HCAI5TML01 – Mathematics of Learning.

Lecture – 01

A Refresher on the Mathematics Behind Machine Learning.

Vector, Matrix and System of Linear Equations.

Siman Giri

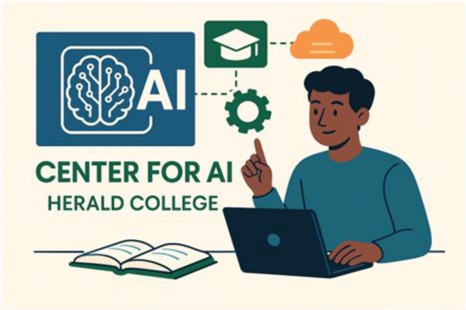


image generated via copilot.

About Center For AI

Siman Giri, Head

Aatiz Ghimire, PI

Ronit Shrestha, RA

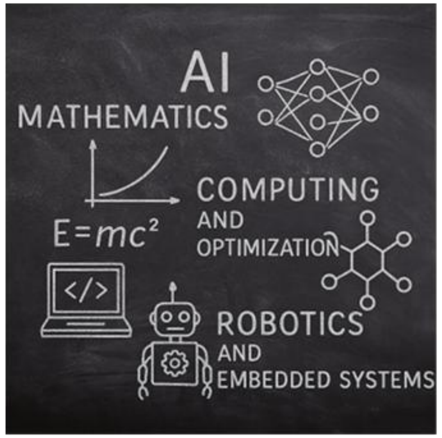
Shiv Kumar Yadav, RA

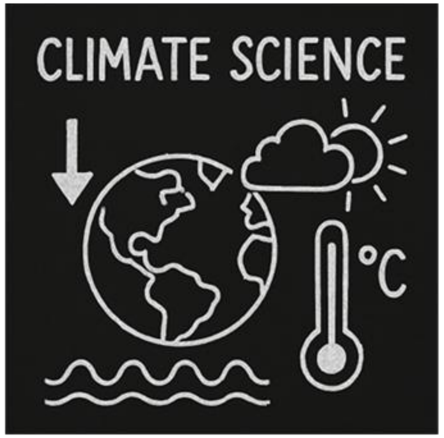
Faculty Researcher

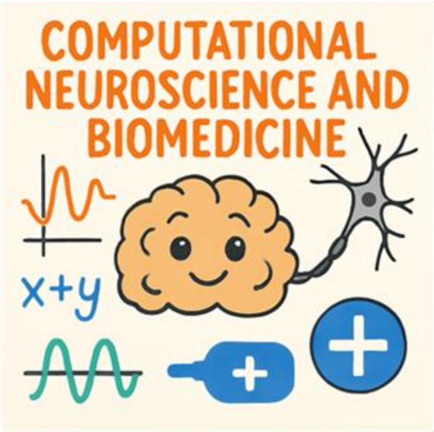
Student Researcher, Thesis and FYP

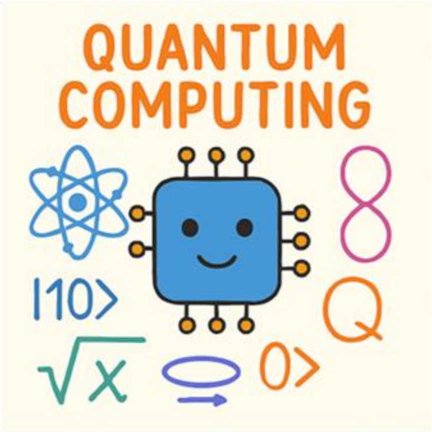
People

Objective









Already Working

Area of Interest …

Coming Soon

A. Why do we need Linear Algebra for ML/DL?

{Why to study Vector and Matrices?}

A.1 What is Linear Algebra?

- **Linear Algebra** is the branch of **mathematics** concerning **linear equations** such as:
  - $\mathbf{a_1 x_1 + \ldots\ldots + a_n x_n = b}$;
  - **linear maps** such as:
    - $(\mathbf{x_1}, \ldots\ldots, \mathbf{x_n}) \mapsto \mathbf{a_1 x_1} + \cdots + \mathbf{a_n x_n}$;
  - and **their representations** in **vector spaces and through matrices.** – **Wikipedia.**
- **Linear algebra** is a branch of mathematics that deals with **vectors, vector spaces** (also **known as linear spaces**),
  - and **linear transformations** between these spaces.
    - It involves operations on matrices and vectors, solving **systems of linear equations**, and understanding geometric concepts like **lines, planes, and subspaces**. – **"chatgpt."**

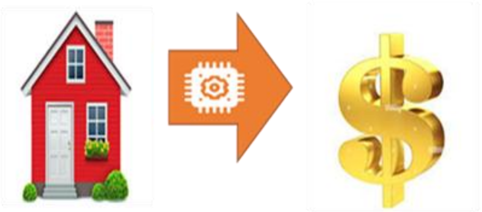
Fig: What is Linear Algebra?

Image: somewhere from web compiled by siman

A.2 Why Linear Algebra for Machine Learning?

- **Representation of Data:**
  - In machine learning, data is typically **represented** as **vectors** and **matrices**. For example, a dataset might be **stored as a matrix** where each row is a data point (vector), and each column is a feature.

Task: House Price Prediction.

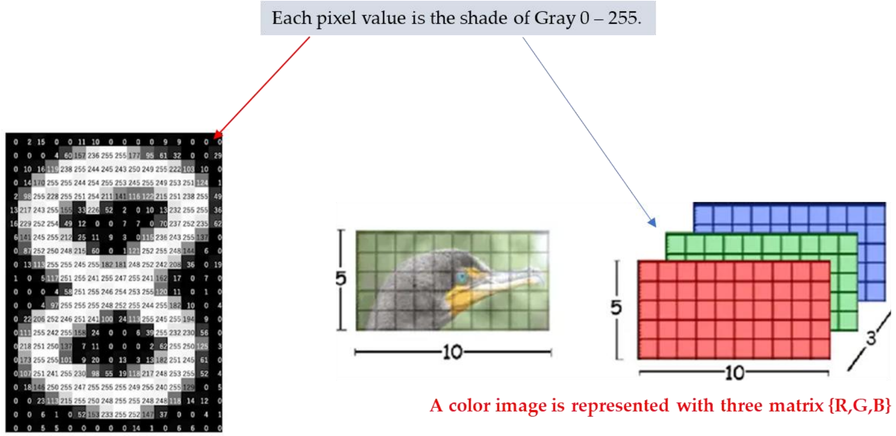


Data: Features/Descriptor of House

Area	Rooms	Price
1080	8	1,00,000.00
1200	10	1,50,000.00

How would you represent this, for computer?
Matrix.

$$\begin{bmatrix} 1080 & 8 \\ 1200 & 10 \end{bmatrix} \begin{bmatrix} 1,00,000 \\ 1,50,000 \end{bmatrix}$$



A color image is represented with three matrix {R,G,B}

A gray scale image is represented with single matrix {R,G,B}

A.2.1 Why Linear Algebra for Machine Learning?

