وزارة الإتصالات
وتكنولوجيا المعلومات

Linear Algebra for Data Science

AI TALENTS HUB

EPITA
ÉCOLE D'INGÉNIEURS EN INFORMATIQUE

Foundation Period

AI Foundation

- Knowledge & AI History
- Data Exploration & Preparation

Mathematical Foundation

- Probability For Machine Learning
- Linear Algebra For Data Science
- Numerical Algorithms
- Optimization For Data Science

Development Foundation

- Introduction To Python
- Java Programming Basics (Online)
- Java & UML Programming
- Algorithm Workshop: Initiation
- Linux Administration

LEARNING OUTCOMES

- Connection between linear algebra and data science
- Linear system of equation and linear transformations: intuition of vectors and matrices
- Matrix/matrix, vector/vector sum and matrix/vector, scalar/vector, scalar/matrix products
- Linear combination, linear independence
- Vector space
- Basis and dimension
- Determinant: definition and properties
- Inversion of matrices
- Eigenvectors and eigenvalues

Structure

1. Linear algebra fundamentals

These concepts are at the core of Data Science.

Datasets are represented as matrices

Many ML algorithms are based on the concept of “gradient” which is a vector.

2. Vector spaces, basis and dimension

Concept of vector space will lead later the idea of what a basis is.

These concepts are useful to understand how PCA works.

3. Determinant, inverse, eigenvalues and eigenvectors of a matrix

Advanced tools and concepts.

Eigenvalues and eigenvectors which can be seen as the DNA of a matrix are a key element in the PCA algorithm and other data science or ML techniques.

Note on Course Structure

- Adjustments can be made depending on the level of the students.
- If all the students are already confident with a lot of content, new topics may be introduced.

Outline – Session 1

- Introduction
  - Relation between linear algebra and data science
  - Examples
- Basics of Linear Algebra
  - Vectors
  - Matrices
- Matrix Row Echelon Form
- System of Linear Equations
- Solving System of Linear Equations using Gaussian Elimination
- Introduction to NumPy

Outline – Session 2

- Linear Combinations
- Vector Space and Subspaces
- Linear Span
- Linear Independence
- Basis of Vector Space
- Orthonormal Basis (Changing Basis)

Outline – Session 3

- Linear Transformation
- Matrix Rank
- Matrix Determinant
- The Properties of the Determinant
- Matrix Inverse
- Orthonormal and Non-Orthonormal Space

Outline – Session 4

- Gram-Schmidt Process
- Transformation in Non-Orthonormal Space
- Eigenvalues and Eigenvectors
- Changing to Eigen Basis (Diagonalization)

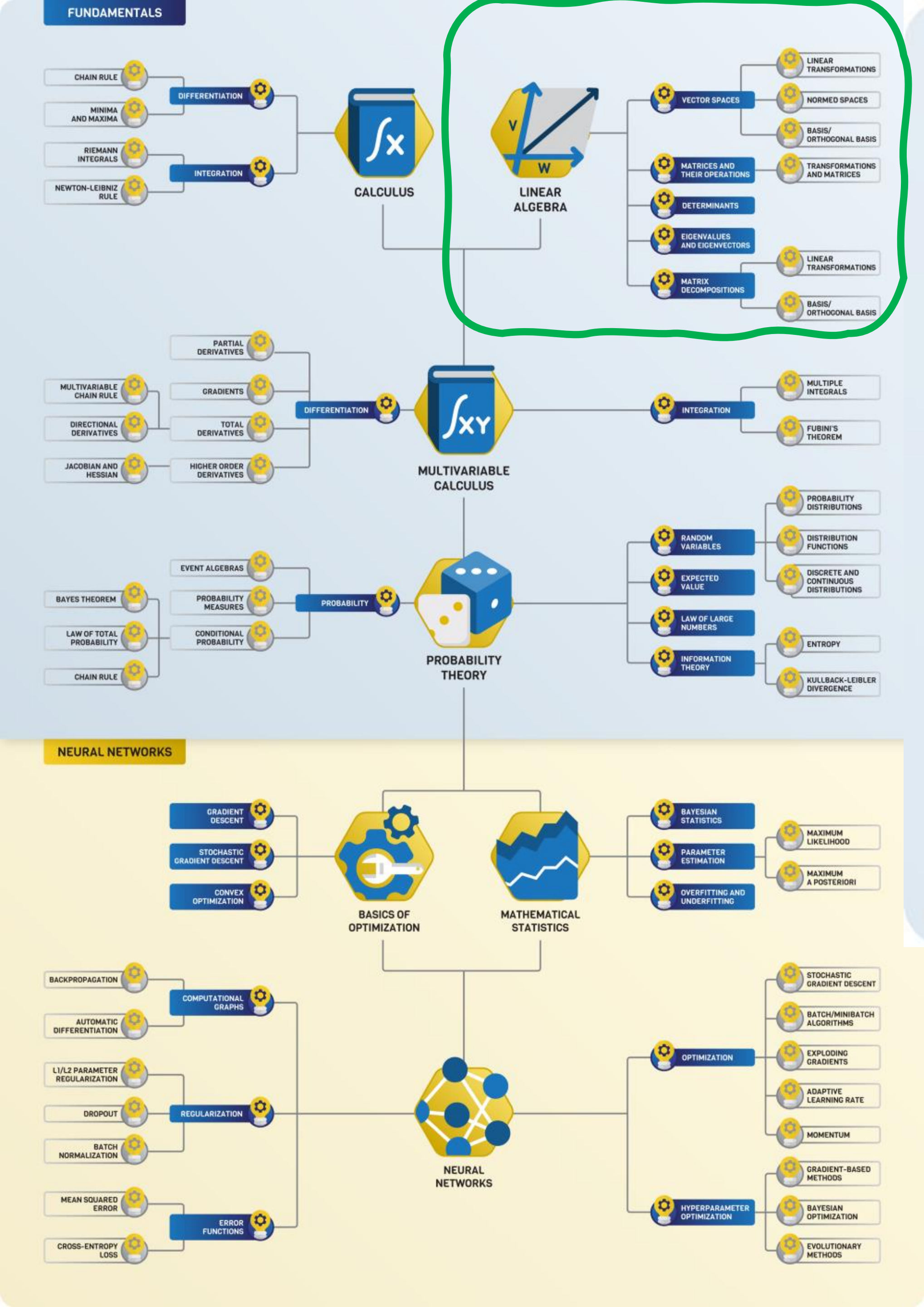
Outline – Session 5

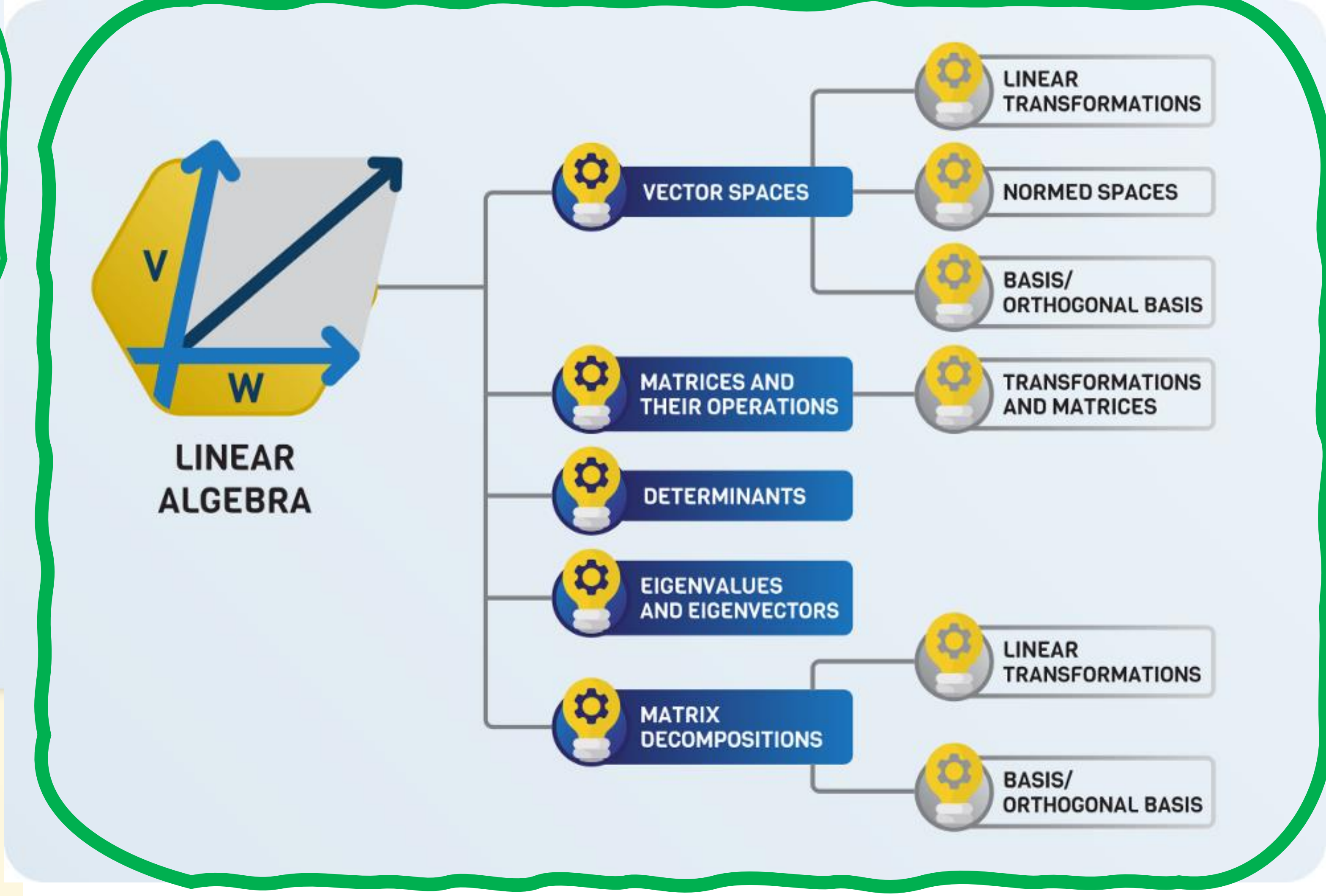
- Singular Value Decomposition (SVD)
- Dimensionality Reduction;
  - Principal Component Analysis (PCA)
- PCA Using Eigenvectors
- PCA Using SVD

Profiling

- How much do you know about:
    - Linear Algebra?
    - Linear transformation and matrices?
    - Dimensionality Reduction?

Rate yourself from 1 (minimum knowledge) to 10.





The roadmap for data science

https://towardsdatascience.com/the-roadmap-of-mathematics-for-deep-learning-357b3db8569b

Why Linear Algebra in data science/ML/AI?

Connection between linear algebra and data science?

- Linear algebra is at the core of data science and is largely used in powerful ML algorithms; (e.g., Gradient Descent & Convex Optimization)
- Core to LLMs, recommendation engines (Netflix, Amazon…), NLP (Alexa, Siri…), computer vision, etc.
- Weights and Inputs in NNs are represented as matrices.
- Vectorization.
- Transformations.
- Regression.
- Dimensionality Reduction (e.g., PCA)
- Computational Efficiency (e.g., in CNNs)
- Graph Neural Networks (GNNs)
- What else?

Introduction

Scalars, Vectors, Matrices, and Tensors

$$5$$

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}$$

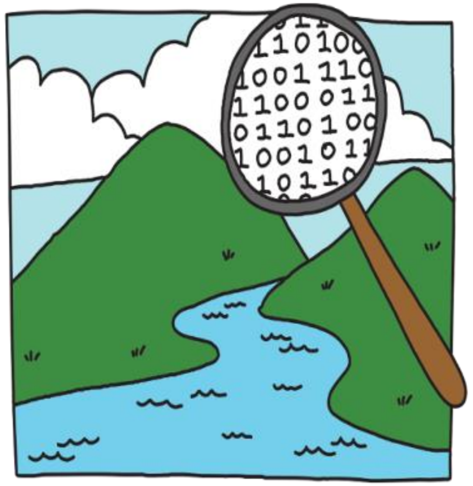
$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}$$

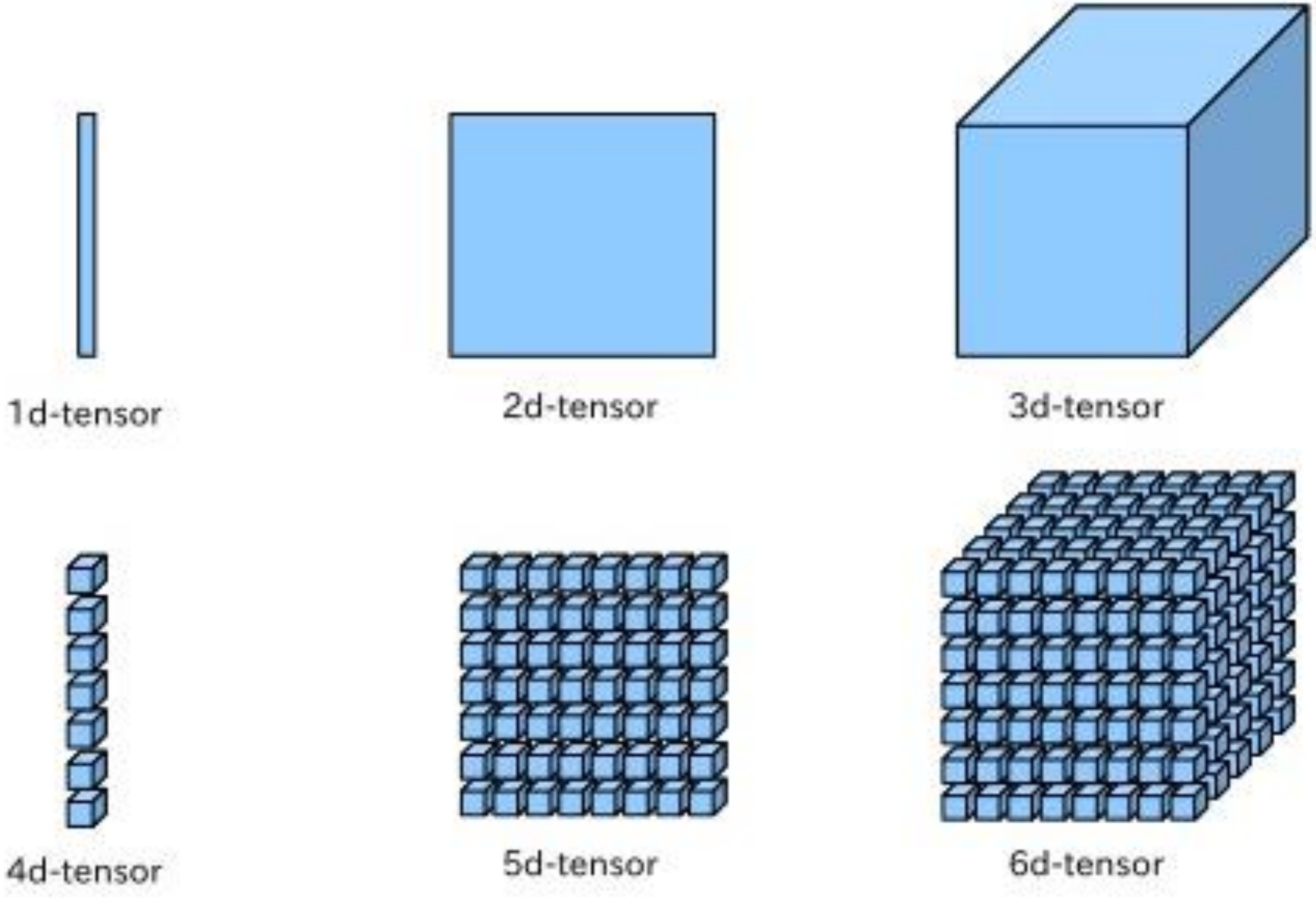
- Scalars are numbers
- Vectors are linear elements in a linear space with very specific properties
- Matrix can be viewed as sequence of vectors
- Matrices map vectors to vectors in a linear manner
- Tensors are generalized multi-dimensional linear maps

