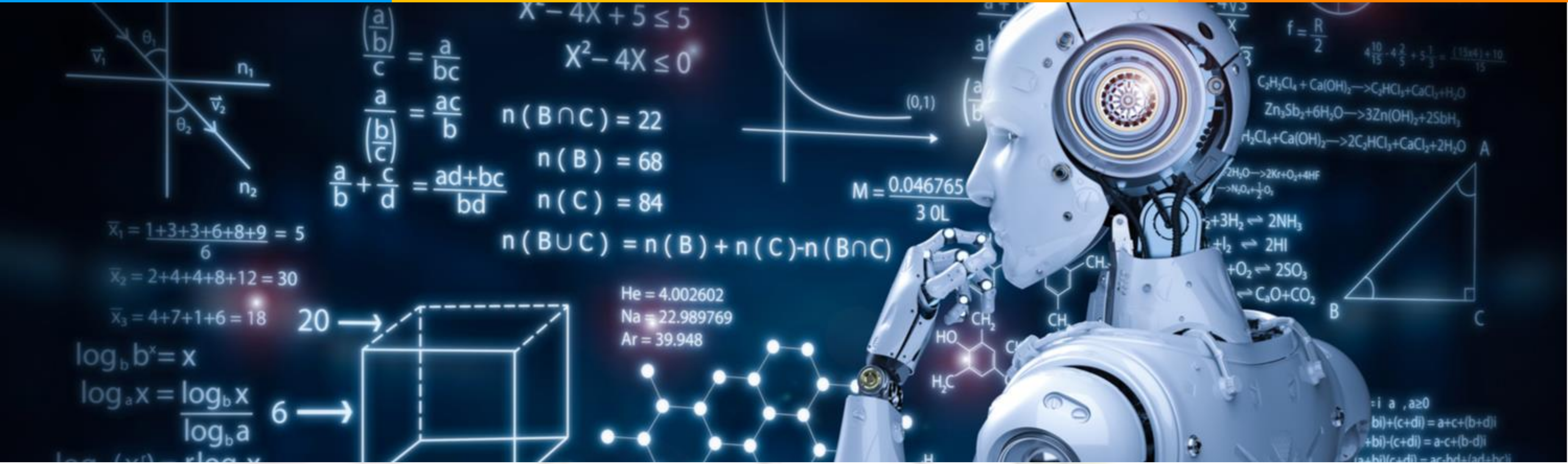
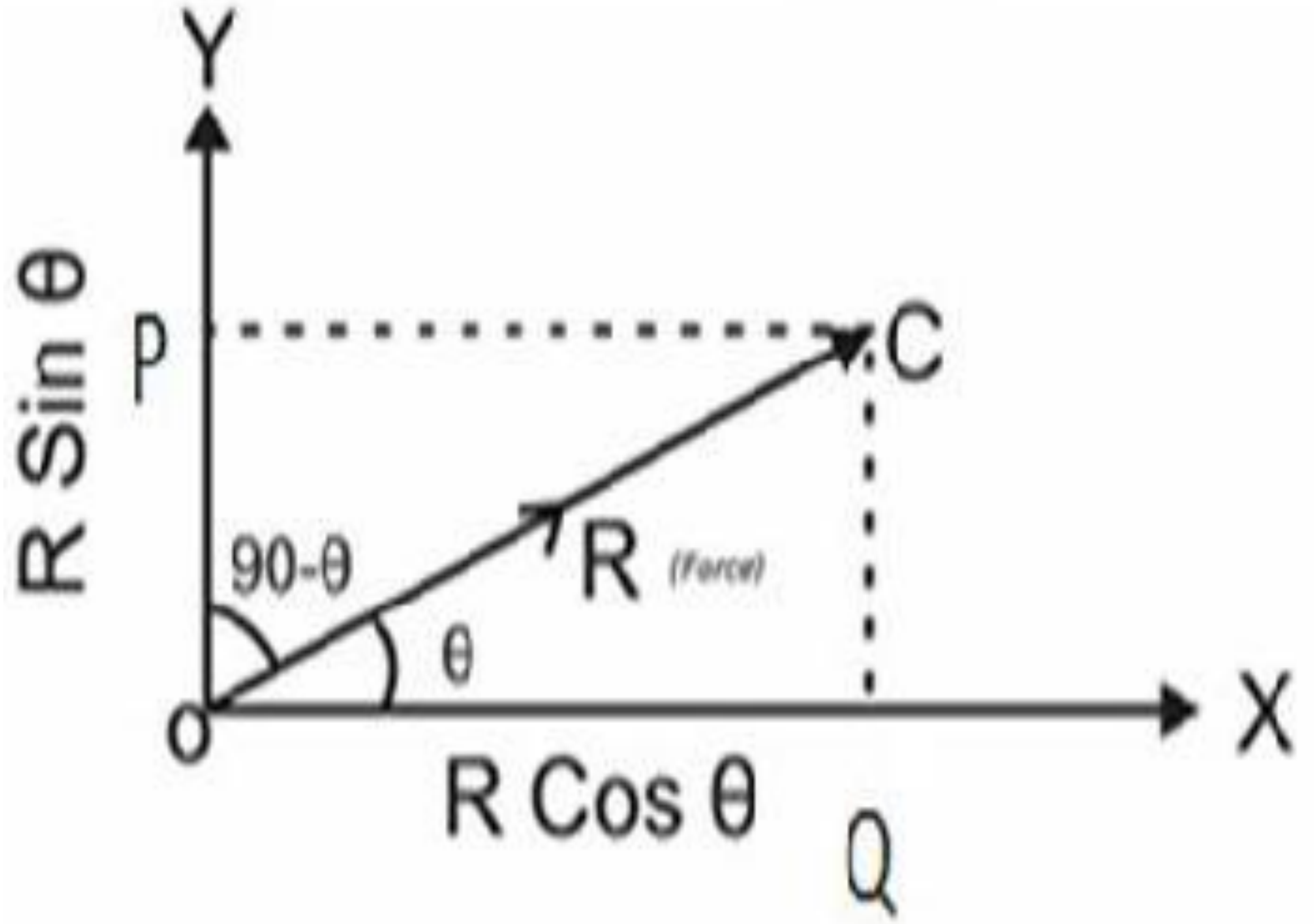


Lecture 02, 03: Vectors – multiple perspectives & basic operations

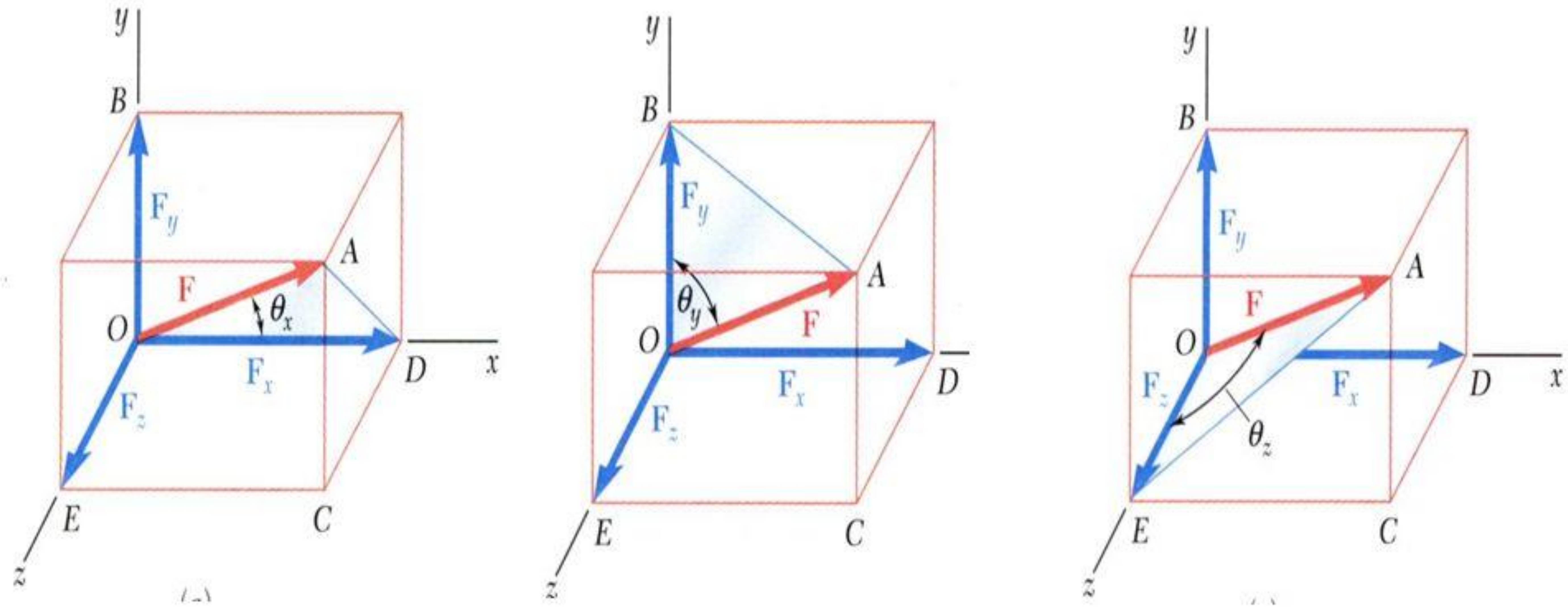
Different ways of looking at vectors & operations

