Lecture 02

Vector II: Norm and Inner Product

Instructor: Ruoyu Sun

香港中文大學(深圳) | 数据科学学院
The Chinese University of Hong Kong, Shenzhen | School of Data Science

Recall

In the last lecture …

- Definition of (column) vectors

scalar $\quad cV+wd = (\cdot \; - \;)$

- Basic vector operations (addition, multiplication, linear combination)

$V+W \qquad \alpha \cdot V = (\alpha_1 v_1, \dots, \alpha_n v_n)$

Learning Goals Today

- **Today**: **More** vector operations

[Sec 1.2 of textbook, 5th edition]
Default: 5th edition textbook

- Outline: (1) Vector Length; (2) Inner Product.

After this lecture, you should be able to:

—calculate the norm of a vector

—calculate the inner product of two vectors

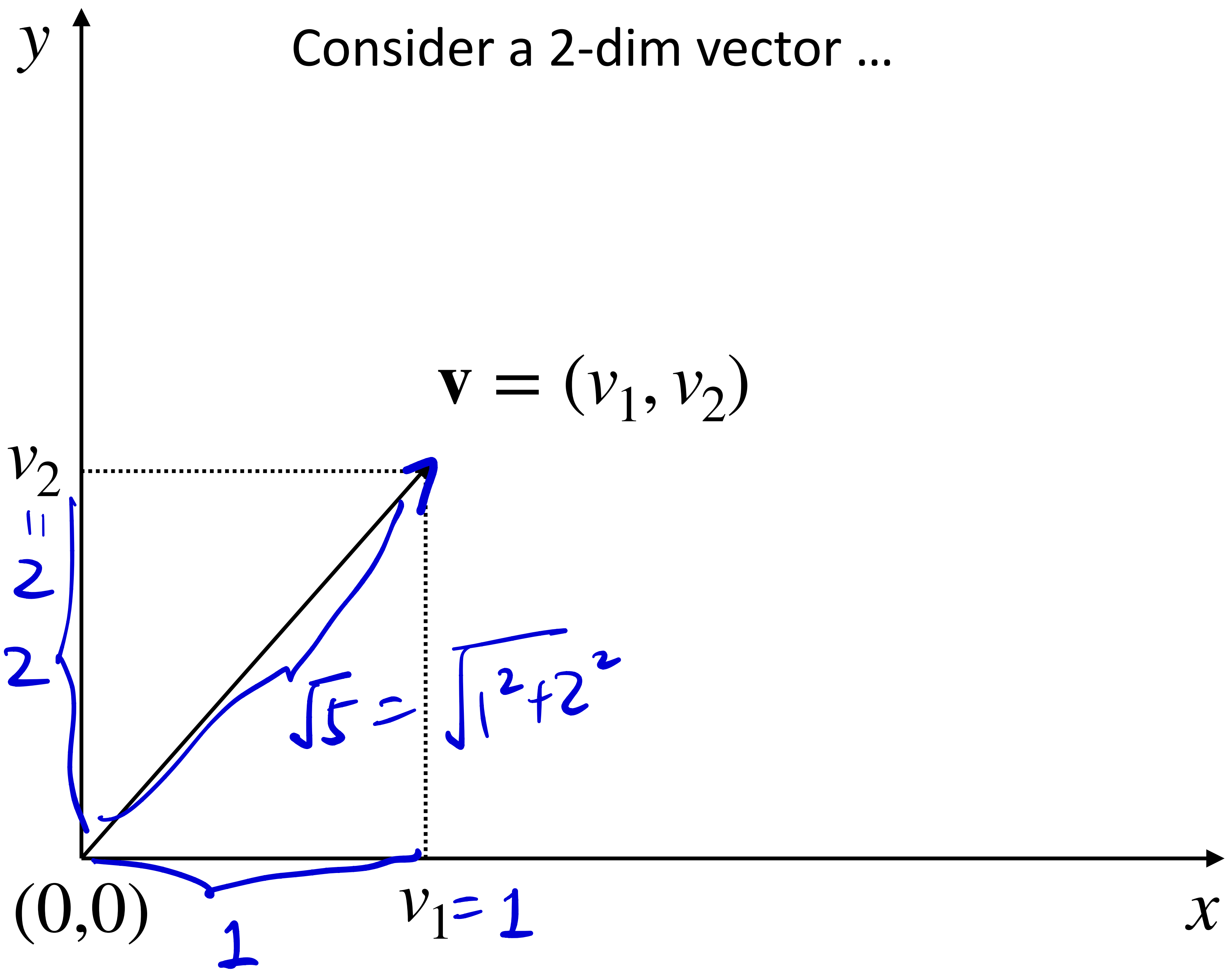
—utilize Cauchy-Schwartz inequality and triangular inequality

—provide real-world applications of "inner product”