Vectors, Matrices, and Tensors

- **In Computer Science:**

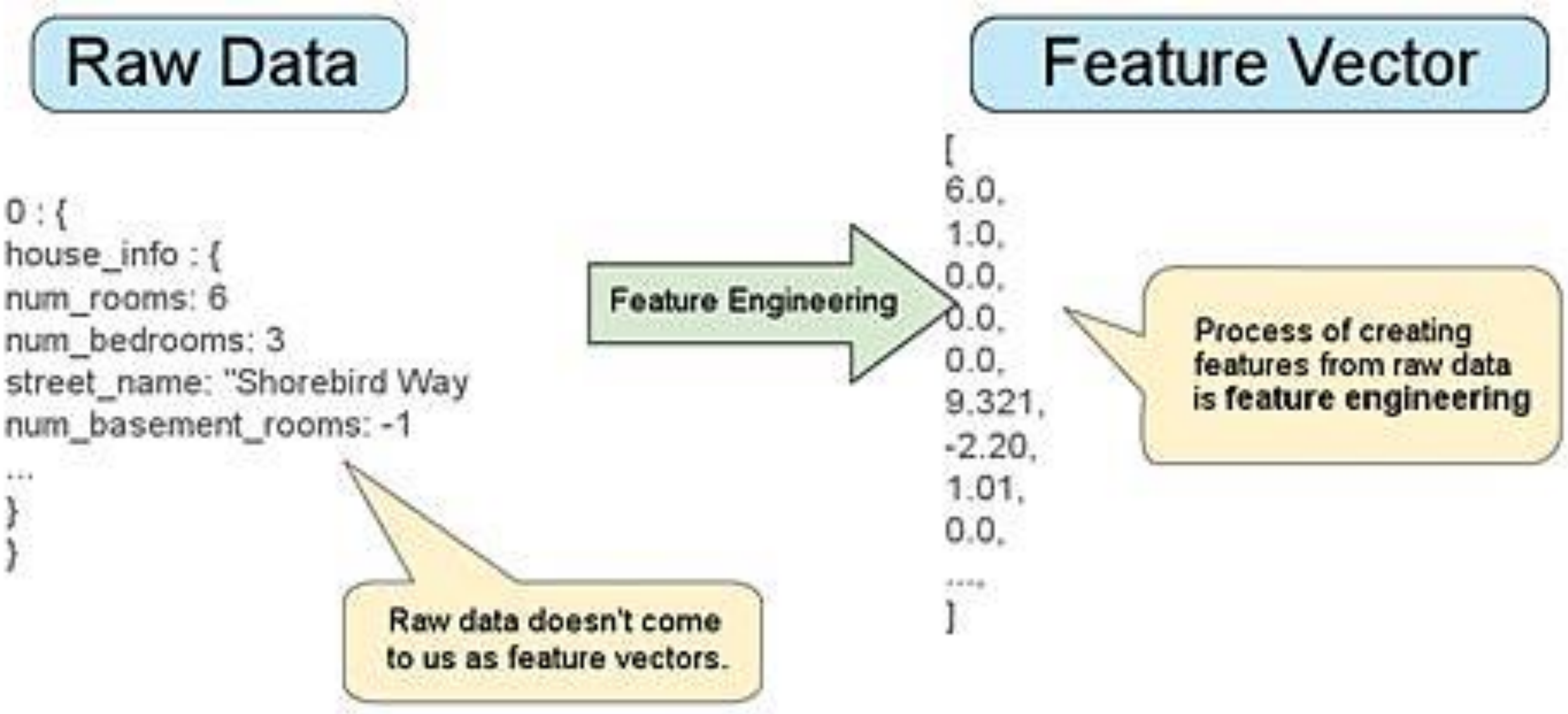
  are collections of related numbers arranged mainly for data representation and manipulation





Uses of Vectors and Matrices

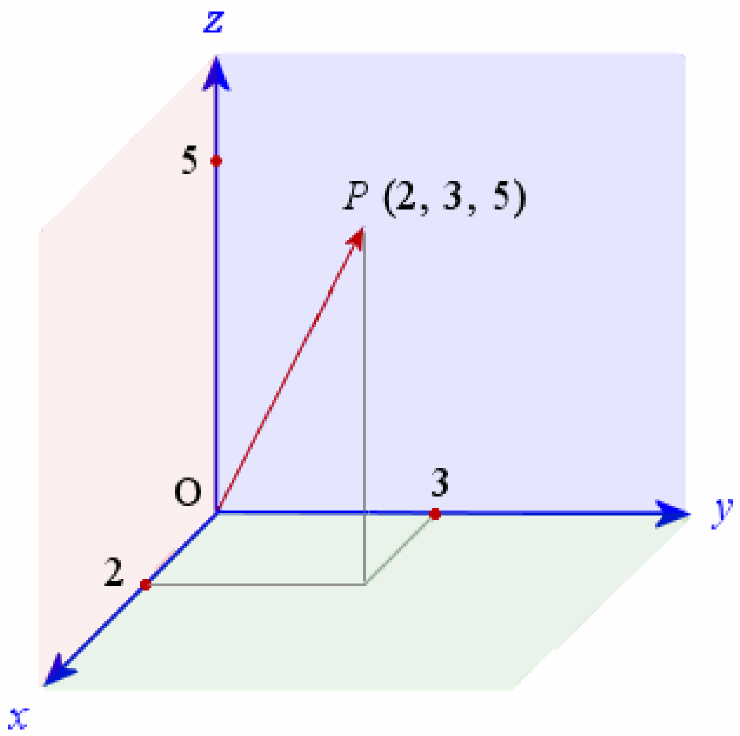
- Vectors can represent:
  - Multi-dimensional feature data.
  - Geometric directions in space.
  - M

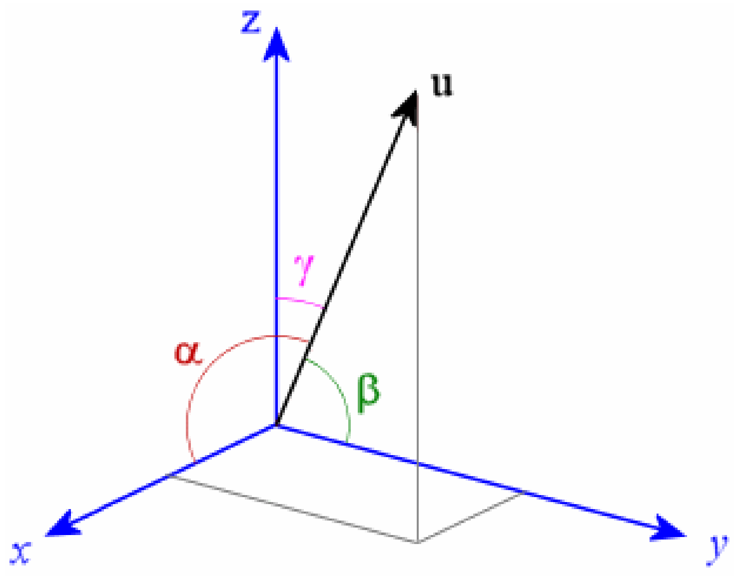


Uses of Vectors and Matrices

- Vectors can represent:
  - Multi-dimensional feature data.
  - Geometric directions in space.
  - Model coefficients.

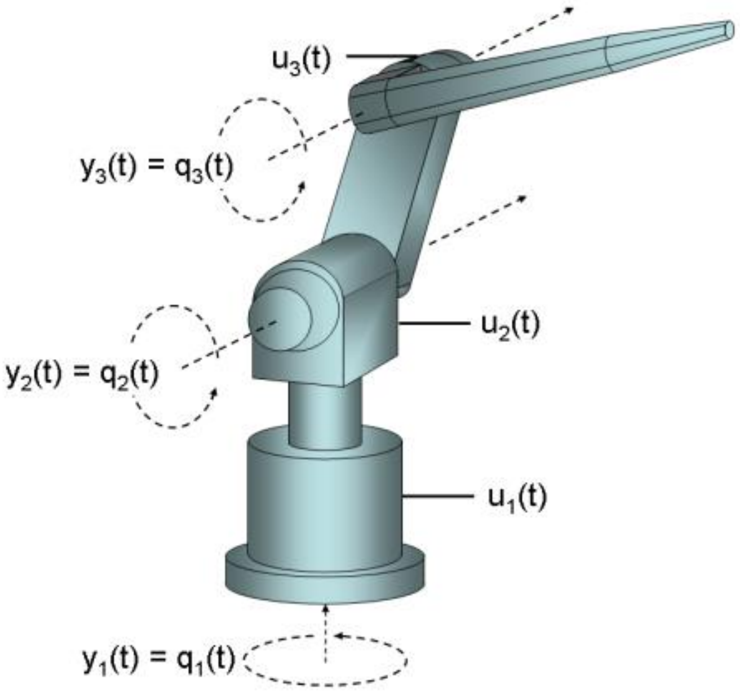
$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

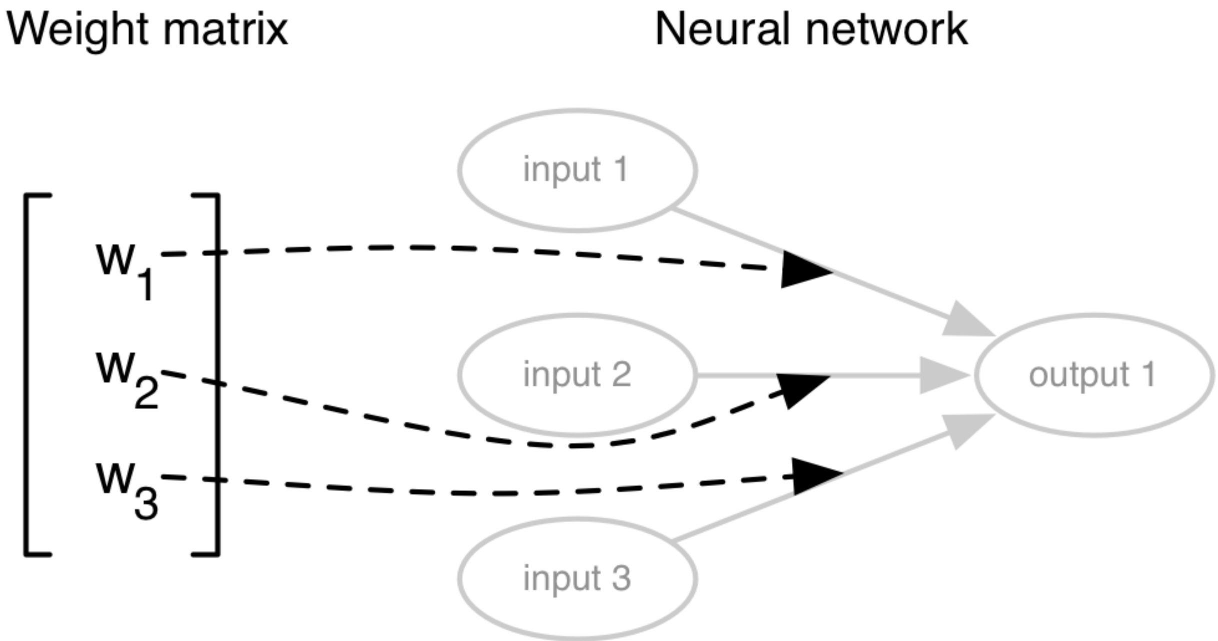




Uses of Vectors and Matrices

- Vectors can represent:
  - Multi-dimensional feature data.
  - Geometric directions in space.
  - Model coefficients (e.g., in robotics, or neural networks).





Uses of Vectors and Matrices

- Matrices can represent:
  - Linear equation systems (see applications in the next slide).
  - Linear Transformations.
  - Images.
  - Graphs.

$$\begin{aligned} x_1 - x_2 + x_3 &= 1 \\ 2x_2 - x_3 &= 1 \\ 2x_1 + 3x_2 \quad &= 1 \end{aligned}$$

$$\underbrace{\begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & -1 \\ 2 & 3 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{B}$$

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 3 & 3 & -1 \\ -2 & -2 & 1 \\ -4 & -5 & 2 \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{B} = \begin{bmatrix} 5 \\ -3 \\ -7 \end{bmatrix}$$

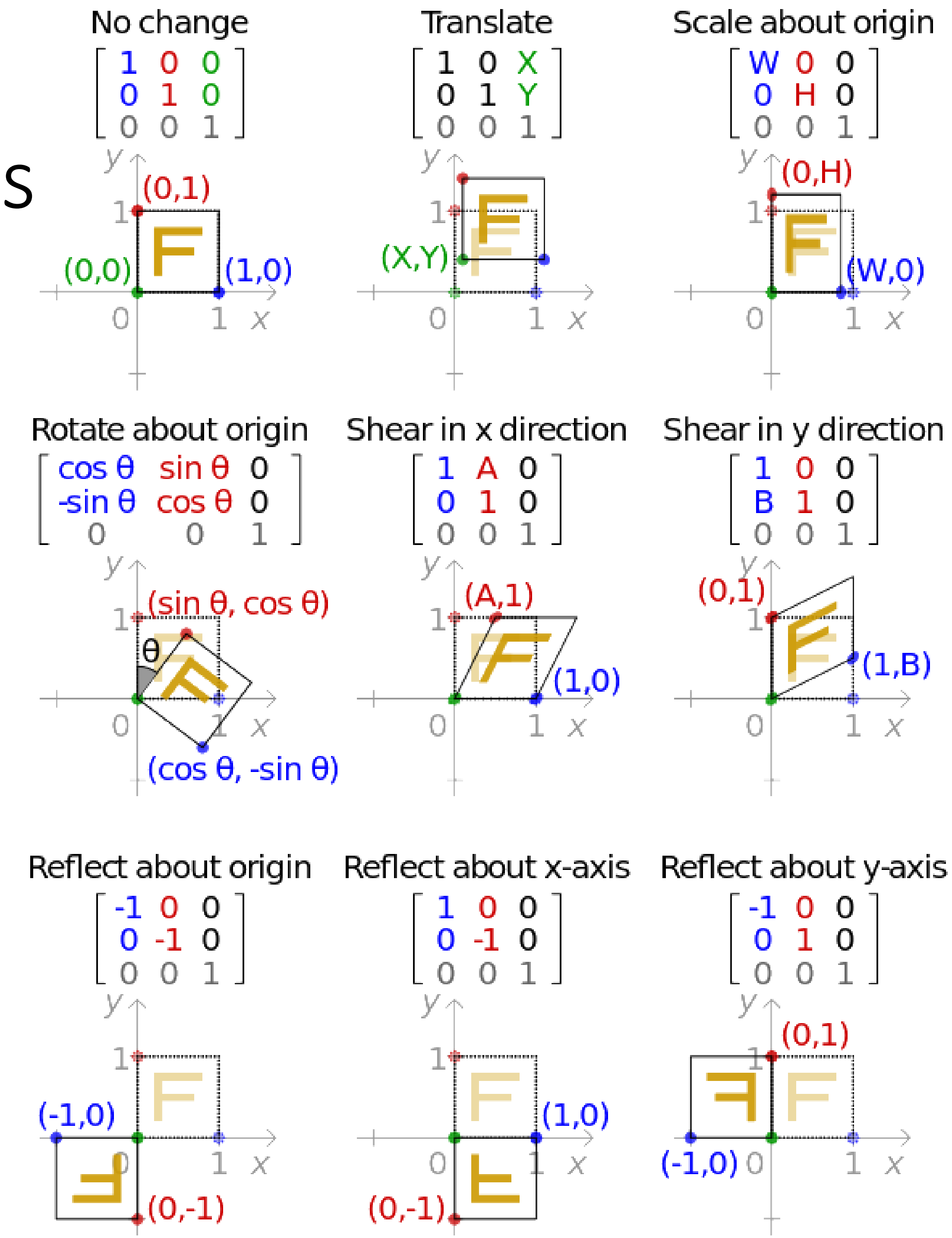
Applications for solving linear systems of equations