- Physical Vectors
- Geometric view
- Algebraic view – Traditional
- Operations on vector with homogeneous data
- Operations on vector with heterogeneous data

1. Physical Vectors


Y
R Sin θ
P
C
90-θ
R
(Force)
θ
O
X
R Cos θ
Q

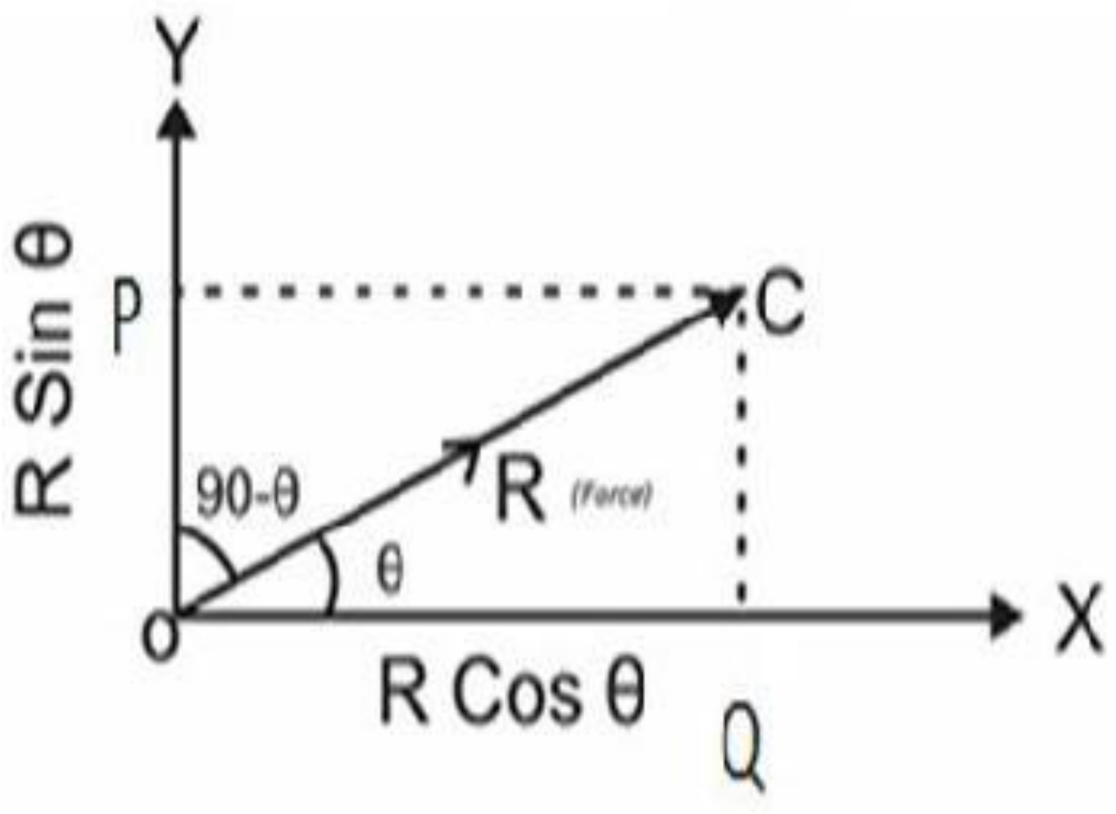
Forces in 3D

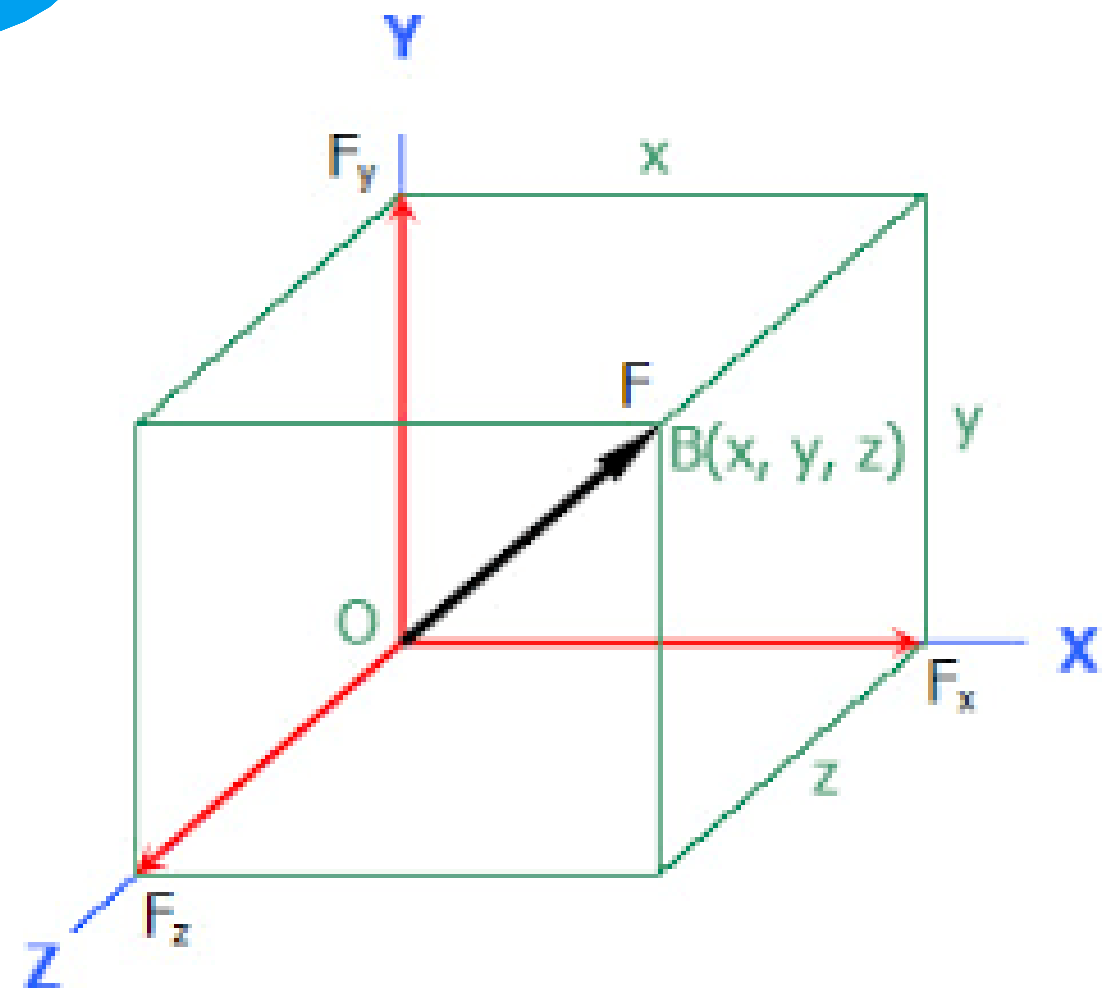


$$\vec{F} = \vec{F}_x + \vec{F}_y + \vec{F}_z = F_x\hat{i} + F_y\hat{j} + F_z\hat{k}$$

$$F_x = F\cos\theta_x; \quad F_y = F\cos\theta_y; \quad F_z = F\cos\theta_z$$

What about higher dimensions?