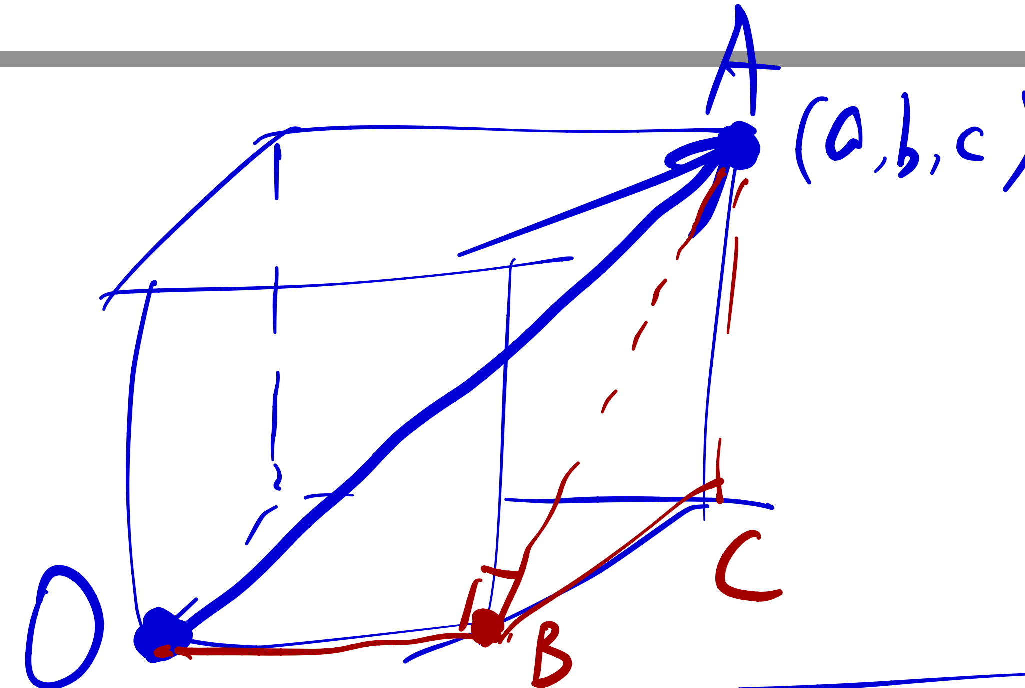
Part I Length and Norm

Vector Lengths

Consider a 2-dim vector …

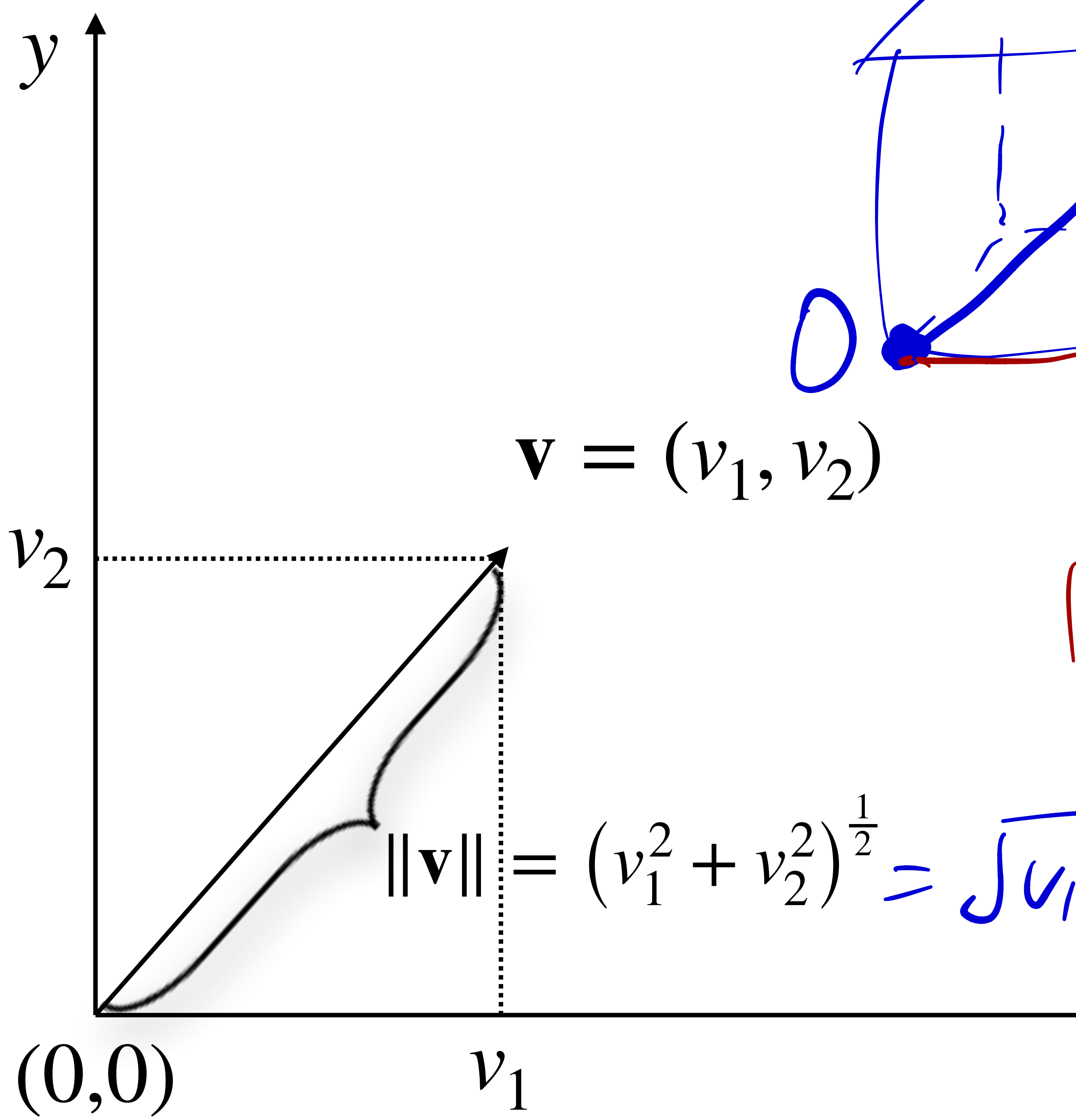


Vector Lengths



$$|OA| = \sqrt{a^2+b^2+c^2}$$

$$|OA|^2 = |OB|^2 + |BC|^2 + |CA|^2$$
$$= a^2+b^2+c^2$$



$$\mathbf{v} = (v_1, v_2)$$

$$\|\mathbf{v}\| = \left(v_1^2 + v_2^2\right)^{\frac{1}{2}} = \sqrt{v_1^2+v_2^2}$$

Vector Lengths

For higher dimensional vector …

A generalized notion of "*vector length*" …

Definition (ℓ_2-norm) 范数.