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 3 & 3 & -1 \\ -2 & -2 & 1 \\ -4 & -5 & 2 \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{B} = \begin{bmatrix} 5 \\ -3 \\ -7 \end{bmatrix}$$

- Linear Programming
- Planning
- Decision Engineering
- Regression
- Forecasting (time series, financial, ….)
- Natural Language processing

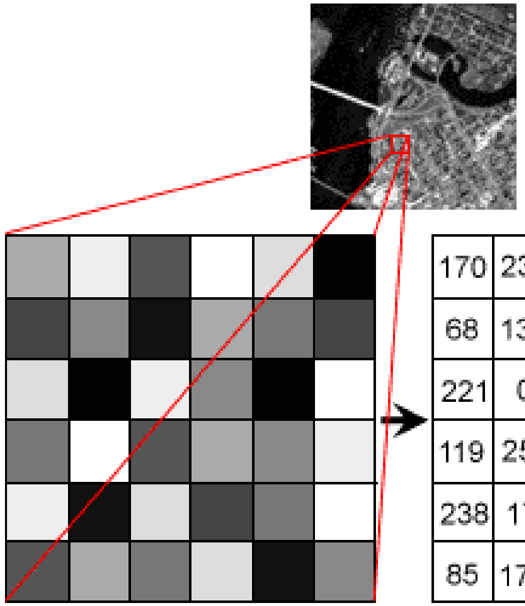
Uses of Vectors and Matrices

- Matrices can represent:
  - Linear equations.
  - Linear Transformations.
  - Images.
  - Graphs.



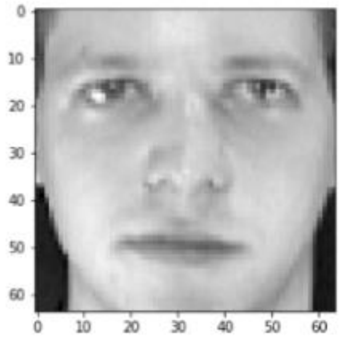
Uses of Vectors and Matrices

- Matrices can represent:
  - Linear equations.
  - Linear Transformations.
  - Images.
  - Graphs.



170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

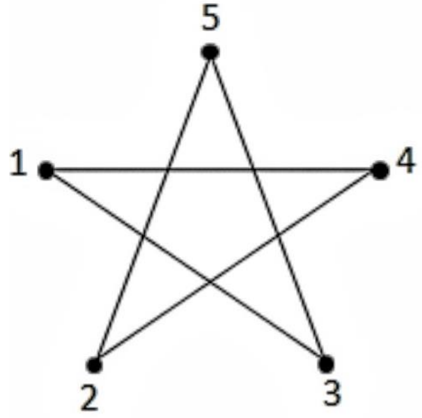
$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

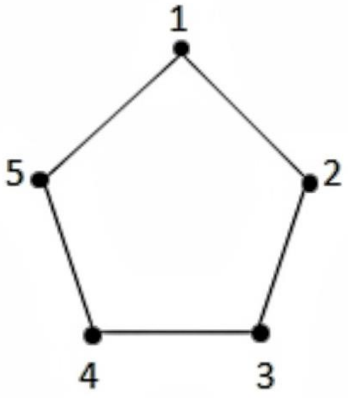


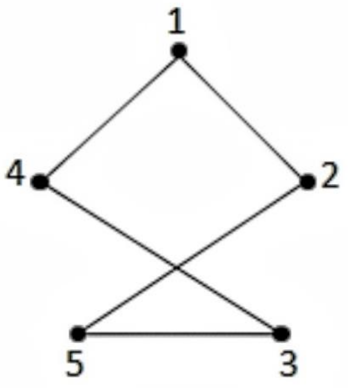
$$= \begin{bmatrix} 0.310 & 0.368 & ... & 0.306 \\ 0.343 & 0.405 & ... & 0.314 \\ \vdots & \vdots & ... & \vdots \\ 0.202 & 0.207 & ... & 0.157 \end{bmatrix}$$

Uses of Vectors and Matrices

- Matrices can represent:
  - Linear equations.
  - Linear Transformations.
  - Images.
  - Graphs.







$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

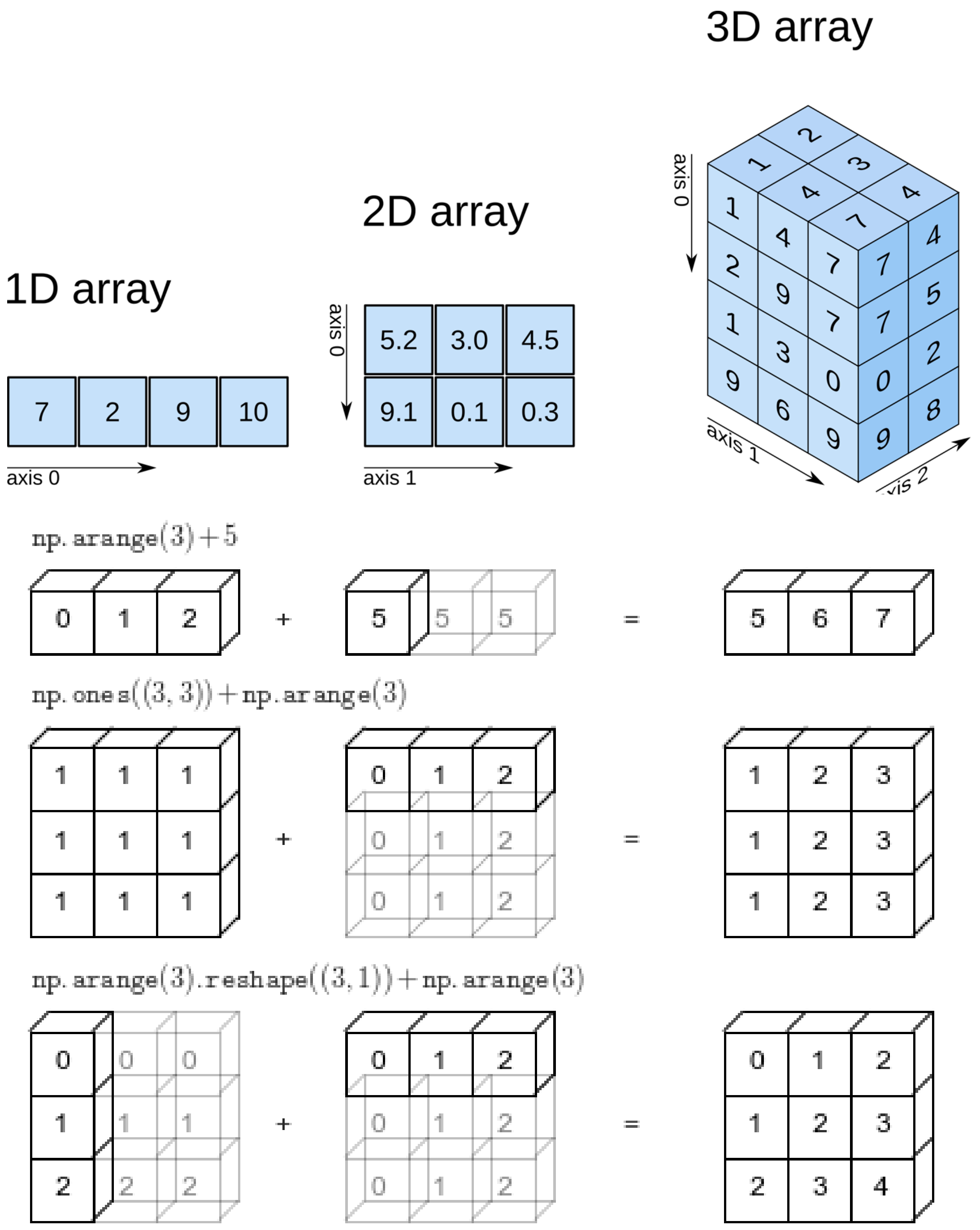
	1	**2**	**3**	**4**	**5**
1			1	1	
2				1	1
3	1				1
4	1	1			
5		1	1		

	1	**2**	**3**	**4**	**5**
1		1			1
2	1		1		
3		1		1	
4			1		1
5	1			1	

	1	**2**	**3**	**4**	**5**
1		1		1	
2	1				1
3				1	1
4	1		1		
5		1	1		

NumPy

- A library for Python
- Support for large, multi-dimensional arrays and matrices
- Fast processing of multidimensional arrays.
- Functions and operators for these arrays.
- Linear algebra and random number generation.
- **Vectorization**
- **Broadcasting?**



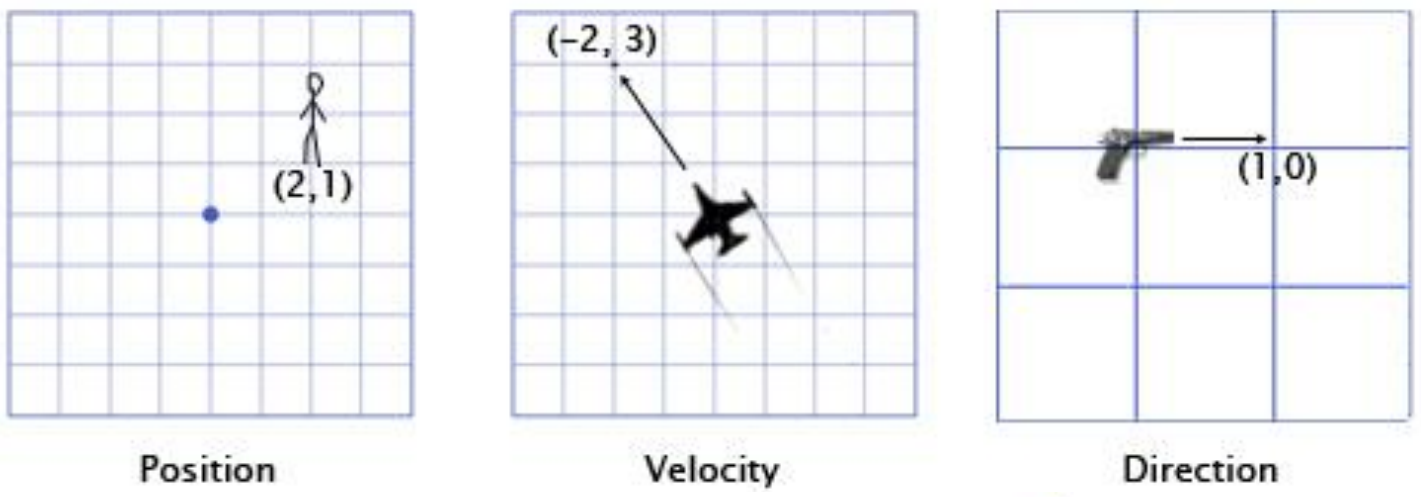
Linear Algebra

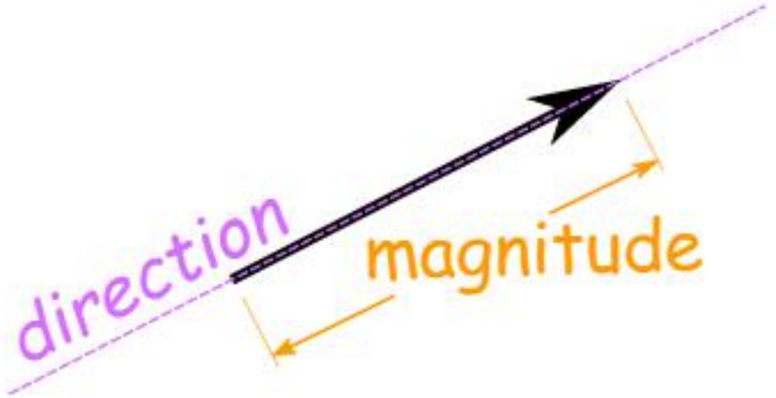
Session 1: Basics of Linear Algebra

Part 1 - Vectors

- Vector representations
- Vector scaling
- Vector addition
- Vector subtraction
- Vector multiplication; dot product
- Vector norms

What is a scalar and what is a vector?





- **Scalar:**
  - one-dimensional vector is a scalar
  - A quantity that has only magnitude and no direction.
    Unlike the vector that has direction and magnitude.
- **Vector:**
  - An array of numbers, either continuous or discrete

Scalar
24

Vector
[1,2,3,4]’

Vectors

- **"You can't add apples and oranges." In a strange way, this is the reason for vectors.**
- We have two separate numbers v_1 and v_2 (e.g., number of apples and number of oranges that a certain box contains).
- That pair produces a two-dimensional vector $\boldsymbol{v}$.