- **Efficient Computing:**
  - Matrix operations allow for efficient computations on large datasets. Libraries like **NumPy**, **TensorFlow**, and **PyTorch** leverage **linear algebra** for operations on large matrices and tensors {**Vectorizations**}, which makes **machine learning models** **faster** and more **scalable**.

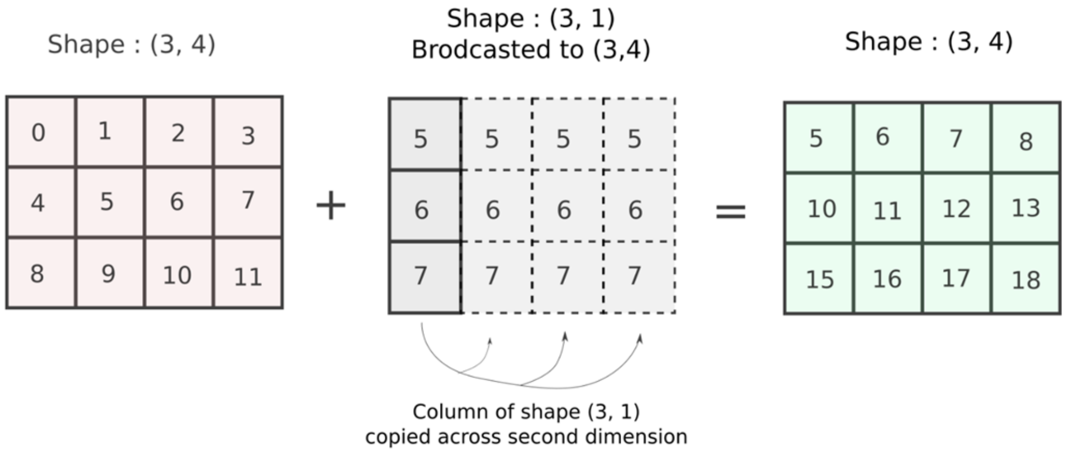


Fig: Idea of Vectorizatons.

A.2.2 Why Linear Algebra for Machine Learning?

- Understanding {Machine or Deep Learning} Algorithms:
  - Training **machine or deep learning models** often involves **solving systems of linear equations**.
    - Linear algebra provides the **necessary tools** to solve these systems efficiently.
  - Many machine learning algorithms are based on linear algebra concepts.
  - For instance:
    - **Linear Regression** involves finding a line (or hyperplane) that best fits the data.
    - **Neural Networks** use matrix multiplication for forward and backward propagation.

B. Summary : Linear Algebra for Machine Learning.

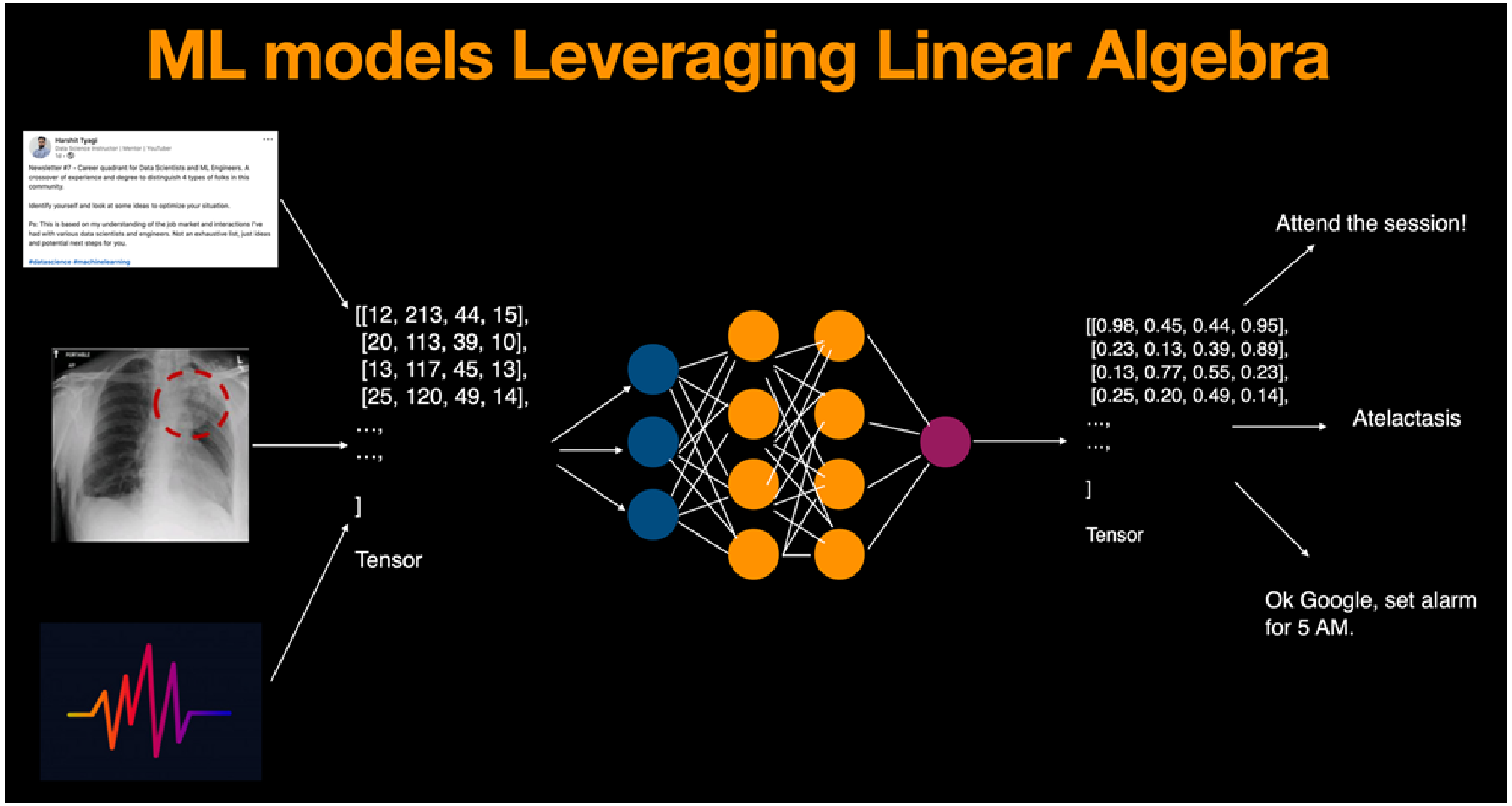


Image By Harshit Tyagi and freeCodeCamp

Understanding Vector and Matrices.

