a_1 b_1 + a_2 b_2 + \cdots a_n b_n = \sum_{i=1}^{n} a_i b_i \tag{9}$$

Geometrically, the inner product can also be expressed as the product of the vectors' lengths and the cosine of the angle θ between them:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \|\boldsymbol{a}\| \|\boldsymbol{b}\| \cos\theta \tag{10}$$

If both vectors are non-zero, this relation can be rearranged to find the cosine of the angle between them:

$$\cos\theta = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \|\boldsymbol{b}\|} \tag{11}$$

As illustrated in Figure 8 (a), given non-zero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the angle between them is θ.

2.5.2 Scalar Projection: Shadows of Vectors

The **scalar projection** of a vector onto another vector represents the "length" of the shadow of one vector along the direction of the other, without carrying directional information. Mathematically, the scalar projection of $\boldsymbol{a}$ onto $\boldsymbol{b}$ is

$$\|\boldsymbol{a}\| \cos\theta = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{b}\|} = \boldsymbol{a} \cdot \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|} = \boldsymbol{a} \cdot \hat{\boldsymbol{b}} \tag{12}$$

where $\hat{\boldsymbol{b}}$ is the unit vector in the direction of $\boldsymbol{b}$.

In Figure 8 (b), this can be interpreted as the coordinate of $\boldsymbol{a}$ along $\boldsymbol{b}$, and the dashed line shows the orthogonal projection perpendicular to $\boldsymbol{b}$.

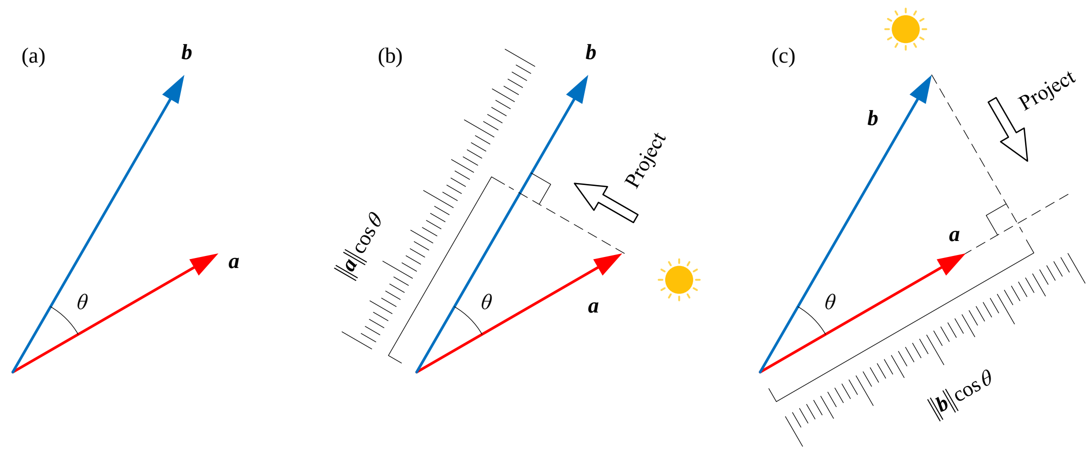

Figure 8. Scalar projection of a vector onto another vector: (a) angle between vectors; (b) projection of ***a*** onto ***b***; (c) projection of ***b*** onto ***a***.

Similarly, the scalar projection of ***b*** onto ***a*** is

$$\|\boldsymbol{b}\|\cos\theta = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\cdot\boldsymbol{b} = \hat{\boldsymbol{a}}\cdot\boldsymbol{b} \tag{13}$$

as shown in Figure 8 (c), representing the component of ***b*** along ***a***.

Figure 9 illustrates how the inner product varies with the angle θ while keeping the vector lengths fixed.

