Vector

Fndt'n of IS & Data Anlys

Contents

What is vector?

- A vector is an ordered list of numbers. Vectors are usually written as vertical arrays:

$$\begin{pmatrix} -1.1 \\ 0.0 \\ 3.6 \end{pmatrix} \text{ or } \begin{bmatrix} -1.1 \\ 0.0 \\ 3.6 \end{bmatrix}$$

It is enclosed by curved bracket or squared bracket.

- The elements (or entries, coefficients, components) of a vector are the values in the array.
- Size (also called dimension or length) of a vector is the number of elements it contains.
- The vector above has size 3.
- Or say the dimension of the vector is 3.
- A vector of size n is called an n-vector, or n-dimensional vector
- A 1-vector is considered to be the same as a number.
- A vector can also be written as numbers separated by commas and surrounded by curved bracket. In this notation style, the vector above is written as (-1.1, 0.0, 3.6).

In advanced concept, a vector may contain infinite elements

What is vector?

CAUTION

DO NOT USE your high school notation to represent a vector

YOU CANNOT use

the so-called $\boldsymbol{i}\,\boldsymbol{j}\,\boldsymbol{k}$ notation to represent a vector in this course.

I do not mark the answers with such notations

Vector notation

- We often use bolded symbols to denote vectors.

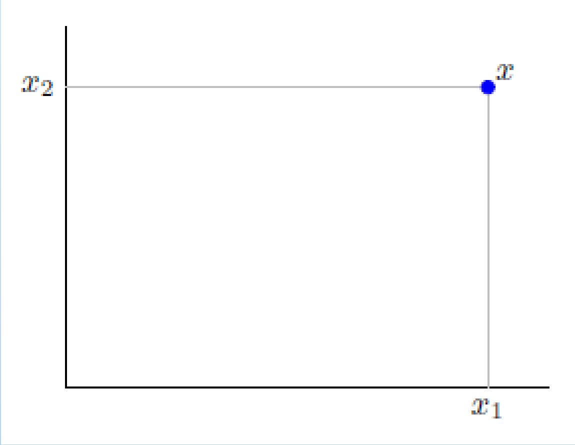
$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, the i-th element of the vector $\boldsymbol{a}$ is denoted as a_i.

$\boldsymbol{a} = \begin{pmatrix} 1.1 \\ 2.2 \\ 2.1 \end{pmatrix}$, then $a_3 = 2.1$

- Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal ($\boldsymbol{a} = \boldsymbol{b}$), if they have the same size and $a_i = b_i$ for all $i = 1, \ldots, n$.
- The numbers or values of the elements in a vector are called scalars. In this course, we consider that the scalars are real numbers.
- Sometimes, we can use $\vec{a}$ to represent a vector.
- **Zero vector** A zero vector is a vector with all elements equal to zero. Sometimes the zero vector of size n is written as $\mathbf{0}$ or $\mathbf{0}_n$.
- **Ones vector** A ones vector is vector with all elements equal to 1. We use the notation $\mathbf{1}_n$ for the n-vector with all its elements equal to 1. . Or we write $\mathbf{1}$ if the size of the vector can be determined from the context.

Vector notation

- Vectors are usually described in terms of their components in a **coordinate system**.
- In high school, given a *xy*-coordinate system, a point in a plane is described by a pair of coordinates (*x*, *y*).
- In a similar fashion, a 2-vector $\boldsymbol{x}$ can be described by a pair of *vector* coordinates (x_1, x_2).
- x_1-axis and x_2-axis



Depend on the context,

Sometimes x_i refers to x_i-axis

Or x_i refers to the i-th element of vector $\boldsymbol{x}$

Vector notation

- **A standard unit vector** is a vector with all elements equal to zero, except one element which is equal to one.
- The i-th standard unit vector (of size n) is the unit vector with i-th element one, and denoted $\boldsymbol{e}_i$.
- For example, for 3 dimensional situation,
- $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\boldsymbol{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Other notations (may be useful in senior year courses)

Block or stacked vectors. It is sometimes useful to define vectors by joining or stacking vectors together.

$\boldsymbol{a} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{c} \\ \boldsymbol{d} \end{pmatrix}$, stack $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ together.

$\boldsymbol{b}$: m-vector, $\boldsymbol{c}$: n-vector, $\boldsymbol{d}$: l-vector

$$\boldsymbol{a} = (b_1, \dots, b_m, c_1, \dots, c_n, d_1, \dots d_l)$$

- Or written as $\boldsymbol{a} = [\boldsymbol{b}; \boldsymbol{c}; \boldsymbol{d}] = (\boldsymbol{b}; \boldsymbol{c}; \boldsymbol{d})$.
- Stacked vectors can include scalars (numbers).
- $\begin{pmatrix} 1 \\ \boldsymbol{a} \end{pmatrix} = \begin{pmatrix} 1 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}$

Other notations (may be useful in senior year courses)

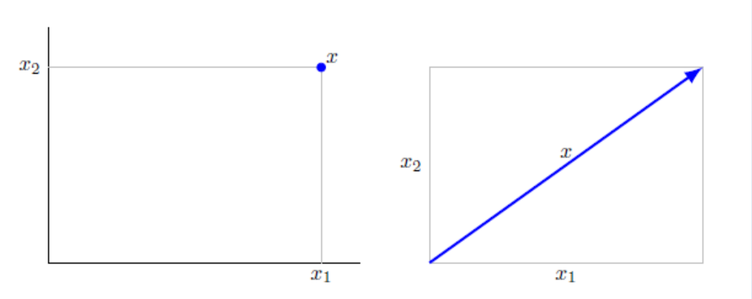
Subvectors

- In the previous slide, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ are subvectors of $\boldsymbol{a}$.
- Colon notation is used to denote subvectors.
- $a_{j:k} = (a_j, a_{j+1}, \dots, a_k)$
- The subscript $j{:}k$ is called the index range.
- Thus, in our example above, we have
- $\boldsymbol{b} = a_{1:m}$; $\boldsymbol{c} = a_{m+1:m+n}$, $\boldsymbol{d} = a_{m+n+1:m+n+l}$
- Concrete example: $\boldsymbol{z} = (1,1,2,0)$
- $z_{2:3} = (1,2)$ and$z_{3:4} = (2,0)$

Representation our real world data using vector

Location and displacement:

- A 2-vector can be used to represent a position or location in a 2-dimensional (2-D) space.



- For example, let $\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ be a vector represent the position of a car. That means, the position of the car at the x_1-axis is 1 and the position of the car at the x_2-axis is 2.
- A vector can also be used to represent a displacement in a 2-D space (or 3-D space), in which case it is typically drawn as an arrow.

Representation our real world data using vector

Location and displacement:

- A vector can also be used to represent the velocity or acceleration, at a given time, of a point that moves in a 2D-space or 3-D space.
- For example, let $\boldsymbol{a} = \binom{1}{2}$ be a vector represent the velocity of a car. Then the velocity along the x_1-direction is 1 and the velocity along the x_2-direction is 2.

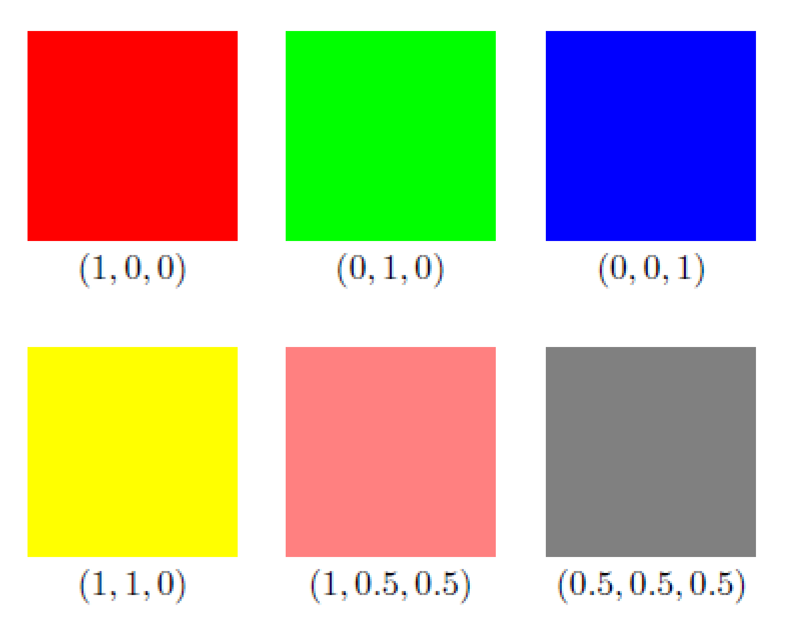
Now you know the context of our mathematics objects in engineering problem are important.

Given $\boldsymbol{a} = \binom{1}{2}$, without the context, we know nothing about the physical meaning of $\boldsymbol{a}$

Representation our real world data using vector

Color

- A 3-vector can represent a color, with its entries giving the Red, Green, and Blue (RGB) intensity values (often between 0 and 1).
- $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ = black, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} =$ a bright pure green color
- $\begin{pmatrix} 1 \\ 0.5 \\ 0.5 \end{pmatrix} =$ a shade of pink.



Representation our real world data using vector

Portfolio.

An n-vector $\boldsymbol{a}$ = a stock portfolio or investment in n different assets

a_i : the number of shares of asset-i held.

For example $\boldsymbol{a} = \begin{pmatrix} 100 \\ 50 \\ 20 \end{pmatrix}$

$\boldsymbol{a}$ represents a portfolio consisting of 100 shares of asset 1, 50 shares of asset 2, and 20 shares of asset 3.

Representation our real world data using vectors

Values across a population

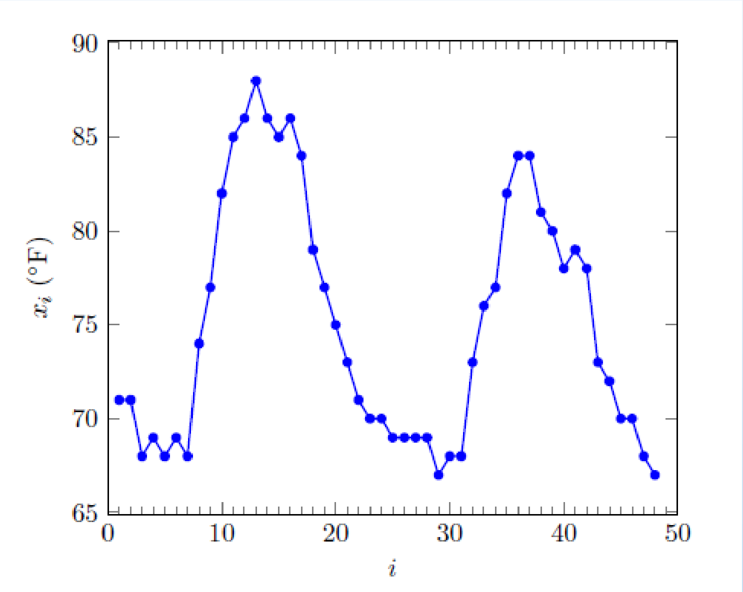
An n-vector $\boldsymbol{a}$ = the values of some quantity across a population of individuals or entities.

For example, an n-vector $\boldsymbol{a}$ can represent the blood pressure values of a collection of n patients,

a_i : the blood pressure of patient-i.

Representation our real world data using vectors

Time series

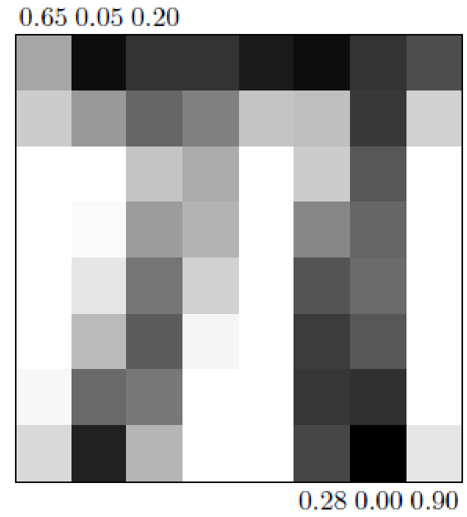


- An n-vector can represent a time series or signal, that is, the value of some quantity at different times.
- The entries in a vector that represents a time series are sometimes called samples, especially when the quantity is something measured.
- A vector might give the hourly rainfall at a location over some time period.
- When a vector $\boldsymbol{a}$ represents a time series, it is natural to plot a_i versus i with lines connecting consecutive time series values.
- The figure is the hourly temperature in downtown Los Angeles over two days.

Representation our real world data using vectors

Image

- A monochrome (grey scalar) image is an array of MN pixels with M rows and N columns.
- Each pixel has a grayscale or intensity value, with 0 corresponding to black and 1 corresponding to bright white. (Other ranges are also used.)
- An image can be represented by a vector of length MN, with the elements giving grayscale levels at the pixel locations, typically ordered column-wise or row-wise.
- A color M by N image is described by a vector of length 3 MN, with the entries giving the R, G, and B values for each pixel, in some agreed-upon order.



8×8 image and the grayscale levels at six pixels.

Representation our real world data using vectors

Document Analysis

- Given n keywords (dictionary), we can use an n-vector $\boldsymbol{a}$ to represent a document.
- a_i represents the number of times the i-th keyword in the dictionary.

Word count vectors are used in computer based document analysis. Each entry of the word count vector is the number of times the associated dictionary word appears in the document.

Consider that we use 6 keywords:{word, in, number horse, the, document}

$$\begin{array}{ll} \text{word} & \\ \text{in} & \\ \text{number} & \\ \text{horse} & \\ \text{the} & \\ \text{document} & \end{array} \quad \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \\ 4 \\ 2 \end{bmatrix}$$

Common practice : count variations of a word as the same word; for example, `rain', `rains', `raining' and 'rained' are counted as `rain'.

Representation our real world data using vectors

- Features or attributes in machine learning
- In many machine learning applications an n-vector $\boldsymbol{a}$ collects together n different quantities that represent a single thing or object.
- The quantities can be measurements, or quantities that can be measured or derived from the object.
- We the vector as Feature vector, and its entries are called the features or attributes.
- For example, a 6-vector $\boldsymbol{a}$ could give the age, height, weight, blood pressure, temperature, and gender of a patient admitted to a hospital.

The last entry of the vector could be encoded as $a_6 = 0$ for male, $a_6 = 1$ for female, $a_6 = 2$ for others)

Addition

- Two vectors with the same size can be added together by adding the corresponding elements, to form another vector of the same size, called the sum of the vectors.
- Vector addition is denoted by the symbol +.

$$\boldsymbol{u} = \begin{pmatrix} 0 \\ 7 \\ 3 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \boldsymbol{u} + \boldsymbol{v} = \begin{pmatrix} 0 \\ 7 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 9 \\ 3 \end{pmatrix}$$

$$\boldsymbol{u} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \boldsymbol{u} + \boldsymbol{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$$

Vector subtraction is similar.