{Basic Concepts, Definition and Notations.}

1.1 What are Vectors?

Interpretation – 1: Point in Space.

- E.g., in **2D{dimension}**
  - we can visualize **the data points** with respect to a **coordinate origin**

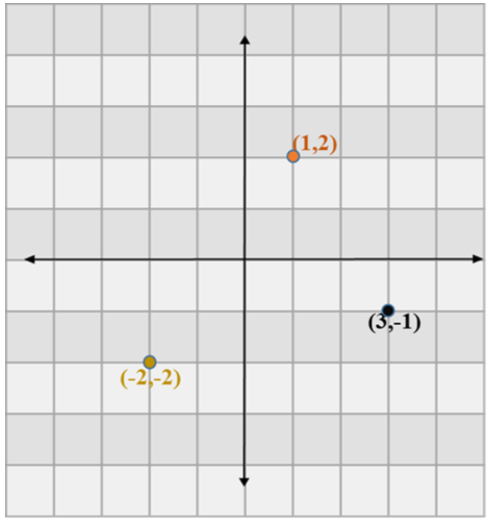


Fig: Vector as a point

Interpretation – 2: Direction in Space.

- E.g., the vector $\vec{\mathbf{v}} = [3, 2]^T$ has a **direction** of **3 steps to the right** and **2 steps up**
- The **notation** $\vec{\mathbf{v}}$ is sometimes used to indicate that the **vectors have a direction**
- All **vectors in the figure** have the **same direction**

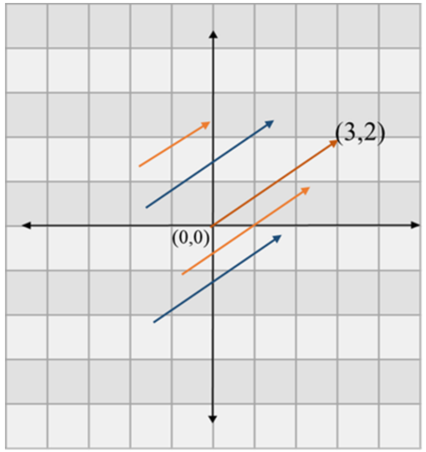


Fig: Vector as Direction

1.2 Vector formal Definition.

- In Linear Algebra and Applied Mathematics, we define vector with in n-dimensional vector space.
- **Vector Space**:
  - If **n** is a positive integer, then an **ordered n-tuple** is a sequence of **n real numbers** $[\mathbf{n_1, n_2, \dots, n_n}]$
  - The set of all **ordered n-tuples** is called **n – space or n – dimensional vector space** and is denoted by $\mathbb{R}^{\mathbf{n}}$.
- **Vectors in** $\mathbb{R}^n$**:**
  - Let $\mathbb{R}^{\mathbf{n}} = \{(\mathbf{x_1, \dots, x_n}) : \mathbf{x_j} \in \mathbb{R} \textbf{ for } \mathbf{j = 1, \dots, n}\}$. Then,
    - $\vec{\mathbf{x}} = [\mathbf{x_1, \dots, x_n}]$ is called **a vector in vector space** $\mathbb{R}^n$.
  - The number $\mathbf{x_j} \rightarrow \mathbf{x_1, \dots x_n}$ are called the **components** of $\vec{\mathbf{x}} \in \mathbb{R}^n$.
- Examples:

$$\mathbf{a} = [\mathbf{a_1, a_2}] \in \mathbb{R}^2$$

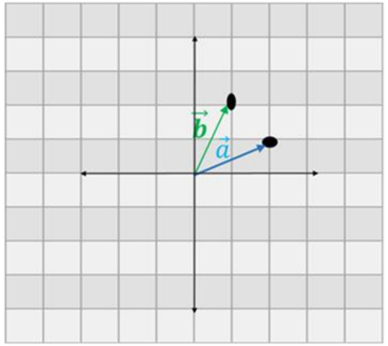


Fig: 2 dimensional vector space

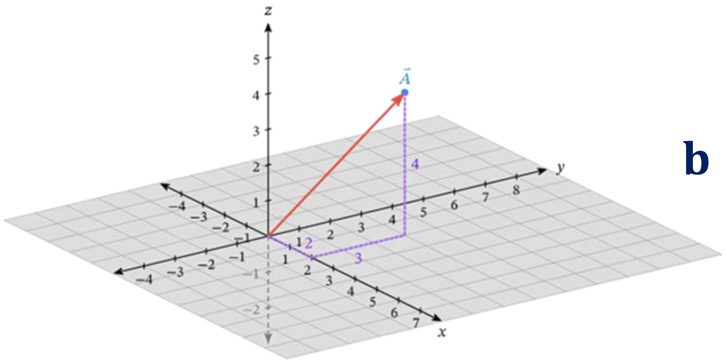


Fig: 3 dimensional vector space

$$\mathbf{b} = [\mathbf{b_1, b_2, b_3}] \in \mathbb{R}^3$$

1.3 Vector in Vector – Space.

- **Vector Space:**
  - A set **V** of ***n*-dimensional vectors** (with a corresponding set of scalars) such that the set of vectors is:
    - **"closed" under vector addition.**
    - **"closed" under scalar multiplication.**
    - **Origins are defined and fixed {0 vector must exist}**
  - In other words:
    - **For addition of two vectors:**
      - takes **two vectors** $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, and it produces the **third vector** $\mathbf{u} + \mathbf{v} \in \mathbb{R}^2$.
      - (addition of vectors – gives another vector in the same set)
    - **For scalar Multiplication:**
      - Takes **a scalar c** $\in$ **F** and a vector $\mathbf{v} \in \mathbb{R}^n$ produces a **new vector cv** $\in \mathbb{R}^n$.
      - (multiplying a vector by a scalar – gives another vector in the same set)