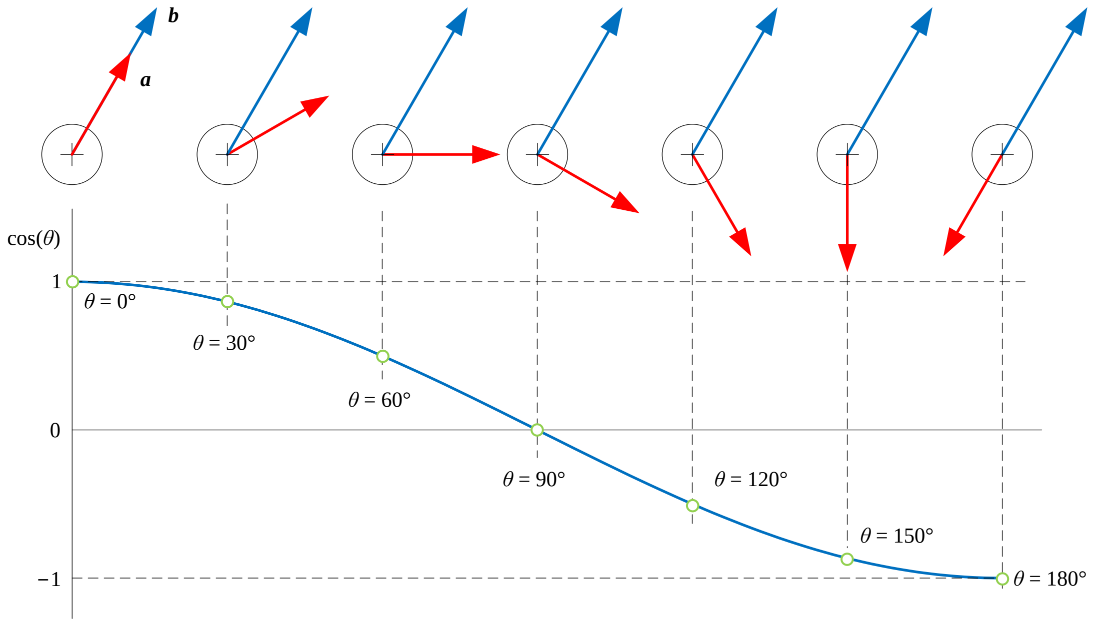

Figure 9. Variation of the inner product with varying angle θ for fixed-length vectors ***a*** and ***b***

2.5.3 Vector Projection: Adding Direction to Shadows

The **vector projection** extends the scalar projection by including direction. It represents the actual vector along the projection direction, not just its length. The vector projection of ***a*** onto ***b*** is

$$\text{proj}_{\boldsymbol{b}}\,\boldsymbol{a} = \underbrace{\|\boldsymbol{a}\|\cos\theta}_{\text{Scalar projection}} \times \underbrace{\hat{\boldsymbol{b}}}_{\text{Direction vector}} = \left(\boldsymbol{a}\cdot\hat{\boldsymbol{b}}\right)\hat{\boldsymbol{b}} \tag{14}$$

Here, $\boldsymbol{a}\cdot\hat{\boldsymbol{b}}$ is the scalar projection, and $\hat{\boldsymbol{b}}$ is the unit vector in the direction of $\boldsymbol{b}$. This is shown in Figure 10 (a).

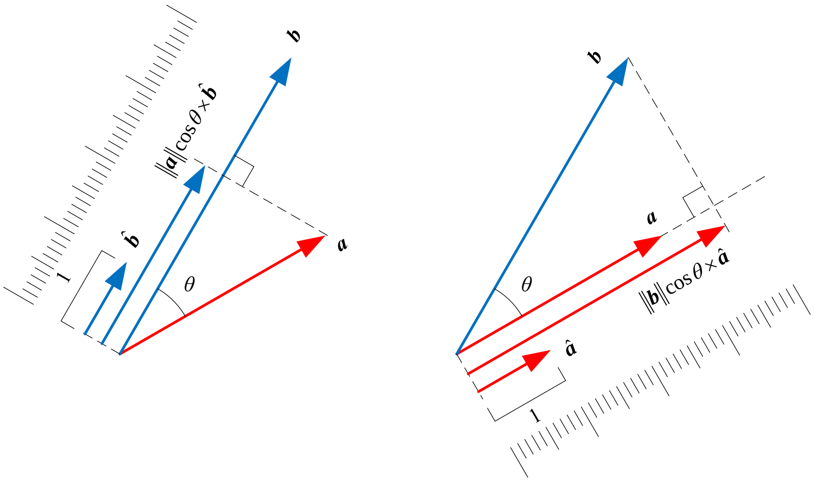

Figure 10. Vector projections: (a) ***a*** onto ***b***; (b) ***b*** onto ***a***.