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{u} - \boldsymbol{v} = \begin{pmatrix} 1 \\ 9 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 8 \end{pmatrix}$$

Actually, we no need to define subtraction. Why?

Addition

Properties

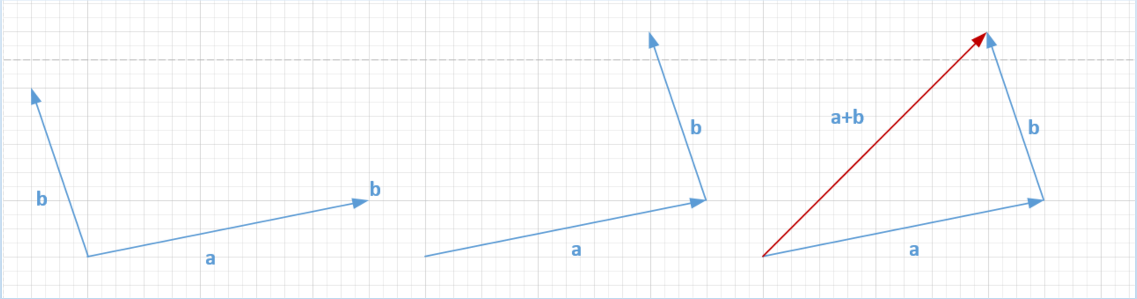
For any vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ of the same size,

- Vector addition is commutative: $\boldsymbol{a} + \boldsymbol{b}$ = $\boldsymbol{b} + \boldsymbol{a}$.
- Vector addition is associative: $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$. Therefor We can therefore write both as $\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c}$. .
- $\boldsymbol{a} + \boldsymbol{0} = \boldsymbol{0} + \boldsymbol{a}$.
- $\boldsymbol{a} - \boldsymbol{a} = \boldsymbol{0}$. Subtracting a vector from itself yields the zero vector.
- Of course, $\boldsymbol{0}$ denotes a zero vector with the size equal to the size of $\boldsymbol{a}$.

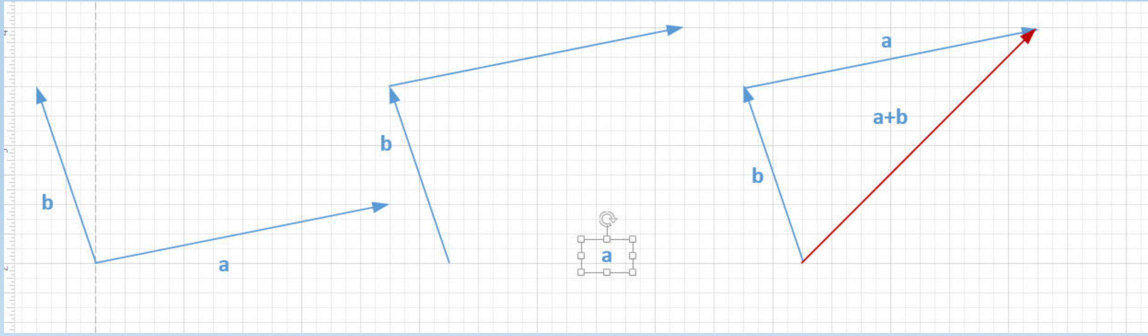
Addition Graphical illustration

$$\boldsymbol{a} = \binom{5}{1}, \boldsymbol{b} = \binom{-1}{3}$$

The *head-to-tail method* is a graphical way to add vectors, described in the following figure. The *tail* of vector **a** is the starting point of vector **b**.



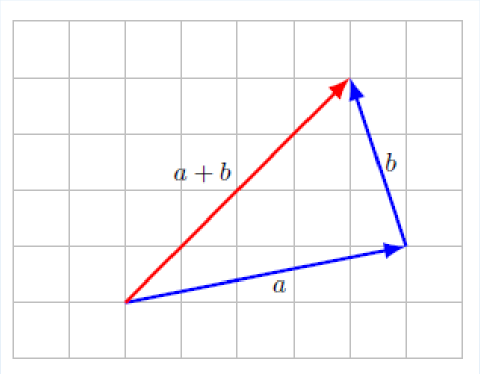
$$\boldsymbol{a} + \boldsymbol{b} = \binom{5}{1} + \binom{-1}{3} = \binom{4}{4}$$



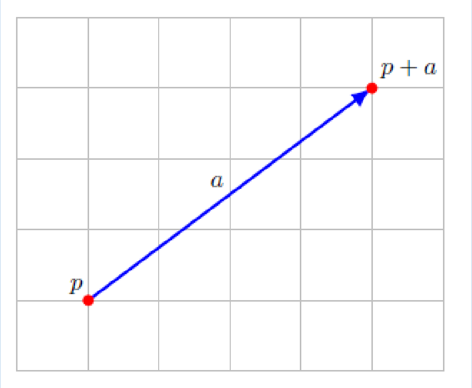
Examples

Displacements

- When vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ represent displacements, the sum $\boldsymbol{a}+\boldsymbol{b}$ is the net displacement found by first displacing by $\boldsymbol{a}$, then displacing by $\boldsymbol{b}$,



- If the vector $\boldsymbol{p}$ represents a position and the vector $\boldsymbol{a}$ represents a displacement, then $\boldsymbol{p}+\boldsymbol{a}$ is the position of the point $\boldsymbol{p}$, displaced by $\boldsymbol{a}$.

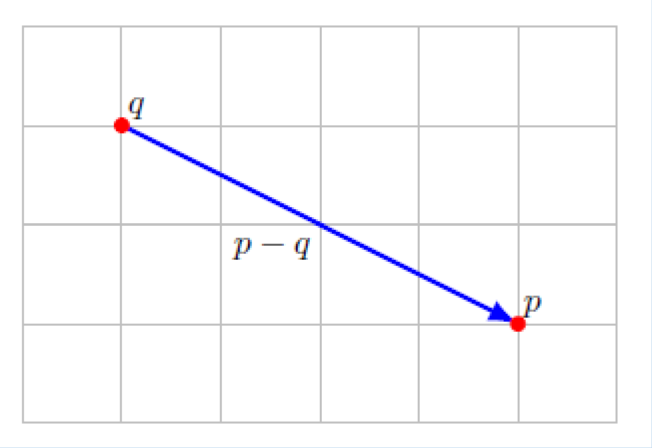


Examples

Displacements between two points.

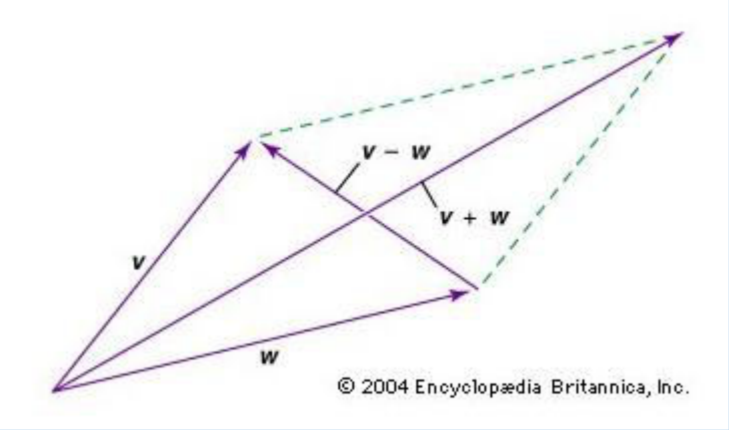
If the vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ represent the positions of two points in 2-D or 3-D space, then $\boldsymbol{p} - \boldsymbol{q}$ is the displacement vector from $\boldsymbol{q}$ to $\boldsymbol{p}$. We can denote it as

$\boldsymbol{d}_{\boldsymbol{qp}}$ the displacement vector from $\boldsymbol{q}$ to $\boldsymbol{p}$.



Examples

Addition examples:



Examples

Word counts

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are word count vectors (using the same dictionary) for two documents, the sum $\boldsymbol{a} + \boldsymbol{b}$ is the word count vector of a new document created by combining the original two (in either order).

Bill of materials

Suppose $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N$ are n -vectors that give the quantities of n different resources required to accomplish N tasks. Then the sum n -vectors

$\boldsymbol{q}_1 + \boldsymbol{q}_2 \ldots + \boldsymbol{q}_N$ gives the bill of materials for completing all N tasks.

Scalar Multiplication

A vector $\boldsymbol{a}$ is multiplied by a scalar α(i.e., number), which is done by multiplying every element of the vector by the scalar

$$\alpha\boldsymbol{a} = \alpha\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_3 \end{pmatrix}, \text{ or } \boldsymbol{a}\alpha = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}\alpha = \begin{pmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_3 \end{pmatrix}$$

$$(-2)\begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix}(1.5) = \begin{bmatrix} 1.5 \\ 13.5 \\ 9 \end{bmatrix}$$

Scalar Multiplication

Associative:

$$\alpha\beta\boldsymbol{a} = (\alpha\beta)\boldsymbol{a} = \alpha(\beta\boldsymbol{a}) = \beta(\alpha\boldsymbol{a})$$

Distributive

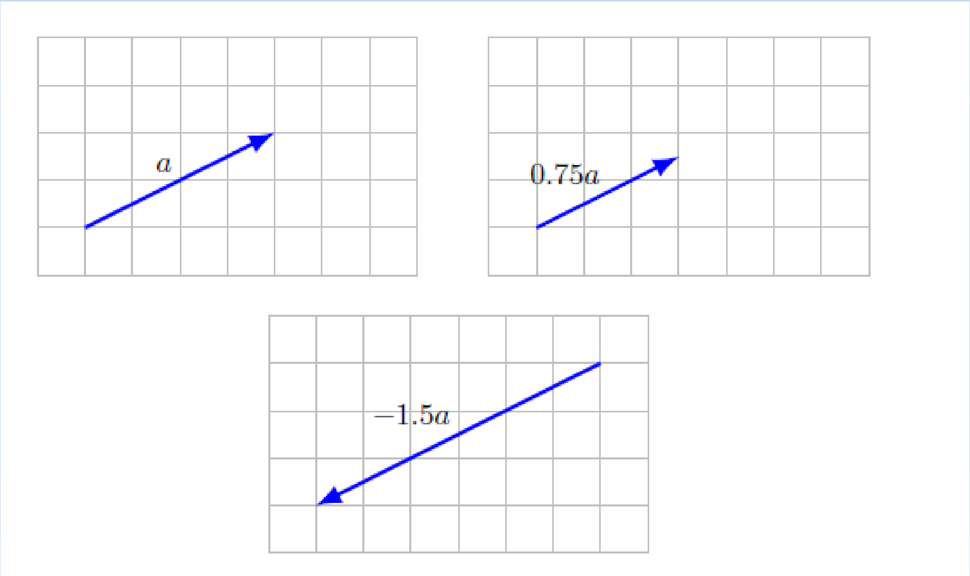
$$(\alpha + \beta)\boldsymbol{a} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}$$

Scalar Multiplication (Example)

Displacements

When a vector $\boldsymbol{a}$ represents a displacement, and $\beta > 0$, $\beta\boldsymbol{a}$ is a displacement vector in the same direction of $\boldsymbol{a}$, with its magnitude scaled by β.

When $\beta < 0$, a represents a displacement vector in the opposite direction of a, with magnitude scaled by β.



Scalar Multiplication (Example)

Materials requirements

Suppose the n-vector $\boldsymbol{q}$ is the bill of materials for producing one unit of some product, i.e., q_i is the amount of raw material required to produce one unit of product.

To produce k units of the product will then require raw materials given by $k\boldsymbol{q}$.

Linear combinations

If $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_m$ are n-vectors, and $1, \dots, m$ are scalars, the n-vector

$$\beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2, \dots, +\beta_m \boldsymbol{a}_m$$

is linear combination of the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_m$.

The scalars $\beta_1, \beta_2, \dots, \beta_m$ are called the coefficients of the linear combination.

$\boldsymbol{a}_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \boldsymbol{a}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\beta_1 = 2,\ \beta_2 = 3, \beta_3 = -1$

$$\boldsymbol{c} = \beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2 + \beta_3 \boldsymbol{a}_3 = \begin{pmatrix} 10 - 3 - 1 \\ 2 + 9 - 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \end{pmatrix}$$

Given $\boldsymbol{a}_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$, and $\boldsymbol{c} = \begin{pmatrix} 5 \\ 0.5 \end{pmatrix}$, whether $\boldsymbol{c}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2$???

Given $\boldsymbol{a}_1 = \begin{pmatrix} 5 \\ 0 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, and $\boldsymbol{c} = \begin{pmatrix} 5 \\ 0.5 \end{pmatrix}$, whether $\boldsymbol{c}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2$???

Special Linear combination

We can write any n -vector $\boldsymbol{b}$ as a linear combination of the standard unit vectors, as

$\boldsymbol{b} = b_1\boldsymbol{e}_1 + b_2\boldsymbol{e}_2 + \ldots, b_n\boldsymbol{e}_n$

$$\begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix} = (-1)\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

In linear algebra, $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ form a basis.

Any n -vector can be written as a linear combination of $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$.

Linear combination of standard unit vectors.

Some linear combinations of the vectors.

have special names.

- The linear combination with $\beta_1 = \beta_2 =, \ldots, \beta_m = 1$, given by $\boldsymbol{a}_1 + \boldsymbol{a}_2, \ldots + \boldsymbol{a}_m$, is the sum of the vectors.

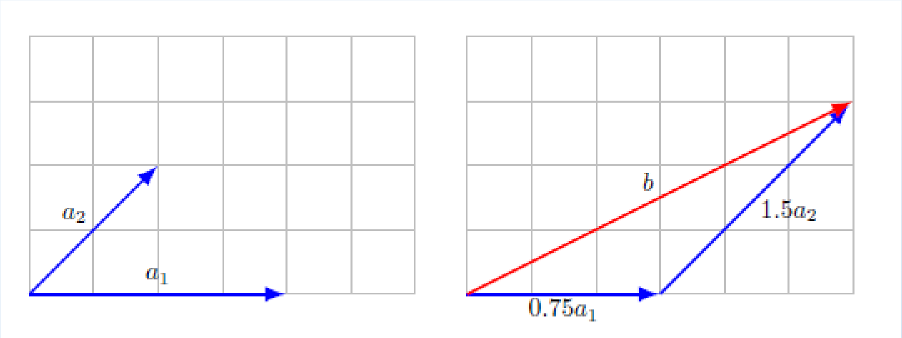
- The linear combination with $\beta_1 = \beta_2 =, \ldots, \beta_m = 1/m$ given by $\frac{1}{m}(\boldsymbol{a}_1 + \boldsymbol{a}_2, \ldots + \boldsymbol{a}_m)$, is the average of the vectors.