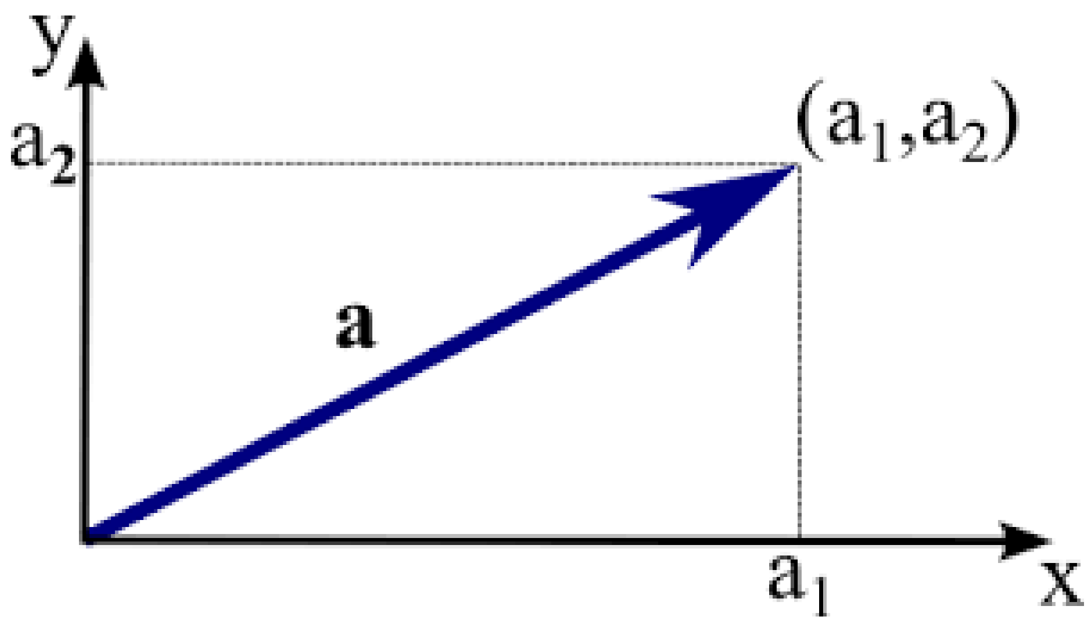
- Angles and magnitudes hard to define beyond 3D
- Coordinates are sufficient to represent in n-d
- Physics vectors are specific. Linear Algebra is generic

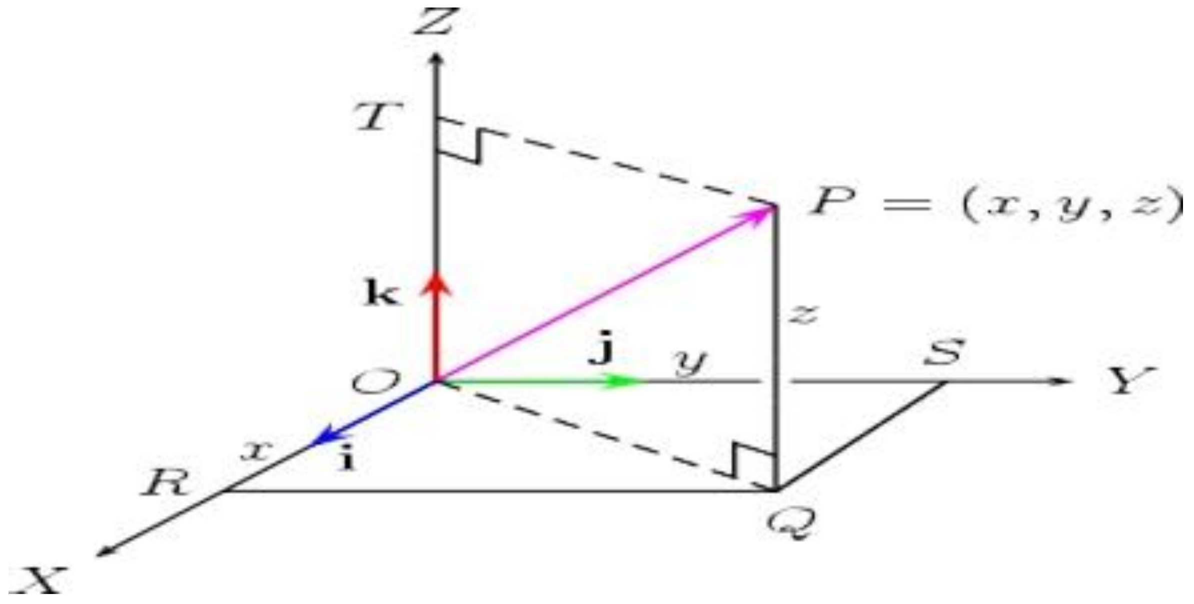
2. Geometric view

Geometric view

- Example of vectors in 2D, 3D, 1D coordinate space

$\mathbb{R}^2$





$\mathbb{R}^3$

- Length (magnitude)

$$\sqrt{{x_1}^2 + {x_2}^2}$$

- Length (magnitude)

$$\sqrt{{x_1}^2 + {x_2}^2 + {x_3}^2}$$

- n-dimensional space $\mathbb{R}^n$
- What is $\mathbb{R}^0$

- Magnitude in n-d

$$\sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2}$$

1. Algebraic view

Algebraic view

- Vector – Just a collection of numbers
- Each number is called entry/component
- First entry $x_1 = 6$ Size = 3
- Vector of size n is called n-vector
- Represented as $x \in \mathbb{R}^n$ - n-d coordinate space
- In Linear Algebra Length means magnitude
- Length or magnitude = $\|x\|_2 = (\sum_{i=1}^{n} {x_i}^2)^{\frac{1}{2}}$
- Caution: In numpy len(x) gives size of vector

$$x = \begin{bmatrix} 6 \\ 11 \\ 3 \end{bmatrix}$$

Vector means column vector

- Notation for mathematical convenience
- Transpose is row vector
- Numpy quirks
  - Expected shape versus numpy shape
  - Failure to transpose
  - Summary: Call reshape(-1,1) for performing vector operations
- sklearn quirks: sklearn expects the shape of y vector as (n,)

$$x = \begin{bmatrix} 6 \\ 11 \\ 3 \end{bmatrix}$$

Vectors with homogeneous & heterogenous data

- Boston House Price Dataset
- How many vectors?
  - 4 feature vectors
  - 1 target vector
  - 5 house vectors

	CRIM	AGE	RM	LSTAT	PRICE
0	0.00632	65.2	6.575	4.98	24.0
1	0.02731	78.9	6.421	9.14	21.6
2	0.02729	61.1	7.185	4.03	34.7
3	0.03237	45.8	6.998	2.94	33.4
4	0.06905	54.2	7.147	5.33	36.2

- Columns are features – homogeneous
- Rows are houses -heterogeneous

Vector examples

- Investment portfolio
  - Entries - GOOG, NVDA, .. MSFT
- Audio signal at t, t+1, ….
- Time series
- Polynomial coefficients
- Features
- Patients

$$x = \begin{bmatrix} 100 \\ 250 \\ .. \\ 250 \end{bmatrix}$$

$$ax_1 + bx_2 + cx_3 \quad = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