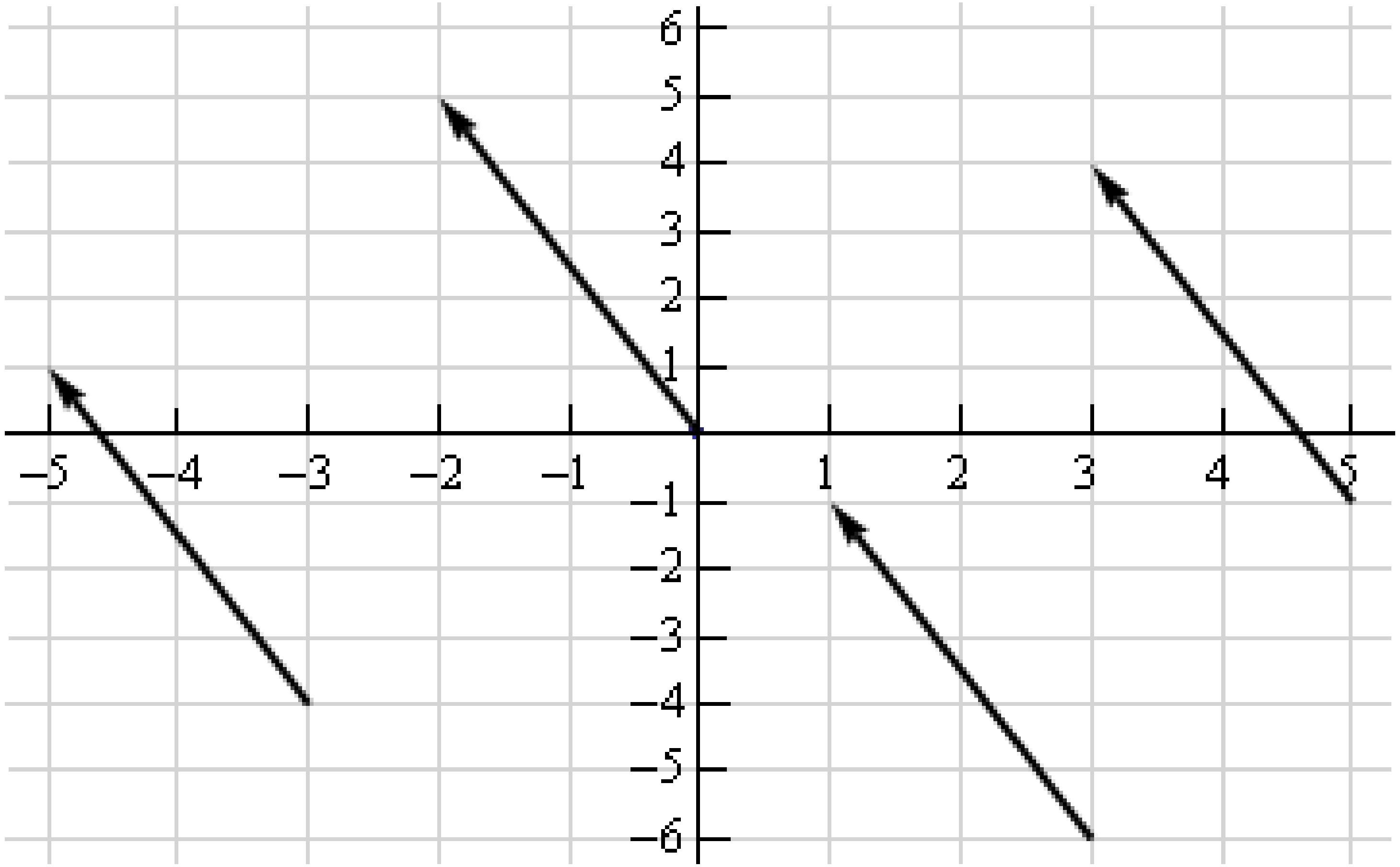
- Column vector $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ $\quad v_1 = first\ component\ of\ \boldsymbol{v}$
  $v_2 = second\ component\ of\ \boldsymbol{v}$

- Other representation of a vector, besides the boldface lowercase letter $\boldsymbol{v}$ include $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = (v_1, v_2)$.

We write $\boldsymbol{v}$ as a $\boldsymbol{column}$, not as a row. The main point so far is to have a single letter (in $\boldsymbol{boldface\ italic}$) for this pair of number v_1 and v_2 (in $lightface\ italic$).

Vector

The vector $\boldsymbol{v} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$ says move left 2 units and up 5 units



Vectors, Examples

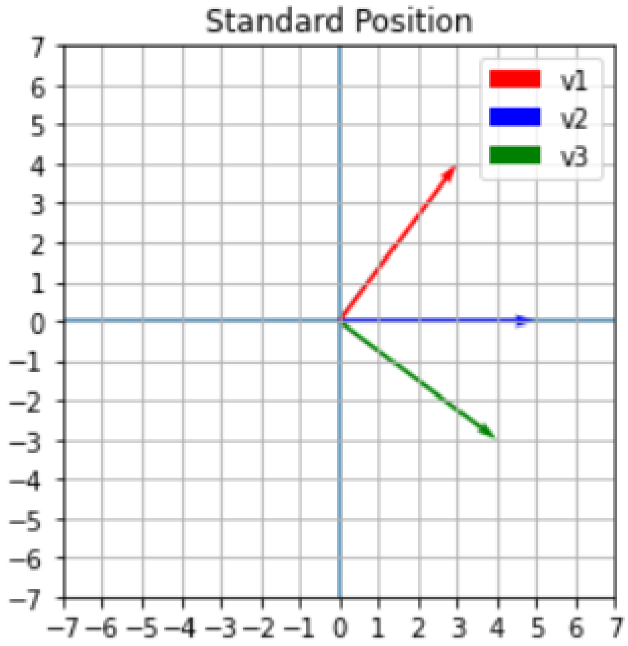
- $\boldsymbol{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} 4 \\ -3 \end{bmatrix}.$

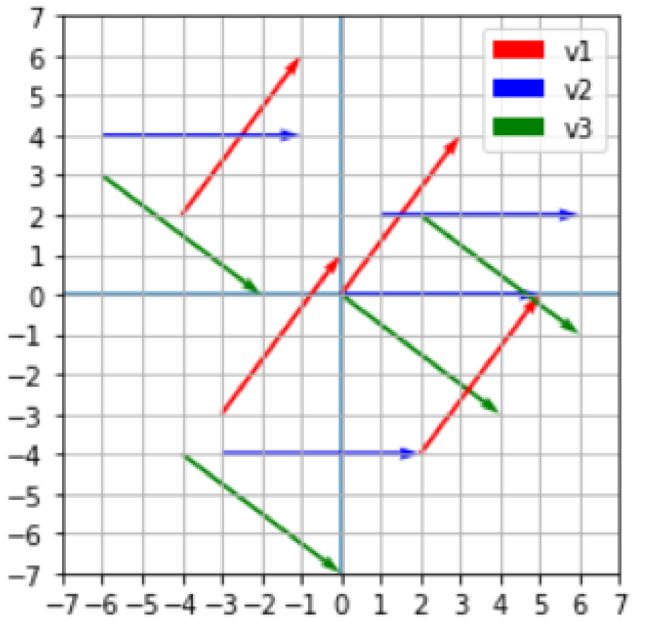
- $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are members of $\mathbb{R}^2$. $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in \mathbb{R}^2$.

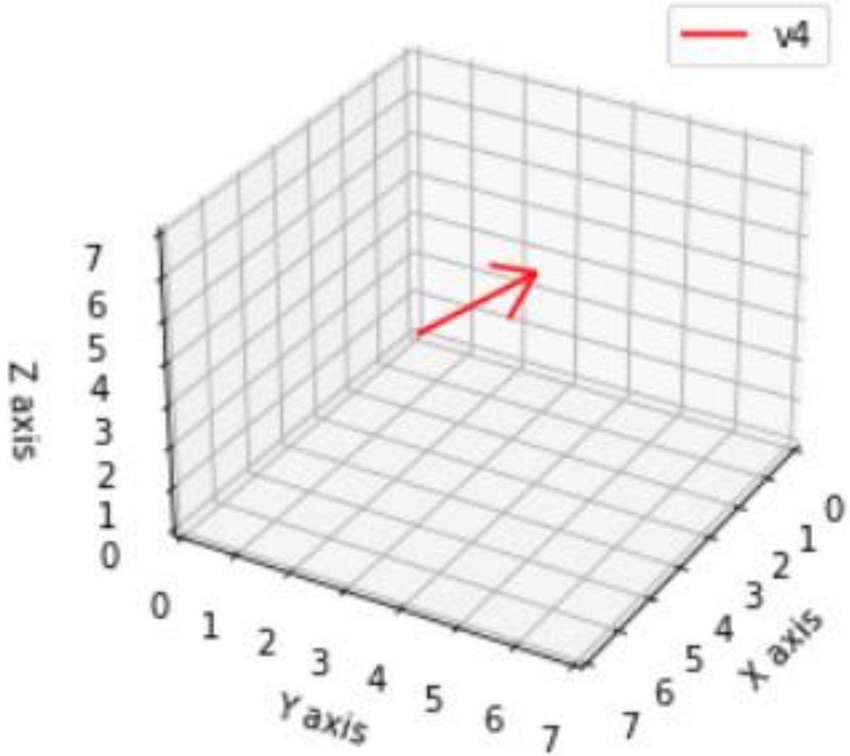
- $\boldsymbol{v}_4 = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \boldsymbol{v}_4 \in \mathbb{R}^3.$

- $\boldsymbol{v} \in \mathbb{R}^n$ has n components, $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

- ($\mathbb{R}^n$is the n-dimensional real coordinate space.)

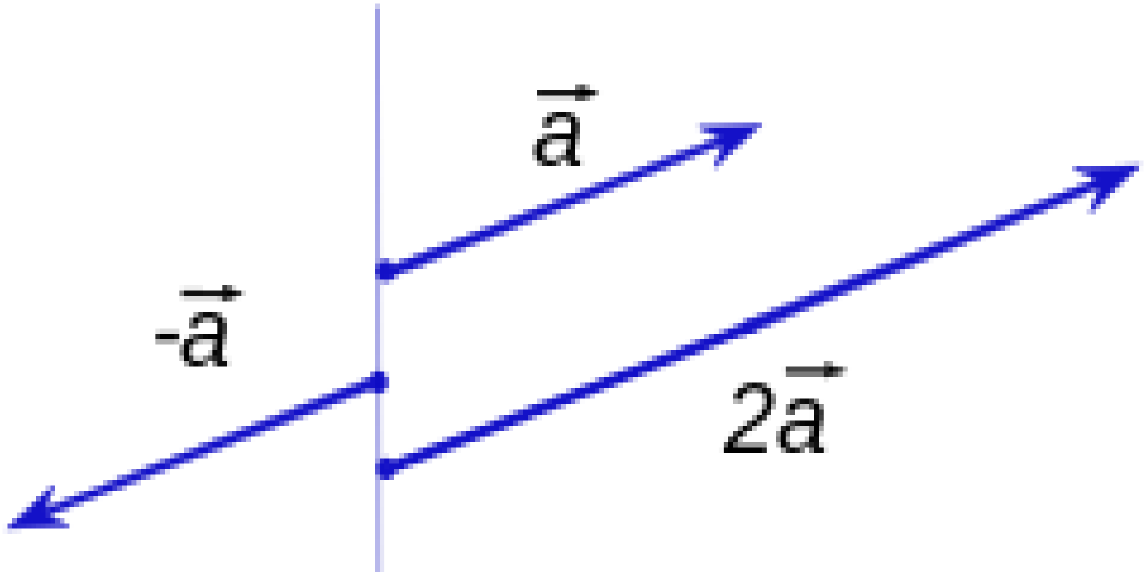






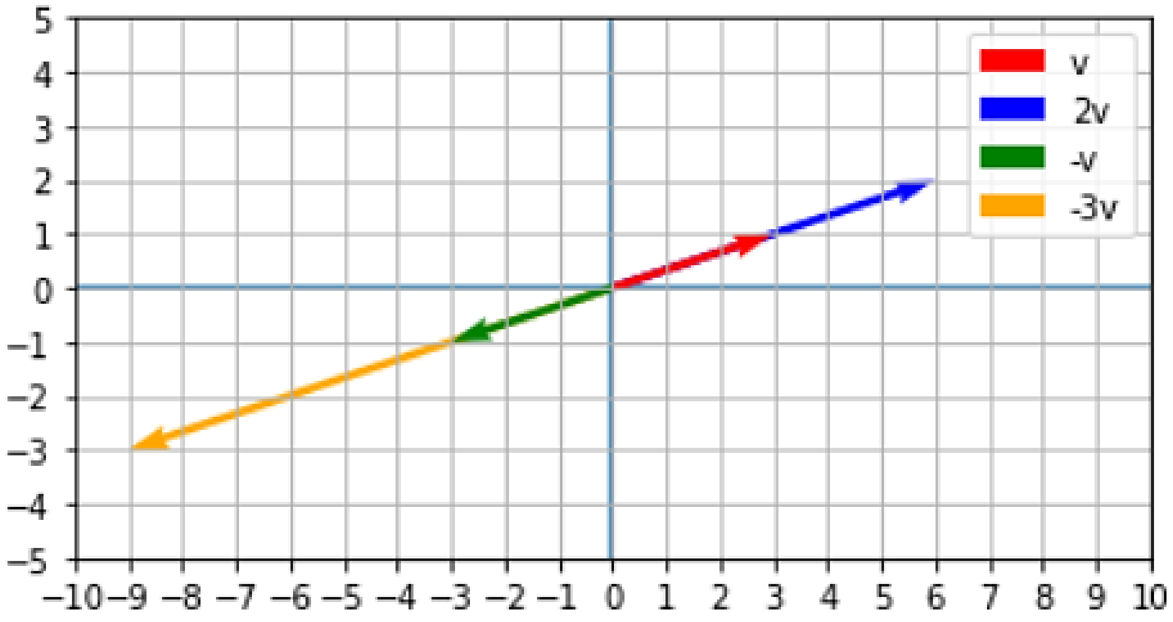
Vector Scaling

- We can multiply vectors by scalars to change the scale



Vector Scaling (Scalar Multiplication), Examples

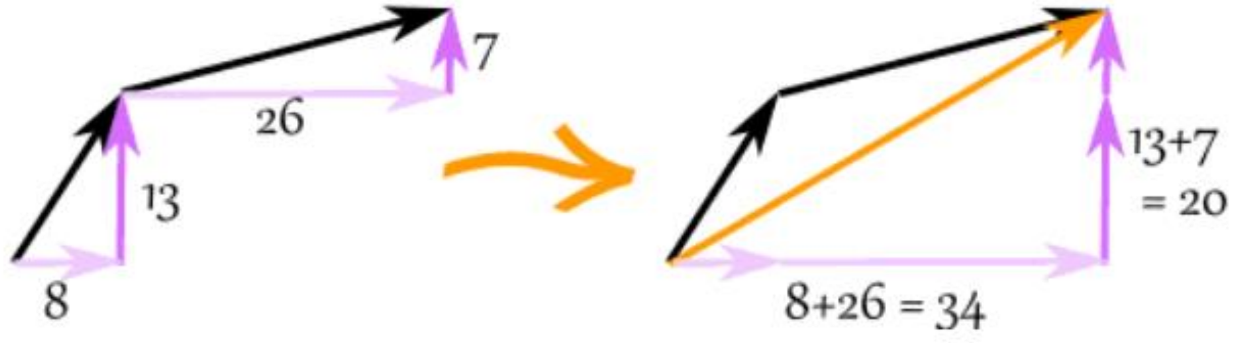
- Let $\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, Calculate $2\boldsymbol{v}$, $-\boldsymbol{v}$, $-3\boldsymbol{v}$?
  - $2\boldsymbol{v} = 2\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 \\ 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$;
  - Same direction, twice magnitude.
  - $-\boldsymbol{v} = -1\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \cdot 3 \\ -1 \cdot 1 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$;
  - flipped in direction – same magnitude.
  - $-3\boldsymbol{v} = -3\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \cdot 3 \\ -3 \cdot 1 \end{bmatrix} = \begin{bmatrix} -9 \\ -3 \end{bmatrix}$;
  - flipped in direction – triple magnitude.