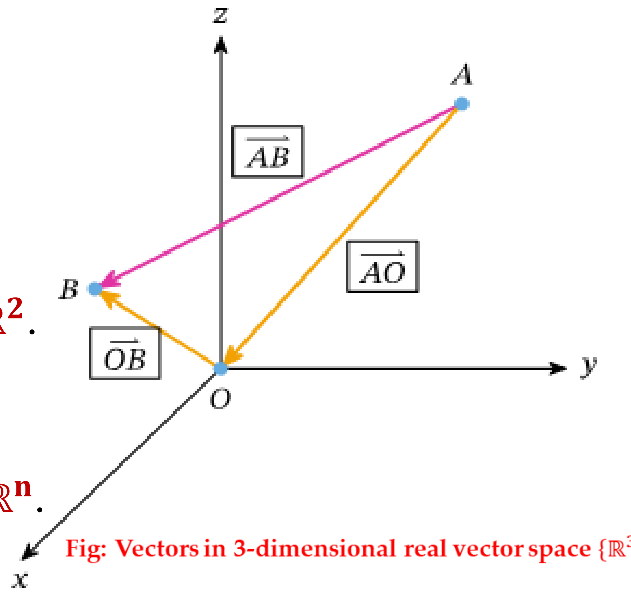


Fig: Vectors in 3-dimensional real vector space $\{\mathbb{R}^3\}$

1.4.1 Axioms of Vector – Space.

- If **V** is a set of vectors satisfying the above definition of a vector space, then it satisfies the following axioms:
  - **Existence of an Additive Identity**: any vector space **V** **must have a zero vector**.
  - **Existence of Negative Vector**: for any vector **v** in **V** its **–ve** must also be in **V**.
  - Has **Athematic /Algebraic Properties** – We can perform valid mathematical operations.

{details in course note}

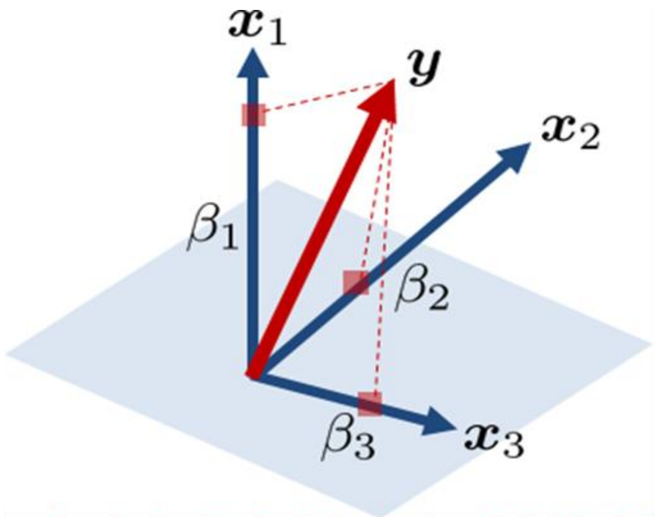


Image from Stanley Chan Book: Introduction to Probability for Data Science.

1.5 Matrices: Introduction.

- In general: A **matrix** is a **rectangular array** of numbers. The **numbers in the array** are called **the entries** in the **matrix**.
    - Array of numbers are an "*ordered collection of vectors*".
    - Like vectors matrices are also fundamentals in machine learning/AI, as matrices are the way computer *interact with data* in practice.
- A **matrix** is represented with *italicized* upper-case letter like "**A**".
    - For two dimensions: we say the matrix **A** has:
        - **m rows** and **n columns**.
        - Each entry/element of **A** is defined as $\mathbf{a_{ij}}$.
    - Thus, a **matrix** $\mathbf{A^{m \times n}}$ is define as:

$$\boxed{A_{m\times n} := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, a_{ij} \in \mathbb{R}}$$

- Overview of notation for discussing matrices:
- Given a set $\mathbf{C} \in \mathbb{R}$, we let $\mathbf{C_{m\times n}}$ denote the set of all matrices of **m rows** and **n columns** consisting of items from set **C**.
    - For matrix: $\mathbf{A} \in \mathbf{C_{m\times n}}$: we let $\mathbf{a_{ij}}$ denote the item at the $\mathbf{i^{th}}$ **row** and $\mathbf{j^{th}}$ **column** of **A**.
    - For matrix $\mathbf{A} \in \mathbf{C_{m\times n}}$: we let $\mathbf{a_{i*}}$ denote the $\mathbf{i^{th}}$ **row vector of A**.
    - For matrix $\mathbf{A} \in \mathbf{C_{m\times n}}$: we let a_{*i} denote the $\mathbf{j^{th}}$ **column vector** of **A**.

1.6 Special Matrices.

- Rectangular Matrix:
  - Matrices are said to be rectangular when the number of rows is ≠ to the number of columns, i.e. $\mathbf{A}^{\mathbf{m}\times\mathbf{n}}$ with $\mathbf{m} \neq \mathbf{n}$. For instance:

$$\boxed{A_{2\times3} := \begin{bmatrix} 1 & 2 & 3 \\ 5 & 5 & 4 \end{bmatrix}}$$

- Square Matrix:
  - Matrices are said to be square when the number of rows = the number of columns, i.e. $\mathbf{A}^{\mathbf{m}\times\mathbf{n}}$. For instance:

$$\boxed{A_{2\times2} := \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}}$$

- Diagonal Matrix:
  - Square matrices are said to be diagonal when each of its non-diagonal elements is zero, i.e. for
    - $\mathbf{D} = (\mathbf{d_{ij}})$,we have $\forall \mathbf{i}, \mathbf{j} \in \mathbf{n}\ \mathbf{i} \neq \mathbf{j} \Rightarrow \mathbf{d_{ij}} = \mathbf{0}$.
  - For instance:

$$\boxed{A_{3\times3} := \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}}$$

- Upper triangular matrix:
  - Square matrices are said to be upper triangular when the elements below the main diagonal are zero i.e. For $\mathbf{D} = (\mathbf{d_{ij}})$, we have $\mathbf{d_{ij}} = \mathbf{0}$, for $\mathbf{i} > \mathbf{j}$. For instance:

$$\boxed{A_{3\times3} := \begin{bmatrix} 9 & 8 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 5 \end{bmatrix}}$$

- Lower triangular matrix:
  - Square matrices are said to be lower triangular when the elements above the main diagonal are zero . i.e. $\mathbf{D} = (\mathbf{d_{ij}})$, we have $\mathbf{d_{ij}} = \mathbf{0}$, for $\mathbf{i} < \mathbf{j}$. For instance:

$$\boxed{A_{3\times3} := \begin{bmatrix} 9 & 0 & 0 \\ 8 & 1 & 0 \\ 4 & 2 & 5 \end{bmatrix}}$$