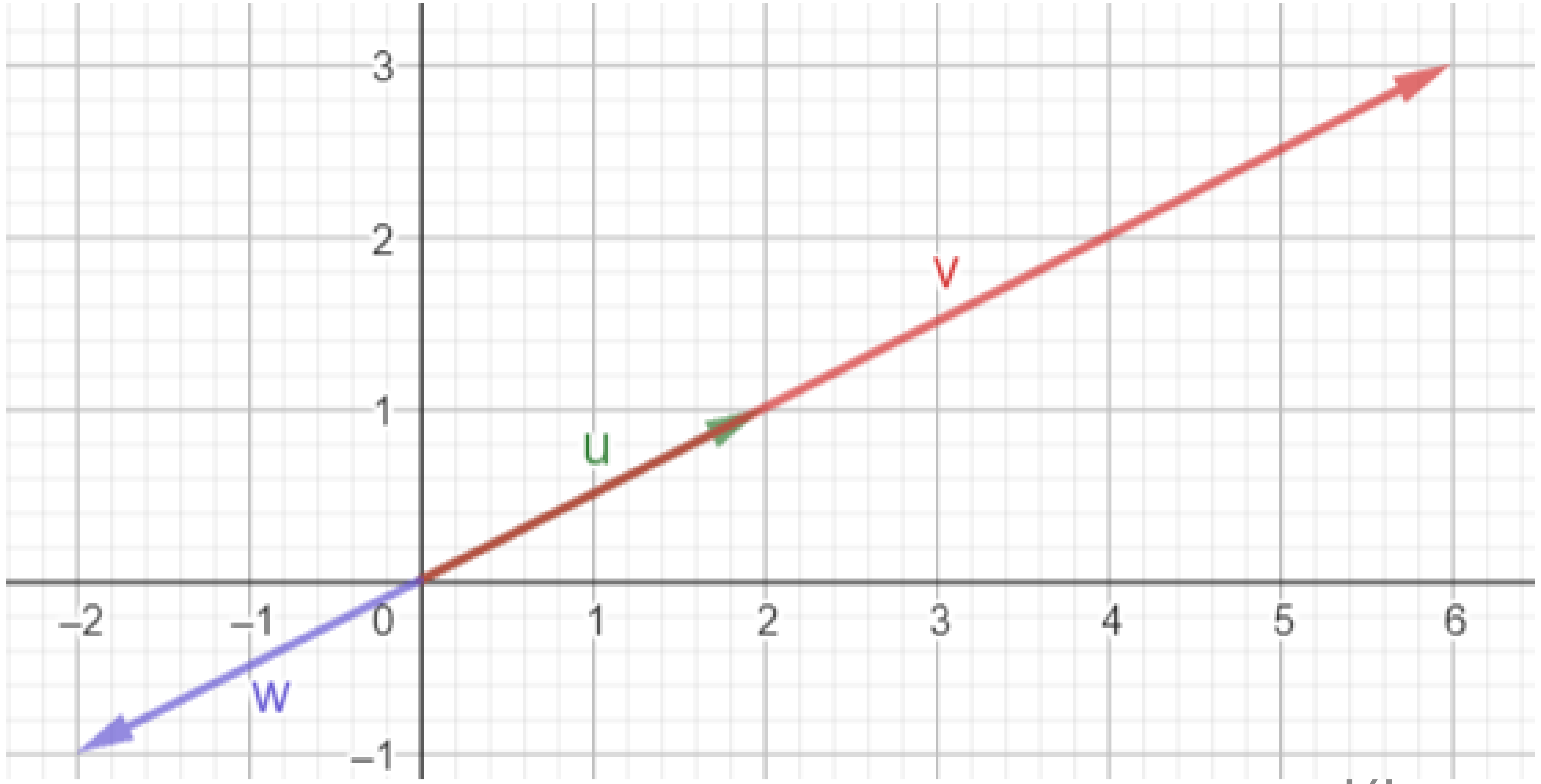
Scalar multiplication

- E.g. Rainfall in the month of July (31-vector)
  - Scalar multiplication – unit conversion inch to cm

$$x = \begin{bmatrix} 2 \\ 2.25 \\ .. \\ 3.5 \\ .. \\ 3 \end{bmatrix}$$

$$y = 2.5x = 2.5 \begin{bmatrix} 2 \\ 2.25 \\ .. \\ 3.5 \\ .. \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 5.63 \\ 2.5 \times .. \\ 8.75 \\ 2.5 \times .. \\ 7.5 \end{bmatrix}$$

- Geometric meaning



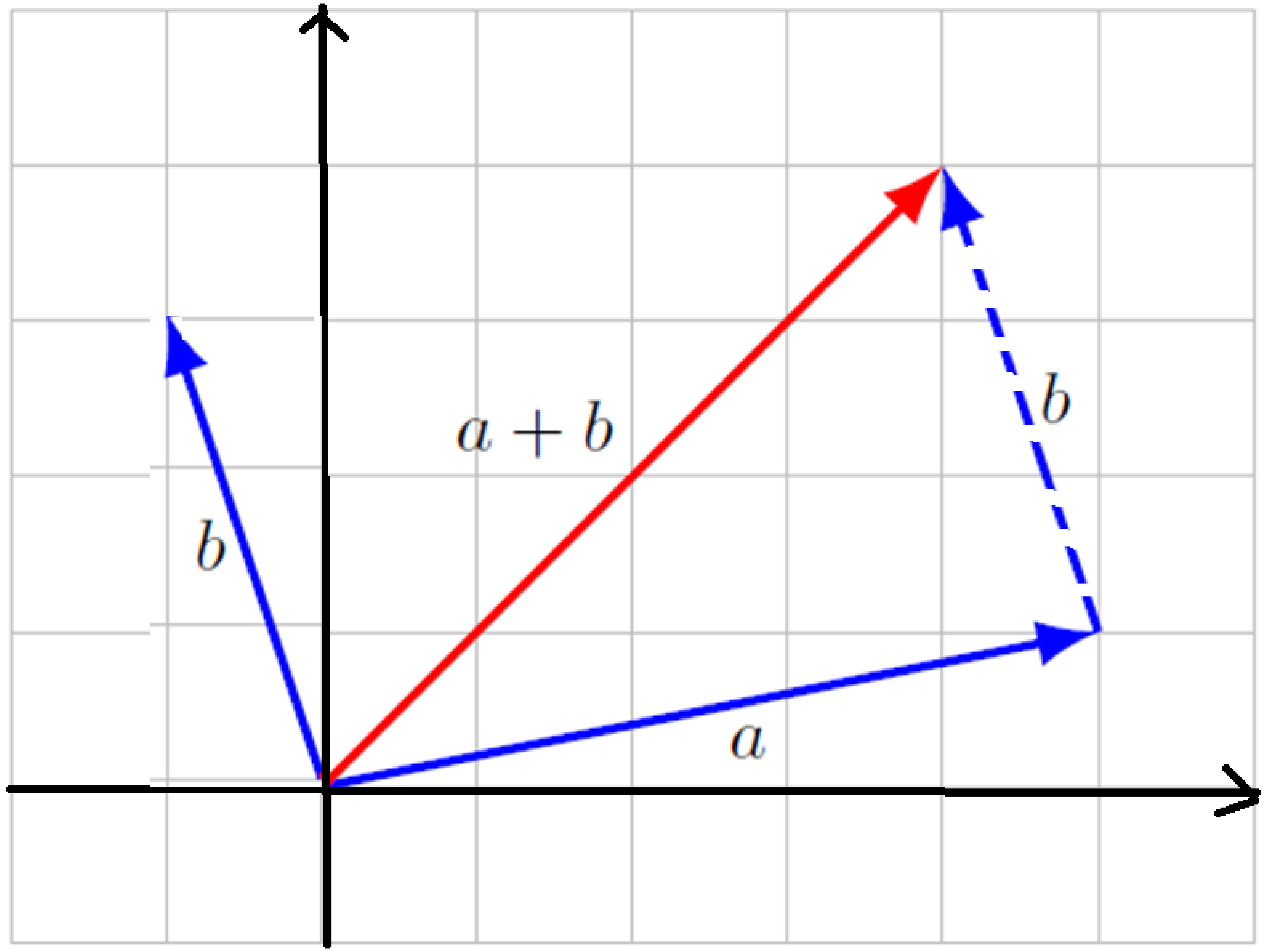
Vector addition

- Rainfall in the month of July (31-vector)
- What is the total rainfall in July for last 5 years?

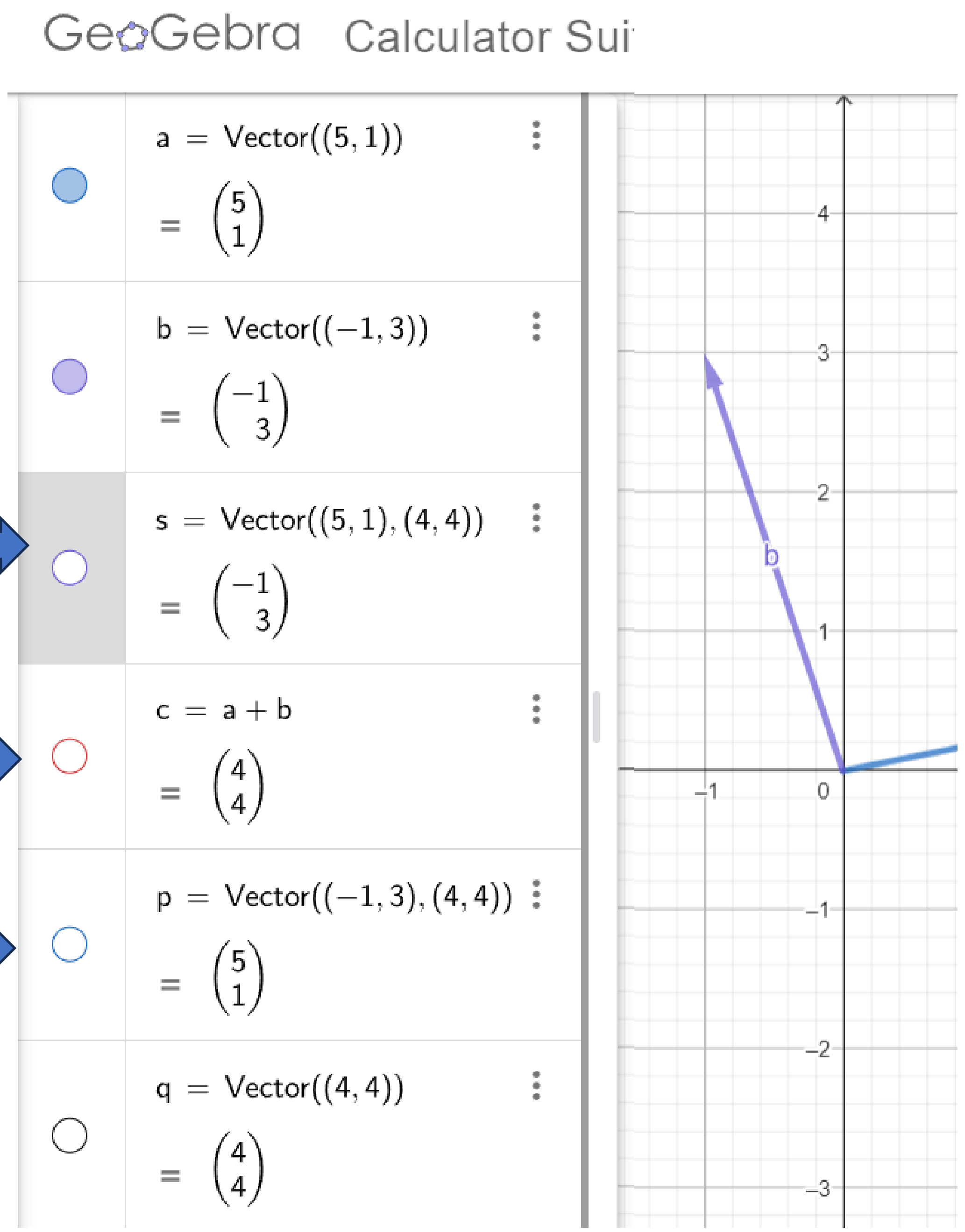
$$x_{19} = \begin{bmatrix} 2 \\ 2.25 \\ .. \\ 3.5 \\ .. \\ 3 \end{bmatrix} \qquad x_{20} = \begin{bmatrix} 2.5 \\ 3 \\ .. \\ 0 \\ .. \\ 3 \end{bmatrix} \qquad ......x_{23} = \begin{bmatrix} 0 \\ 1.5 \\ .. \\ 2 \\ .. \\ 3 \end{bmatrix}$$