- Also, given m n-vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ with $m < n$, we can use a linear combination to define a linear sub-space, given by $\beta_1\boldsymbol{a}_1 + \beta_2\boldsymbol{a}_2, \ldots + \beta_m\boldsymbol{a}_m$, where β_i's are real numbers. This is $\boldsymbol{c} = \beta_1\boldsymbol{a}_1 + \beta_2\boldsymbol{a}_2, \ldots + \beta_m\boldsymbol{a}_m$ for real β_i's . Any vector in this subspace is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$.

Examples

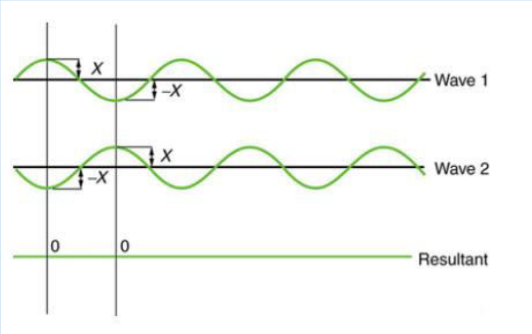
Displacements

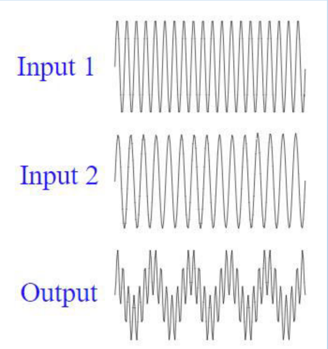
- When the vectors represent displacements, a linear combination is the sum of the scaled displacements.



Audio mixing

When $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ are vectors representing audio signals (over the same period of time, for example, simultaneously recorded), they are called tracks. The linear combination $\beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2, .. + \beta_m \boldsymbol{a}_m$ is perceived as a mixture of the audio tracks.

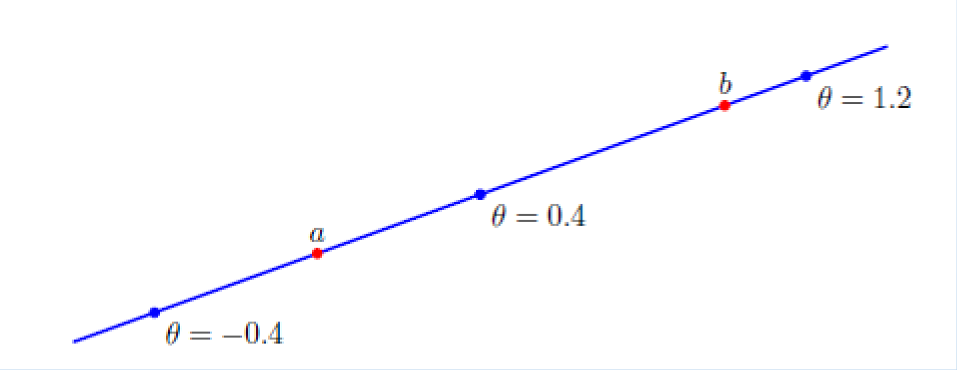




Example

Line and segment

When $\boldsymbol{a}$ and $\boldsymbol{b}$ are different n-vectors, the combination $\boldsymbol{c} = (1-\theta)\boldsymbol{a} + \theta\boldsymbol{b}$, where θ $(0 \le \theta \le 1)$is a scalar, describes a point on the line passing through $\boldsymbol{a}$ and $\boldsymbol{b}$.

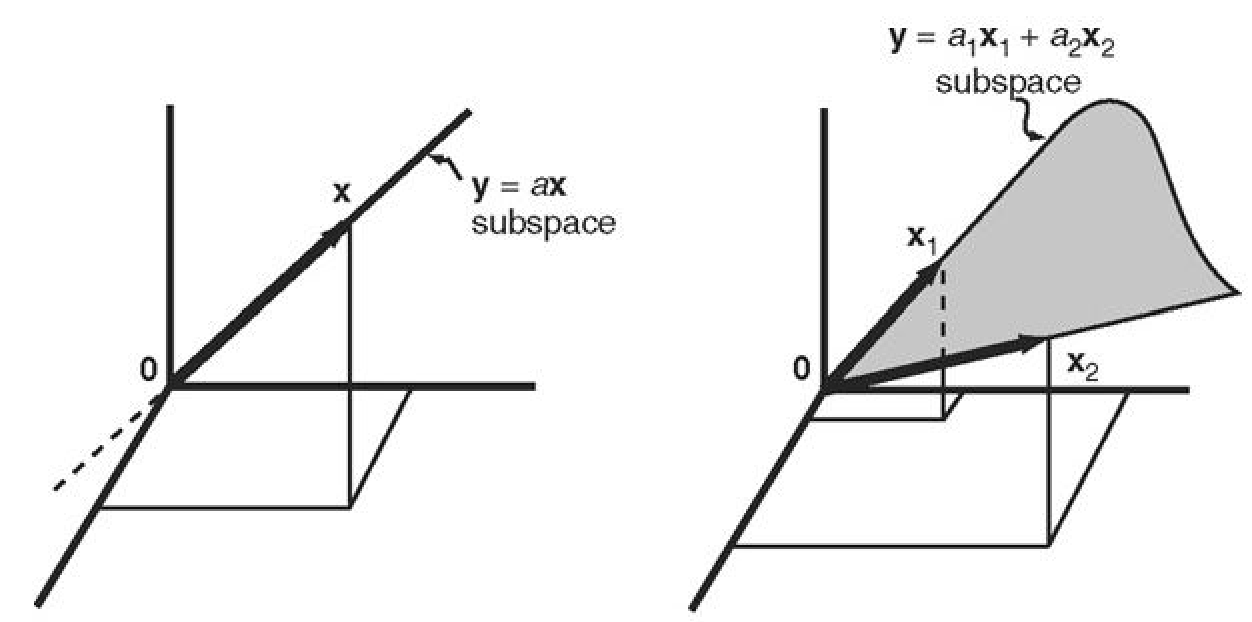


These points are on the line passing through $\boldsymbol{a}$ and $\boldsymbol{b}$. For $0 \le \theta \le 1$, the points are on the line segment between $\boldsymbol{a}$ and $\boldsymbol{b}$.

Example

Given $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m$,

- we can use $\boldsymbol{y} = a\boldsymbol{x}_1$ to define a subspace
- we can use $\boldsymbol{y} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2$ to define a subspace



Inner product

- The inner product of two n-vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined as
$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = a_1 b_1 + \cdots, + a_n b_n.$$
- Other notation: $\boldsymbol{a} \cdot \boldsymbol{b}$.
- The notation of ‘T′ will be explained later.

$$\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7.$$

Inner product

Properties

Commutativity : $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ = $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{a}$. The order of the two vector arguments in the

Associativity with scalar multiplication:

$(\beta\boldsymbol{a})^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{a}^{\mathrm{T}}(\beta\boldsymbol{b})$

Distributivity with vector addition:

$(\boldsymbol{a} + \boldsymbol{b})^{\mathrm{T}}\boldsymbol{c} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}$

Inner product

Block vectors. If the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are block vectors,

$$\begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \boldsymbol{a}_3 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \boldsymbol{b}_3 \end{pmatrix} = \boldsymbol{a}_1{}^{\mathrm{T}} \boldsymbol{b}_1 + \boldsymbol{a}_2{}^{\mathrm{T}} \boldsymbol{b}_2 + \boldsymbol{a}_3{}^{\mathrm{T}} \boldsymbol{b}_3.$$

Example

Portfolio value

Suppose $\boldsymbol{s}$ is an n-vector representing the holdings in shares of a portfolio of n different assets, with negative values meaning short positions.

If $\boldsymbol{p}$ is an n-vector giving the prices of the assets, then $\boldsymbol{p}^{\mathrm{T}}\boldsymbol{s}$ is the total (or net) value of the portfolio.

Portfolio return. Suppose $\boldsymbol{r}$ is the vector of (fractional) returns of n assets over some time period:

$$r_i = \frac{p_i^{final} - p_i^{init}}{p_i^{init}}$$

If $\boldsymbol{s}$ is an n -vector giving our portfolio, with s_i denoting the dollar value of asset i, then the inner product $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{s}$ is the total return of the portfolio, in dollars, over the period.

Norm and Distance

The Euclidean norm of an n-vector $\boldsymbol{x}$:

- $\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2, \ldots, +x_n^2} = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$

$$\left\| \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} \right\| = \sqrt{9} = 3, \qquad \left\| \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\| = 1.$$

- **Unit vector:** A unit vector $\boldsymbol{x}$ is a vector with $\|\boldsymbol{x}\| = 1$.

Norm and Distance

Properties of norm

- Nonnegative homogeneity: $\|\beta \boldsymbol{x}\| = |\beta|\|\boldsymbol{x}\|$
- Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

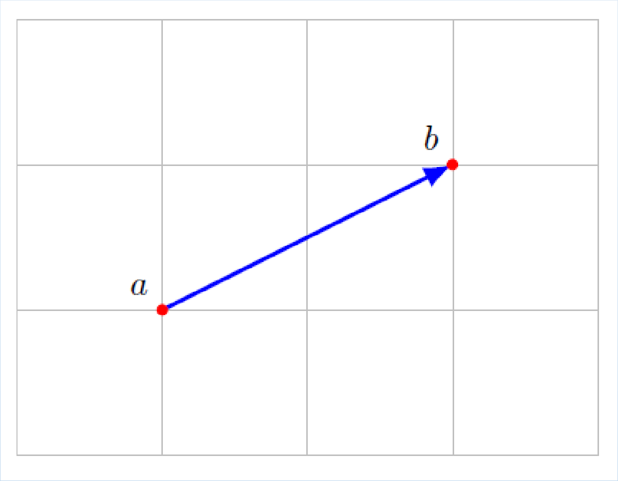
  (Can You prove it ???)
- Nonnegativity. $\|\boldsymbol{x}\| \geq 0.$
- Definiteness. $\|\boldsymbol{x}\|$ = 0 if and only if $\boldsymbol{x}$ = 0 .

Norm and Distance (distance)

We can use the norm to define the Euclidean distance between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as the norm of their difference:

$$\text{dist}(\boldsymbol{a}, \boldsymbol{b}) = \|\boldsymbol{a} - \boldsymbol{b}\|$$

$$\text{dist}(\boldsymbol{a}, \boldsymbol{b}) = \text{dist}(\boldsymbol{b}, \boldsymbol{a})$$



$$\boldsymbol{u} = \begin{bmatrix} 1.8 \\ 2.0 \\ -3.7 \\ 4.7 \end{bmatrix}, \boldsymbol{v} = \begin{bmatrix} 0.6 \\ 2.1 \\ 1.9 \\ -1.4 \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} 2.0 \\ 1.9 \\ -4.0 \\ 4.6 \end{bmatrix}$$

$\|\boldsymbol{u} - \boldsymbol{v}\| = \mathbf{8.368}$, $\|\boldsymbol{u} - \boldsymbol{w}\| = \mathbf{0.387}$, $\|\boldsymbol{v} - \boldsymbol{w}\|$=8.533

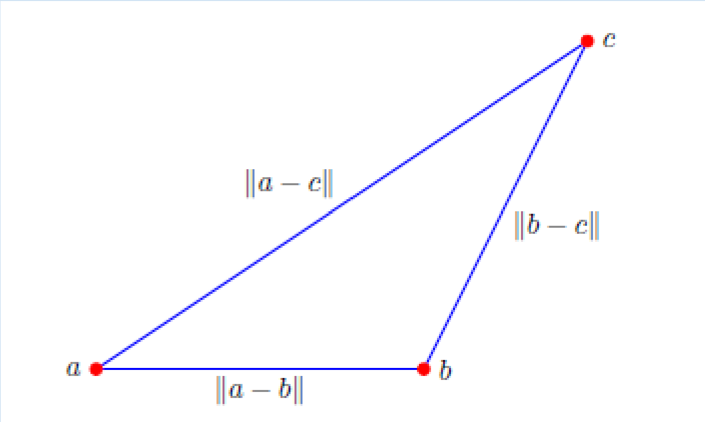
Norm and Distance (distance)

Triangle inequality

Given three vectors: $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$

$\|\boldsymbol{a}-\boldsymbol{c}\| \leq \|\boldsymbol{a}-\boldsymbol{b}\| + \|\boldsymbol{c}-\boldsymbol{b}\|$

Can you prove it ?



Norm and Distance (example)

Feature distance

If $\boldsymbol{x}$ and $\boldsymbol{y}$ represent vectors of n features of two objects, the quantity $\|\boldsymbol{x} - \boldsymbol{y}\|$ is called the feature distance, and gives a measure of how different the objects are (in terms of their feature values).

Similarity of two patients

Each patient associates a feature vector with entries such as weight, age, presence of chest pain, difficulty breathing, and the results of tests. We can use feature vector distance to say that one patient case is near another one (at least in terms of their feature vectors).