Vector Addition

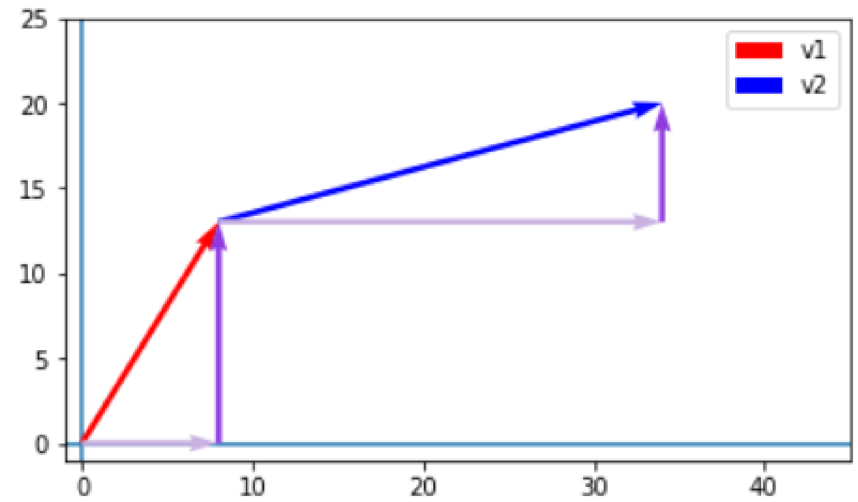
- **Algebraically (by using the components):**
  - Adding the vectors **[8, 13]’** and **[26, 7]**
  - Adding the x components 8+ 26 = 34
  - Adding the y components 7 + 13 = 20
  - The result is vector **(34,20)**

- **Or, Graphically as shown.**



The vector (8, 13) and the vector (26, 7) add up to the vector (34, 20)

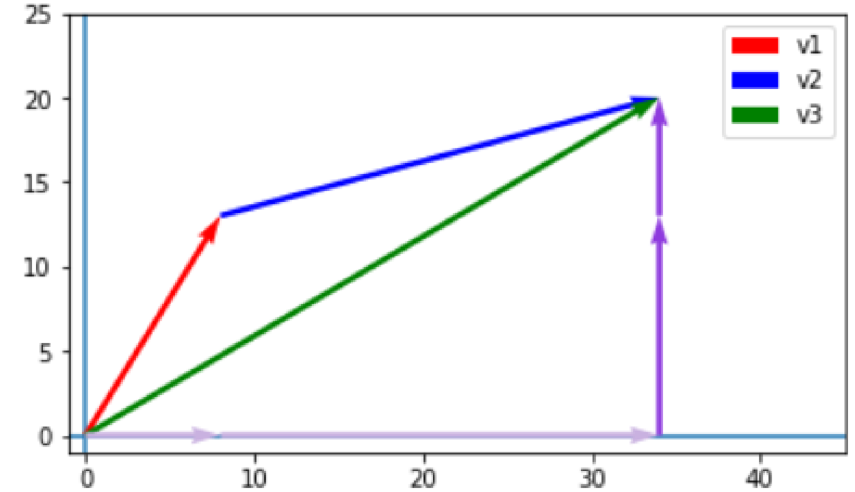
Vector Addition, Examples



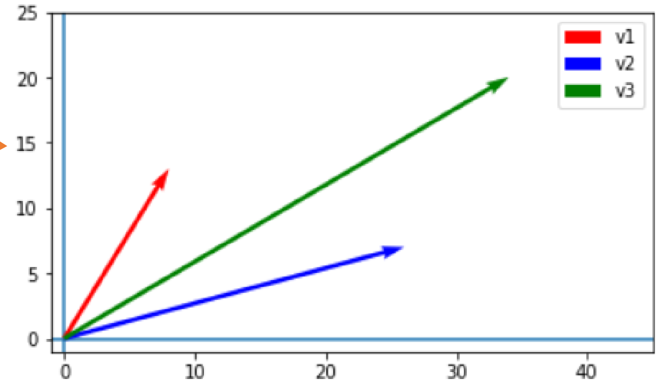
- Let $\boldsymbol{v}_1 = \begin{bmatrix} 8 \\ 13 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} 26 \\ 7 \end{bmatrix}$
- Find $\boldsymbol{v}_3 = \boldsymbol{v}_1 + \boldsymbol{v}_2$?

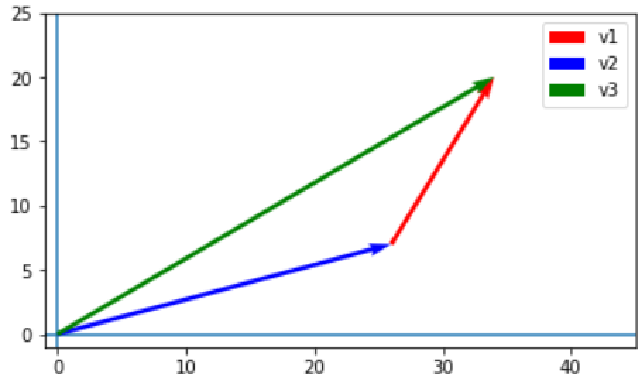
Solution:

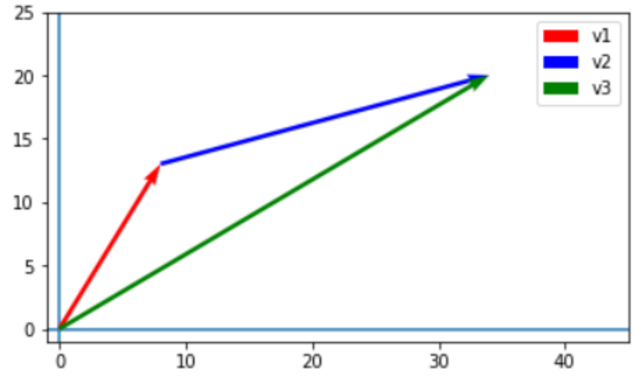
- Algebraically: $\boldsymbol{v}_3 = \boldsymbol{v}_1 + \boldsymbol{v}_2 = \begin{bmatrix} 8+26 \\ 13+7 \end{bmatrix} = \begin{bmatrix} 32 \\ 20 \end{bmatrix}$



- Graphically:







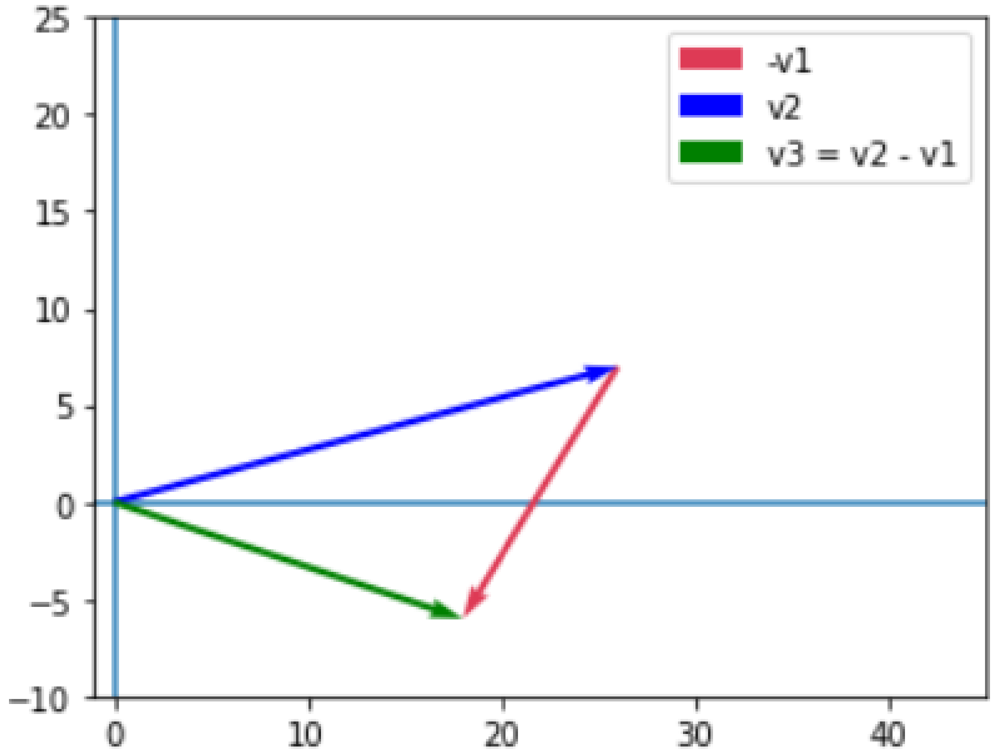
Vector Subtraction, Examples

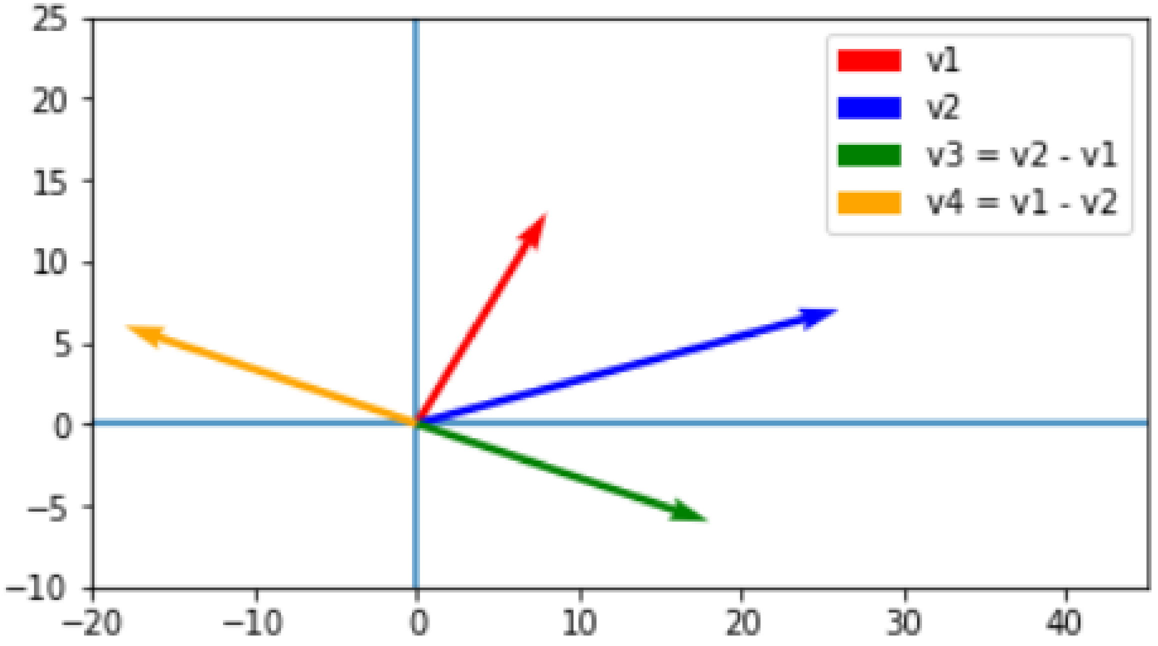
- Let $\boldsymbol{v}_1 = \begin{bmatrix} 8 \\ 13 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} 26 \\ 7 \end{bmatrix}$
- Find $\boldsymbol{v}_3 = \boldsymbol{v}_2 - \boldsymbol{v}_1$ and $\boldsymbol{v}_4 = \boldsymbol{v}_1 - \boldsymbol{v}_2$?

Solution:

- $\boldsymbol{v}_3 = \boldsymbol{v}_2 - \boldsymbol{v}_1 = v2 + (-v1) = \begin{bmatrix} 26-8 \\ 7-13 \end{bmatrix} = \begin{bmatrix} 18 \\ -6 \end{bmatrix}$

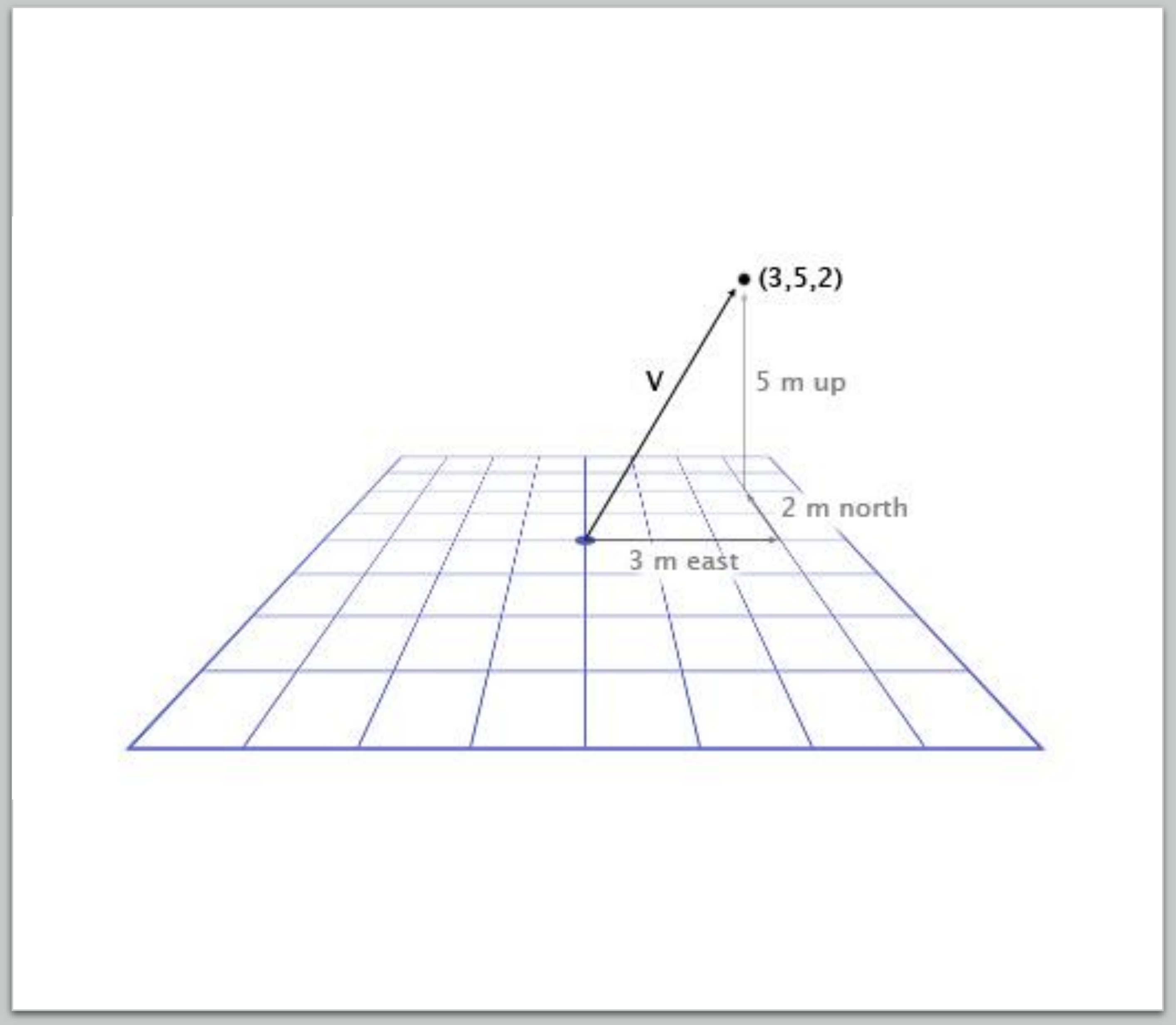
- $\boldsymbol{v}_4 = \boldsymbol{v}_1 - \boldsymbol{v}_2 = \begin{bmatrix} 8-26 \\ 13-7 \end{bmatrix} = \begin{bmatrix} -18 \\ 6 \end{bmatrix}$





Vector Addition in 3D space

- The displacement (shortest path between two points) the origin (0,0,0) and (3,5,2)
- To reach this point, we will have to go through three main vectors according to our coordinate system (x,y,z):
  first: (3,0,0),
  second: (3,0,2),
  third: (3,5,2)
- Or: (3,0,0) + (0,5,0) + (0,0,2) = (3,5,2)

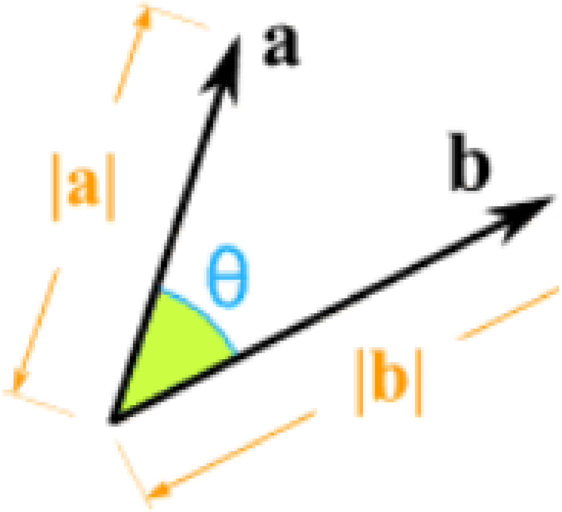


Summary: Vectors Addition and Scaling

- Vector Addition $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ add to $\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}$