Let $\mathbf{v} = (v_1, \ldots, v_n)$ Be a vector. The ℓ_2**-norm** of $\mathbf{v}$, denoted by $\|\mathbf{v}\|_2$, is defined as

$$\triangleq$$

$$\|\mathbf{v}\|_2 := \left(v_1^2 + \cdots + v_n^2\right)^{\frac{1}{2}}$$

is defined as (appear in the first time).

Vector Lengths

For higher dimensional vector …

A generalized notion of "*vector length*" …

Definition (ℓ_2-norm)

Let $\mathbf{v} = (v_1, \ldots, v_n)$ be a vector. The ℓ_2**-norm** of $\mathbf{v}$, denoted by $\|\mathbf{v}\|_2$ is defined as

$$\|\mathbf{v}\|_2 := \left(v_1^2 + \cdots + v_n^2\right)^{\frac{1}{2}}$$

$\|v\| = \|v\|_2$

Vector Lengths

For higher dimensional vector …

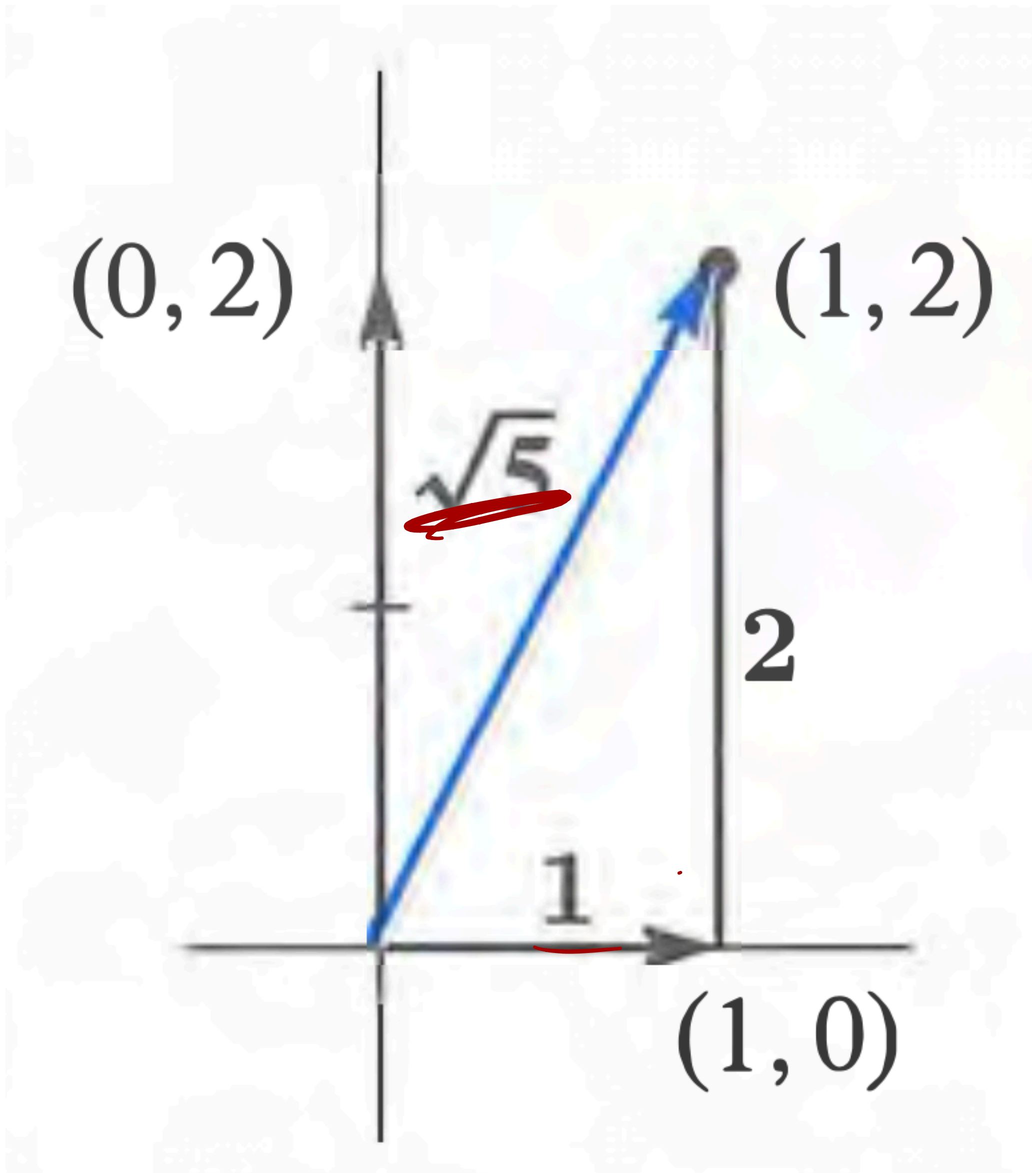
A generalized notion of "*vector length*" …