Similarly, the vector projection of ***b*** onto ***a*** is

$$\text{proj}_{\boldsymbol{a}}\,\boldsymbol{b} = \underbrace{\|\boldsymbol{b}\|\cos\theta}_{\text{Scalar projection}} \times \underbrace{\hat{\boldsymbol{a}}}_{\text{Direction vector}} = \left(\boldsymbol{b}\cdot\hat{\boldsymbol{a}}\right)\hat{\boldsymbol{a}} \tag{15}$$

as illustrated in Figure 10 (b).

2.5.4 Orthogonal Components: Decomposing into Independent Directions

The **orthogonal component** of a vector with respect to another vector represents the part that cannot be expressed along the other vector. For non-zero vectors ***a*** and ***b***, the orthogonal component of ***a*** relative to ***b*** is

$$\text{oproj}_{\boldsymbol{b}}\,\boldsymbol{a} = \boldsymbol{a} - \text{proj}_{\boldsymbol{b}}\,\boldsymbol{a} \tag{16}$$

This component is perpendicular to ***b***, as shown in Figure 11 (a).

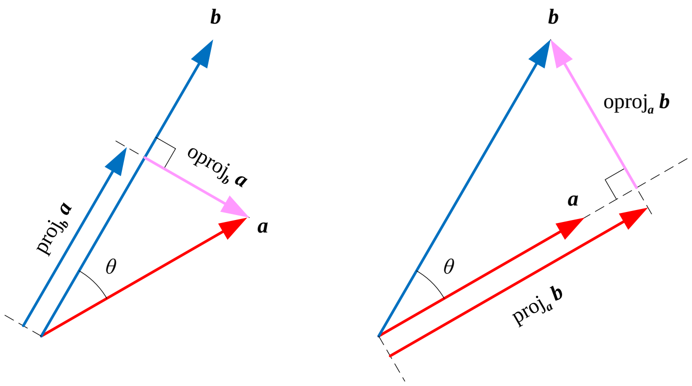

Figure 11. Orthogonal components: (a) $\boldsymbol{a}$ orthogonal to $\boldsymbol{b}$; (b) $\boldsymbol{b}$ orthogonal to $\boldsymbol{a}$.

Conversely, the orthogonal component of $\boldsymbol{b}$ with respect to $\boldsymbol{a}$ is

$$\text{oproj}_{\boldsymbol{a}} \boldsymbol{b} = \boldsymbol{b} - \text{proj}_{\boldsymbol{a}} \boldsymbol{b} \tag{17}$$

which is perpendicular to $\boldsymbol{a}$. Orthogonal components are particularly important in machine learning, physics, and geometry, as they allow decomposition of vectors into independent directions.

2.6 Vector Size Matters: Norms and Lengths

2.6.1 What is a Norm? Measuring Overall Vector Size

When comparing vectors, it is not enough to look at individual components; we need a systematic way to measure their overall size or magnitude. This is where **vector norms** come in. A norm is a function that assigns a non-negative length to a vector, providing a consistent way to quantify its size.

The most commonly used norms are the L1 norm, L2 norm, and L∞ norm, which are special cases of the more general Lp norm.

2.6.2 The General Lp Norm: Definition and Formula

The Lp norm of an n-dimensional vector $\boldsymbol{x}$ is defined as:

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \tag{18}$$

Please note that $p \geqslant 1$.

For a two-dimensional vector $\boldsymbol{x} = [x_1, x_2]^{\mathrm{T}}$, the Lp norm becomes

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_p = \left(|x_1|^p + |x_2|^p \right)^{\frac{1}{p}} \tag{19}$$

By varying the value of p, we obtain different norms with distinct geometric interpretations, as illustrated in Figure 12.