$$x_{19} + x_{19} + ... + x_{23} = \begin{bmatrix} 2 + 2.5 + ... + 0 \\ 2.25 + 3 + ... + 1.5 \\ .. \\ 3.5 + 0 + ... + 2 \\ .. \\ 3 + 3 + ... + 3 \end{bmatrix}$$

Vector Addition: Geometric meaning



Click on these left choices one after another to see the flow



- Geometric meaning of vector addition
  - Parallelogram diagonal
  - https://www.geogebra.org/calculator/m6rjgfgq

Problem 1: Average rainfall

- Calculate average rainfall for 5 years
- Add 5 vectors. Scalar multiply with 1/5

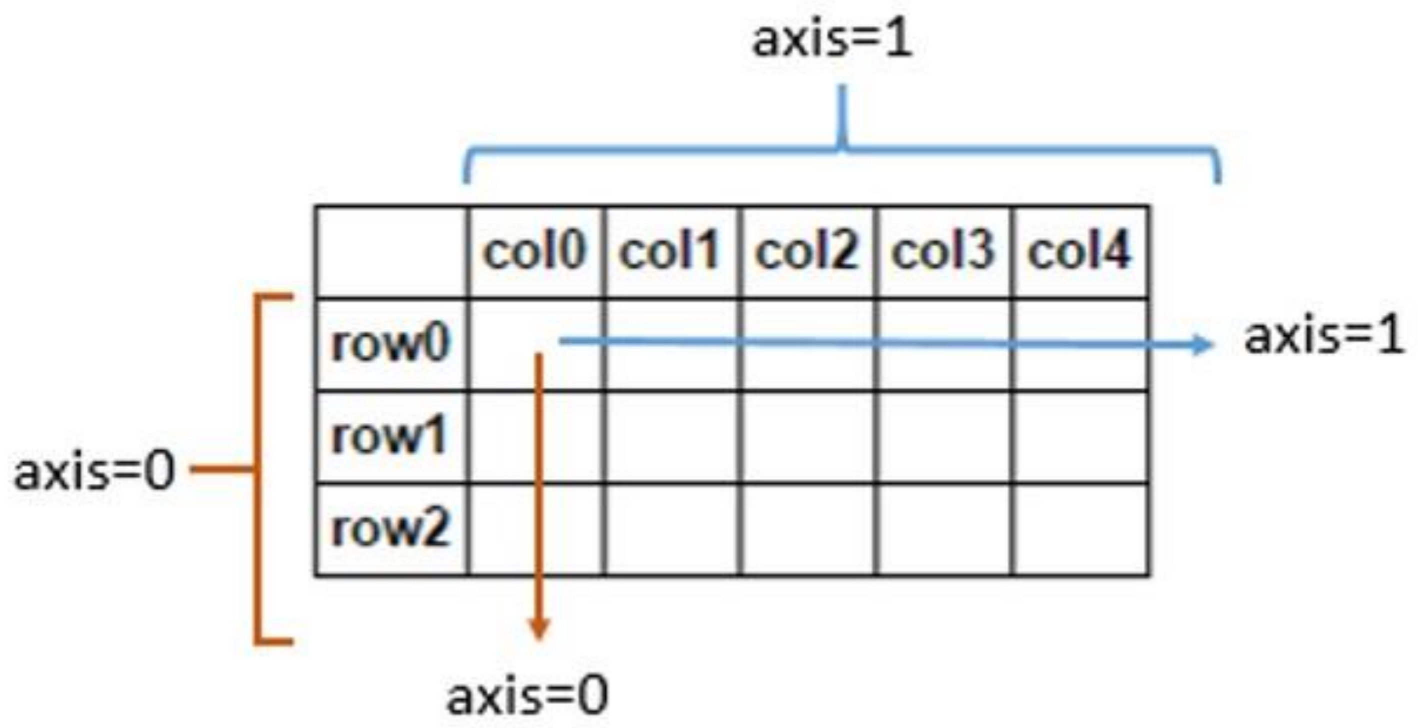
$$\bar{x} = \frac{1}{5}(x_{19} + x_{20} + \ldots + x_{23}) = \frac{1}{5}\left(\begin{bmatrix} 2 + 2.5 + \ldots + 0 \\ 2.25 + 3 + \ldots + 1.5 \\ .. \\ 3.5 + 0 + \ldots + 2 \\ .. \\ 3 + 3 + \ldots + 3 \end{bmatrix} \right)$$

Problem 2: Find the average patient

- Combine vector addition and scalar multiplication to calculate average patient

$$\bar{x} = \frac{1}{4}\left(x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)}\right) \qquad \bar{x} = \begin{bmatrix} 75 \\ 125 \\ 37.625 \end{bmatrix}$$

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0




```
avg_patient = (1/patients.shape[0]) * np.sum(patients, axis=0)
```

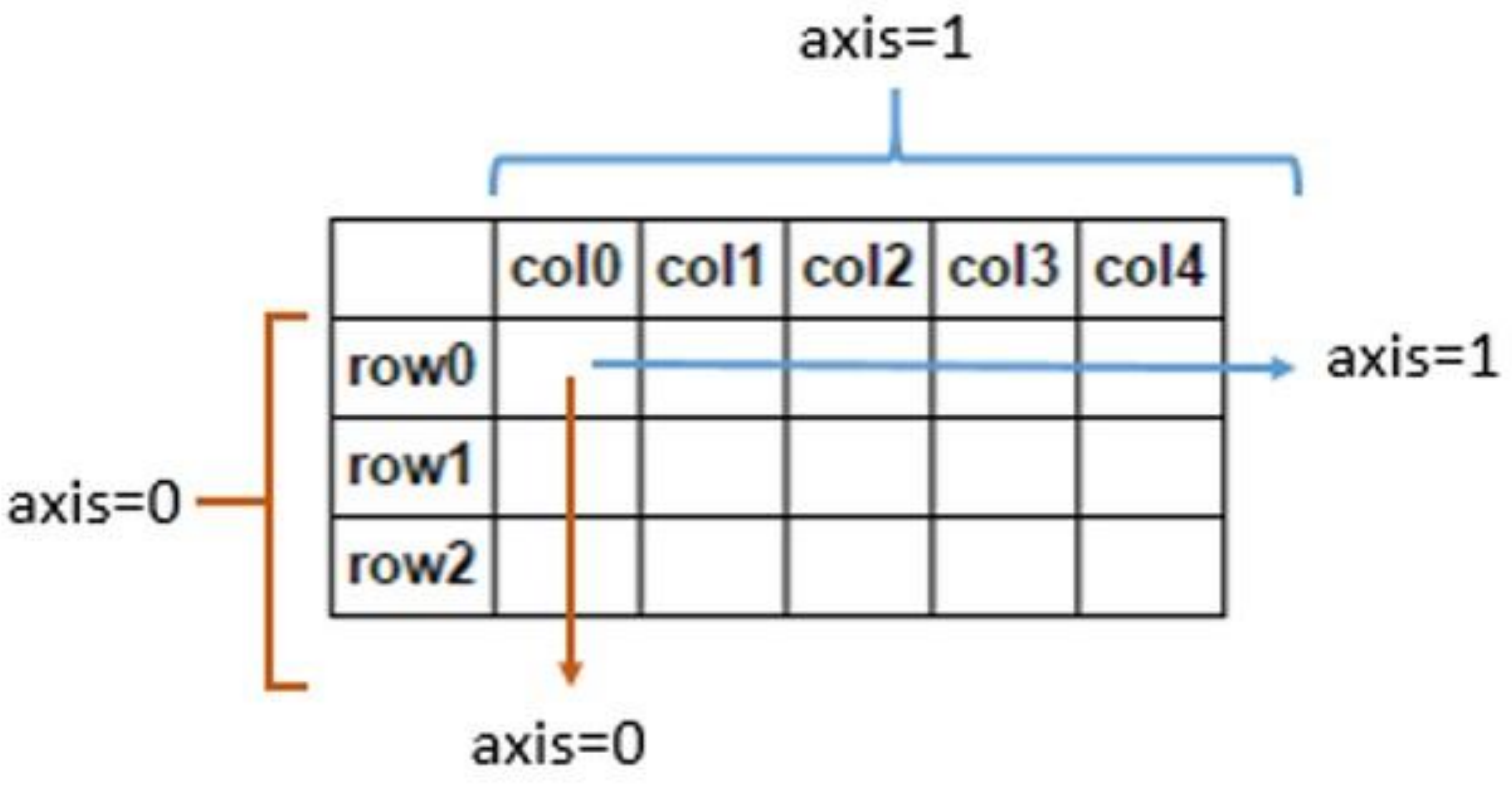
```
avg_patient = np.mean(patients, axis=0)
```