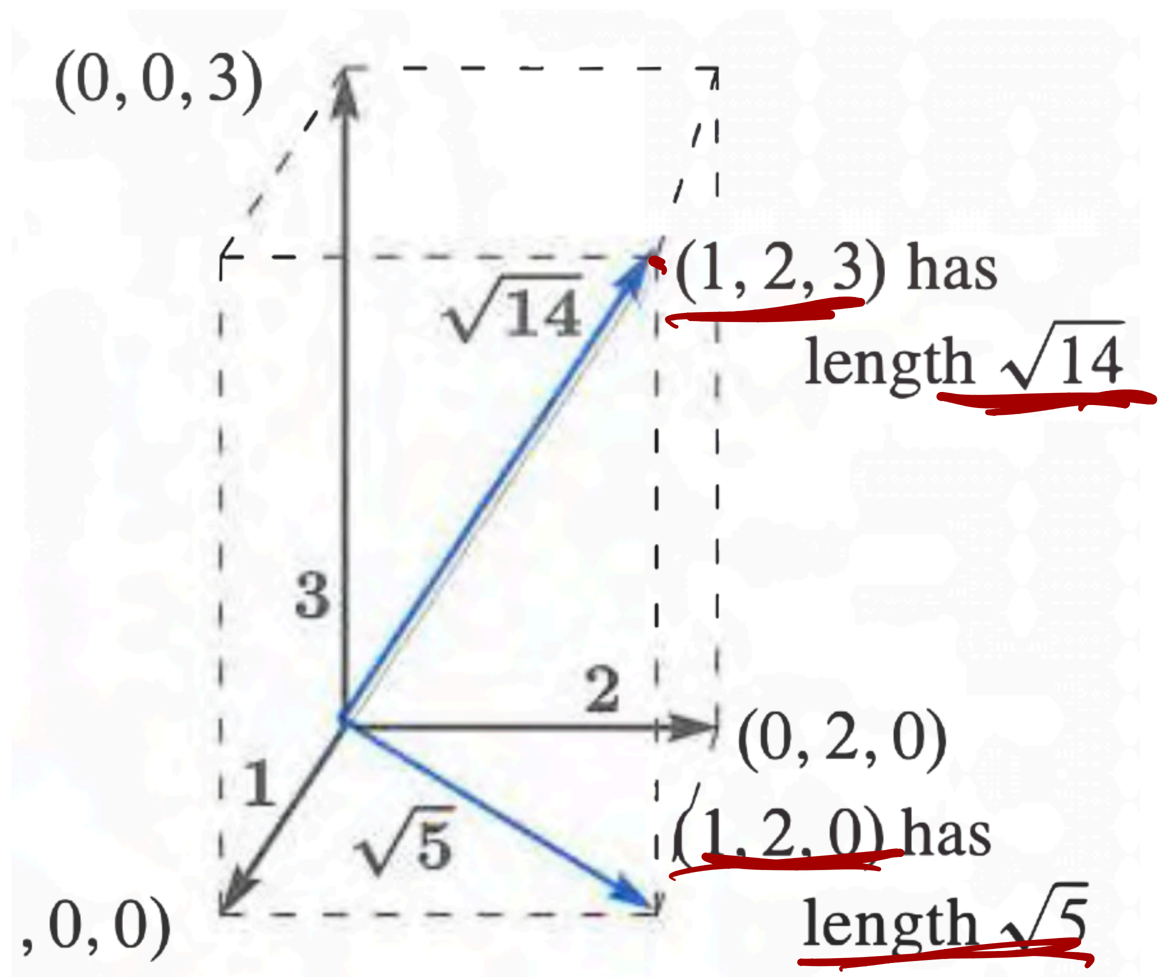
欧几里德.

The ℓ_2-norm is also called an *Euclidean norm* or *length*

We often abbreviate $||\cdot||_2$ as $||\cdot||$

Vector Lengths



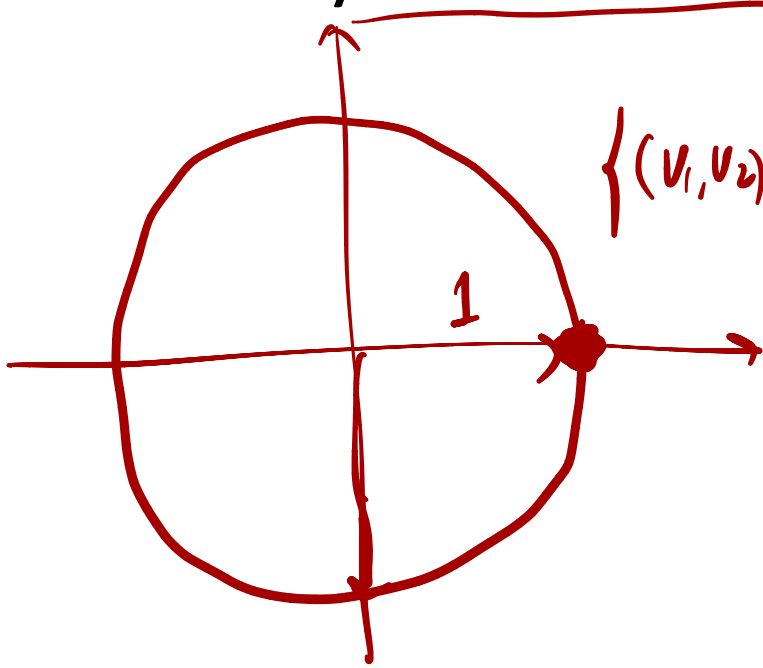


Unit Vector: extension of number 1 to vector

Definition (Unit Vector)

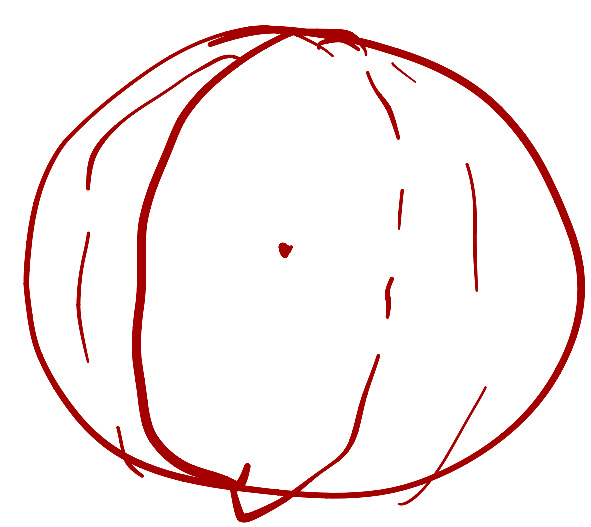
A vector $\mathbf{v}$ is called a **unit vector** if $\|\mathbf{v}\| = 1$

For any non-zero vector $\mathbf{v}$, $\dfrac{\mathbf{v}}{\|\mathbf{v}\|}$ is a unit vector



$\{(v_1, v_2) : \|\mathbf{v}\| = 1\}$

3-dim space



sphere
球面.
(not ball).