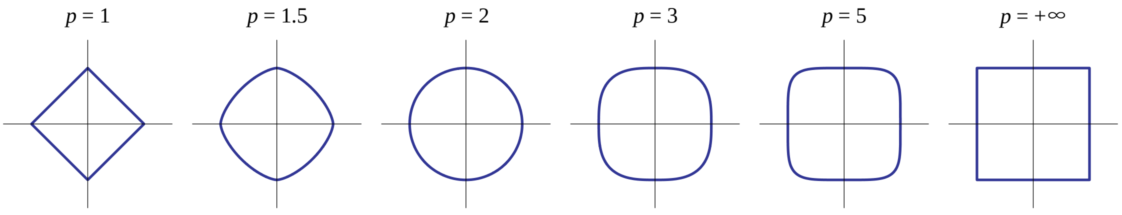

Figure 12. Geometric shapes of Lp norms in the plane for different values of p, illustrating how the notion of “vector size” changes with the norm.

2.6.3 Special Cases of Lp Norms: L1, L2, and L∞

Case a) L1 Norm

When p = 1, we have the **L1 norm**, also called the Manhattan distance or city block distance. This norm sums the absolute values of the vector components:

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i| \tag{20}$$

For a two-dimensional vector, this reduces to

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_1 = |x_1| + |x_2| \tag{21}$$

As shown in Figure 13, the L_1 *norm* of a two-dimensional vector $\boldsymbol{x}$ measures its "taxicab" or "Manhattan" distance from the origin.

If we fix the norm value at some constant c > 0, all vectors that satisfy

$$|x_1| + |x_2| = c \tag{22}$$

lie on the same *iso-distance line* (or *level curve*). In the two-dimensional plane, these curves form a square rotated by 45°, which can also be viewed as a diamond shape centered at the origin.

This geometric interpretation helps us see that the L1 norm measures distance along grid-like paths rather than straight lines, which explains why it is often used in settings that encourage sparsity or align with coordinate axes—such as LASSO regression in machine learning.

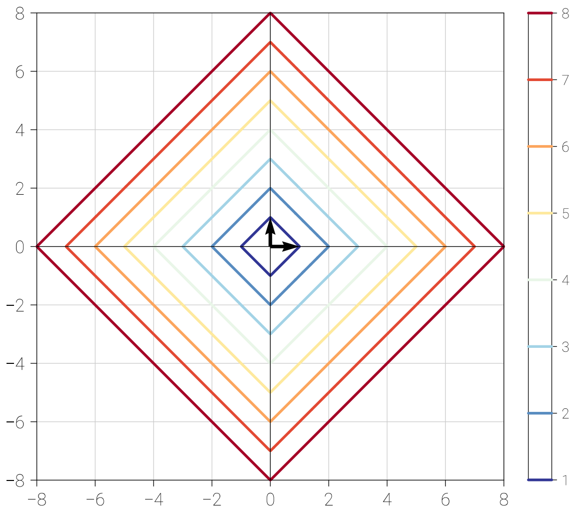

Figure 13. Iso-distance lines of the L1 norm for a two-dimensional vector $\boldsymbol{x}$. Each diamond-shaped contour represents vectors with the same L1 norm value. Figure generated by Ch02_03_L1_Norm.ipynb.

Case b) L2 Norm

When p = 2, we obtain the **L2 norm**, also known as the Euclidean norm or simply the vector's magnitude. It measures the straight-line distance from the origin to the vector's endpoint in space:

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{23}$$

For a two-dimensional vector, the L2 norm is

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_2 = x_1^2 + x_2^2 \tag{24}$$

This is the most familiar measure in the physical world, as it corresponds to the direct distance between points in space.

This norm is the most intuitive measure of distance because it matches our physical understanding of how far one point is from another in ordinary space. It is the same distance you would measure using a ruler between two points on a plane.