- Identity Matrix:
  - A diagonal matrix is said to be the identity when the elements along its main diagonal are equal to one. For instance:

$$\boxed{A_{3\times3} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}$$

1.6.1 Special Matrices.

- **Symmetric Matrix:**
  - Square matrices are said to be symmetric its equal to its transpose, i.e. $\mathbf{A} = \mathbf{A}^{\mathbf{T}}$. For instance:

$$A_{3\times3} := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 6 \\ 3 & 6 & 1 \end{bmatrix}$$

- **Scalar Matrix**:
  - Diagonal matrices are said to be scalar when all the elements along its main diagonal are equal, i.e. $\mathbf{D} = \boldsymbol{\alpha}\mathbf{I}$. For instance:

$$A_{3\times3} := \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- **Null or Zero Matrix:**
  - Matrices are said to be null or zero matrices when all its elements equal to zero, which is denoted as $\mathbf{0}_{\mathbf{m}\times\mathbf{n}}$. For instance:

$$A_{3\times3} := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- **Equal Matrix:**
  - Two matrix are said to be equal if
    - $\mathbf{A}(\mathbf{a}_{\mathbf{ij}}) = \mathbf{B}(\mathbf{b}_{\mathbf{ij}})$.
  - For instance:

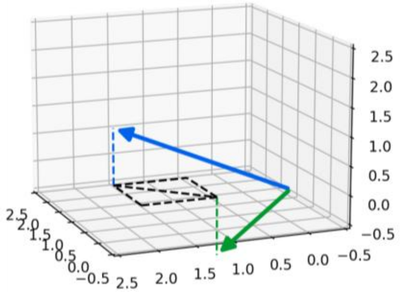
$$B_{2\times2} := \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$$

$$A_{2\times2} := \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$$

1.7 Interpretation of a Matrix: Collection of Vectors.

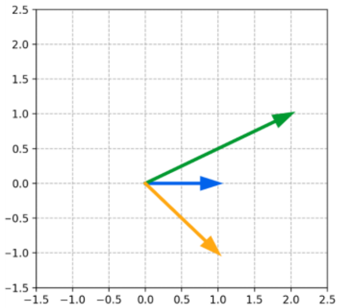
- A matrix can be thought of as a set of vectors.
- For example, for the following matrix:
  - $\mathbf{A} \coloneqq \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & -\mathbf{1} \end{bmatrix}$ can be thought of as
    - a **two** **three-dimensional** row vectors i.e.
    - $\boldsymbol{a}_{1*} \coloneqq [\mathbf{1} \quad \mathbf{2} \quad \mathbf{1}]$ and $\boldsymbol{a}_{2*} \coloneqq [\mathbf{0} \quad \mathbf{1} \quad -\mathbf{1}]$

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



    - Or as a **three** **two-dimensional** column vectors:
    - $\mathbf{a}_{*1} \coloneqq \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}$; $\mathbf{a}_{*2} \coloneqq \begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix}$ and $\mathbf{a}_{*3} \coloneqq \begin{bmatrix} \mathbf{1} \\ -\mathbf{1} \end{bmatrix}$

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



1.7.1 Interpretation of Matrix: As a table of data.

- The simplest interpretation of matrix is as a two – dimensional array of values.
- For example:
  - A numerical dataset represented as a matrix.

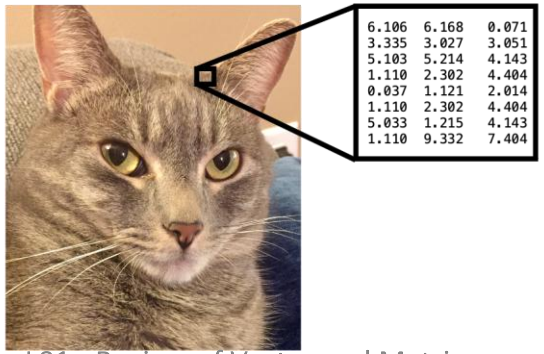
Spreadsheet

	A	B	C	D
1	sepal_length	sepal_width	petal_length	petal_width
2	5.1	3.5	1.4	0.2
3	4.9	3	1.4	0.2
4	7	3.2	4.7	1.4
5	6.5	2.8	4.6	1.5
6	5.8	2.7	5.1	1.9
7	7.1	3	5.9	2.1

Matrix

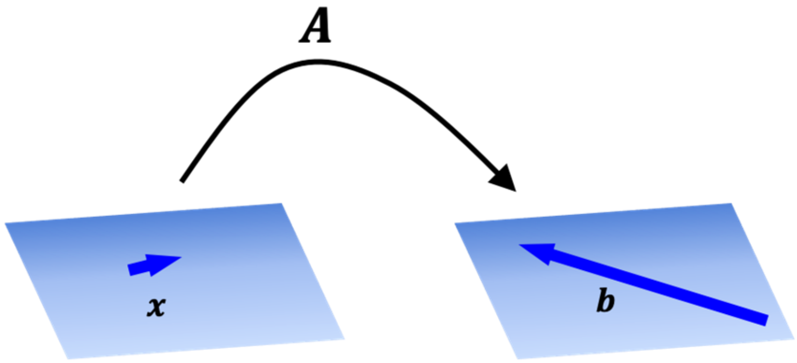
$$\begin{bmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 7.0 & 3.2 & 4.7 & 1.4 \\ 6.5 & 2.8 & 4.6 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \\ 7.1 & 3.0 & 5.9 & 2.1 \end{bmatrix}$$

  - The **pixels** of an image can be represented as a matrix.
  - Let's say we have an image of **m × n pixels**.
    - Let **X** be a **matrix** representing this image where $x_{i,j}$ represents the **intensity of the pixel** at row **i** and **j**.



1.7.2 Interpretation of Matrix: As a Function.

- A matrix can also be viewed as a function **that maps**
  - **vectors in one vector space** to **vectors in another vector space.**
- These special kind of **matrix – defined function** are also called
  - **Linear Transformation** and written as:
    - $\mathbf{T(x) \coloneqq Ax}$
- A very simple visualization of such function is **matrix – vector multiplication**.