- Scalar Multiplication c$\boldsymbol{v} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}$, $2\boldsymbol{v} = \begin{bmatrix} 2v_1 \\ 2v_2 \\ \vdots \\ 2v_n \end{bmatrix} = \boldsymbol{v} + \boldsymbol{v}$, $-\boldsymbol{v} = \begin{bmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{bmatrix}$

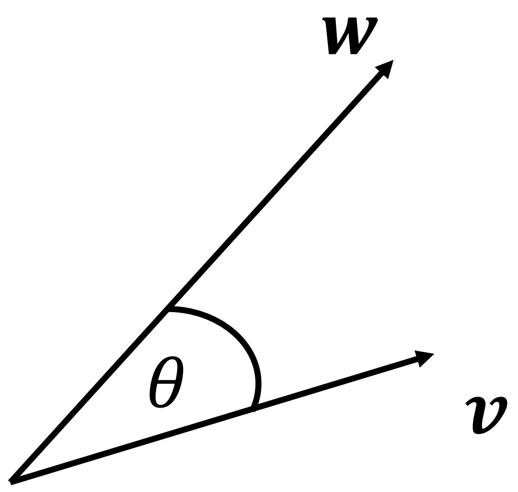
Vector Multiplication



Dot Product (Inner Product) : The result is Scalar

Cross Product: The result is Vector (not covered in the module)

Inner (Dot) Products and Cosines



- The **dot product** of the two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ is **a scalar number** $\boldsymbol{v} \cdot \boldsymbol{w} = \|\boldsymbol{v}\|\|\boldsymbol{w}\| \cos\theta$

where θ is the angle between the two non-zero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$

and $\|\boldsymbol{vector}\|$ is the norm of the vector (its length or magnitude)

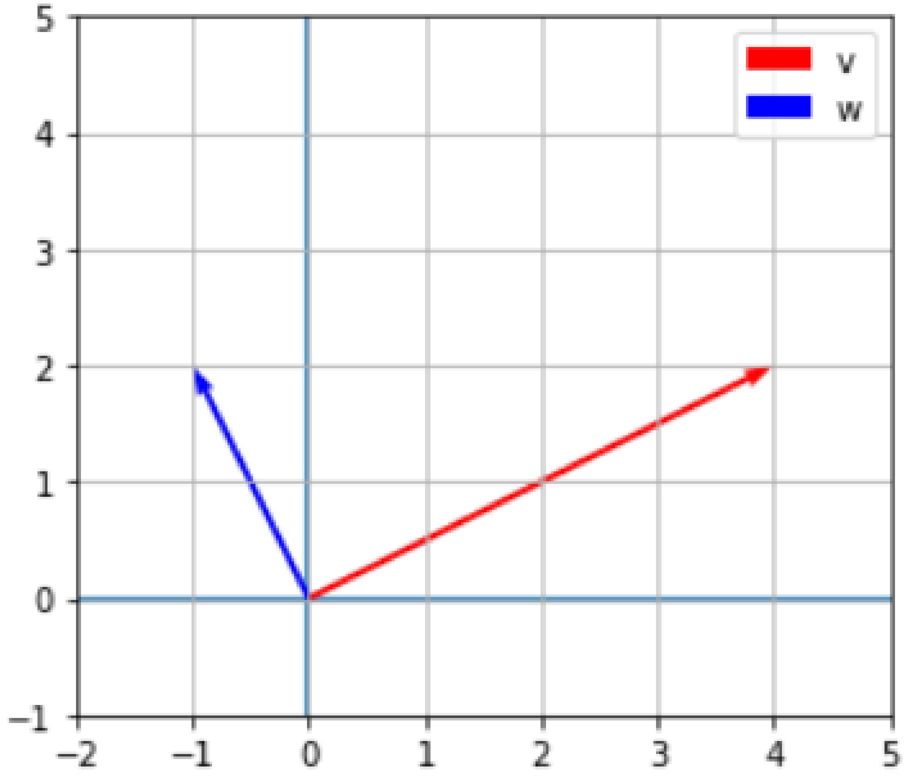
- The **dot product** of $\boldsymbol{v}$ and $\boldsymbol{w}$ can be written in terms of their components
- $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$ → $\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{k=1}^{n} v_k w_k$
- $\cos\theta = \frac{\boldsymbol{v}.\boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$

Dot Product Example

- The vectors $\boldsymbol{v} = (4,2)$ and $\mathbf{w} = (-1,2)$ have a '*zero*' dot product:

$$\begin{bmatrix}4\\2\end{bmatrix}.\begin{bmatrix}-1\\2\end{bmatrix} = -4 + 4 = 0$$

- (note that the angle between v, w is 90 degrees, they are orthogonal vectors).



Dot Product Example

- Two vectors in 3D space

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} v_2 = \begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix}$$

$$v_1 . v_2 = v_1^T v_2 = 1 \times 3 + 2 \times 5 - 3 \times 1 = 10$$

Inner Products and Cosines

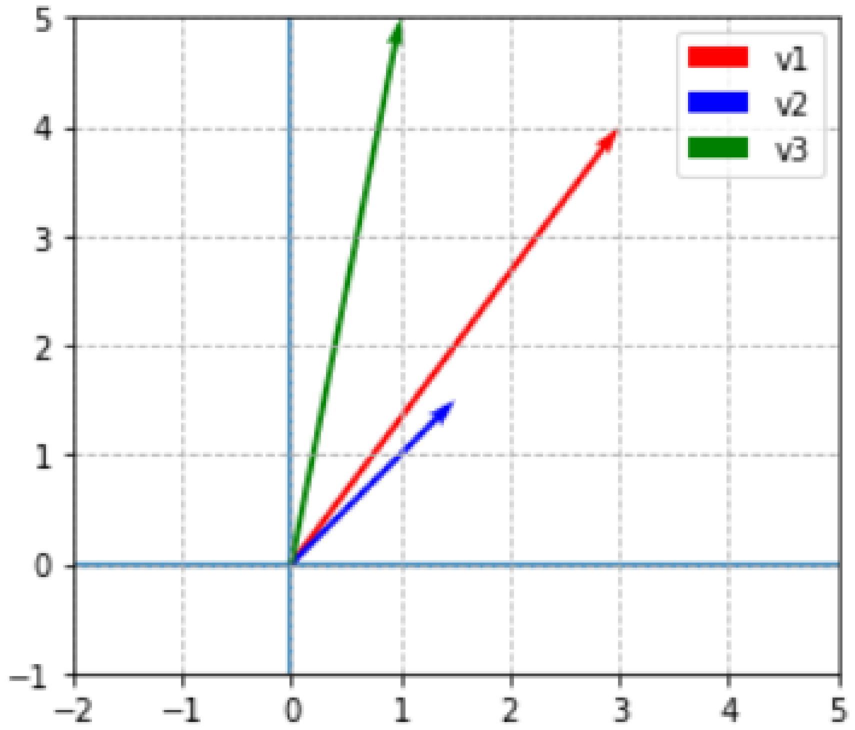
- **Cosine similarity** is a measure of similarity between two non-zero vectors irrespective of their length.

- **Example:** $\boldsymbol{v}_1 = (3,4), \boldsymbol{v}_2 = (1.5,1.5), and\ \boldsymbol{v}_3 = (1,5)$

  Similarity($\boldsymbol{v}_1, \boldsymbol{v}_2$) = 0.9899

  Similarity($\boldsymbol{v}_1, \boldsymbol{v}_3$) = 0.9021

  Similarity($\boldsymbol{v}_2, \boldsymbol{v}_3$) = 0.8321



Vectors: Dot Product Properties

1. Commutative property

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$$

2. Distributive property

$$(\boldsymbol{v} + \boldsymbol{w}) \cdot \boldsymbol{x} = \boldsymbol{v} \cdot \boldsymbol{x} + \boldsymbol{w} \cdot \boldsymbol{x}$$

3. Associative property

$$(\mathrm{c}\boldsymbol{v}) \cdot \boldsymbol{w} = c(\boldsymbol{v} \cdot \boldsymbol{w})$$

   where c is a scalar value.

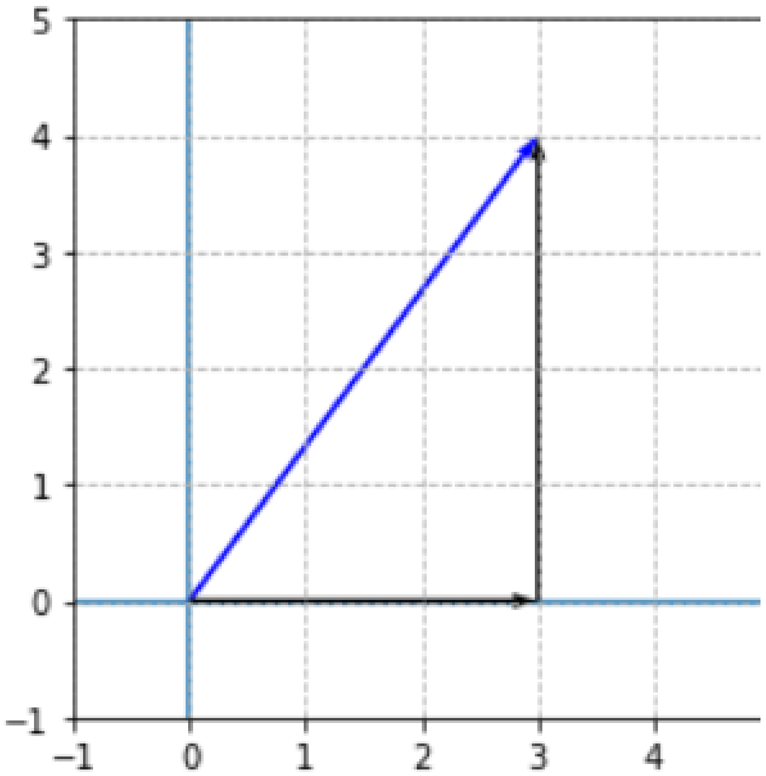
Vectors: Lengths and Norms

- An important case is the dot product of a vector with itself.
- The dot product $\boldsymbol{v} \cdot \boldsymbol{v} = ||\boldsymbol{v}||^2$ is the length squared.

Example; $\boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, ||\boldsymbol{v}||^2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} . \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 9 + 16 = 25$

- **Definition:** The **length** $||\boldsymbol{v}||$ of a vector $\boldsymbol{v}$ is the square root of $\boldsymbol{v} \cdot \boldsymbol{v}$:

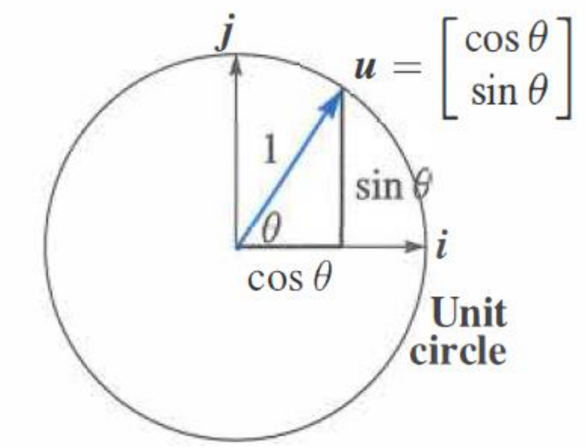
$$length = ||\boldsymbol{v}|| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = (v_1^2 + v_2^2 + \cdots + v_n^2)^{\frac{1}{2}}$$

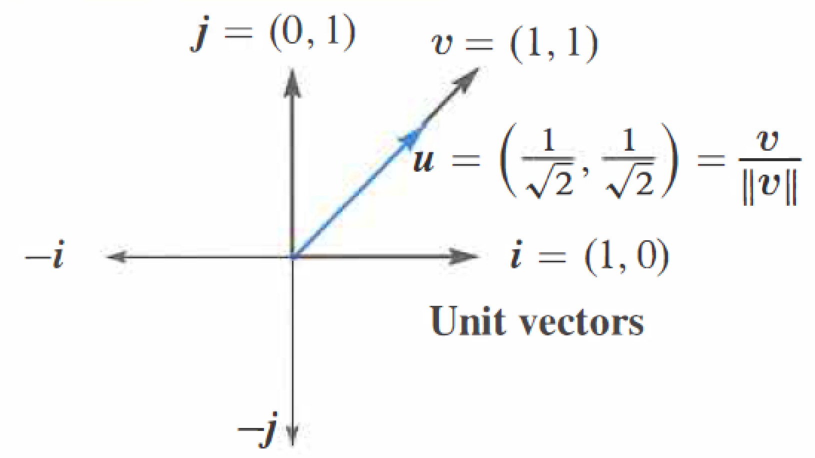


Vectors: Lengths and Norms, the unit vectors

$$\boldsymbol{i} = \begin{bmatrix}1\\0\end{bmatrix} \; and \;\; \boldsymbol{j} = \begin{bmatrix}0\\1\end{bmatrix} \; and \;\; \boldsymbol{u} = \begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$$

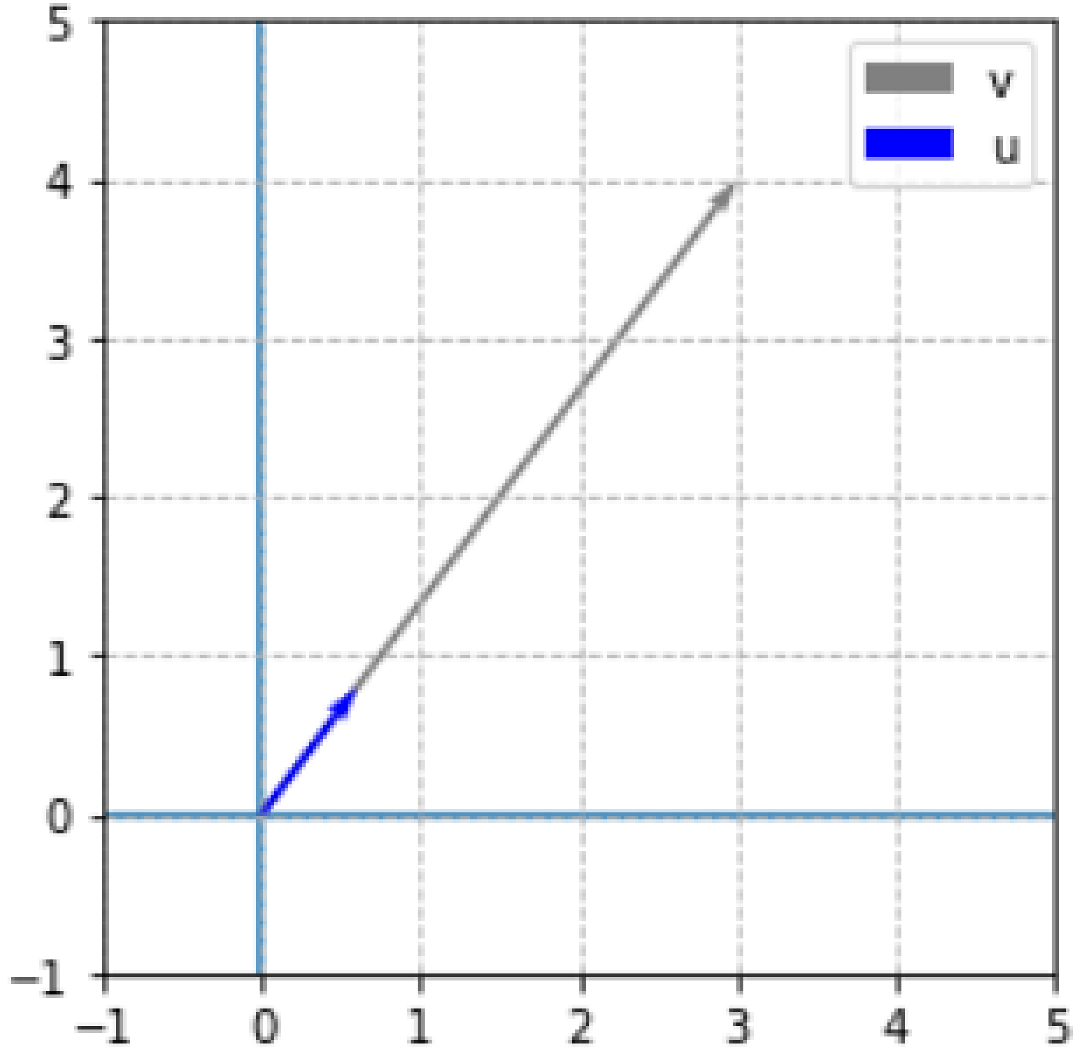
- A unit vector is a vector whose length equals one unit.
- The standard unit vectors along the x and y axes are written $\boldsymbol{i}$ and $\boldsymbol{j}$.
- In the xy plane, the unit vector that makes an angle "theta" with x axis is $(\cos\theta, \sin\theta)$, because $\cos^2\theta + sin^2\theta = 1$
- These vectors reach out the unit circle. Thus, $cos\,\theta$ and $sin\,\theta$ are simply the coordinates of that point at angle θ on the unit circle.