Geometrically, all points that have the same L_2 norm value form a circle (or more generally, a hypersphere in higher dimensions) centered at the origin. In Figure 14, the equation

$$\sqrt{x_1^2 + x_2^2} = c \tag{25}$$

represents such an *iso-distance curve* for a constant $c > 0$. Each circle corresponds to a set of points that are equally distant from the origin.

Because of its geometric simplicity and physical meaning, the L_2 norm is the most commonly used distance measure in mathematics, physics, and machine learning. Many algorithms—such as least squares regression, principal component analysis (PCA), and support vector machines (SVMs)—are built upon this Euclidean notion of distance.

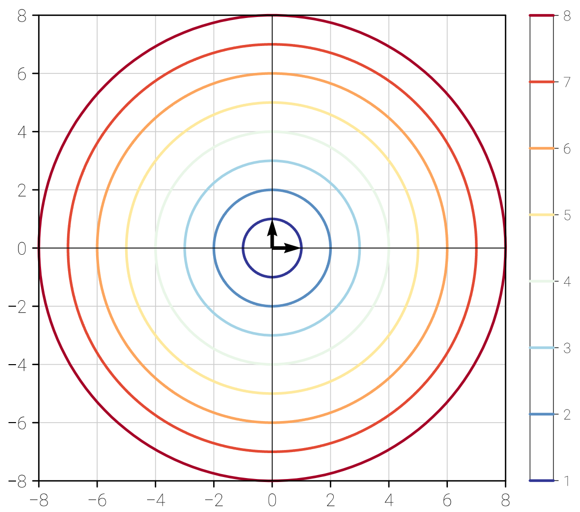

Figure 14. Iso-distance lines of the L_2 norm for a two-dimensional vector. Each circle represents vectors that are equally distant from the origin in the Euclidean sense.

Case c) L∞ Norm

Finally, as p approaches infinity, we obtain the **L∞ norm**, also called the Chebyshev norm. It is defined as the largest absolute value among the vector's components:

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i| \tag{26}$$

For a two-dimensional vector, this simplifies to

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_\infty = \max\left(|x_1|, |x_2|\right) \tag{27}$$

This norm measures distance in terms of the greatest deviation along any coordinate axis. In other words, if you imagine moving from the origin to the point $\boldsymbol{x} = [x_1, x_2]^{\mathrm{T}}$ using axis-aligned steps, the L∞ norm tells you the longest single step needed in either direction.

Geometrically, the set of points with the same L∞ norm value forms a square aligned with the coordinate axes. In Figure 15, all points satisfying

$$\max\left(|x_1|, |x_2|\right) = c \tag{28}$$

lie on the boundary of a square centered at the origin, where $c > 0$ is the chosen norm value.

This measure often appears in optimization problems and machine learning tasks where the goal is to control the *maximum* error or deviation rather than the total or average error. For example, the L∞ norm is useful in settings that prioritize robustness, ensuring no single dimension exceeds a certain bound.

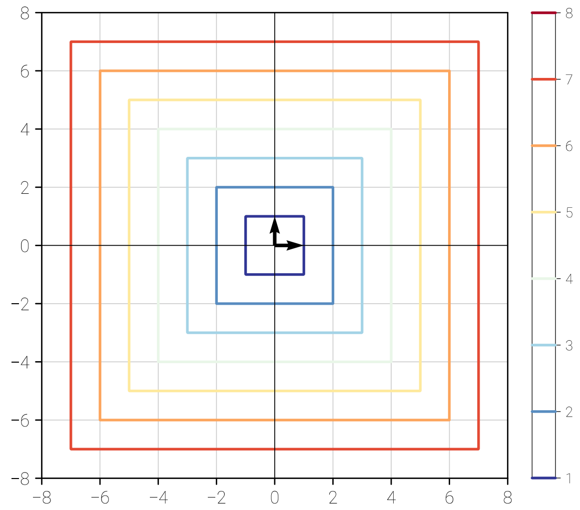

Figure 15. Iso-distance lines of the L∞ norm for a two-dimensional vector. Each square represents points that are equally distant from the origin under the Chebyshev (maximum) distance.