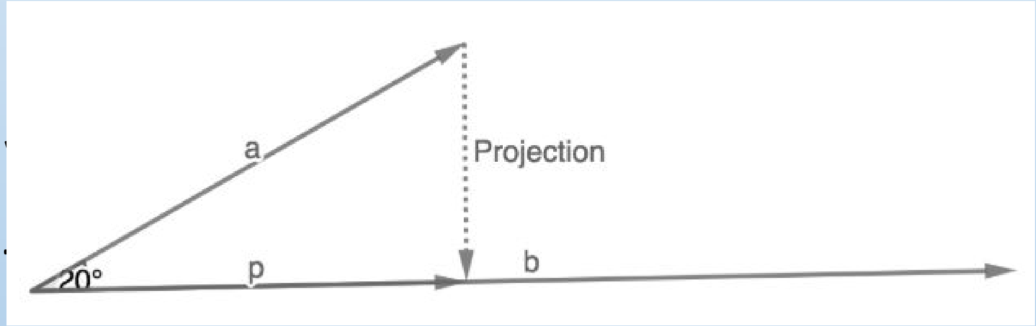
Norm and Distance (example)

Interpretation:

we can visualize the dot product as we project $\boldsymbol{a}$ to $\boldsymbol{b}$,

the projected vector is given by

$$\boldsymbol{p} = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{b}\|\|\boldsymbol{a}\|)\|\boldsymbol{a}\|\ \boldsymbol{b}/\|\boldsymbol{b}\| = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{b}\|)\ \boldsymbol{b}/\|\boldsymbol{b}\|$$



For projecting $\boldsymbol{b}$ to $\boldsymbol{a}$,

the projected vector is given by

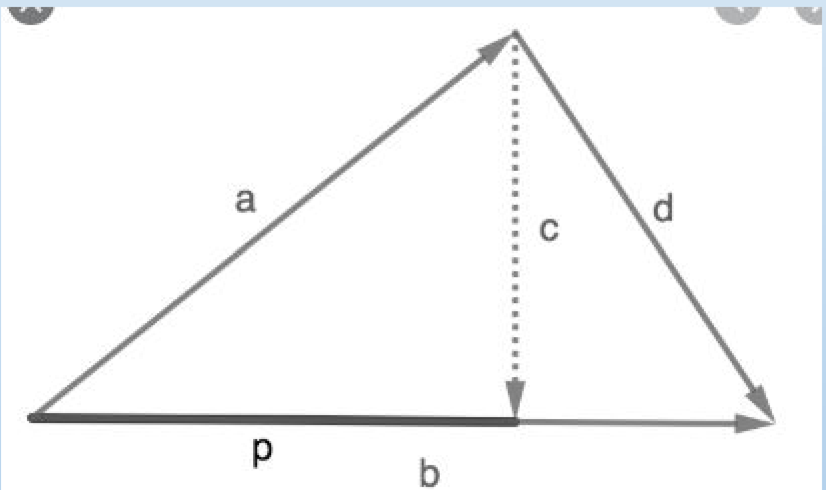
$$\boldsymbol{q} = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{a}\|)\boldsymbol{a}/\|\boldsymbol{a}\|$$

Norm and Distance (example)

Interpretation:

we project $\boldsymbol{a}$ to $\boldsymbol{b}$, the projected vector is given by

$$\boldsymbol{p} = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{b}\|\|\boldsymbol{a}\|)\|\boldsymbol{a}\|\ \boldsymbol{b}/\|\boldsymbol{b}\| = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{b}\|)\ \boldsymbol{b}/\|\boldsymbol{b}\|$$



What is $\boldsymbol{c}$?

$\boldsymbol{a} + \boldsymbol{c} = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{b}\|)\ \boldsymbol{b}/\|\boldsymbol{b}\|$, then $\boldsymbol{c}$=?

Norm and Distance (example)

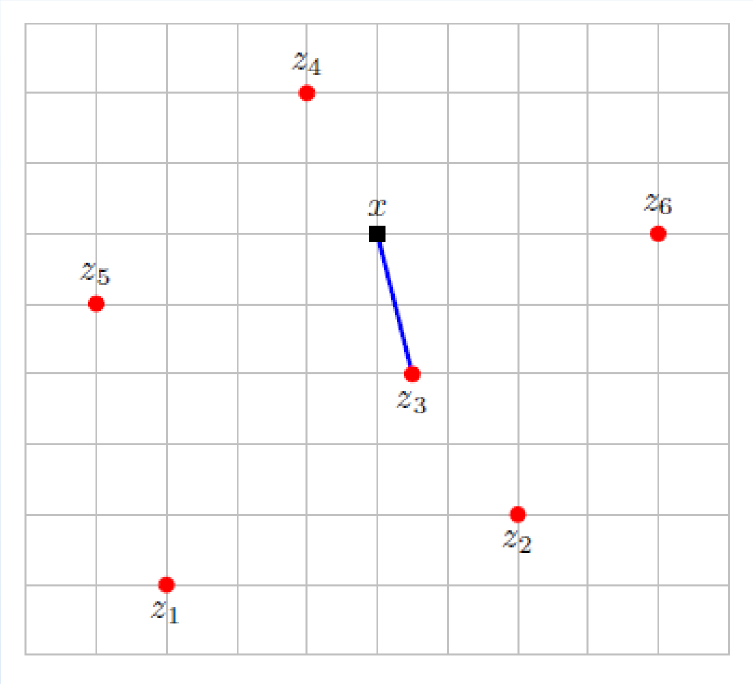
Nearest neighbor

- $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m$: a collection of m n-vectors
- $\boldsymbol{x}$: another n-vector.

If $\left\|\boldsymbol{x} - \boldsymbol{z}_j\right\| \leq \|\boldsymbol{x} - \boldsymbol{z}_i\|$ for all $i \neq j$,

$\boldsymbol{z}_j$ is the nearest neighbor of $\boldsymbol{x}$ (among $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m$).

Or saying, $\boldsymbol{z}_j$ is the closest vector (among $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m$) to $\boldsymbol{x}$.



Norm and Distance (example)

Document dissimilarity

- Two n-vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ represent the histograms of keyword occurrences for two documents. (assume that we a common set of keywords)
- Then $\|\boldsymbol{x} - \boldsymbol{y}\|$ represents a measure of the dissimilarity of the two documents

Pairwise word count histogram distances between five Wikipedia article

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	0.095	0.130	0.153	0.170
Memorial Day	0.095	0	0.122	0.147	0.164
Academy A.	0.130	0.122	0	0.108	0.164
Golden Globe A.	0.153	0.147	0.108	0	0.181
Super Bowl	0.170	0.164	0.164	0.181	0

Angle

Cauchy-Schwarz inequality

An important inequality that relates norms and inner products is the Cauchy-Schwarz inequality:

$$\left|\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}\right| \leq \|\boldsymbol{a}\|\|\boldsymbol{b}\| \text{for any } n\text{-vectors } \boldsymbol{a} \text{ and } \boldsymbol{b}.$$

$$|a_1 b_1 +, \dots, + a_n a_n| \leq \ (a_1^2 +, \dots, + a_n^2)^{1/2}(b_1^2 +, \dots, + b_n^2)^{1/2}$$

The proof :

Let $\alpha = \|\boldsymbol{a}\|$ and $\beta = \|\boldsymbol{b}\|$

Consider

$$\begin{aligned} 0 &\leq \|\beta\boldsymbol{a} - \alpha\boldsymbol{b}\|^2 \\ &= \|\beta\boldsymbol{a}\|^2 - 2(\beta\boldsymbol{a})^{\mathrm{T}}\alpha\boldsymbol{b} + \|\alpha\boldsymbol{b}\|^2 \\ &= \beta^2\|\boldsymbol{a}\|^2 - 2\alpha\beta(\boldsymbol{a})^{\mathrm{T}}\boldsymbol{b} + \alpha^2\|\boldsymbol{b}\|^2 \\ &= \|\boldsymbol{b}\|^2\|\boldsymbol{a}\|^2 - 2\|\boldsymbol{a}\|\|\boldsymbol{b}\|(\boldsymbol{a})^{\mathrm{T}}\boldsymbol{b} + \|\boldsymbol{a}\|^2\|\boldsymbol{b}\|^2 \\ &= \|\boldsymbol{b}\|^2\|\boldsymbol{a}\|^2 - \|\boldsymbol{a}\|\|\boldsymbol{b}\|(\boldsymbol{a})^{\mathrm{T}} = \|\boldsymbol{a}\|\|\boldsymbol{b}\| - (\boldsymbol{a})^{\mathrm{T}}\boldsymbol{b} \end{aligned}$$

Angle

The angle θ between two nonzero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\angle(\boldsymbol{a}, \boldsymbol{b}) = \theta = \cos^{-1}\left(\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|}\right)$$

where $\cos^{-1}$ denotes the inverse cosine.

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = \|\boldsymbol{a}\|\,\|\boldsymbol{b}\| \cos\theta$$

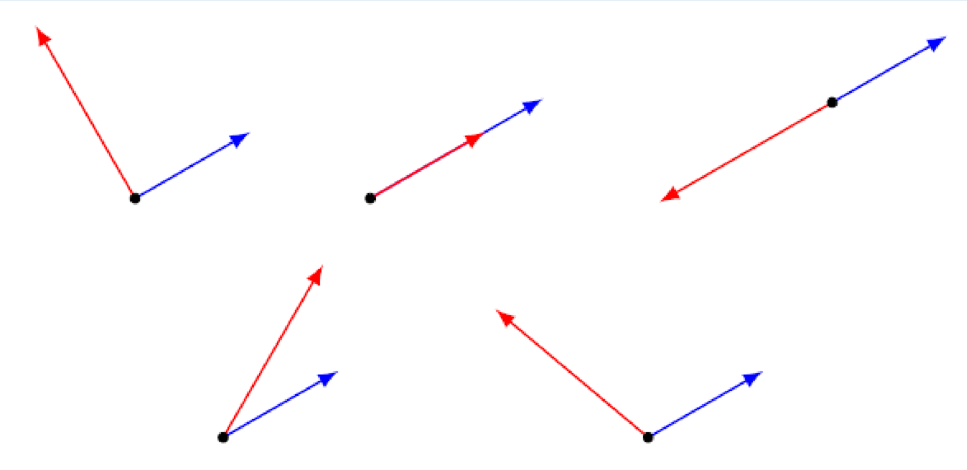
Example:

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix}$$

$$\arccos\left(\frac{5}{\sqrt{6}\,\sqrt{13}}\right) = \arccos(0.5661) = 0.9690 = 55.52^\circ$$

Acute and obtuse angles

- Angles are classified according to the sign of $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$.
- Suppose the two vectors nonzero n-vectors.
- Orthogonal: $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = 0$, $\theta = 90^{\circ}$, $\theta = \pi/2$
- Alighted:$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = \|\boldsymbol{a}\|\|\boldsymbol{b}\|, \theta = 0^{\circ}, \theta = 0$
- Anti-alighted:$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = -\|\boldsymbol{a}\|\|\boldsymbol{b}\|, \theta = \pi$
- Acute angle: $\theta < \frac{\pi}{2} = 90^{\circ}$
- Obtuse angle $\theta > \frac{\pi}{2} = 90^{\circ}$



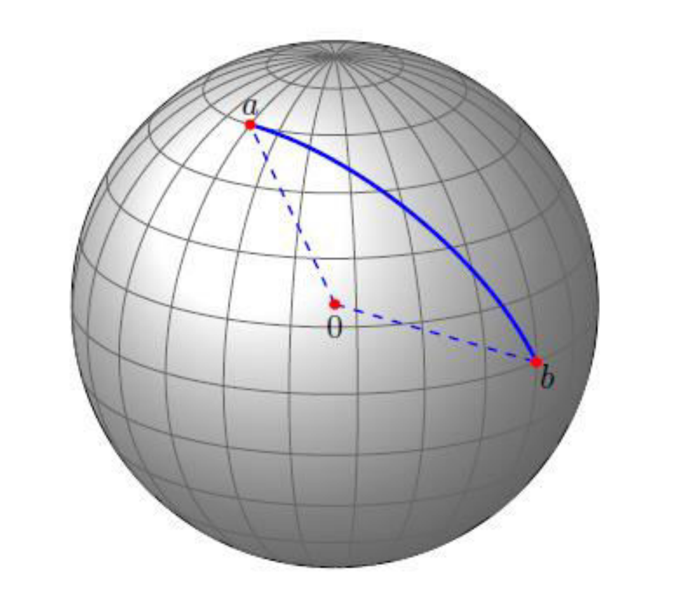
Angle of unit vectors

Unit vectors

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are unit vectors, we can use of $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ or the angle between them to define the similarity between the two vectors.

Spherical distance

$\boldsymbol{a}$ and $\boldsymbol{b}$ are 3-vectors that represent two points on the sphere of radius 1. The spherical distance between them, measured along the sphere, is given by $\angle(\boldsymbol{a}, \boldsymbol{b})$.



Examples

Document similarity via angles.

$\boldsymbol{a}$ and $\boldsymbol{b}$ are n-vectors represent the word counts of two documents, their angle can be used as a measure of document dissimilarity.

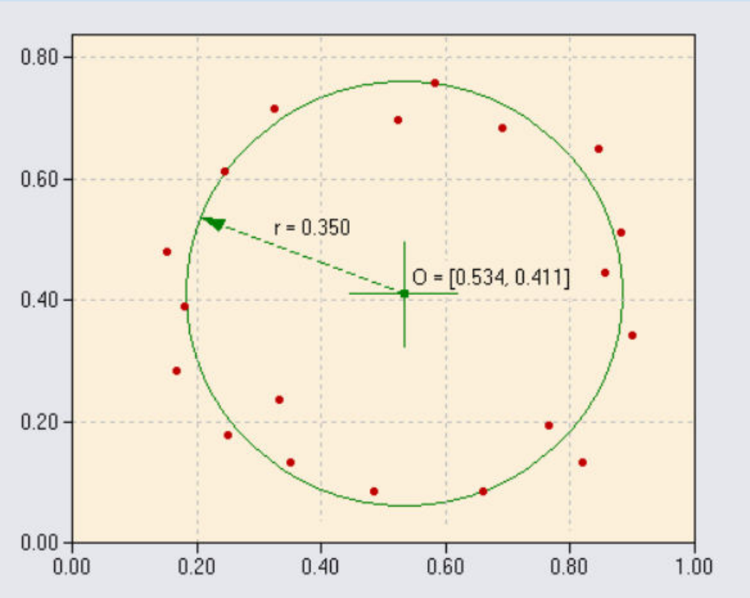
	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	60.6	85.7	87.0	87.7
Memorial Day	60.6	0	85.6	87.5	87.5
Academy A.	85.7	85.6	0	58.7	85.7
Golden Globe A.	87.0	87.5	58.7	0	86.0
Super Bowl	87.7	87.5	86.1	86.0	0

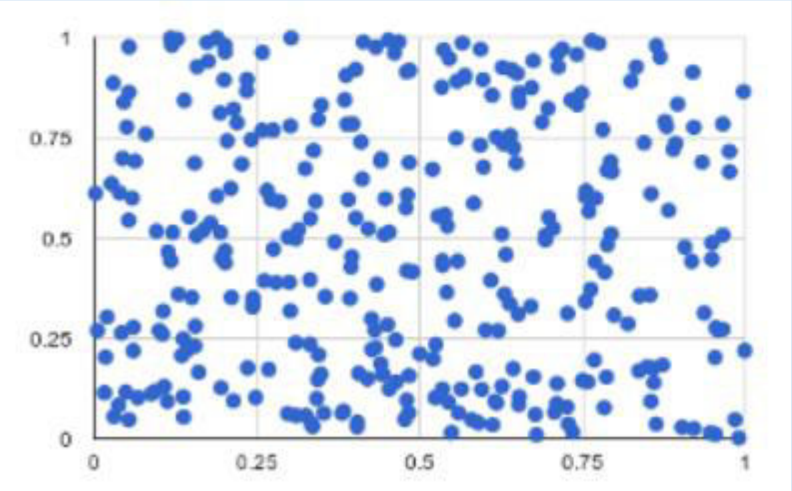
Angle or Distance

Use angle $\angle(\boldsymbol{a}, \boldsymbol{b})$ or distance $\|\boldsymbol{a} - \boldsymbol{b}\|$