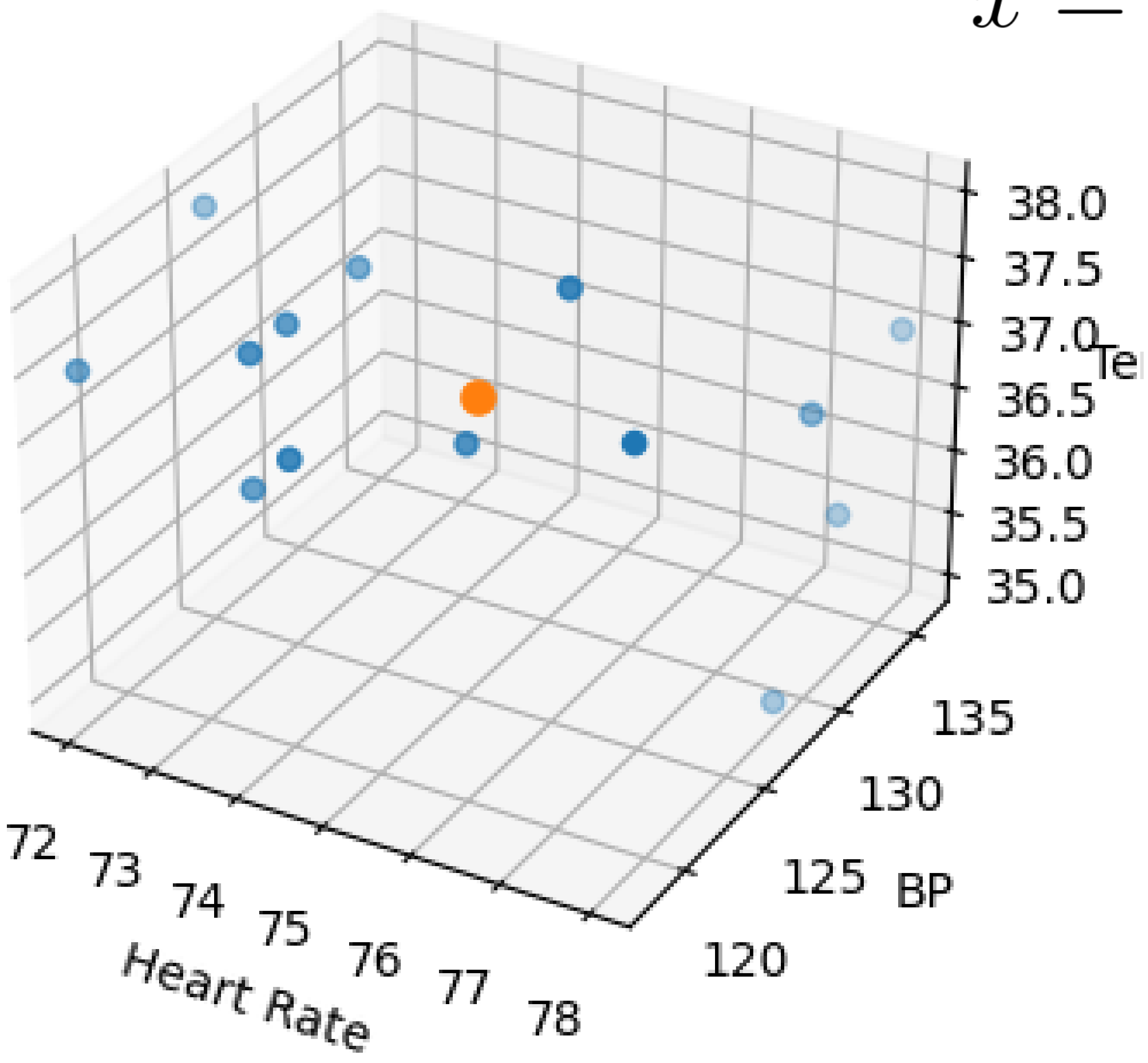
Problem 2: Find the average patient

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

$$\bar{x} = \frac{1}{4}\left(x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)}\right)$$

$$\bar{x} = \begin{bmatrix} 75 \\ 125 \\ 37.625 \end{bmatrix}$$





Vector subtraction

- Rainfall in the month of July 23 and July 22
- How much **excess** rainfall did July 23 have?

$$x_{23} = \begin{bmatrix} 0 \\ 1.5 \\ .. \\ 2 \\ .. \\ 3 \end{bmatrix} \qquad x_{22} = \begin{bmatrix} 2 \\ 2.25 \\ .. \\ 3.5 \\ .. \\ 3 \end{bmatrix} \qquad x_{23} - x_{22} = \begin{bmatrix} 0-2 \\ 1.5-2.25 \\ .. \\ 2-3.5 \\ .. \\ 3-3 \end{bmatrix} = \begin{bmatrix} -2 \\ -0.75 \\ .. \\ -1.5 \\ .. \\ 0 \end{bmatrix}$$

- Question: What do the negative values indicate?

What is word count vector?

- Imagine 2 sentences
  - Fraud identification uses anomaly detection
  - Anomaly detection is based on machine learning
- How can I represent these sentences as vectors using word count?
  - Demonstrate with example (Next Page)

What is word count vector (contd)?

- Sentence 1: Fraud identification uses anomaly detection
- Sentence 2: Anomaly detection is based on machine learning

Why are anomaly and Anomaly different?

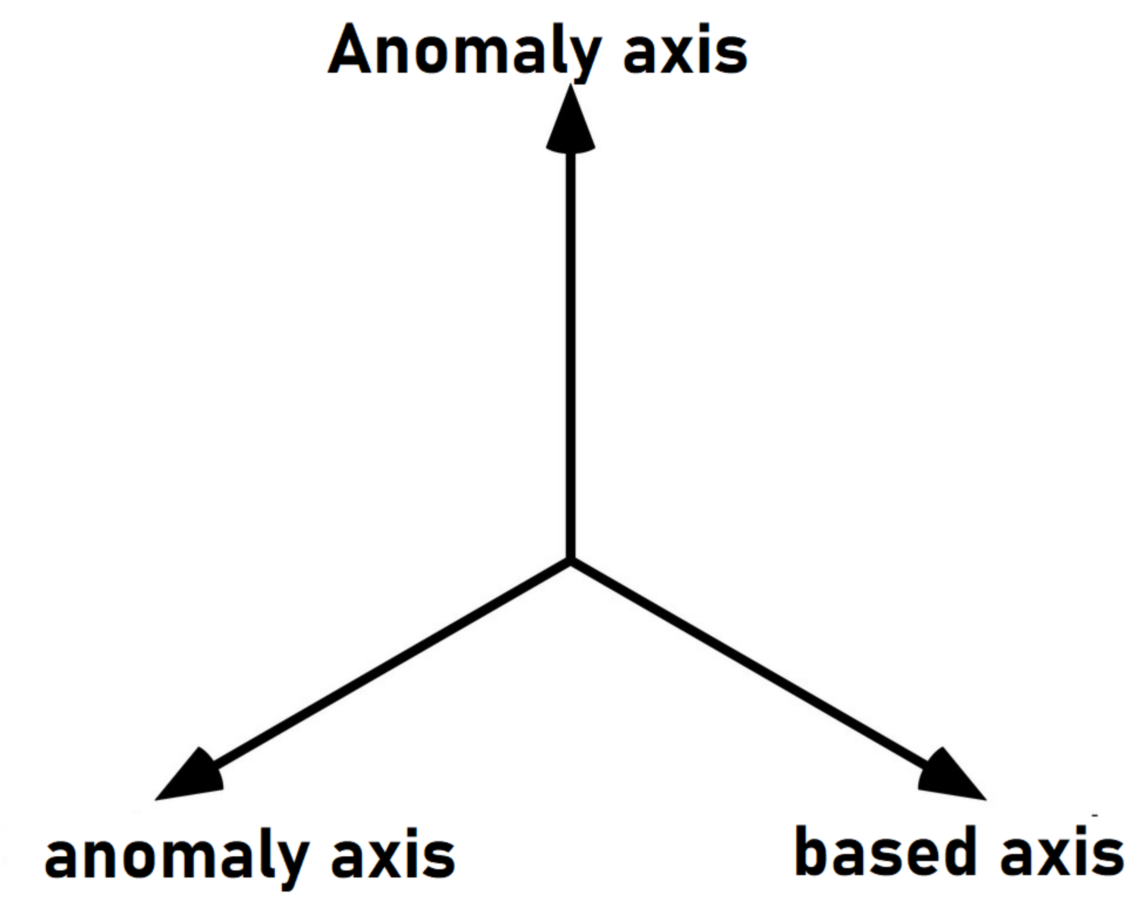
Vocabulary	Sentence 1	Sentence 2
Anomaly	0	1
anomaly	1	0
based	0	1
detection	1	1
Fraud	1	0
identification	1	0
is	0	1
learning	0	1
machine	0	1
on	0	1
uses	1	0