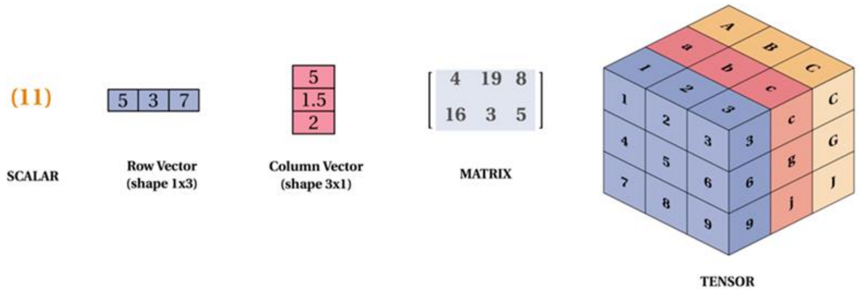


$$A.X \coloneqq \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_{mn} \end{bmatrix}$$

Fig: What happens if we Multiply Matrix A with vector x?

Good to Know!!!

- A tensor is a **multidimensional array** and a **generalization** of the concepts of a **vector and a matrix.**



- Tensors can have many axes, here is a tensor with three axes:

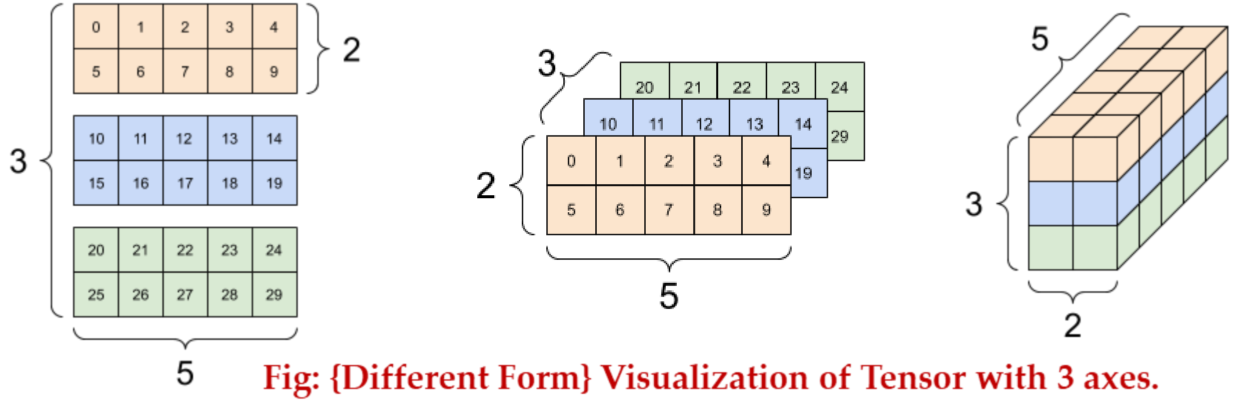
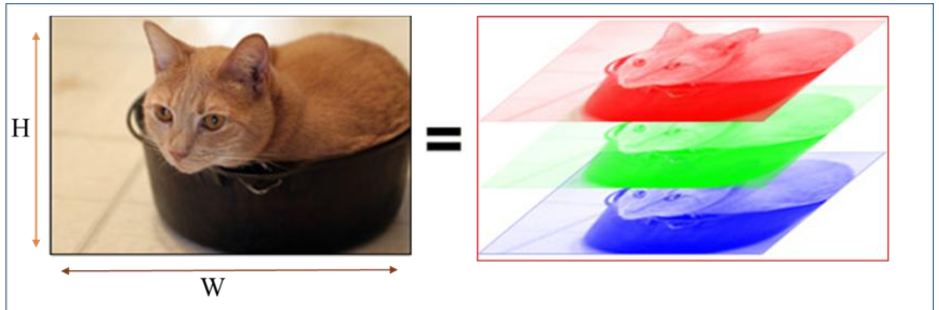


Fig: {Different Form} Visualization of Tensor with 3 axes.

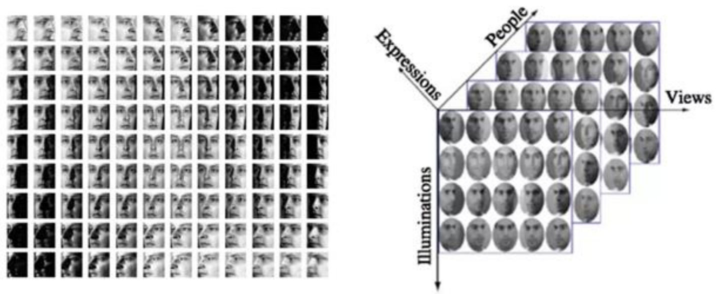
Tensor → Example.

- Tensors in DL are Used to represent an image.
  - image_shape := Height × Width × Color Channel (RGB)



A color image is composed of 3 channels and is 3rd rank tensors.