Problem dependent

See if the norm of your data vectors are close to a constant or not



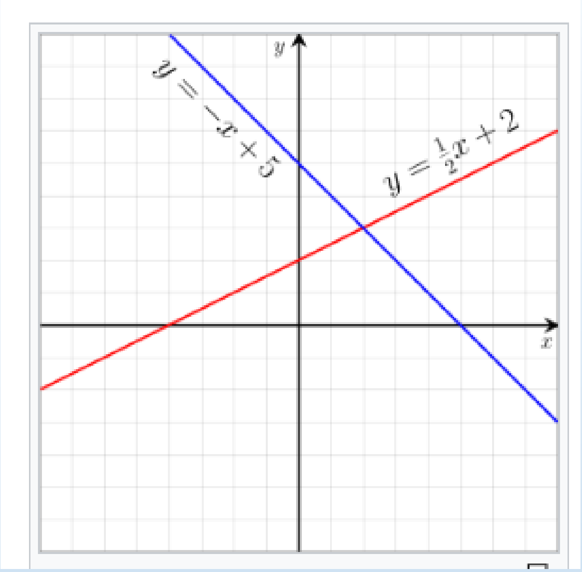


Linear function

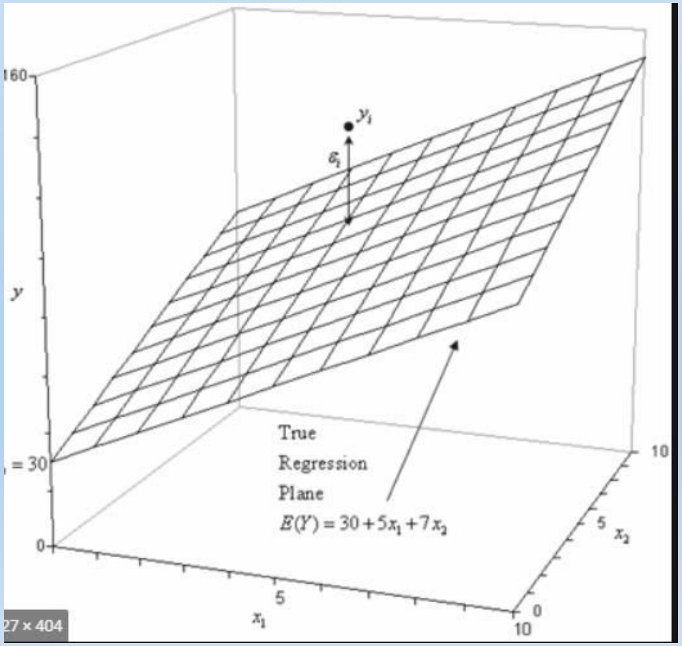
A linear function is mapping that maps a real n-vector $\boldsymbol{x}$ to a real number y.

$$y = f(\boldsymbol{x}) = a_0 + a_1 x_1 +, \dots, + a_n x_n$$

- Only have first order term of x_i's.
- $x_1, \dots, x_n$: independent variables.
- $a_0, \dots, a_n$: coefficients.



Note that we have more advanced defined of linear function in some senior courses.



Linear function

In some applications, we may define a linear using the inner product

$$y = f(\boldsymbol{x}) = a_0 + a_1 x_1 +, \dots, + a_n x_n = a_0 + \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}$$

A special form of linear function is the inner product function.

$$y = f(\boldsymbol{x}) = a_1 x_1 +, \dots, + a_n x_n = \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}$$

Superposition and linearity:

$$f(\alpha \boldsymbol{x} + \beta \boldsymbol{z}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{z})$$

for the inner product function.

Linear function

Example

$$y = f(\boldsymbol{x}) = 1 + 2x_1 + 3x_2$$

Satisfies **Superposition and linearity?**

$$y = f(\boldsymbol{x}) = 2x_1 + 3x_2$$

Satisfies **Superposition and linearity?**

Linear function

For general linear function:

$y = f(\boldsymbol{x}) = a_0 + a_1 x_1 +, \dots, +a_n x_n = a_0 + \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}$

If we have constraint on α and β: $\alpha + \beta = 1$

The we have

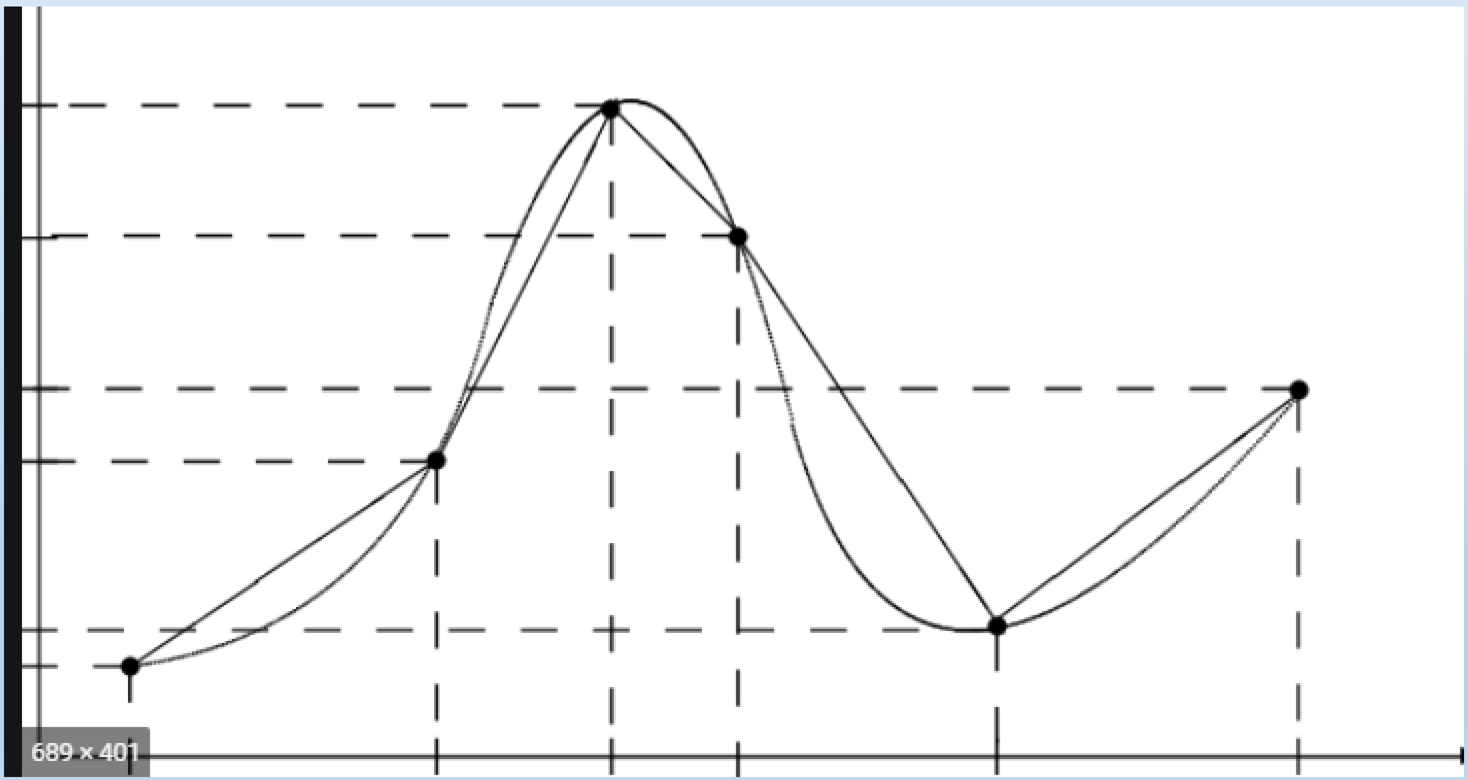
Superposition and linearity:

$$f(\alpha \boldsymbol{x} + \beta \boldsymbol{z}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{z})$$

Approximation

How about nonlinear function:

Using approximation:



Linear Regression

In machine learning or data analysis

We have many the n-vector $\boldsymbol{x}$ represents feature vectors.

$$y = f(\boldsymbol{x}) = a_0 + a_1 x_1 +, \dots, +a_n x_n = a_0 + \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}$$

a regression model. In this context,

The entries of $\boldsymbol{x}$ are called the independent variables and y is called the prediction (output), since the regression model is typically an approximation or prediction of some true value y.

Linear Regression

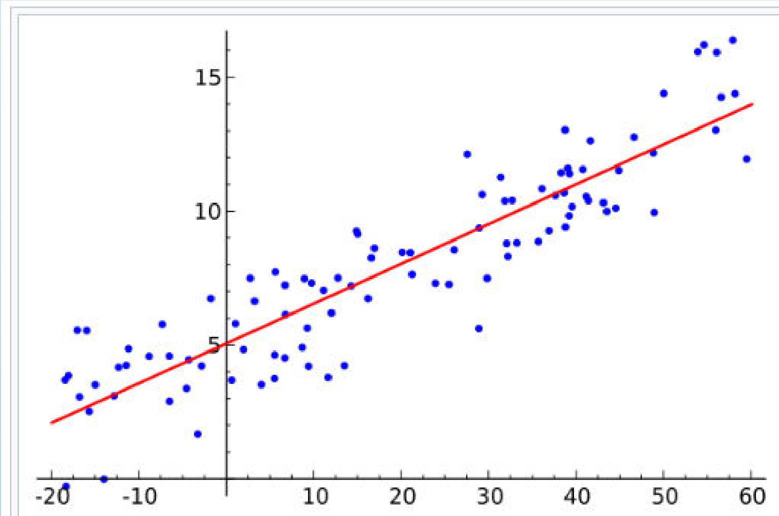
In machine learning or data analysis

Suppose we have unknown system:

We assume the system output is

$y = f(\boldsymbol{x}) = a_0 + a_1 x_1 +, \dots, + a_n x_n = a_0 + \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}$

where x_i's.



We do not know the exact equation

But we have some training pairs

$\{(y_1, \boldsymbol{x}_1), \dots, (y_N, \boldsymbol{x}_N)\}$

where $\boldsymbol{x}_i$'s are n-vectors.

Our task is to find $a_0, a_1, \dots, a_n$ based on the training pairs.

Linear Regression

House price regression model

- y : selling price of a house (thousands of dollars)
- x_1 is the house area (in 1000 square feet),
- x_2 is the number of bedrooms
- If y represents the selling price of the house, in thousands of dollars, the regression model

$$y = a_0 + a_1 x_1 + a_2 x_2$$

Collect some real data and then find a_0, a_1, a_2.

Linear Regression

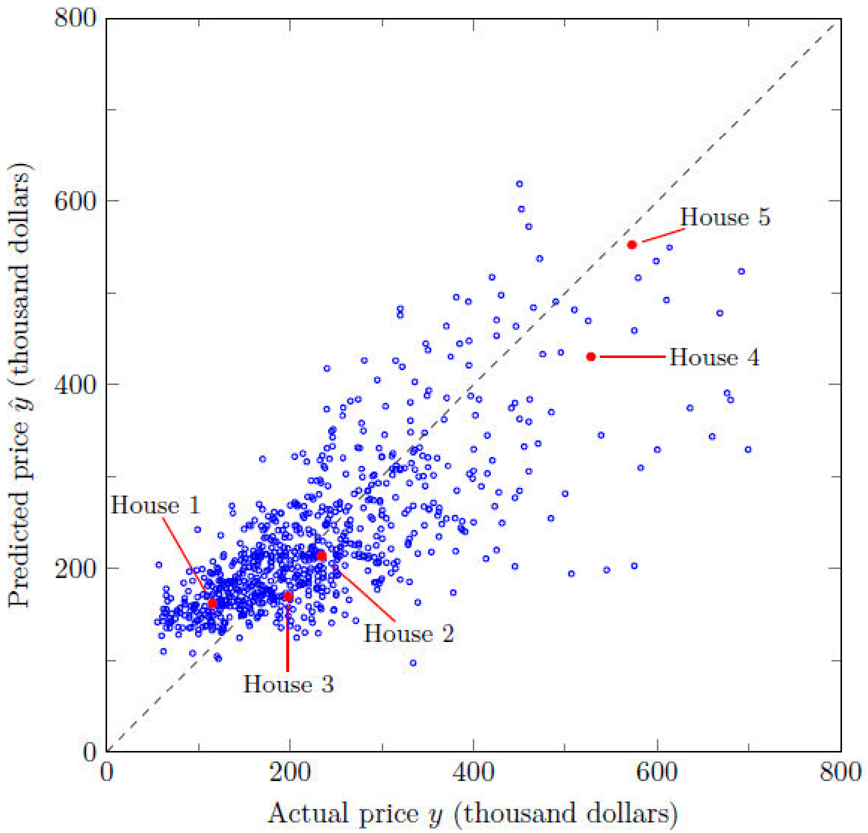


Figure 2.4 Scatter plot of actual and predicted sale prices for 774 houses sold in Sacramento during a five-day period.

House	x_1 (area)	x_2 (beds)	y (price)	$\hat{y}$ (prediction)
1	0.846	1	115.00	161.37
2	1.324	2	234.50	213.61
3	1.150	3	198.00	168.88
4	3.037	4	528.00	430.67
5	3.984	5	572.50	552.66

$a_1 = 148.73$, $a_2 = 18.85$, $a_0 = 54.40$

Clustering

Suppose we have N n-vector $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$. The goal of clustering is to group or partition the vectors (if possible) into k groups or clusters, with the vectors in each group close to each other.