Vectors: Lengths and Norms, the unit vectors

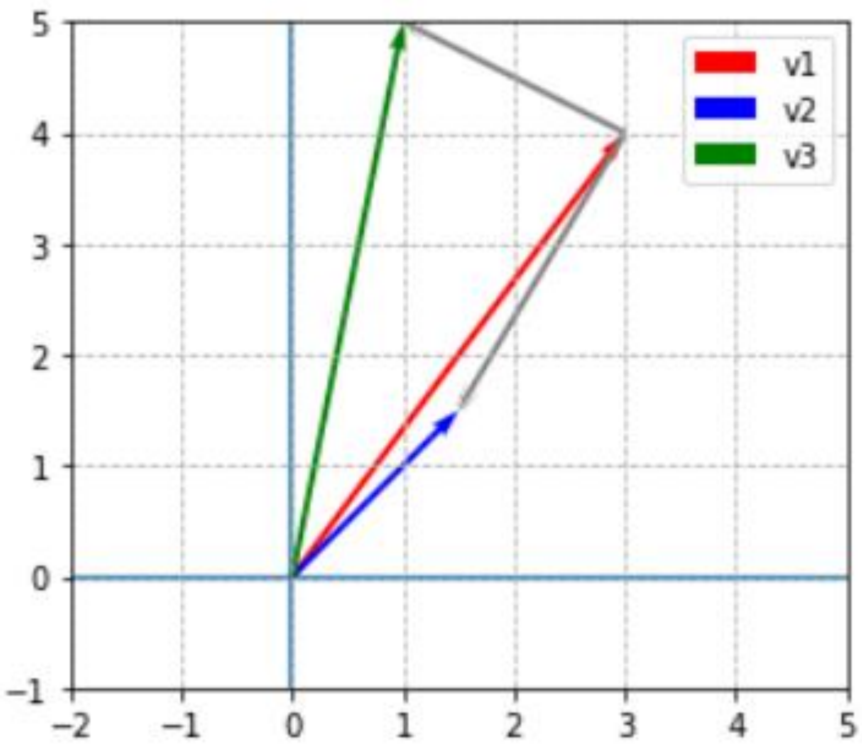
- Any non-zero vector can be converted to a unit vector by scaling it by its norm. This is called "normalization".
- $\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ is a unit vector in the same direction as $\boldsymbol{v}$.



Vectors: Lengths and Norms

- The norm we discussed so far is the **Euclidean norm (the L^2 norm, 2-norm, or ℓ_2 norm)**.
- The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points.
- When computed for a vector difference, e.g., for vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^n$ where $\boldsymbol{v}_1 = (v_{11}, v_{12}, \dots, v_{1n})$ and $\boldsymbol{v}_2 = (v_{21}, v_{22}, \dots, v_{2n})$, the Euclidean distance is defined as:

$$\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2 = \|\boldsymbol{v}_1 - \boldsymbol{v}_2\| = \sqrt{\sum_{i=1}^{n} (v_{1i} - v_{2i})^2}$$



Example: Calculate the Euclidean distance between $\boldsymbol{v}_1$ & $\boldsymbol{v}_2$ and $\boldsymbol{v}_1$ & $\boldsymbol{v}_3$

$\boldsymbol{v}_1 = (3,4), \boldsymbol{v}_2 = (1.5,1.5),$
$and\ \boldsymbol{v}_3 = (1,5)$

$$Distance(\boldsymbol{v}_1, \boldsymbol{v}_2) = 2.915$$
$$Distance(\boldsymbol{v}_1, \boldsymbol{v}_3) = 2.236$$

Vectors: Norms of a vector

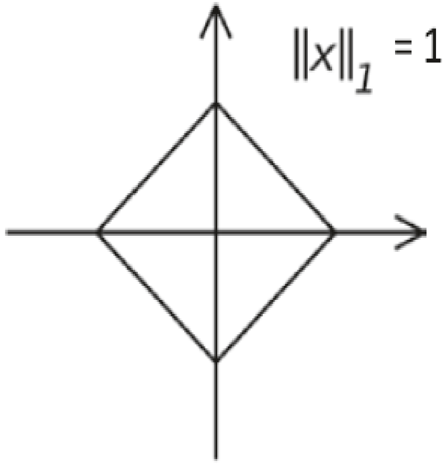
- The norm of a vector is a measure of its magnitude.
- There are several kinds of such norms.
- The most familiar is the Euclidean norm.
- In general, the l^p norm of a vector can be defined when 1 < p < ∞
- the l^1 norm has important applications in machine learning.

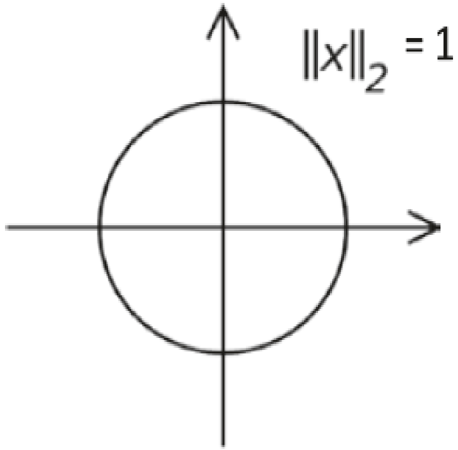
$$\|x\|_2 = \left(|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2\right)^{1/2} = (x.x)^{1/2} = \left(x^T x\right)^{1/2}$$

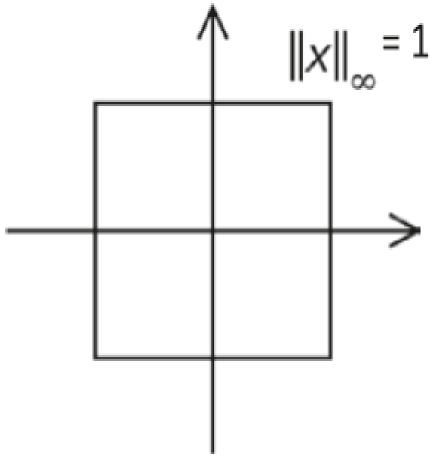
$$\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$

$$\lim_{p\to\infty} \|x\|_p = \lim_{p\to\infty}\left(|x_1|^p + |x_2|^p + \ldots + |x_n|^p\right)^{1/p}$$

Norms of a vector and ML







- Generally, for machine learning we use both l^1 and l^2 norms for several purposes. For instance, the least square cost function that we use in linear regression is the l_2 norm of the error vector; i.e., the difference between the actual target-value vector and the predicted target-value vector.

- Very often we would have to use regularization for our model, with the result that the model doesn’t fit the training data very well and fails to generalize to new data.

- To achieve regularization, we generally add the square of either the l^1 or l^2 norm of the parameter vector for the model as a penalty in the cost function for the model. When the l^2 norm of the parameter vector is used for regularization, it is generally known as Ridge Regularization, whereas when the l^1 norm is used instead it is known as Lasso Regularization.

Part 2 - Matrices

- Matrix representations
- Matrix addition
- Matrix subtraction
- Matrix-Vector multiplication
- Matrix-Matrix multiplication
- Row Echelon form
- Reduced Row-Echelon form (RREF)
- Linear equations

A Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- A rectangular array of numbers is called an $m \times n$ matrix, where m and n represents the number of rows and columns, respectively.

- A can be written as $A = [a_{ij}]$, where a_{ij} is the (i^{th}, j^{th}) element of matrix A. The element of a matrix can be real or complex numbers.

- Matrix A can be considered a matrix that contains n number of column vectors $\in \mathbb{R}^m$ stacked side-by-side. We represent the matrix as $A_{m\times n} \in \mathbb{R}^{m\times n}$.

Matrices: Transpose of a Matrix

- The transpose of matrix A, denoted by A^T, is the result of flipping the rows and columns of the matrix A. When we take the transpose, element (i, j) goes to position (j, i)

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Addition of Two Matrices

The addition of two matrices *A* and *B* implies their element-wise addition. We can only add two matrices, provided their dimensions match. If *C* is the sum of matrices *A* and *B*, then

$$c_{ij} = a_{ij} + b_{ij} \quad \forall\, i \in \{1,2,..m\}, \forall\, j \in \{1,2,..n\}$$

$$where\; a_{ij} \in A, b_{ij} \in B, c_{ij} \in C$$

Example:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \text{ then } A + B = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

Subtraction of Two Matrices

The subtraction of two matrices A and B implies their element-wise subtraction. We can only subtract two matrices provided their dimensions match.

If C is the matrix representing A - B, then

$$c_{ij} = a_{ij} - b_{ij} \quad \forall\, i \in \{1,2,..m\}, \forall\, j \in \{1,2,..n\}$$

$$where\ a_{ij} \in A, b_{ij} \in B, c_{ij} \in C$$

Example:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \text{ then } A - B = \begin{bmatrix} 1-5 & 2-6 \\ 3-7 & 4-8 \end{bmatrix} = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$$

Matrix–Vector Product

To multiply a matrix A by a vector $\boldsymbol{x}$, the number of columns in A must equal the number of rows in $\boldsymbol{x}$.

If A is an m×n matrix, the product A$\boldsymbol{x}$ is defined for n×1 column vectors $\boldsymbol{x}$.

If we let A$\boldsymbol{x}$ = $\boldsymbol{b}$, then $\boldsymbol{b}$ is an m×1 column vector.

The number of rows of A determines the number of rows in the product $\boldsymbol{b}$.

Performing the multiplication can be explained in two different ways:

<u>1. Multiplication a row at a time.</u>

Example:

$$A\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} (a_{11}, a_{12}, \dots, a_{1n}) \,.\, (x_1, x_2, \dots, x_n) \\ (a_{21}, a_{22}, \dots, a_{2n}) \,.\, (x_1, x_2, \dots, x_n) \\ \vdots \\ (a_{m1}, a_{m2}, \dots, a_{mn}) \,.\, (x_1, x_2, \dots, x_n) \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Matrix–Vector Product

2. <u>**Combination of columns**</u>: The matrix A acts on $\boldsymbol{x}$ vector . The output $A\boldsymbol{x}$ is a **combination $\boldsymbol{b}$ of the columns** of A :

$$A\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$A\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

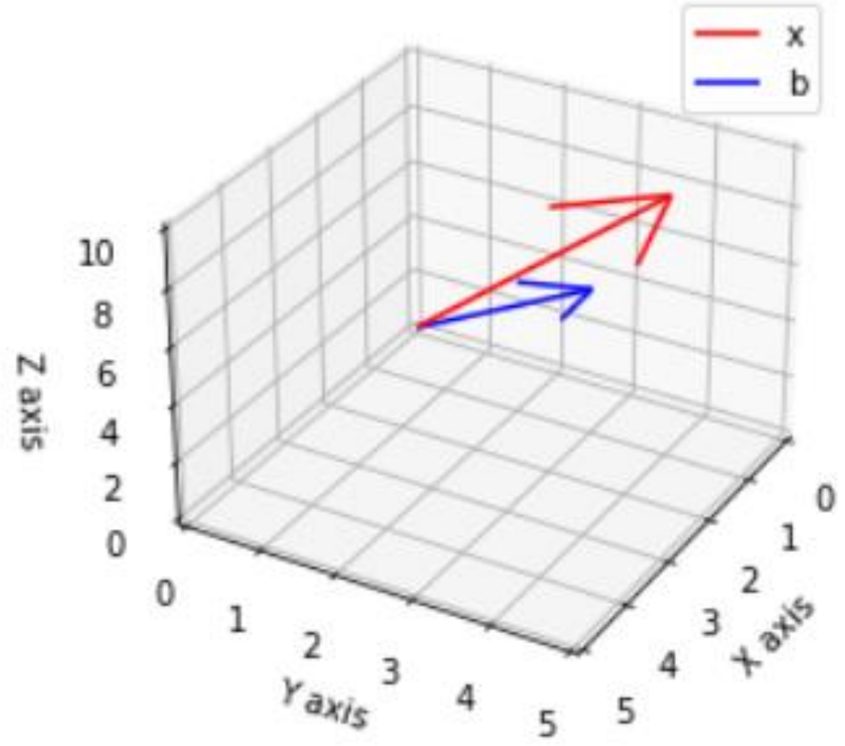
Example: Matrix–Vector Product

Example:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

The product:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 + x_2 \\ -x_2 + x_3 \end{bmatrix}$$



For any input vector $\boldsymbol{x}$, the output of the operation "multiplication by A" some vector $\boldsymbol{b}$ ($\boldsymbol{b}$ is a transformed version of $\boldsymbol{x}$):

$$A \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

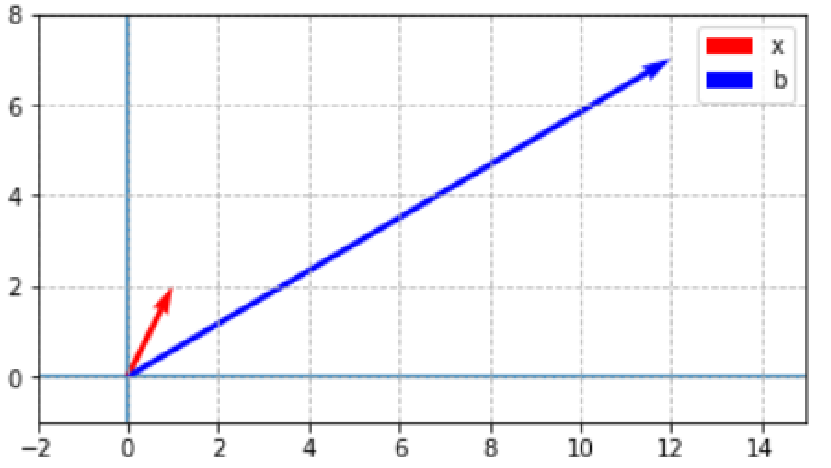
Example: Matrix–Vector Product

Example: Multiply matrix $A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$ by vector $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = ?$$

One method is to think of the entries of $\boldsymbol{x}$ as the coefficient of a linear combination of the column vectors of the matrix:

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$



This technique shows that $A\boldsymbol{x}$ is a linear combination of the columns of A.