facial images database is 6th-order tensor



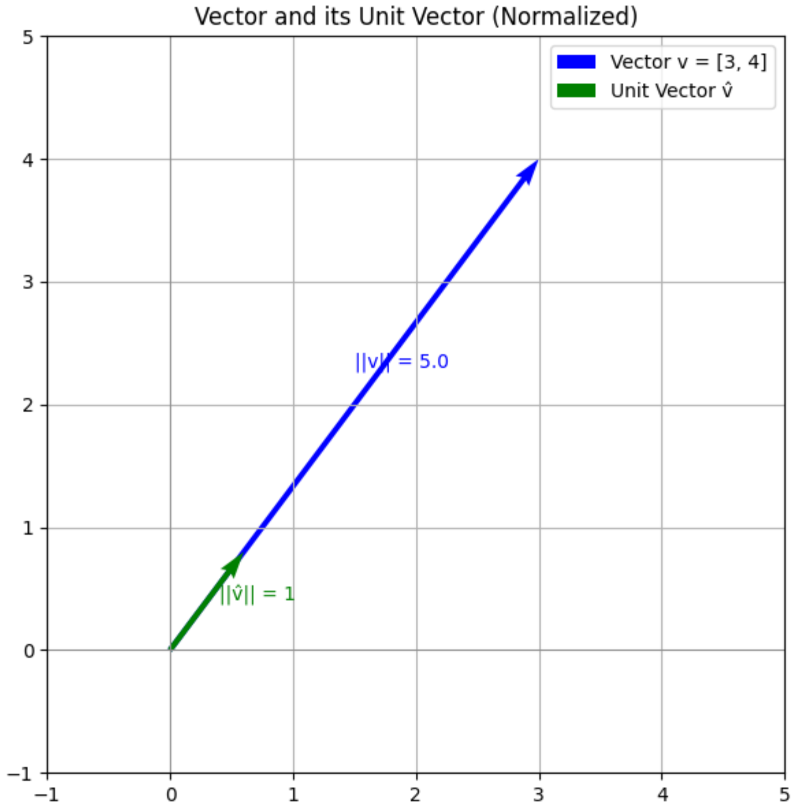
color video is 4th-order tensor

2. The Geometry of Vectors.

{Operations, Linear Dependence, and Basis}

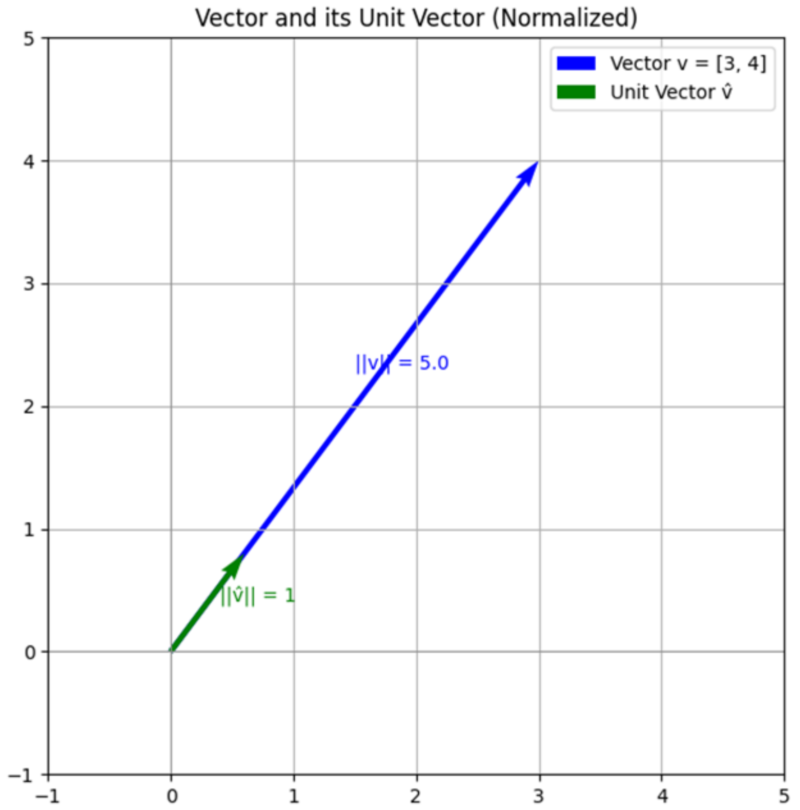
2.1 Norm – "Length" of a vector.

- The **norm** of a **vector v,** often written as **$\|\mathbf{v}\|$,** is a measure of its magnitude or length.
- For vectors in $\mathbb{R}^n$, the most common is the Euclidean norm aka L2 Norm:
  - $\|\mathbf{v}\|_2 = \sqrt{\mathbf{v}_1^2 + \mathbf{v}_2^2 + \cdots + \mathbf{v}_n^2} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$



2.2 Unit Vector.

- A **unit vector** is any vector **whose norm is exactly 1:**
  - $\|\mathbf{u}\| = 1$
- Why unit vectors matter?
  - They capture direction only, no magnitude.
  - They are building blocks of basis vectors.
  - Useful in normalizations ,projections and angle calculations.
- How to Get a unit vector from any vector – **Normalization?**
  - You can normalize a vector by dividing it by its own norm:
    - $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$
  - This gives a unit vector in the same direction as v.



2.3 Inner Product (General Concept).

- An inner product is a general mathematical concept that defines a **way to compute an angle – like and length – like** relationship between two vectors.
  - Formally, an inner product on a **vector space V** is a function:
    - $\langle . , . \rangle : \mathbf{V} \times \mathbf{V} \rightarrow \mathbb{R}$
  - that satisfies these properties:
  - **Linearity in the first argument:**
    - $\langle \alpha \mathbf{u} + \mathbf{b}\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \mathbf{b} \langle \mathbf{v}, \mathbf{w} \rangle$
    - Taking the inner product of a linear combination of vectors u and v with another vector w, is the same as taking the same linear combination of their inner products with w.
  - **Symmetry:**
    - $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
  - **Positive – definiteness:**
      - $\langle \mathbf{v}, \mathbf{v} \rangle \geq \mathbf{0}$, **and equals 0 only if** $\mathbf{v} = \mathbf{0}$
    - When you take the inner product of a vector with itself, you always get a **non-negative number until and unless itself is the zero vector.**
    - This property ensures that the inner product defines a **valid notion of length** (or norm), because:
      - $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$

2.3.1 Why Non – Negativity Matters?

- The inner product $\langle \mathbf{v}, \mathbf{v} \rangle$ is used to define the norm (length) of a vector: $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
- Now imagine what would happen if $\langle \mathbf{v}, \mathbf{v} \rangle < \mathbf{0}$**:**
  - You'd have to take the square root of a **negative number**.
  - That would **break the geometry** of the space
    - distances and lengths would **become imaginary or undefined in** $\mathbb{R}$**.**
  - Concepts like distance, angle, and orthogonality would stop making sense
    - in the real – valued world of machine learning.
- Therefore:
  - The **positive – definiteness condition** ensures that the **norm or length of any vector is a real**, non – negative number something we can safely interpret geometrically.

2.3.2 Understanding Dot Products.

- **Dot product:**
  - Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the quantity $\mathbf{u}^T\mathbf{v}$, sometimes called the **inner product** or **dot product** of the vectors, is a real number given by:
  - $\mathbf{u}^T\mathbf{v} \in \mathbb{R} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] \cdot \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \ldots \\ \mathbf{v}_n \end{bmatrix} = \sum_{i=1}^{n} \mathbf{u}_i \times \mathbf{v}_i$
- It satisfies all the inner product properties - so the dot product is **special case of inner product**
  - but not all **inner products are dot products.**
- **Orthogonal Vectors:**
  - A pair of vectors **u and v** are **orthogonal** if their **dot product** is zero
    - i.e. $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{0}.$
  - Notation for a pair of orthogonal vectors is $\mathbf{u} \perp \mathbf{v}$ {i.e. **Vector are perpendicular to each other**}.
  - In the $\mathbb{R}^n$; this is equal to pair of vector forming a $\mathbf{90^0}$ angle.