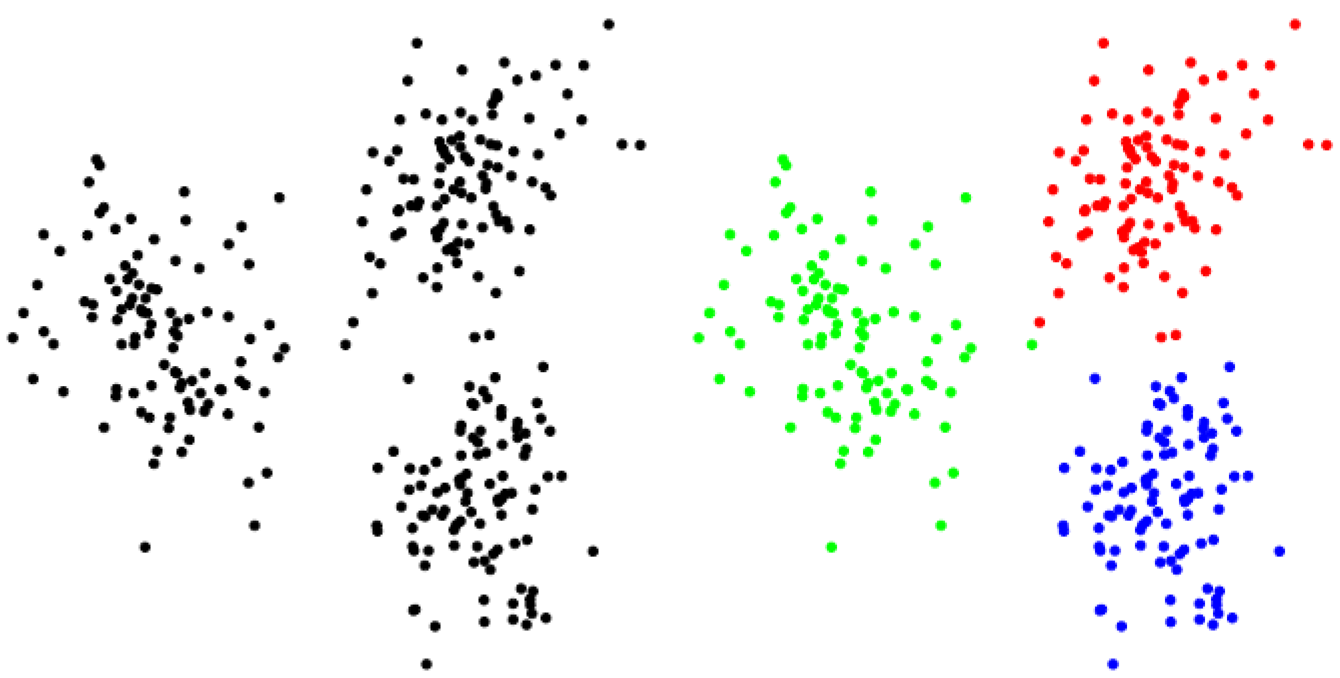


Figure 4.1 300 points in a plane. The points can be clustered in the three groups shown on the right.

Clustering Applications

Topic discovery

Suppose $\boldsymbol{x}_i$'s are word histograms associated with N documents. A clustering algorithm partitions the documents into k groups, which typically can be interpreted as groups of documents with the same or similar topics.

Patient clustering

Suppose $\boldsymbol{x}_i$'s are feature vectors associated with N patients admitted to a hospital, a clustering algorithm clusters the patients into k groups of similar patients.

Customer market segmentation

Suppose $\boldsymbol{x}_i$'s give the quantities (or dollar values) of n items purchased by customers over some period of time. A clustering algorithm will group the customers into k market segments, which are groups of customers with similar purchasing patterns.

Clustering

N n-vector $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$

The goal of clustering is to group or partition the vectors (if possible) into k groups or clusters.

Each group is represented a representative vector is

$$\boldsymbol{z}_1, \dots, \boldsymbol{z}_k$$

$$\boldsymbol{z}_j = \frac{1}{\text{number of vectors in Group j}} \sum_{\boldsymbol{x}_i \text{ in Group j}} \boldsymbol{x}_i$$

K-mean

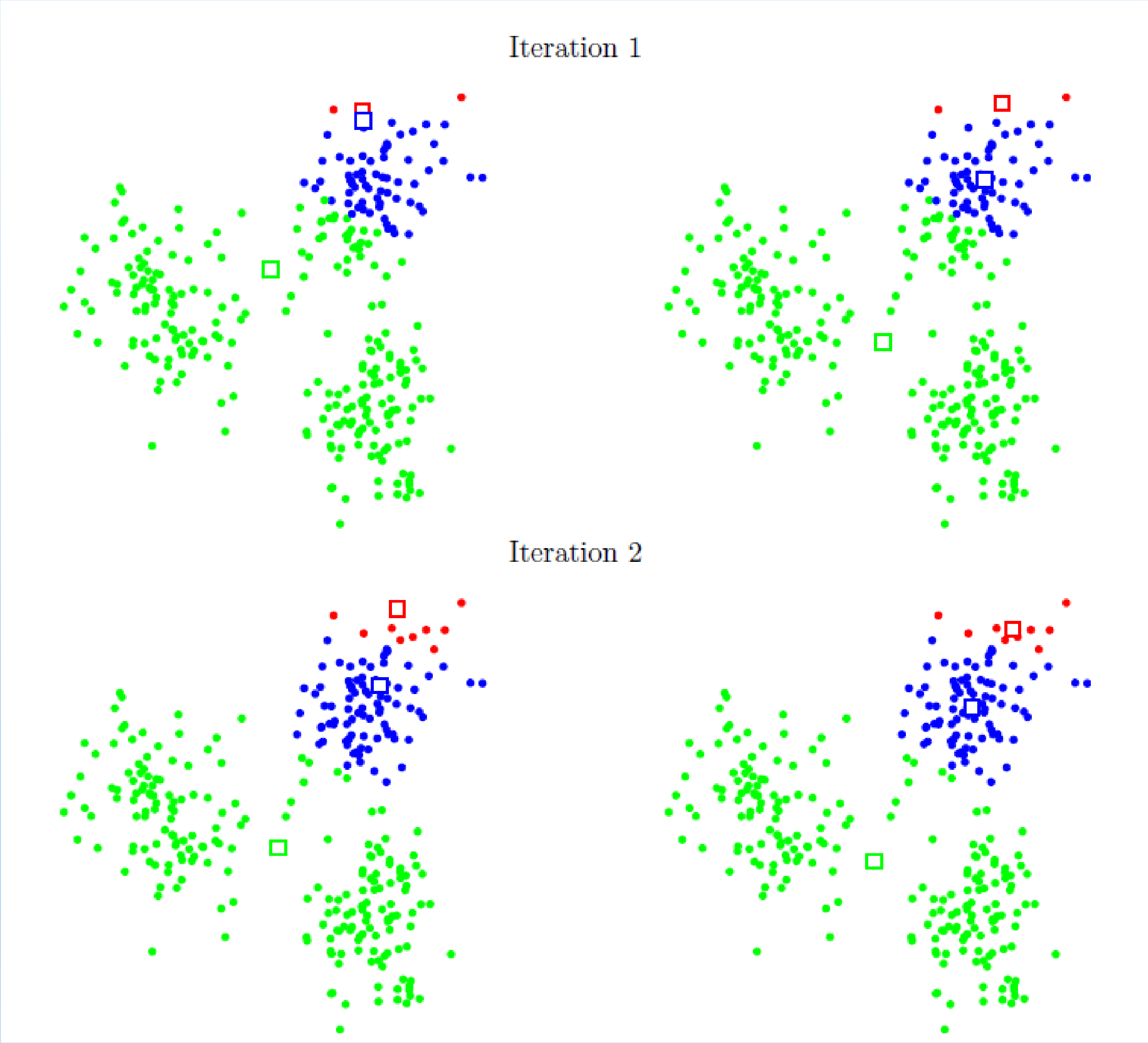
N n-vector $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ and k initial representative vectors $\boldsymbol{z}_1, \dots, \boldsymbol{z}_k$

Repeat until convergence

1. Partition the vectors into k groups. For each vector $i = 1, \dots, N$, assign to the group j if $\|\boldsymbol{x}_i - \boldsymbol{z}_j\| < \|\boldsymbol{x}_i - \boldsymbol{z}_{j'}\|$, for all j' not equal to j. That is $\boldsymbol{z}_j$ is the nearest representative.
2. Update representatives. For each group $j = 1, \dots, k$, set $\boldsymbol{z}_j$ to be the mean of the vectors in group j.

$$\boldsymbol{z}_j = \frac{1}{\text{number of vectors in Group j}} \sum_{\boldsymbol{x}_i \text{ in Group j}} \boldsymbol{x}_i$$

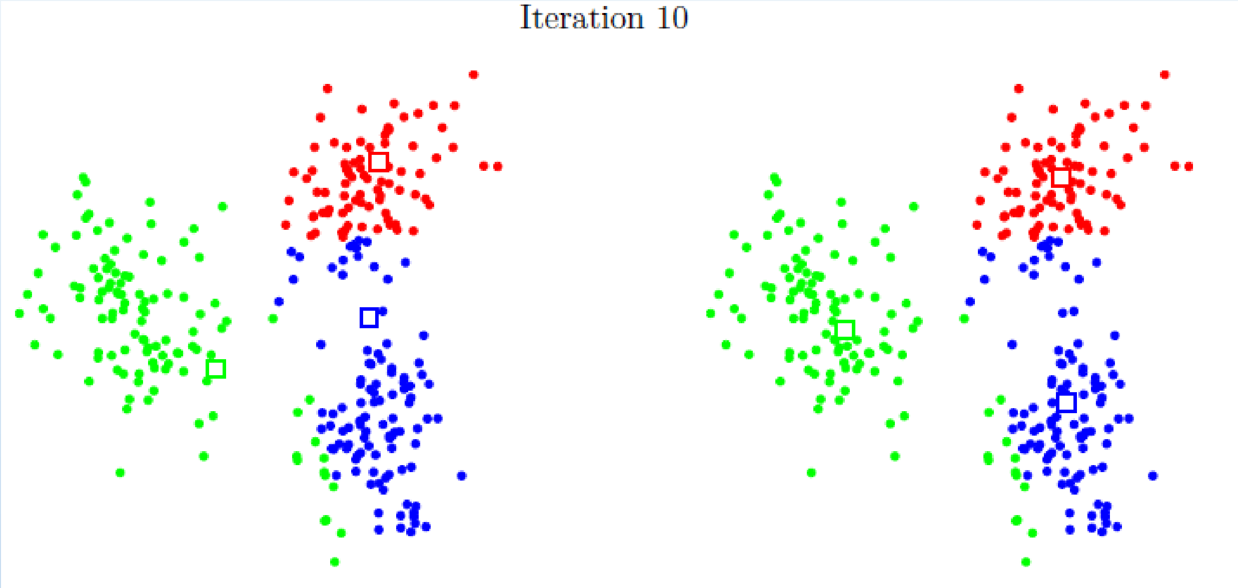
Clustering



Left: before the update

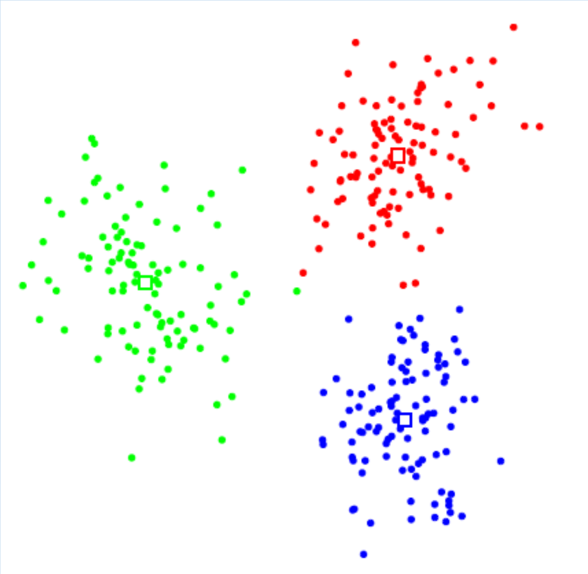
right: after the update

Clustering



Left: before the update

right: after the update



K-mean

4 vectors $\boldsymbol{x}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix}2\\1\end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix}4\\3\end{bmatrix}, \boldsymbol{x}_4 = \begin{bmatrix}5\\4\end{bmatrix}$

2 representative with initial $\boldsymbol{z}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \boldsymbol{z}_2 = \begin{bmatrix}2\\1\end{bmatrix}$

First iteration:

Partition the vectors into k groups.

Distances:

$\|\boldsymbol{x}_1 - \boldsymbol{z}_1\| = 0, \|\boldsymbol{x}_1 - \boldsymbol{z}_2\| = 1$, => $\boldsymbol{x}_1$ in Group 1

$\|\boldsymbol{x}_2 - \boldsymbol{z}_1\| = 1, \|\boldsymbol{x}_2 - \boldsymbol{z}_2\| = 0$, => $\boldsymbol{x}_2$ in Group 2

$\|\boldsymbol{x}_3 - \boldsymbol{z}_1\| = 3.6, \|\boldsymbol{x}_3 - \boldsymbol{z}_2\| = 2.83$, => $\boldsymbol{x}_3$ in Group 2

$\|\boldsymbol{x}_4 - \boldsymbol{z}_1\| = 5, \|\boldsymbol{x}_4 - \boldsymbol{z}_2\| = 4.24$, => $\boldsymbol{x}_4$ in Group 2

$$\boldsymbol{z}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \boldsymbol{z}_2 = \frac{1}{3}\left(\begin{bmatrix}2\\1\end{bmatrix} + \begin{bmatrix}4\\3\end{bmatrix} + \begin{bmatrix}5\\4\end{bmatrix}\right) = \begin{bmatrix}11/3\\8/3\end{bmatrix}$$

K-mean

Second iteration:

Partition the vectors into k groups.

$$\boldsymbol{z}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \boldsymbol{z}_2 = \begin{bmatrix} 11/3 \\ 8/3 \end{bmatrix}$$

Distances:

$\|\boldsymbol{x}_1 - \boldsymbol{z}_1\| = 0, \|\boldsymbol{x}_1 - \boldsymbol{z}_2\| = 3.14$, => $\boldsymbol{x}_1$ in Group 1

$\|\boldsymbol{x}_2 - \boldsymbol{z}_1\| = 1, \|\boldsymbol{x}_2 - \boldsymbol{z}_2\| = 2.36$, => $\boldsymbol{x}_2$ in Group 1

$\|\boldsymbol{x}_3 - \boldsymbol{z}_1\| = 3.6, \|\boldsymbol{x}_3 - \boldsymbol{z}_2\| = 0.47$, => $\boldsymbol{x}_3$ in Group 2

$\|\boldsymbol{x}_4 - \boldsymbol{z}_1\| = 5, \|\boldsymbol{x}_4 - \boldsymbol{z}_2\| = 1.89$, => $\boldsymbol{x}_4$ in Group 2

$$\boldsymbol{z}_1 = \begin{bmatrix} (1+2)/2 \\ (1+1)/2 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}, \boldsymbol{z}_2 = \begin{bmatrix} (4+5)/2 \\ (3+4)/2 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 3.5 \end{bmatrix}$$

K-mean

Third iteration:

Partition the vectors into k groups.

$$\boldsymbol{z}_1 = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}, \boldsymbol{z}_2 = \begin{bmatrix} 4.5 \\ 3.5 \end{bmatrix}$$

Distances:

$\|\boldsymbol{x}_1 - \boldsymbol{z}_1\| = 0.5$, $\|\boldsymbol{x}_1 - \boldsymbol{z}_2\| = 4.30$, =>$\boldsymbol{x}_1$ in Group 1

$\|\boldsymbol{x}_2 - \boldsymbol{z}_1\| = 0.5$, $\|\boldsymbol{x}_2 - \boldsymbol{z}_2\| = 3.54$, =>$\boldsymbol{x}_2$ in Group 1

$\|\boldsymbol{x}_3 - \boldsymbol{z}_1\| = 3.2$, $\|\boldsymbol{x}_3 - \boldsymbol{z}_2\| = 0.71$, =>$\boldsymbol{x}_3$ in Group 2

$\|\boldsymbol{x}_4 - \boldsymbol{z}_1\| = 4.61$, $\|\boldsymbol{x}_4 - \boldsymbol{z}_2\| = 0.71$, =>$\boldsymbol{x}_4$ in Group 2

No more change in $\boldsymbol{z}_1$, and $\boldsymbol{z}_2$

Vector Space

Definition: A vector space is a set $\mathbb{V}$ on which two operations + and · are defined, called vector addition and scalar multiplication.