You may also calculate the product $A\boldsymbol{x}$ by taking the dot product of each row of A with the vector of $\boldsymbol{x}$:

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 5 \cdot 2 \\ 1 \cdot 1 + 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

Example: Matrix–Vector Product

Distributive property: For matrices $A, B \in \mathbb{R}^{m\times n}$ and a vector $\boldsymbol{x} \in \mathbb{R}^n$

$$(A+B)\boldsymbol{x} = A\boldsymbol{x} + B\boldsymbol{x}$$

Example: $A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}, B = \begin{bmatrix} 4 & 2 \\ 1 & -2 \end{bmatrix}, and\ \boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$A+B = \begin{bmatrix} 6 & 7 \\ 2 & 1 \end{bmatrix}, (A+B)\boldsymbol{x} = \begin{bmatrix} 6 & 7 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 20 \\ 4 \end{bmatrix}$$

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}, \qquad B\boldsymbol{x} = \begin{bmatrix} 4 & 2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ -3 \end{bmatrix}, \qquad A\boldsymbol{x} + B\boldsymbol{x} = \begin{bmatrix} 20 \\ 4 \end{bmatrix}$$

Matrix-Matrix Product

And its properties

Matrices: Matrix–Matrix Product

There are different ways of thinking about the product $AB = C$ of two matrices. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then C is an $m \times p$ matrix. We use c_{ij} to denote the entry in row i and column j of matrix C.

Standard (row times column)

The standard way of describing a matrix product is to say that c_{ij} equals the dot product of row i of matrix A and column j of matrix B. In other words,

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Matrix–Matrix Product

Columns

The product of matrix A and column j of matrix B equals column j of matrix C. This tells us that the columns of C are combinations of columns of A.

Rows

The product of row i of matrix A and matrix B equals row i of matrix C. So the rows of C are combinations of rows of B.

Column times row

A column of A is an $m \times 1$ vector and a row of B is a $1 \times p$ vector. Their product is a matrix:

$$\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 12 \\ 3 & 18 \\ 4 & 24 \end{bmatrix}.$$

Matrices: Multiplication Properties

Matrix multiplication is **ASSOCIATIVE**:

$$\mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{A}\mathbf{B})\mathbf{C}$$

Matrix operations are **DISTRIBUTIVE**:

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}$$

$$(\mathbf{B}+\mathbf{C})\mathbf{D} = \mathbf{B}\mathbf{D} + \mathbf{C}\mathbf{D}$$

Matrix multiplication is **NOT COMMUTATIVE**:

(usually)

$$AB \neq BA$$

Reduced-Row Echelon Form

RREF

Matrix Row Echelon Form

- Gauss-Jordan Elimination is an algorithm that can be used to solve systems of linear equations and to find the inverse of any invertible matrix. It relies upon three elementary row operations one can use on a matrix:
    1. Swap the positions of two of the rows
    2. Multiply one of the rows by a nonzero scalar.
    3. Add or subtract the scalar multiple of one row to another row.
- Row echelon form: examples of matrices in row echelon form

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 & 7 \\ 0 & 1 & 5 & 8 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Reduced Row Echelon Form (RREF)

- Examples of matrices in reduced row echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Systems of Linear Equations

And their solutions

System of Linear Equations

The center problem of linear algebra is a system of equations. For example:

$$\begin{array}{cc} Two\ equations & x - 2y = 1 \\ Two\ unkonws & 3x + 2y = 11 \end{array}$$

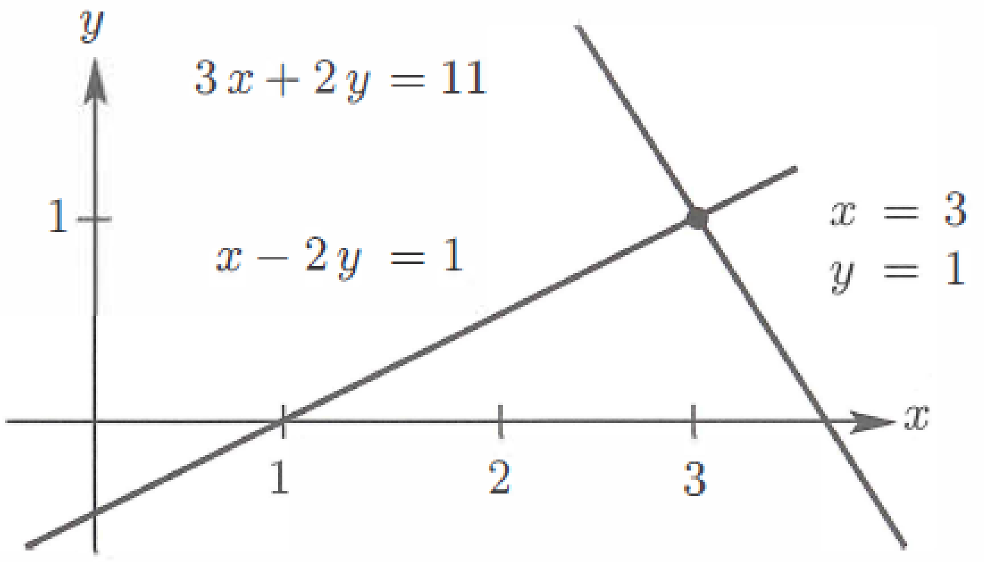
These equations are linear, which means that the unknowns are only multiplied by numbers – we never see $x\ times\ y$.

We can view this problem in three ways:

1. Row picture
2. Column picture
3. Matrix picture

System of Linear Equations:
1. Row Picture

- **The row picture**: Each equation in $A\boldsymbol{x} = \boldsymbol{b}$ gives a line $(n = 2)$ or a plane $(n = 3)$ or a "hyperplane" $(n > 3)$.
- They intersect at the solution or solutions, if any.
- **Example:** two lines meeting at a single point (the solution).

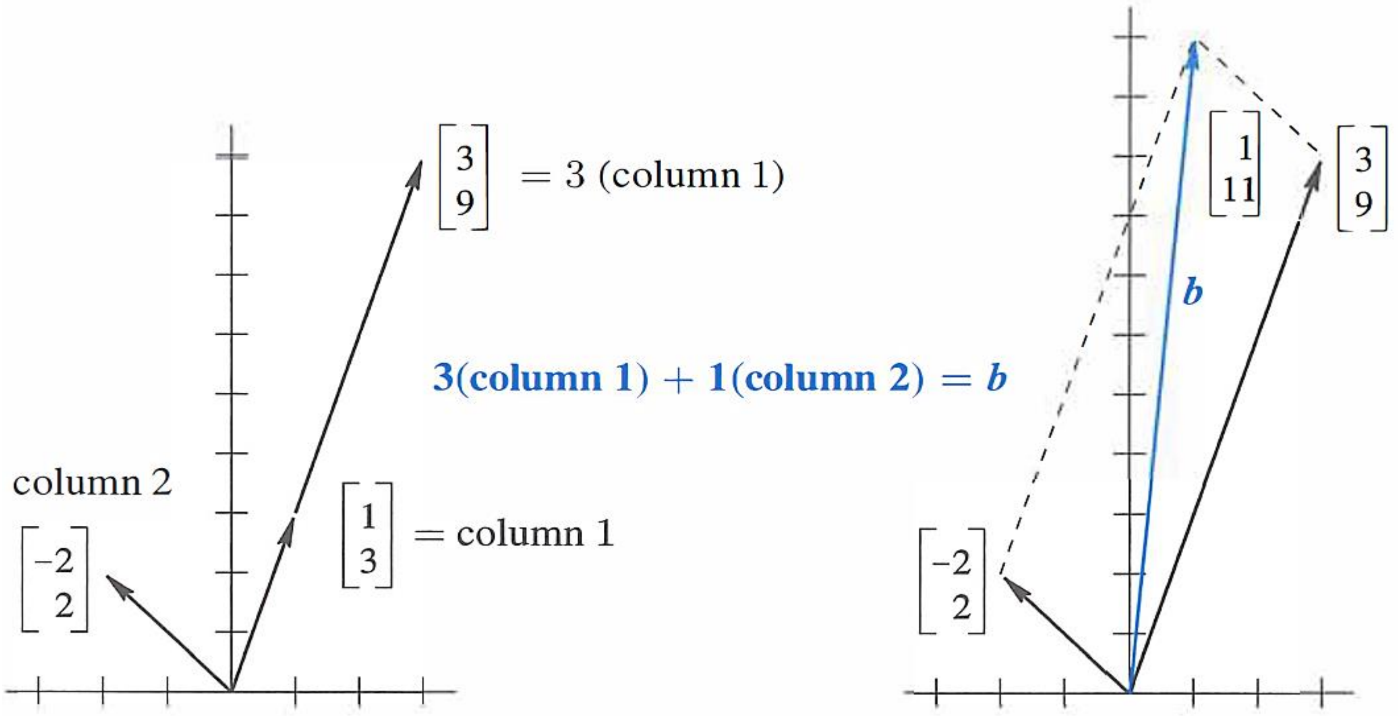


System of Linear Equations: 2. Column Picture

- **Combine the column vectors** on the left side to produce the vector $\boldsymbol{b}$ on the right side.

$$x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix} = \boldsymbol{b}$$

- Geometrically, we want to find numbers x and y so that x copies of vector $\begin{bmatrix}1\\3\end{bmatrix}$ added to y copies of vector $\begin{bmatrix}-2\\2\end{bmatrix}$ equals the vector $\begin{bmatrix}1\\11\end{bmatrix}$.
- As we see from the figure, $x = 3\ and\ y = 1$, agreeing with the row picture.



System of Linear Equations: 3. Matrix Picture

- **The matrix picture:** as a single equation by using matrices and vectors. Its rows give the row picture, and its columns give the column picture. Same numbers, different pictures, same equations. We combine those equations into a matrix problem $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

- The matrix $\mathrm{A} = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}$ is called the *coefficient matrix*. The vector $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ is the vector of unknowns. The values on the right-hand side of the equation form the vector $\boldsymbol{b}$:

$$A\boldsymbol{x} = \boldsymbol{b}$$

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

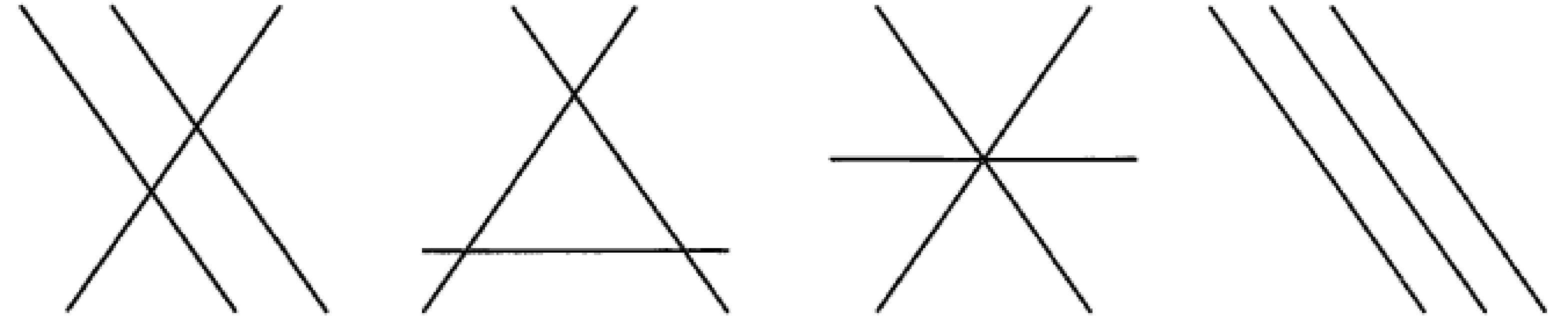
Introduction to Overdetermined and Underdetermined Systems

Can they be solved? To be discussed in detail in later sessions

The Singular Case

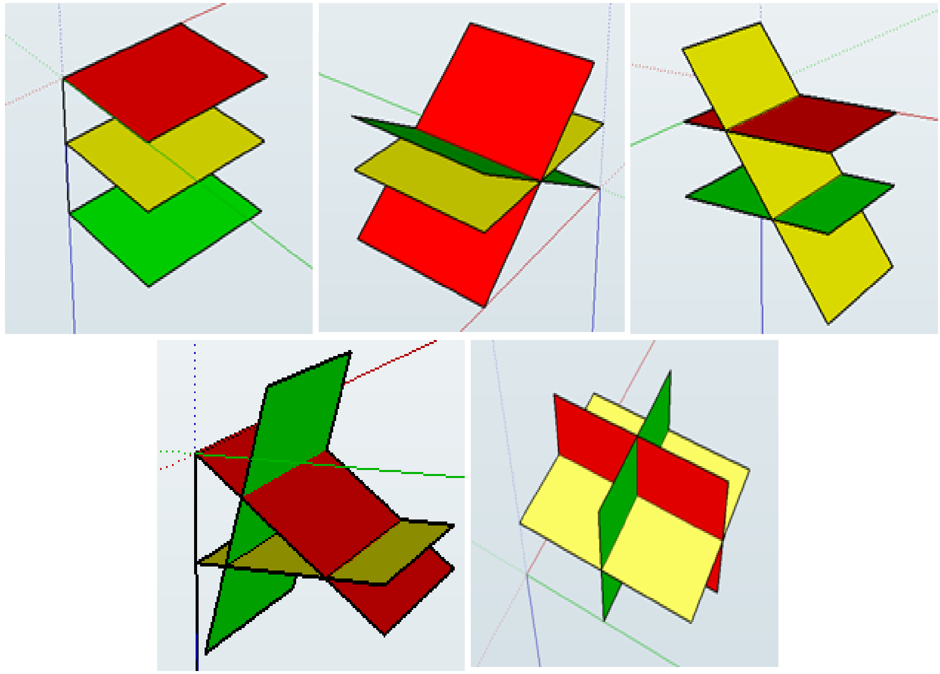
Suppose we are again in three dimensions, and the three planes in the row picture *do not intersect*. What can go wrong? One possibility is that two planes may be parallel. The equations $2u + v + w = 5$ and $4u + 2v + 2w = 11$ are inconsistent—and parallel planes give no solution (Figure 1.5a shows an end view). In two dimensions, parallel lines are the only possibility for breakdown. But three planes in three dimensions can be in trouble without being parallel.

The most common difficulty is shown in Figure 1.5b. From the end view the planes form a triangle. Every pair of planes intersects in a line, and those lines are parallel. The



two parallel planes (a) | no intersection (b) | line of intersection (c) | all planes parallel (d)

Figure 1.5 Singular cases: no solution for (a), (b), or (d), an infinity of solutions for (c)



Four possibilities:

- Unique solution: all the three planes intersect at a point.
- No intersection at all.
- Planes intersect in a line.
- They can intersect in a plane.


Thank You!

Python-based Examples

- Kindly check the provided notebooks.

Resources

Reading?

"Linear Algebra Explained in four Pages"

https://courses.engr.illinois.edu/ece498rc3/fa2016/material/linearAlgebra_4pgs.pdf

Textbook:

Gilbert Strang, Introduction to Linear Algebra, 3rd Edition (or any later edition).

Questions for the instructor?