Each norm provides a different way to quantify vector size, and choosing the right norm depends on the context. For example, L1 is often used in optimization problems that promote sparsity, L2 is the standard for measuring geometric distance, and L∞ is useful when the maximum deviation is most important.

2.7 Special Vectors: Unit, Zero, and All-Ones

2.7.1 Unit Vectors: Standardizing Direction

Certain types of vectors are particularly important in mathematics and machine learning due to their special properties. One of the most common operations is vector normalization, also called vector unitization. Normalization adjusts the length of a non-zero vector to one while preserving its original direction. The resulting vector is called a unit vector, or sometimes a direction vector when emphasizing its orientation. Normalization is essentially a special form of scalar multiplication applied to the vector.

For any non-zero n-dimensional column vector $\boldsymbol{a}$, the normalized vector is obtained by dividing the vector by its length (magnitude, Euclidean norm, or L2 norm):

$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} = \frac{1}{\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{29}$$

This produces a vector with length equal to 1, pointing in the same direction as the original vector.

In a two-dimensional Cartesian coordinate system, the unit vectors along the positive horizontal and vertical axes are:

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{2\times 1}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{2\times 1} \tag{30}$$

As illustrated in Figure 16, a normalized 2D vector originating at the origin has its endpoint lie exactly on the **unit circle**, demonstrating that all unit vectors in the plane have length 1.

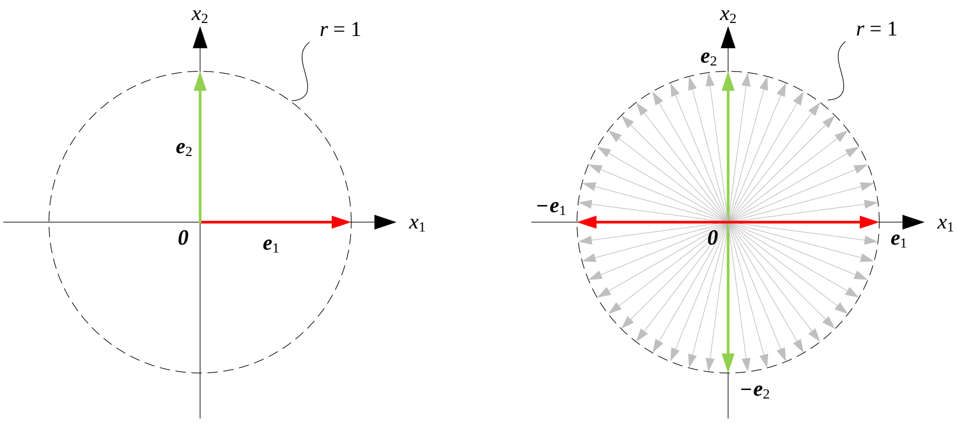

Figure 16. A 2D unit vector originating at the origin; its endpoint lies on the unit circle.

Similarly, in three dimensions, the positive unit vectors along the x_1, x_2, and x_3 axes are:

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{3\times 1}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_{3\times 1}, \quad \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_{3\times 1} \tag{31}$$

Any normalized 3D vector originating from the origin has its endpoint on the **unit sphere**, as shown in Figure 17.