Practice. Prove $\frac{u}{\|u\|}$ is unit vector, if $u \neq 0$.

30% get correct proof.

Wrong proof. 1 $\quad \underbrace{u = \|u\|}_{\text{vector} \quad \text{scalar}}$, so $\frac{u}{\|u\|} = \frac{u}{u} = 1$.

only true if u is scalar and $u \neq 0$.

Incomplete proof $\quad$ Let $\alpha = \|u\|$ ①

$$\left\|\frac{u}{\|u\|}\right\| \overset{①}{=} \left\|\frac{u}{\alpha}\right\| \overset{?}{=} \frac{\|u\|}{\alpha} = 1$$

take it outside

major step

$$\left|\frac{u}{\alpha}\right| = \frac{|u|}{\alpha}$$

General Advice for "proof"

* **Use Definition.**

Want to prove $\frac{u}{\|u\|} = v$ is <u>unit vector</u>

$$\iff \quad \|v\| = 1.$$

$$\overset{\text{Def of norm}}{\iff} \quad \sqrt{v_1^2 + \cdots + v_n^2} = 1.$$

figure out what to prove by definition.

Define $\alpha = \|u\|$, then $v = \frac{u}{\alpha} = \underset{\text{scalar}}{\frac{1}{\alpha}} \cdot \underset{\text{vec}}{u}$

Next, $v = \left(\frac{1}{\alpha}u_1, \frac{1}{\alpha}u_2, \dots, \frac{1}{\alpha}u_n\right)$.

Thus $\|u\| = \sqrt{\left(\frac{1}{\alpha}u_1\right)^2 + \cdots + \left(\frac{1}{\alpha}u_n\right)^2}$

$$= \frac{\sqrt{u_1^2 + \cdots + u_n^2}}{\sqrt{\alpha^2}} = \frac{\|u\|}{|\alpha|} \overset{\alpha = \|u\|}{=} 1.$$

Property For scalar α, vector v,

$$\|\alpha v\| = |\alpha| \cdot \|v\|.$$

Unit Vector: Examples

Examples (Unit Vector)

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$$

$$\begin{pmatrix} 1/\sqrt{2} \\ -1/2 \\ 1/2 \end{pmatrix}$$

Part II Inner Product

Vector Operations

Question1: Can you think about any other vector operations?

Vector Operations

Question1: Can you think about any other vector operations?

Vector Addition $$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}$$

Scalar · vector $\underset{\text{scalar}}{\alpha} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix}$.

$$\boxed{\text{Q: What should "vector times vector" be? How to DEFINE?}}$$

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

≥ 3 answers.

not a formal notation

Answer 1 $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \otimes \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} v_1 w_1 \\ v_2 w_2 \end{pmatrix}$

Answer 2 $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \otimes \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \|v\| \cdot \|w\|$

Answer 3 $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = v_1 v_2 w_1 w_2$

Answer 4 $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = v_1 w_1$

~~Answer 5~~ $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \sin(v_1 w_1) \cdot e^{v_2 + w_2}$

$n=1$ should be consistent with product.

~~What is correct?~~

What is GOOD ?

$n=1$, $v_1 \cdot w_1 = \sin(v_1 w_1)$, not consistent

Notation

use this one

$\|v\|$ v.s. $|v|$ → extending def for $|v|$ absolute value

Both are fine.

→ remind you, v is vector

arrow helps remind you it's an arrow

Write $\vec{v}$ or v, are both fine.

Vector Operations

Question1: Can you think about any other vector operations?

Vector Addition $\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}$

How about $\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 w_1 \\ v_2 w_2 \\ \vdots \\ v_n w_n \end{bmatrix}$?

This is "point-wise product".

Dot Product

Definition (Dot Product)

A *dot product* between two vectors (of the same size)

$$\mathbf{v} = (v_1, \ldots, v_n)$$
$$\mathbf{w} = (w_1, \ldots, w_n)$$

is defined as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i w_i$$

(v_1, v_2)
(w_1, w_2)
$\langle v, w \rangle = v_1 w_1 + v_2 w_2$
$v \cdot w$

Example: Calculate the dot product of (-1,2,2) and (1,0,-3)

$= (-1) \cdot 1 + 2 \cdot 0 + 2 \cdot (-3)$

$= -7$

Dot Product is also called Inner Product

~~don't write $v \times w$~~

Sometime we also write a dot product as $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

Recommend writing as $\langle \mathbf{v}, \mathbf{w} \rangle$ to avoid confusion.

A *dot product* is called an *inner product* in more general settings

内积

commutative rule

Property: $\langle \mathbf{v}, \mathbf{v} \rangle = \|v\|^2$ and $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ $\quad \sum_i v_i w_i = \sum_i w_i v_i$

$v \cdot v = v_1 \cdot v_1 + \cdots + v_n \cdot v_n = \sum_{i=1}^{n} v_i^2 = \|v\|^2$

Reading: Inner product naturally induces a norm and every "inner product space" is a "normed vector space"

Associative rule. (结合律.)

$$\langle v, w+u \rangle = \langle v, w \rangle + \langle v, u \rangle$$

or

$$v \cdot (w+u) = v \cdot w + v \cdot u .$$

Exercise

Multi-Choice: The dot product of two vectors is:

(A) always a scalar ✓

(B) sometimes a vector, sometimes a scalar

Judge: True or False.