The operation + (vector addition) must satisfy the following conditions:

Closure: If $\boldsymbol{x}$ and $\boldsymbol{z}$ are any vectors in $\mathbb{V}$, then the sum $\boldsymbol{x}+\boldsymbol{z}$ belongs to $\mathbb{V}$.

Example: $\boldsymbol{x}$ and $\boldsymbol{z}$ are 2-vectors, then $\boldsymbol{x}+\boldsymbol{z}$ in $\mathbb{V}$

Commutative law: For all vectors $\boldsymbol{x}$ and $\boldsymbol{z}$ in $\mathbb{V}$, $\boldsymbol{x}+\boldsymbol{z}$= $\boldsymbol{z}+\boldsymbol{x}$

Associative law: For all vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ in $\mathbb{V}$,

$(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$= $\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})$

Additive identity: The set $\mathbb{V}$ contains an additive identity element, denoted by **0 (zero vector)**, such that for any vector $\boldsymbol{x}$ in in $\mathbb{V}$, $\boldsymbol{x}+\mathbf{0}$= $\boldsymbol{x}$ and $\mathbf{0}+\boldsymbol{x}$= $\boldsymbol{x}$

Additive inverses: For each vector $\boldsymbol{x}$ in $\mathbb{V}$, the equations

$\boldsymbol{x}+\boldsymbol{v}=\mathbf{0}$ and $\boldsymbol{v}+\boldsymbol{x}=\mathbf{0}$ have a solution $\boldsymbol{v}$ in $\mathbb{V}$, called an additive inverse of $\boldsymbol{x}$.

Vector Space

The operation · (scalar multiplication) is defined between real numbers (or scalars) and vectors, and must satisfy the following conditions:

Closure: If $\boldsymbol{v}$ is any vector in $\mathbb{V}$, and α is any real number, then the product $\alpha\boldsymbol{v}$ belongs to $\mathbb{V}$.

Distributive law: For all real numbers α and all vectors $\boldsymbol{u}, \boldsymbol{v}$ in $\mathbb{V}$, $\alpha \cdot (\boldsymbol{u} + \boldsymbol{v}) = \alpha \cdot \boldsymbol{v} + \alpha \cdot \boldsymbol{u}$

Distributive law: For all real numbers α, β and all vectors $\boldsymbol{v}$ in $\mathbb{V}$, $(\alpha + \beta) \cdot \boldsymbol{v} = \alpha \cdot \boldsymbol{v} + \beta \cdot \boldsymbol{v}$

Associative law: For all real numbers α, β and all vectors $\boldsymbol{v}$ in $\mathbb{V}$,

$$\alpha \cdot \beta \cdot \boldsymbol{v} = (\alpha \cdot \beta) \cdot \boldsymbol{v} = \alpha \cdot (\beta \cdot \boldsymbol{v}) = \beta \cdot (\alpha \cdot \boldsymbol{v})$$

Unitary law: For all vectors $\boldsymbol{v}$ in $\mathbb{V}$, $1 \cdot \boldsymbol{v} = \boldsymbol{v}$

Vector Space Example

For the set of 2-vectors, where the elements in the vectors are real number. We define vector addition $+$ as

$$\boldsymbol{u} + \boldsymbol{v} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix},$$

We also define scalar multiplication $\cdot$ as

$$c \cdot \boldsymbol{u} = \begin{pmatrix} cu_1 \\ cu_2 \end{pmatrix}$$

Does the set of 2-vectors with the two operators define a vector space?

Vector Space Example

Given two 3-vectors, $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Define a set: $\mathbb{A} = \{\boldsymbol{x} = \alpha\boldsymbol{u} + \beta\boldsymbol{v}\}$, where α, β are any real numbers, with vector addition and scalar multiplication.

Does the set $\mathbb{A}$ with the two operators define a vector space?

Linear Dependence and Linear Independence

Linear Dependence

- A collection or list of n vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ (with $k \geq 1$) is called linearly dependent

If there are some $\beta_1, \beta_2, \beta_k$ that are not all zero such that $\beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \ldots + \beta_k \boldsymbol{x}_k = \boldsymbol{0}$.

- The zero vector as a linear combination of the vectors, with coefficients that are not all zero.
- Linear dependence of a list of vectors does not depend on the ordering of the vectors in the list.
- In other words, if any vector in $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ is a linear combination of the other vectors then the collection of vectors is linearly dependent.

Linear Dependence and Linear Independence

Linear Independence

- A collection or list of n vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ (with $k \geq 1$) is called linearly independent

  If it is not linearly dependent

- This is "$\beta_1, \beta_2, \beta_k$ are all zero" is the only solution of

$$\beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \ \ldots + \beta_k \boldsymbol{x}_k = \boldsymbol{0}.$$

- linear independence is an attribute of a collection of vectors, and not individual vectors.

Linear Dependence and Linear Independence

Examples

- A list consisting of a single vector is linearly dependent only if the vector is zero. It is linearly independent only if the vector is nonzero.
- Any list of vectors containing the zero vector is linearly dependent.

  Why?

  $$\beta_1\mathbf{0} + \beta_2\boldsymbol{x}_2 + \ \dots + \beta_k\boldsymbol{x}_k$$

  So we can select any non-zero β_1 and $\beta_2 = \cdots \beta_k = 0$, then

  $\beta_1\mathbf{0} + \beta_2\boldsymbol{x}_2 + \ \dots + \beta_k\boldsymbol{x}_k = \mathbf{0}$
- A list of two vectors is linearly dependent if and only if one of the vectors is a multiple of the other one. More generally, a list of vectors is linearly dependent if any one of the vectors is a multiple of another one.

  If multiple of another one => $\beta_1\boldsymbol{x}_1 = \beta_2\boldsymbol{x}_2$ => $\beta_1\boldsymbol{x}_1 - \beta_2\boldsymbol{x}_2 = \mathbf{0}$

  If not multiple of another one => $\beta_1\boldsymbol{x}_1 \neq \beta_2\boldsymbol{x}_2$ => "$\beta_1 = \beta_2 = 0$" is the only solution for $\beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 = \mathbf{0}$

Linear Dependence and Linear Independence

Examples

- Show that $\boldsymbol{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{a}_3 = \begin{pmatrix} 4 \\ 5 \\ 1 \end{pmatrix}$ are linearly dependent.

  since $2\boldsymbol{a}_1 + \boldsymbol{a}_2 - \boldsymbol{a}_3 = \boldsymbol{0}$ **=>** linearly dependent

- Show that $\boldsymbol{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$, are linearly independent.

Suppose $\beta_1 \boldsymbol{a}_1 + \beta_{\boldsymbol{2}} \boldsymbol{a}_2 = \boldsymbol{0}$ **=>** $\beta_1 = 0$ **and** $-\beta_{\boldsymbol{2}} = 0$

- Show that $\boldsymbol{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{a}_3 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ are linearly independent.

Suppose $\beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2 + \beta_3 \boldsymbol{a}_2 = \boldsymbol{0}$ **=>**

$\beta_1 - \beta_3 = 0$ **,** $-\beta_{\boldsymbol{2}} + \beta_3 = 0,\ \beta_2 + \beta_3 = 0$

Adding the last two equations, => $\beta_3 = 0$

Hence $\beta_1 = 0$ and $\beta_2 = 0$

Linear Dependence and Linear Independence

Examples

The standard unit n-vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are linearly independent.

Suppose:

$$\beta_1 \boldsymbol{e}_1 + \beta_2 \boldsymbol{e}_2 + \ \ldots + \beta_n \boldsymbol{e}_n = \boldsymbol{0}$$

=> $\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \boldsymbol{0}$

=> $\beta_i = 0$ for all i.

Linear Dependence and Linear Independence

Examples

The standard unit n-vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are linearly independent.

Suppose:

$$\beta_1 \boldsymbol{e}_1 + \beta_2 \boldsymbol{e}_2 + \ldots + \beta_n \boldsymbol{e}_n = \boldsymbol{0}$$

$$\Rightarrow \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \boldsymbol{0}$$

=> $\beta_i = 0$ for all i.

Linear combinations of linearly independent vectors

Suppose a vector $\boldsymbol{x}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$, This is $\boldsymbol{x} = \beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2, \ldots, +\beta_k \boldsymbol{a}_k$.

When the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ are linearly independent, the coefficients that form $\boldsymbol{x}$ are unique.

Proof:

Suppose there are another set of coefficients

$\boldsymbol{x} = \gamma_1 \boldsymbol{a}_1 + \gamma_2 \boldsymbol{a}_2, \ldots, +\gamma_k \boldsymbol{a}_k$.

$\Rightarrow \mathbf{0} = (\gamma_1 - \beta_1)\boldsymbol{a}_1 + (\gamma_2 - \beta_2)\boldsymbol{a}_2, \ldots, +(\gamma_k - \beta_k)\boldsymbol{a}_k$.

Since $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ are linearly independent,

$(\gamma_i - \beta_i) = 0$ for all i. Contradiction for another set of coefficients.

Span and Basis

Given a set of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$, one create a vector space (subspace) by forming all linear combinations of that set of vectors.

The **span** of the set of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$is the vector space consisting of all linear combinations of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$.

Example:

$$\boldsymbol{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{a}_3 = \begin{pmatrix} -2 \\ -3 \\ 0 \end{pmatrix}$$

We can use these vectors to define a subspace $\mathbb{V}$. That is

$\boldsymbol{x} = \gamma_1 \boldsymbol{a}_1 + \gamma_2 \boldsymbol{a}_2 + \gamma_3 \boldsymbol{a}_k$ for all possible real $\gamma_1, \gamma_2, \gamma_3$.

$\mathbb{V} = \{\boldsymbol{x} = \gamma_1 \boldsymbol{a}_1 + \gamma_2 \boldsymbol{a}_2 + \gamma_3 \boldsymbol{a}_k$ for all possible real $\gamma_1, \gamma_2, \gamma_3 .\}$

Example:

In the previous example, we do not need to use all the

three vectors to define the subspace $\mathbb{V}$ because the three vectors are linearly dependent.

The smallest set of vectors needed to span this vector space $\mathbb{V}$ forms a basis for $\mathbb{V}$.

$$\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix}\right\},\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}-2\\-3\\0\end{pmatrix}\right\},\left\{\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}-2\\-3\\0\end{pmatrix}\right\}$$

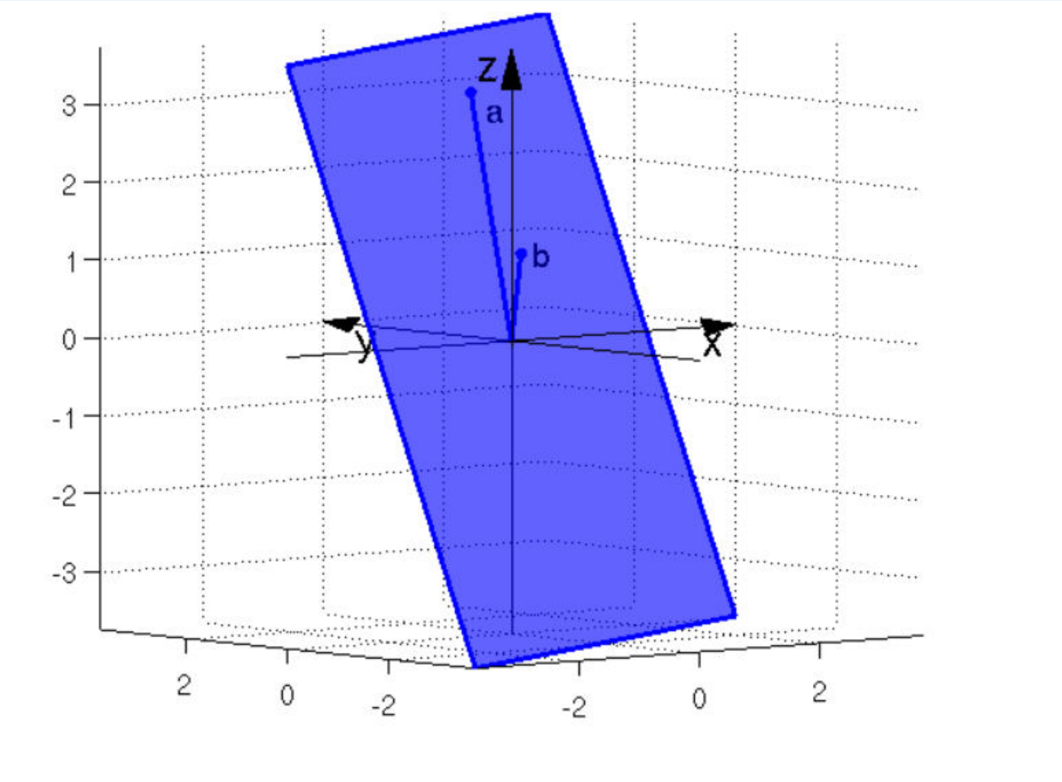
We have three possible choices (Three base)

Each of the choices spans the vector space $\mathbb{V}$

- Basis: A collection of linearly independent vectors is called a basis.

Example:

- Basis: A collection of linearly independent vectors is called a basis.



Two 3-vectors span a subspace (2D).

Span and basis

Suppose $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$, are linearly independent.

Let $\mathbb{V} = \{\boldsymbol{x} = \beta_1\boldsymbol{a}_1 + \beta_2\boldsymbol{a}_2, \ldots, +\beta_k\boldsymbol{a}_k, \text{for all possible real } \beta_i$'s}

be the space spanned by these vectors.

Prove that any vector in $\mathbb{V}$ is written in a unique way as a linear combination $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$.

Proof:

Let $\boldsymbol{x} = \beta_1\boldsymbol{a}_1 + \beta_2\boldsymbol{a}_2, \ldots, +\beta_k\boldsymbol{a}_k$. Suppose there are another set of coefficients: $\boldsymbol{x} = \gamma_1\boldsymbol{a}_1 + \gamma_2\boldsymbol{a}_2, \ldots, +\gamma_k\boldsymbol{a}_k$.