Geometric meaning of word count vector

- Word count vector is a vector in a coordinate space where each axis is one word from vocabulary and word count is the value along such axis

Vocab Sent1

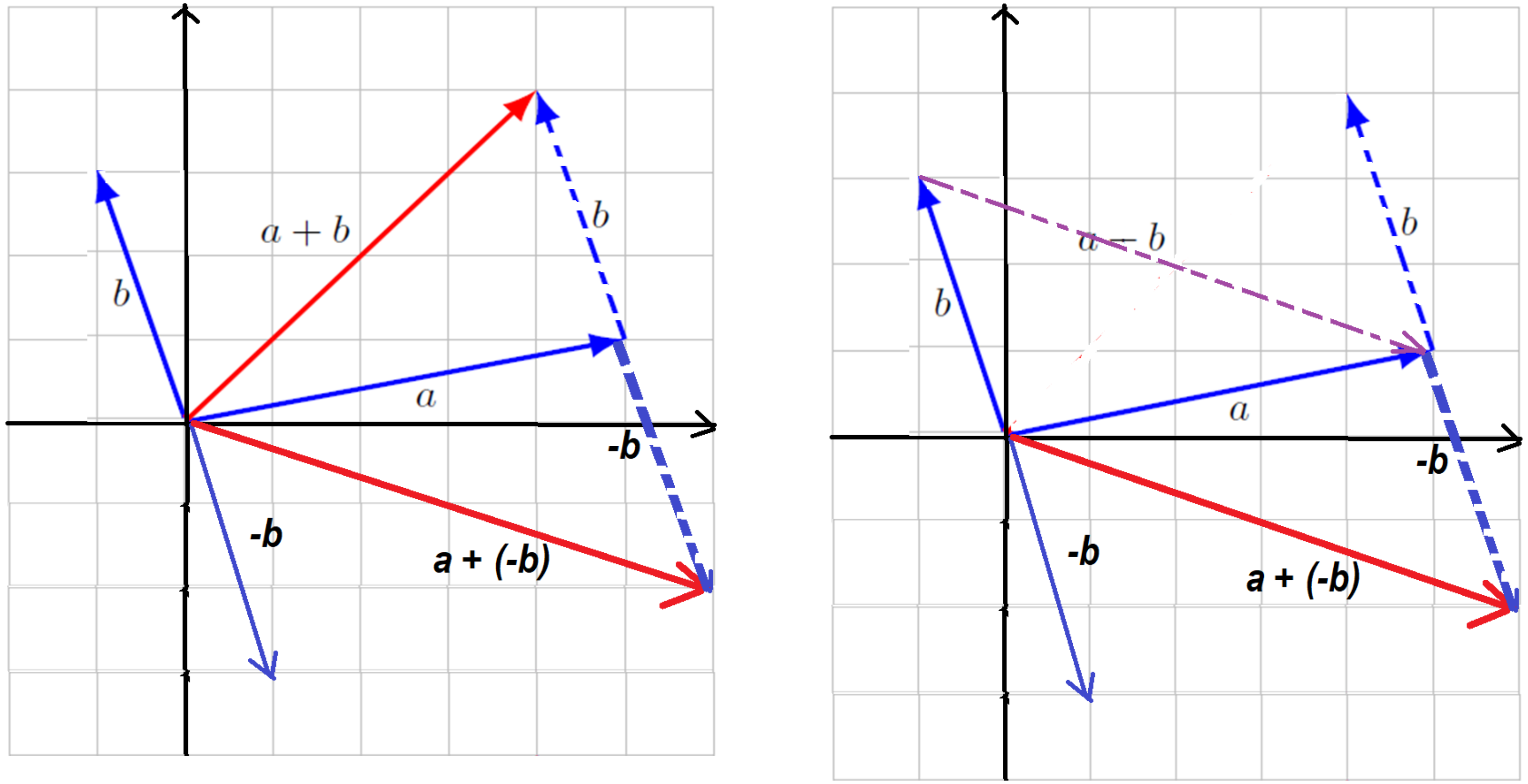
$$\begin{array}{c} Anomaly \\ anomaly \\ based \\ detection \\ Fraud \\ identification \\ is \\ learning \\ machine \\ on \\ uses \end{array} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$



Problem 3

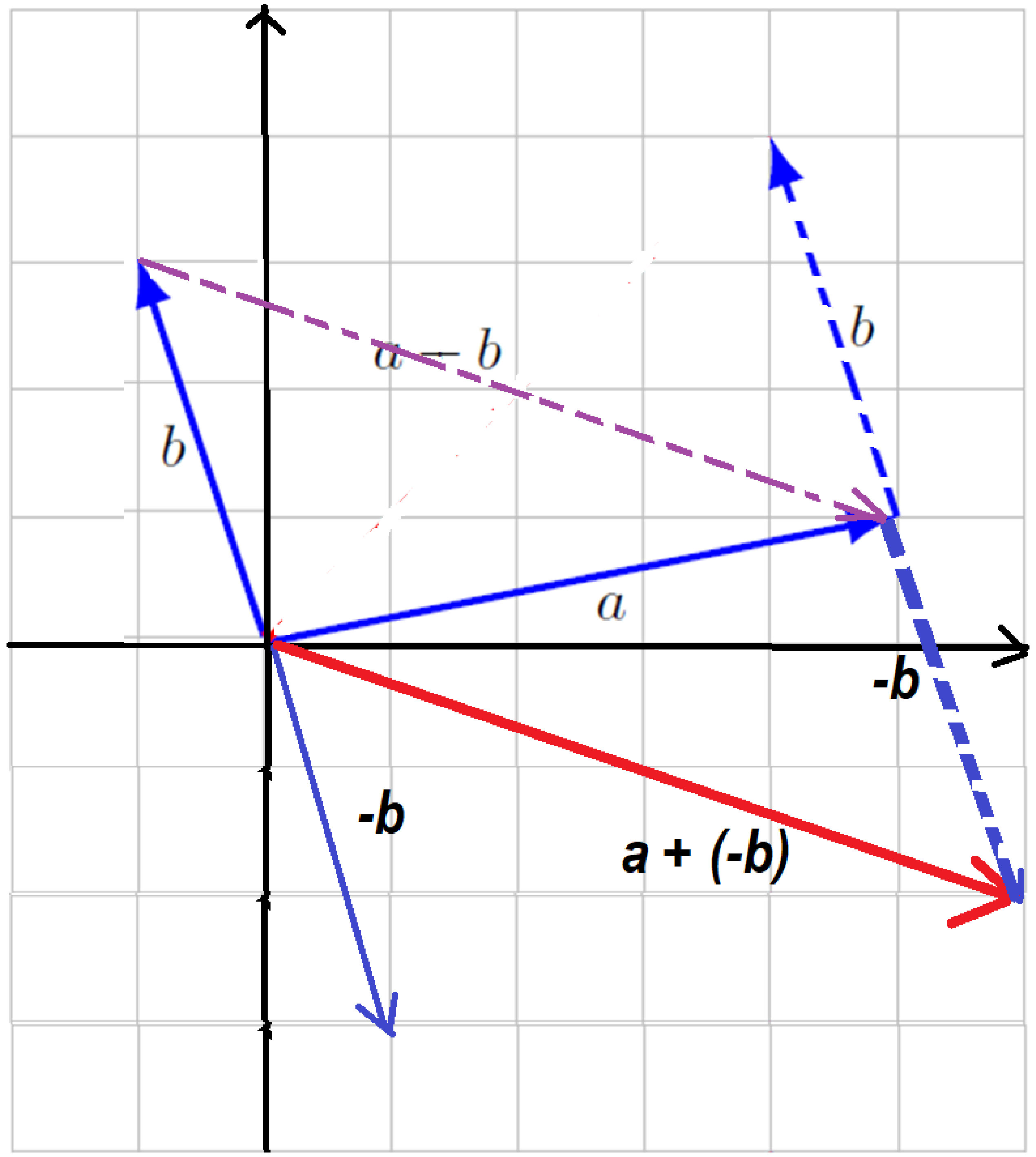
- Suppose you wrote your second year M.E. thesis as two documents and combined them into one doc before submission
- Suppose a and b are the word count vectors for doc 1 & 2 respectively
- What do these represent?
  - a+b
  - a-b

Vector subtraction: Geometric meaning



- Geometric meaning of vector subtraction
  - “Other” parallelogram diagonal
- https://www.geogebra.org/calculator/qmayzjm9

Vector subtraction: Geometric meaning



- Very Important geometric meaning of vector subtraction
  - $a-b$ is distance vector between two vectors
- What is the magnitude of distance vector?