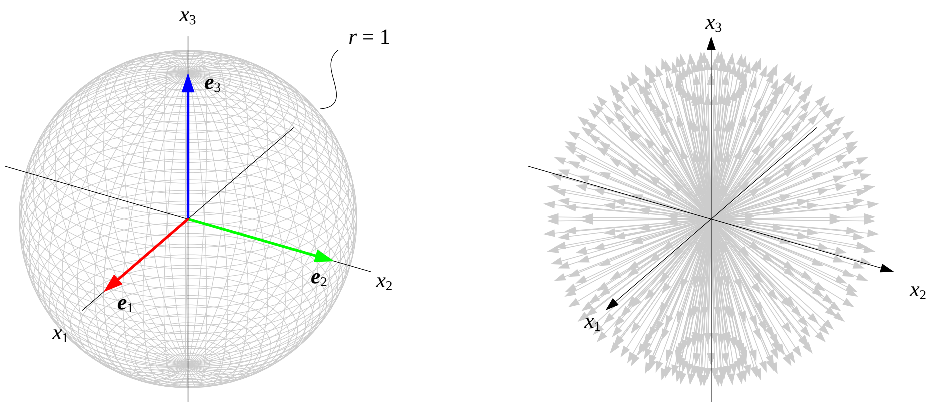

Figure 17. A 3D unit vector originating at the origin; its endpoint lies on the unit sphere.

2.7.2 Zero Vectors: The Origin Point

The **zero vector**, denoted $\mathbf{0}$, is a vector in which all elements are zero. For example, the zero vector in the plane, in three-dimensional space, or in n-dimensional space can be written as:

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}_{2\times 1}, \quad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3\times 1}, \quad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} \tag{32}$$

The zero vector has both its start and end at the origin, giving it a length of zero and no well-defined direction. Some references suggest that the zero vector can be considered to point in any direction.

2.7.3 All-Ones Vectors: Diagonals in Space

Another important special vector is the **all-ones vector**, denoted $\mathbf{1}$, in which every component equals one. For example, in two, three, or n dimensions:

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{2\times 1}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{3\times 1}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1} \tag{33}$$

Note that an all-ones vector generally does not have a length of 1; its magnitude depends on the number of dimensions.

As shown in Figure 18 (a), in the plane, the all-ones column vector corresponds to the diagonal of the unit square in the first quadrant. As shown in Figure 18 (b), in three-dimensional space, the all-ones column vector corresponds to the body diagonal of the unit cube in the first octant.

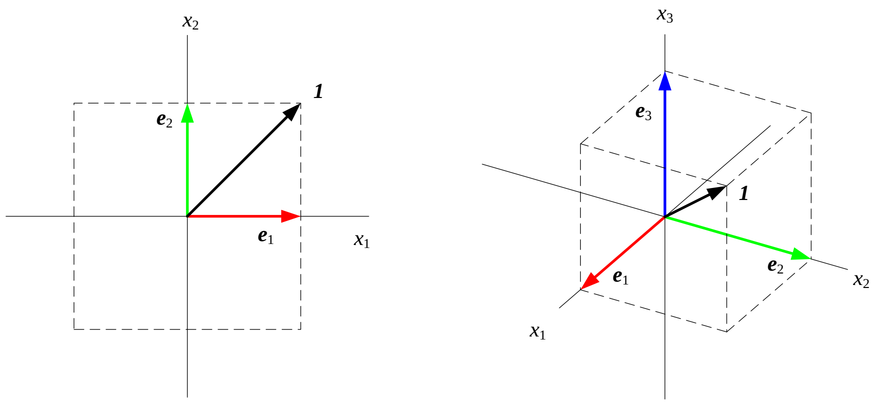

Figure 18. All-Ones Column Vectors in 2D and 3D Space

These special vectors—unit vectors, zero vectors, and all-ones vectors—form the foundation for vector operations, coordinate representation, and many algorithms in machine learning.

2.8 Conclusion

This chapter introduces the essential concepts of vectors and their operations, forming a mathematical foundation for machine learning. It begins with the distinction between scalars, which have only magnitude, and vectors, which have both magnitude and direction.

Vectors can be represented in row or column form, and their arrangement can be converted using the transpose operation. The chapter then explores Cartesian coordinate systems, extending from one-dimensional number lines to two- and three-dimensional spaces, providing a framework to visualize vectors geometrically.

Fundamental vector operations, including addition, subtraction, scalar multiplication, inner products, projections, and orthogonal components, are explained with geometric intuition and visual examples. Vector norms, such as L1, L2, and L-infinity, offer ways to measure vector size in different contexts.

Finally, the chapter discusses special vectors, including unit vectors, zero vectors, and all-ones vectors, highlighting their significance in representing directions, origins, and standardized reference values in machine learning applications.