$\Rightarrow \mathbf{0} = (\gamma_1 - \beta_1)\boldsymbol{a}_1 + (\gamma_2 - \beta_2)\boldsymbol{a}_2, \ldots, +(\gamma_k - \beta_k)\boldsymbol{a}_k$.

Since $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ *are linearly independent,*

$(\gamma_i - \beta_i) = 0$ *for all* i. *Contradiction for another set of coefficients.*

Span and basis

Examples:

- The n standard unit n-vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ form a basis. Any n-vector $\boldsymbol{x}$ can be written as the linear combination of $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$.
- $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}, \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}-2\\-3\\0\end{pmatrix}\right\}, \left\{\begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}-2\\-3\\0\end{pmatrix}\right\}$

The above three sets of two vectors form the same subspace. Which space is better?

Of course, from computational point of view

$\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$. This is because this set is orthogonal.

Orthogonal and Orthonormal

- A collection of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ is orthogonal or mutually orthogonal, if $(\boldsymbol{a}_i)^{\mathrm{T}}\boldsymbol{a}_j = 0$ for any i, j and $i \neq j$. (The angle of between $\boldsymbol{a}_j$ and $\boldsymbol{a}_i$ is $\frac{\pi}{2}$ (90°))
- A collection of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ is orthonormal, if it is orthogonal and $(\boldsymbol{a}_i)^{\mathrm{T}}\boldsymbol{a}_i = 1$.

- $(\boldsymbol{a}_i)^{\mathrm{T}}\boldsymbol{a}_j = \begin{cases} 0, \text{for } i \neq j \\ 1, \text{for } i = j \end{cases}$

Orthogonal and Orthonormal

Examples:

- The standard unit n -vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ are orthonormal.

- $$\boldsymbol{a}_1 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{a}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{a}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

are orthonormal.

Orthogonal and Orthonormal

Orthonormal vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ in $\mathbb{R}^n$ in n dimensional space are linearly independent.

Proof

Consider

$$\beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2, \dots, + \beta_k \boldsymbol{a}_n = \boldsymbol{0}$$

Taking the inner product of this equality with $\boldsymbol{a}_i$ yields

$$\beta_1 (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{a}_1 + \beta_2 (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{a}_2, \dots, + \beta_k (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{a}_n = (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{0}$$
$$\beta_i (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{a}_i = 0$$
$$\beta_i = 0$$

for all i.

That means $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ form a basis.

Orthogonal and Orthonormal

Given a n –vector $\boldsymbol{x}$ and a set of orthonormal vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$.

The vector $\boldsymbol{x}$ can be written as a linear combination of the orthonormal vectors. And the coefficients are unique.

$$\boldsymbol{x} = \beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2, \ldots, + \beta_n \boldsymbol{a}_n$$

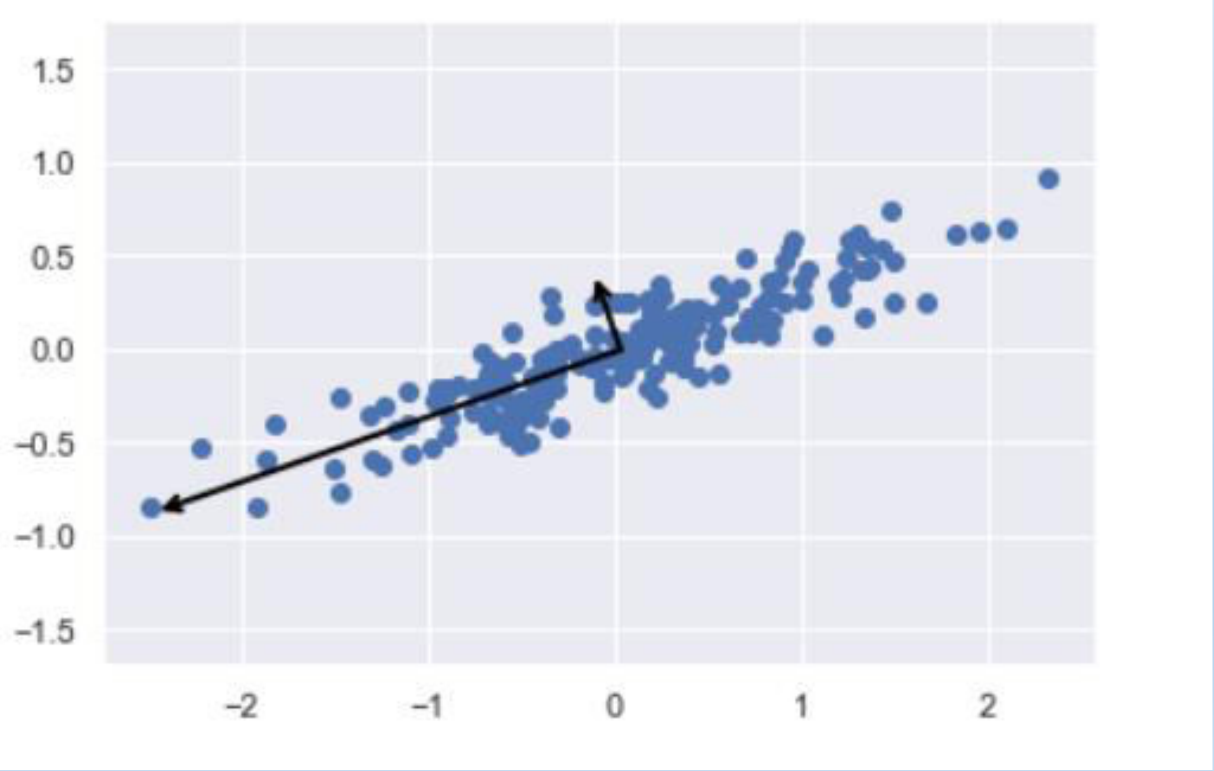
Taking the inner product of this equality with $\boldsymbol{a}_i$ yields

$$\beta_i (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{a}_i = (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{x}$$

$$\beta_i = (\boldsymbol{a}_i)^{\mathrm{T}} \boldsymbol{x}$$

Applications

Given many 2D data points, if we the original basis, we need two elements to represent a data. However, if we can find another basis, then a data point can be approximated by one element only.



Gram-Schmidt algorithm

Given a set of linearly independent vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$, the Gram-Schmidt algorithm produces an orthonormal collection of vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_k$.

Gram-Schmidt algorithm

Given n-vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$

$\boldsymbol{q}_1 = \boldsymbol{a}_1 / \|\boldsymbol{a}_1\|$

For $i = 2, \dots, k$

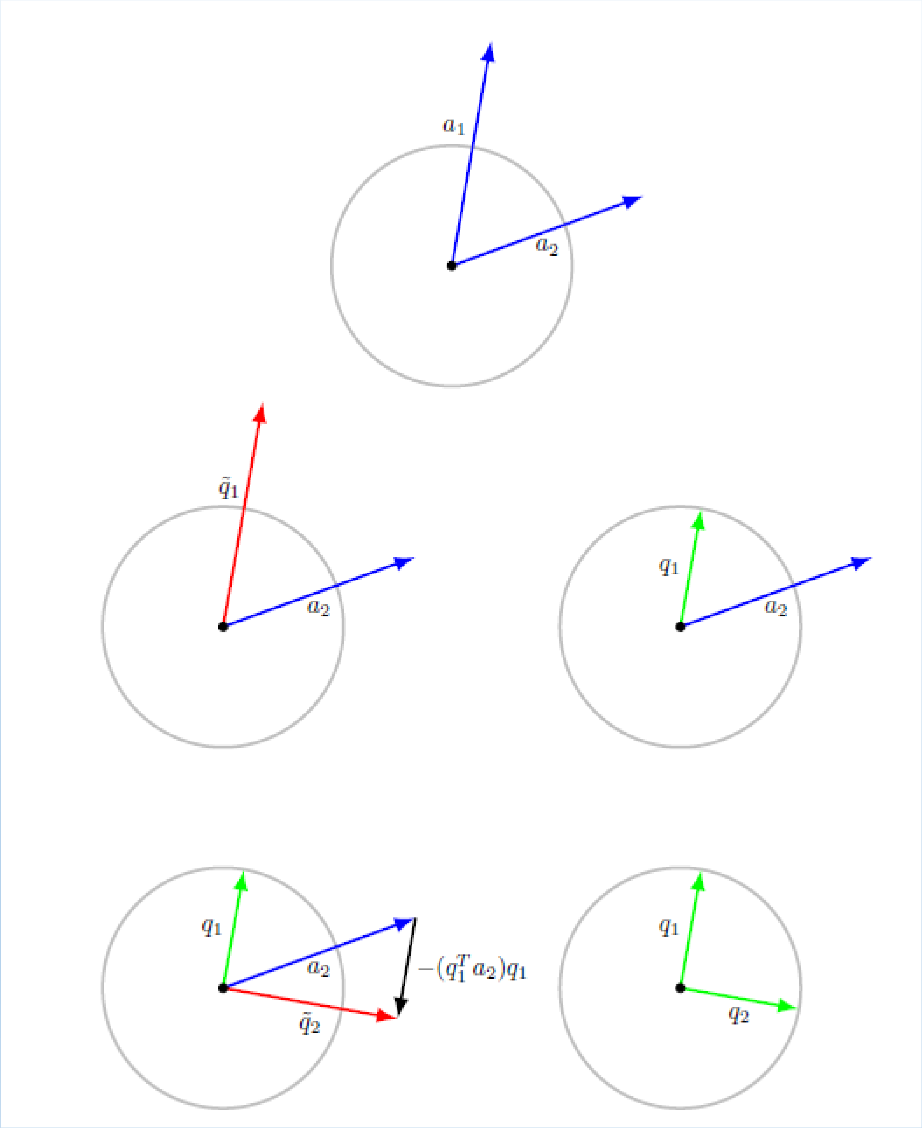
1. Orthogonalization:

$$\tilde{\boldsymbol{q}}_i = \boldsymbol{a}_i - (\boldsymbol{q}_1^T \boldsymbol{a}_i)\boldsymbol{q}_1 - \cdots \left(\boldsymbol{q}_{i-1}^T \boldsymbol{a}_i\right)\boldsymbol{q}_{i-1}$$

2. If $\tilde{\boldsymbol{q}}_i$=**0, quit (**$\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ are dependent)

3 $\boldsymbol{q}_i = \tilde{\boldsymbol{q}}_i / \|\tilde{\boldsymbol{q}}_i\|$

Example



$(\boldsymbol{q}_1^T \boldsymbol{a}_2)\boldsymbol{q}_1$ is the component of $\boldsymbol{a}_2$ along $\boldsymbol{q}_1$. (project $\boldsymbol{a}_2$ to $\boldsymbol{q}_1$)

After we subtract this component, $\boldsymbol{a}_2 - (\boldsymbol{q}_1^T \boldsymbol{a}_2)\boldsymbol{q}_1$ does not have any component along $\boldsymbol{q}_1$ so $\boldsymbol{q}_2$ is orthonormal to $\boldsymbol{q}_1$

Gram-Schmidt algorithm

Gram-Schmidt algorithm

Given n-vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$

$\boldsymbol{q}_1 = \boldsymbol{a}_1 / \|\boldsymbol{a}_1\|$

For $i = 2, \dots, k$

1. Orthogonalization:

$$\tilde{\boldsymbol{q}}_i = \boldsymbol{a}_i - (\boldsymbol{q}_1^T \boldsymbol{a}_i)\boldsymbol{q}_1 - \cdots (\boldsymbol{q}_{i-1}^T \boldsymbol{a}_i)\boldsymbol{q}_{i-1}$$

$(\boldsymbol{q}_1^T \boldsymbol{a}_i)\boldsymbol{q}_1 + \cdots (\boldsymbol{q}_{i-1}^T \boldsymbol{a}_i)\boldsymbol{q}_{i-1}$ is the component of $\boldsymbol{a}_i$ in the subspace formed by $\boldsymbol{q}_1$ to $\boldsymbol{q}_{i-1}$.

Now $\tilde{\boldsymbol{q}}_i$ does not have any components along $\boldsymbol{q}_1$ to $\boldsymbol{q}_{i-1}$)

2. If $\tilde{\boldsymbol{q}}_i$=**0, quit (**$\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ are dependent)

3 $\boldsymbol{q}_i = \tilde{\boldsymbol{q}}_i / \|\tilde{\boldsymbol{q}}_i\|$ normaization

Gram-Schmidt algorithm

Example:

$$\boldsymbol{a}_1 = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} -1 \\ 3 \\ -1 \\ 3 \end{pmatrix}, \boldsymbol{a}_3 = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \end{pmatrix}$$

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|} = \begin{pmatrix} -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{pmatrix}$$

$$\boldsymbol{q}_1^T\boldsymbol{a}_2 = 4,\ \tilde{\boldsymbol{q}}_2 = \boldsymbol{a}_2 - (\boldsymbol{q}_1^T\boldsymbol{a}_2)\boldsymbol{q}_1 = \begin{pmatrix} -1 \\ 3 \\ -1 \\ 3 \end{pmatrix} - 4\begin{pmatrix} -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$$

$$\boldsymbol{q}_1^T\boldsymbol{a}_3 = 2, \boldsymbol{q}_2^T\boldsymbol{a}_3 = 8,,\ \tilde{\boldsymbol{q}}_3 = \boldsymbol{a}_3 - (\boldsymbol{q}_1^T\boldsymbol{a}_3)\boldsymbol{q}_1 - (\boldsymbol{q}_2^T\boldsymbol{a}_3)\boldsymbol{q}_2$$

$$\tilde{\boldsymbol{q}}_3 = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \end{pmatrix} - 2\begin{pmatrix} -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{pmatrix} - 8\begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{q}_3 = \begin{pmatrix} -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$$

Floating point operations

- When computers carry out addition, subtraction, multiplication, division, or other arithmetic operations on numbers represented in floating point format, the result is rounded to the nearest floating point number. For 64-bit, the number of significant decimal digits (after the decimal point) is around 11 ($1.3333333333x10^{10}$). For 32-bit, the number of significant decimal digits is around 4-5 after the decimal point ($1.3333x10^{10}$).
- These operations create very small error in the computed result is called (floating point) round-off error. In most applications, these very small errors have no practical effect.
- But you add a lot of number together with different ranges, your results may have large error.
- How to design error-aware algorithm?

Gram-Schmidt algorithm

Flop counts and complexity

- Operations, likes scalar multiplication, vector addition, and the inner product are elementary operation.
- How quickly these operations can be carried out by a computer depends very much on the computer hardware and software, and the size of the vector.
- In algorithm design, we should reduce the number of operations (save time, save energy, and save $$$).