The dot product of any two vectors is well-defined.

False.

$(2,3,x)$

$(2,3) \cdot (1,4,5) = ?$

possible way: $2 \cdot 1 + 3 \cdot 4 + 0 \cdot 5$ NOT in our definition

Application: Evaluation and Ranking

Evaluate an object: how?

e.g. how much to pay for a soccer/basketball player?
—e.g. Messi’s salary? Neymar’s salary?

Rank (compare) two objects: how?

e.g. is University A better than University B?

e.g. is soccer player A better than soccer player B?

Use data

quantitative [illegible]

Application: Comparison and Evaluation

Evaluate an object: how?

e.g. how much to pay for a soccer/basketball player?
—e.g. Messi's salary? Neymar's salary?

(4, 10)

Rank (compare) two objects: how?

> (1, 2)

e.g. is University A better than University B?

e.g. is soccer player A better than soccer player B?

Short answer:
Step 1: "vectorize" object
—(then want to compare two vectors)
Step 2: compare weighted sum of two vectors

M → (2, 3)
L → (1, 10)

Example: Sports Player Comparison

Example: You want to compare two soccer players

65%

Player 1

	Assists	Shoots	intercepts
Value	7	8	10

Vectorize

(7, 8, 10)

Player 2

	Assists	Shoots	intercepts
Value	10	10	6

(10, 10, 6)

Vectorize

Example: Sports Player Comparison

Example: You want to compare two soccer players

Player 1

	Assists	Shoots	intercepts	Total
Weight	20%	50%	30%	
Value	7	8	10	
contribution	1.4	4	3	8.4

"weight"

"Feature"

Weights

Vectorize

$$(0.2, 0.5, 0.3) \cdot (7, 8, 10) = 8.4$$

Example: Sports Player Comparison

Example: You want to compare two soccer players

10m/s

Player 1

"weight"

"Feature"

	Assists	Shoots	intercepts	Total
Weight	20%	50%	30%	
Value	7	8	10	
contribution	1.4	4	3	8.4

Weights

Vectorize

"Score": weighted sum

inner product

12 m/s

Player 2

"weight"

"Feature"

	Assists	Shoots	intercepts	Total
Weight	20%	50%	30%	
Value	10	10	6	
contribution	2	5	1.8	8.8

"Score"

Example: ~~University Comparison~~ Evaluation

Application 1: (general: score computation)

v represents a set of "features" of an object,

w is a vector of the same size (often called a weight vector),

Score: inner product $w^T v$ is a weighted sum of the feature values.

Examples:

	Soccer player	University ranking	Your example?
"Feature"	Contribution type	Univ. indicator	?
"Weight"	Manual weight	Manual weight	?
"Score"	salary	Evaluation score	?

Application: Evaluation

Sub-application: evaluation.
(Useful for ranking, comparing, etc.)

Goal: to evaluate a city, a university, an employee, a basketball player, a movie director

Step 1: Set up "features" (indicators)
Step 2: Evaluate each feature of the object, obtaining a feature vector

Vectorize object

Step 3: Provide weights to the features; get weight vector

Setting weights

Step 4: Compute the inner product, to get the "score"

"scalarize" object

Remark: "Score" can be used in other areas, e.g. machine learning

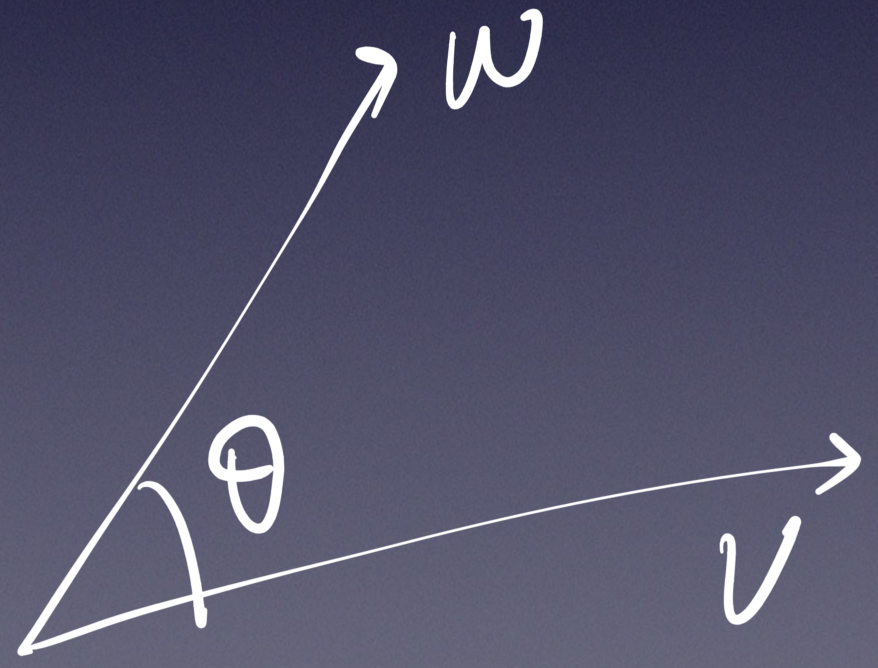
Application: Ranking (via Evaluation)