Recall that

$$\|x\|_2 = (\sum_{i=1}^{n} {x_i}^2)^{\frac{1}{2}}$$

$$\|a-b\|_2 = (\sum_{i=1}^{n} (a_i - b_i)^2)^{\frac{1}{2}}$$

Problem 4 – (Assignment 2)

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

- Solve this and upload the solution to your github ALA assignments repo

- Which patient is farthest from the rest?
- Which two patients are nearest?
- What is the time complexity?
- Given a new patient, which is the closest patient?

Questions: Does it make any semantic sense to add

- Homogeneous data (e.g. monthly rainfall)
  - Rainfall data – Geometrically what is this?
  - Two BP vectors at different times for patient dataset
- Heterogeneous data
  - Adding Patient 1 to Patient 2?
- What about averaging?

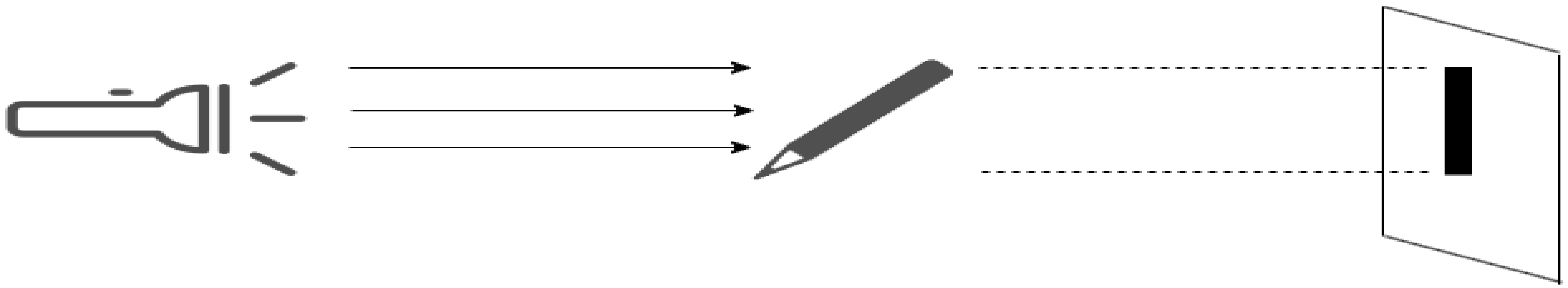
Dot Product (Inner product) definition

- Dot product of two vectors a and b

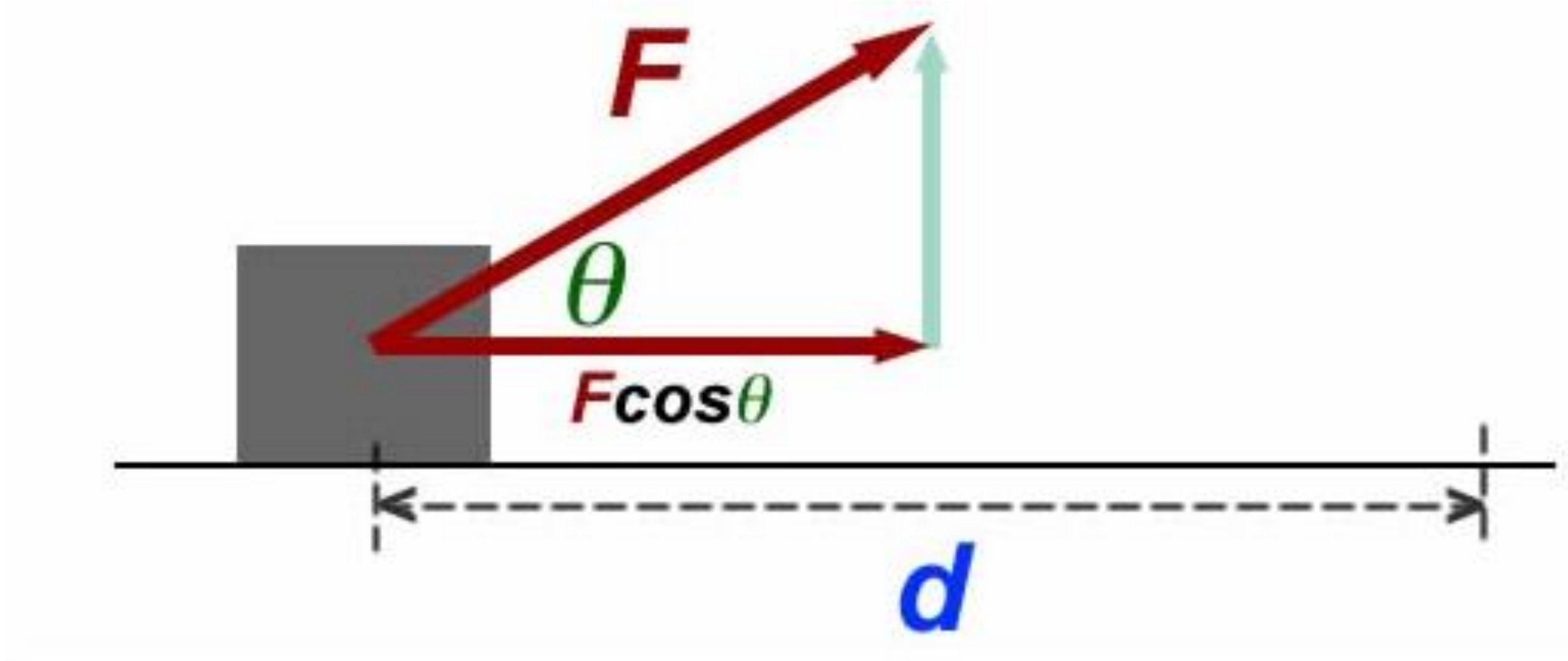
$$a = \begin{bmatrix} a_1 \\ a_2 \\ .. \\ a_n \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ .. \\ b_n \end{bmatrix} \qquad a^T b = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$$

- Very useful as a single measure of separation of two vectors in higher dimensions
- Demo in Geogebra – Check orientation of 3 sentences
- https://www.geogebra.org/calculator/chnjqzq5

Dot Product (Inner product) as projection



- Projection of something onto something



Work done =

$$F\ cos\theta\ d = F\ d\ cos\theta$$

$$f = \begin{bmatrix} Fcos\theta \\ Fsin\theta \end{bmatrix} \qquad s = \begin{bmatrix} d \\ 0 \end{bmatrix} \qquad f^T s = Fcos\theta\ d + Fsin\theta\ 0 = Fcos\theta\ d$$

Dot Product (Inner product) definition

- Dot product of two vectors a and b

$$a = \begin{bmatrix} a_1 \\ a_2 \\ .. \\ a_n \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ .. \\ b_n \end{bmatrix}$$

$$a^T b = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$$

$$a^T b = \|a\| \; \|b\| \; cos\theta$$

We will derive this in later lectures

- Angle between two vectors $\theta = arccos\left(\frac{a^T b}{\|a\| \; \|b\|}\right)$

- Very useful as a single measure of separation of two vectors in higher dimensions

Assignment 3

- You will get a movie review data set
  - name,review
- Write a program that identifies top 3 pairs of movies that are very much alike
- Hint:
  - Calculate movie review vector
  - Find similarity using
    - dot product formula followed by
    - angle between two vector formula
- Pandas for data read, numpy for coding.

Takeaways

- Geometric & algebraic meaning of a n-vector
- Formula for magnitude of a vector
- Vector is always column vector for convenience
- Scalar multiplication, vector addition, & subtraction
- Averaged vector is centroid vector
- Subtraction gives difference vector.
- Magnitude of difference vector is distance between vectors
- Word count vector
- Dot product – projection link
- Leads to single number for similarity in n-dimensions


QUESTIONS


Thank You!