Ranking: Given n subjects, rank them（排名 or 排序）

Example: Rank soccer players

Example: Rank universities

Part III Angle Between Two Vectors



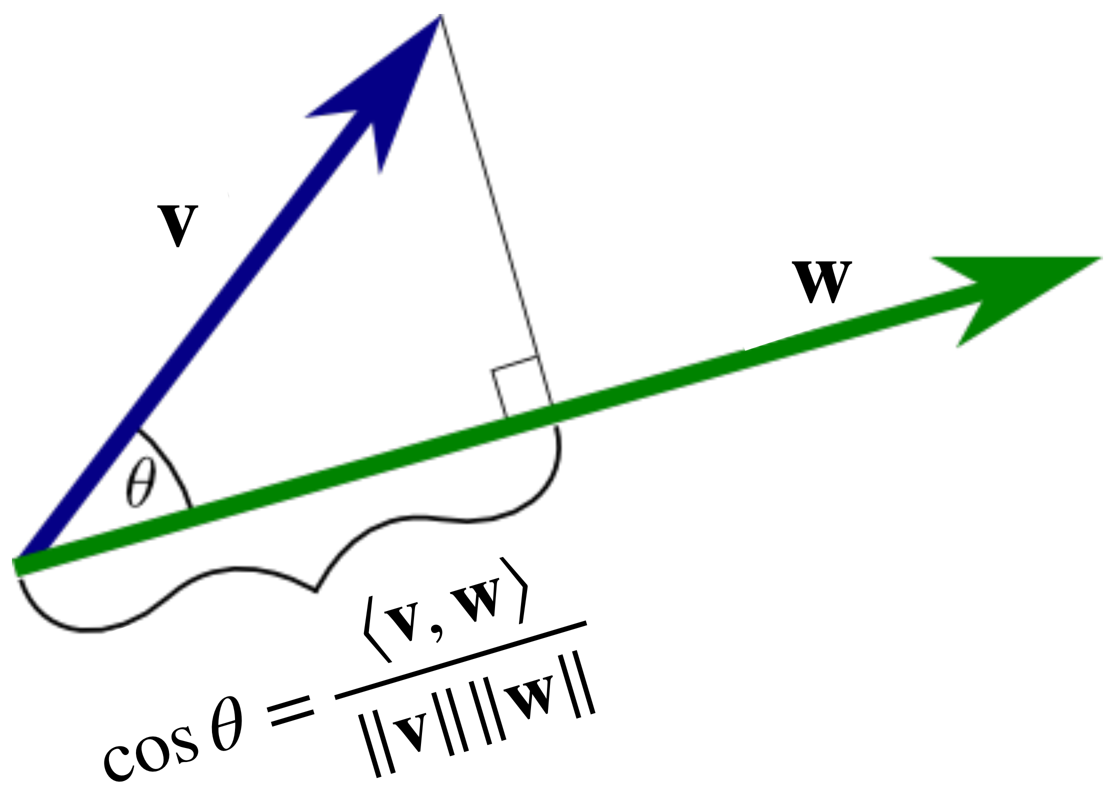
Cosine Formula

Cosine Formula: If v, w are nonzero vectors, and Angle(v, w) = θ, then

$$\cos\theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\|\,\|\mathbf{w}\|}$$

skip proof.

Also called: cosine similarity of **v** and **w**



Cauchy-Schwartz Inequality
Relation of Dot Product and Norm

柯西不等式.

Cauchy–Schwarz Inequality

or $|v \cdot w|$

$$|\langle \mathbf{v}, \mathbf{w}\rangle| \le \|\mathbf{v}\|\|\mathbf{w}\|$$

$\theta = \angle(v, w)$

Intuitively: $|\cos\theta| = \dfrac{|\langle \mathbf{v}\cdot\mathbf{w}\rangle|}{\|\mathbf{v}\|\|\mathbf{w}\|} \le 1$

$$\underbrace{\left(\sum_{i=1}^{n} a_i^2\right)}_{= \|a\|^2}\underbrace{\left(\sum_{i=1}^{n} b_i^2\right)}_{= \|b\|^2} \ge \underbrace{\left(\sum a_i b_i\right)^2}_{= (\langle a, b\rangle)^2}$$

A pure algebraic proof:

$n=2$: $(a_1^2+a_2^2)(b_1^2+b_2^2)-(a_1b_1+a_2b_2)^2$

$= (a_1b_2 - a_2b_1)^2 \ge 0.$

(NOT required) check yourself

General n: $\left(\sum_i a_i^2\right)\left(\sum_i b_i^2\right) - \left(\sum_i a_ib_i\right)^2 = \sum_{1\le i\ne j\le n} (a_ib_j - a_jb_i)^2$

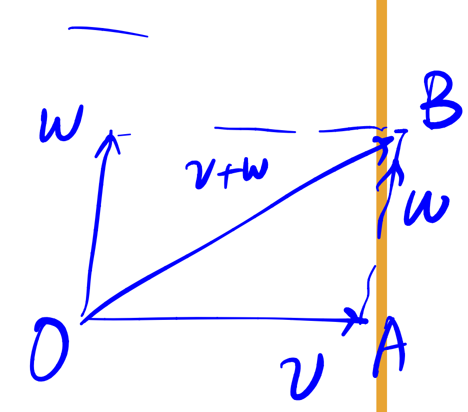
Triangle Inequality

length of One side of triangular ≤ sum of lengths of two sides.

Example (~~Triangle Inequality~~)

$$\|\mathbf{v}+\mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\|$$

$$|OB| \qquad |OA| \qquad |AB|$$



Proof:

$$(\|v\|+\|w\|)^2 - \|v+w\|^2$$

$\langle a,a\rangle = \|a\|^2 \rightarrow$ trick.

$$= \|v\|^2 + \|w\|^2 + 2\|v\|\,\|w\| - (v+w)\cdot(v+w)$$

associative rule

$$= \cancel{\|v\|^2} + \cancel{\|w\|^2} + 2\|v\|\,\|w\| - (\cancel{v\cdot v} + 2\,v\cdot w + \cancel{w\cdot w})$$

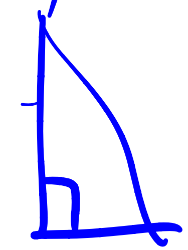
$$= 2\|v\|\,\|w\| - 2\,v\cdot w \ge 0.$$

Pythagoras Law

Pythagoras Law (毕达哥拉斯定律, i.e. 勾股定理)

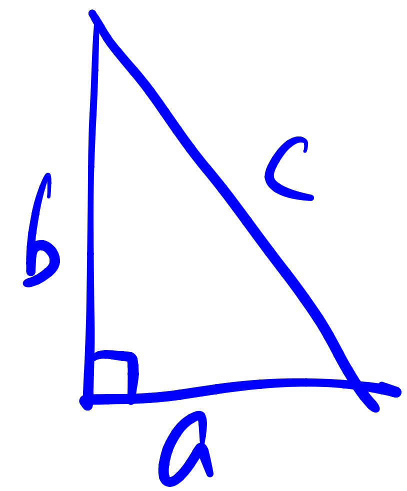
$\theta = 90^\circ$

$\cos\theta = 0$



$$\|v\|^2 + \|w\|^2 = \|v - w\|^2 \quad \text{iff} \quad \langle \mathbf{v}, \mathbf{w} \rangle = 0$$

Proof: exercise.



$a^2 + b^2 = c^2$.

Practice Problem

Problem (Dot Product)

Can we write find three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ such that in a 2D plane

$$\langle \mathbf{u}, \mathbf{v} \rangle < 0 \qquad \langle \mathbf{u}, \mathbf{w} \rangle < 0 \qquad \langle \mathbf{v}, \mathbf{w} \rangle < 0?$$

$$\mathbf{u} = (1,0) \qquad \mathbf{v} = (-1,4) \qquad \mathbf{w} = (-1,-4)$$

How about four vectors?

Application: Searching (搜索)

Searching: Given a query, find the most 10 relevant entities e.g. videos; websites

This is a ranking problem. Rank entities in a search (Will talk about Google's PageRank in ~ Lec 20)

Practically important!
Millions of people work on this problem

Application: Searching (搜索)

Searching: Given a query, find the most 10 relevant entities
e.g. videos; websites

This is a ranking problem: Rank entities in a search
(Will talk about Google's PageRank in ~ Lec 20)

Practically important!
Millions of people work on this problem

Method 1: Ignore query, just rank all entities by their evaluation scores
e.g. when searching restaurants, just rank restaurants by scores

Method 2: Find 10 most similar entities to the query

How to make coffee?

Challenge: How do computers know target "sentence" and query "sentence" are "similar"?

Application: Searching in a Vector Database

Idea: **Vectorize + compute cosine similarity**

there's way

Query Sentence: "How to make coffee?" ⟶ $(0.8, 1.1, 0.5) = \mathbf{q}$

Candidate 1: "Procedure for preparing Latte" ⟶ $(2.5, 4.3, 3.7) = \mathbf{u}$

Candidate 2: "Story of Car Maker" ⟶ $(-1.5, 2.5, -0.3) = \mathbf{v}$

Application: Searching in a Vector Database

Idea: **Vectorize + compute cosine similarity**

major challenges

Query Sentence: "How to make coffee?" ⟶ (0.8, 1.1, 0.5) = **q**

0.95

Candidate 1: "Procedure for preparing Latte" ⟶ (2.5, 4.3, 3.7) = **u**

0.33

Candidate 2: "Story of Car Maker" ⟶ (-1.5, 2.5, -0.3) = **v**

Cosine similarity of q and u is ~ 0.95;

Cosine similarity of q and v is ~ 0.33

Compare cosine similarity: 0.95 > 0.33.

so candidate 1 is more similar to the query.

Calculation of Cosine Similarity

Just list the computation procedure for the first pair here.

First Pair: $(0.8, 1.1, 0.5)$ **and** $(2.5, 4.3, 3.7)$

1. **Dot Product:**

$$\mathbf{A} \cdot \mathbf{B} = 0.8 \times 2.5 + 1.1 \times 4.3 + 0.5 \times 3.7 = 2 + 4.73 + 1.85 = 8.58$$

1. **Euclidean Norms:**

$$\|\mathbf{A}\| = \sqrt{0.8^2 + 1.1^2 + 0.5^2} = \sqrt{0.64 + 1.21 + 0.25} = \sqrt{2.1} \approx 1.4491$$

$$\|\mathbf{B}\| = \sqrt{2.5^2 + 4.3^2 + 3.7^2} = \sqrt{6.25 + 18.49 + 13.69} = \sqrt{38.43} \approx 6.2006$$

1. **Cosine Similarity:**

$$\text{Cosine Similarity} = \frac{8.58}{1.4491 \times 6.2006} \approx \frac{8.58}{8.9848} \approx 0.9549$$

Related Academic Talk

Prof. Ping Li (formally Cornell professor, and Baidu researcher) gave a talk on vector database and search in ~Aug 28, 2023

Similarity-score based search is still an active area of research.

Reading [Not Required to Know]: Questions to Explore

Read

The applications listed today are somewhat simple.

Nevertheless, if you are a deep thinker, you may realize: the described methods are NOT perfect.

There are many issues!

e.g. What weights to pick for evaluation?

What issues do you notice?

Answering each question can lead to a large area…

Summary Today

Today, we have learned:

Math	Application

—**Norm** of vector

$$\|\mathbf{v}\| = \|\mathbf{v}\|_2 := \left(v_1^2 + \cdots + v_n^2\right)^{\frac{1}{2}}$$

a.k.a. (also known as) ℓ_2-norm

—**Inner product** of vector: $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i w_i$

—**Angle** of two vectors θ satisfies $\cos\theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|}$

—**Properties**:

Cauchy-Schwartz inequality

Triangular inequality

—**Application 1**:

~~Evaluation~~

Ranking (via evaluation)

—**Application 2**:

Ranking (via similarity)