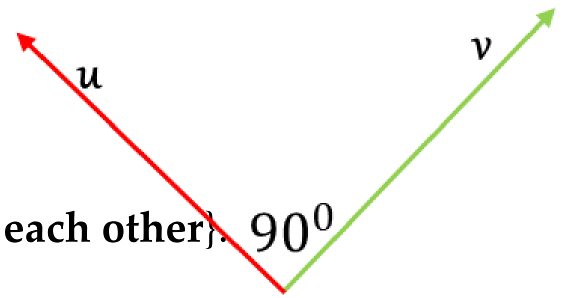
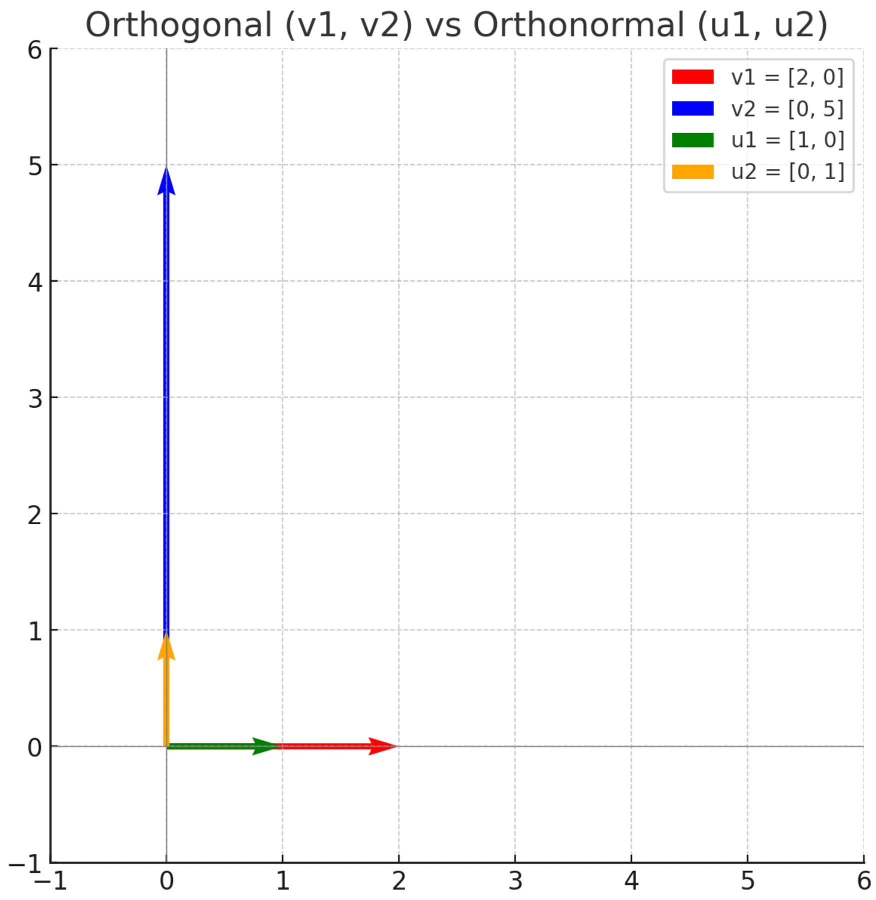


Fig: Orthogonal Vectors

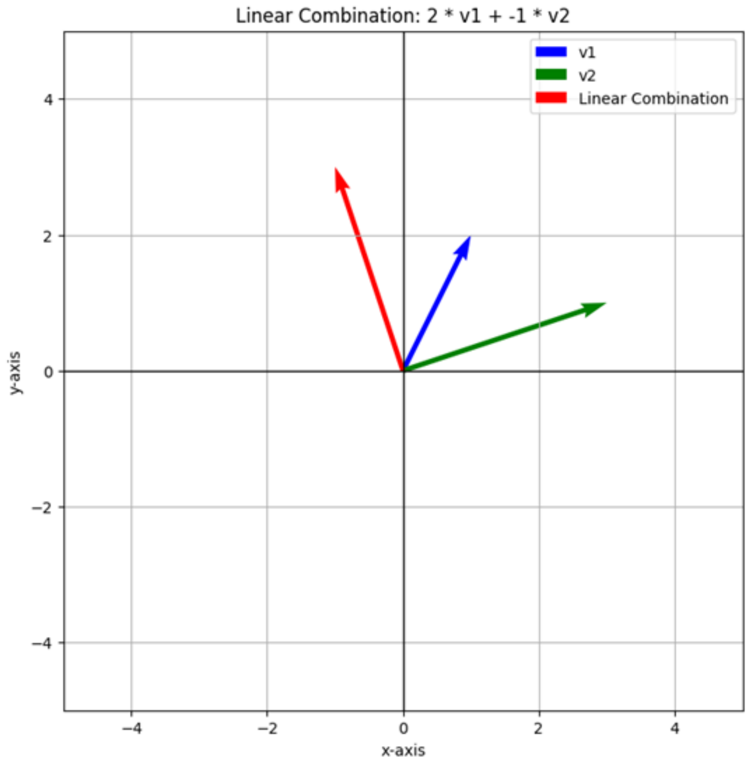
2.3.3 Orthogonal vs. Orthonormal.

- A set of vectors is orthogonal if: $\langle \mathbf{v_i}, \mathbf{v_j} \rangle = \mathbf{0}$ **for all i** $\neq$ **j**
  - They are **perpendicular**, But not **necessarily unit length**.
- **Orthonormal:**
  - A set of vectors is orthonormal if:
    - **They are orthogonal i.e.** $\langle \mathbf{v_i}, \mathbf{v_j} \rangle = \mathbf{0}$ **for all i** $\neq$ **j**
    - **Each vector has unit length:** $\|\mathbf{v_i}\| = \mathbf{1}$ **for all i**
  - **Example in** $\mathbb{R}^2$:
    - **Orthogonal but not orthonormal:**
      - $\mathbf{v_1} = [\mathbf{2}, \mathbf{0}]$
 $\mathbf{v_2} = [\mathbf{0}, \mathbf{5}]$
        - $\langle \mathbf{v_1} \cdot \mathbf{v_2} \rangle = \mathbf{0}$ → **orthogonal**
        - **But** $\|\mathbf{v_1}\| = \mathbf{2}, \|\mathbf{v_2}\| = \mathbf{5}$ → **not unit length**
    - **Orthonormal:**
      - $\mathbf{u_1} = [\mathbf{1}, \mathbf{0}]$ $\quad \mathbf{u_2} = [\mathbf{0}, \mathbf{1}]$
      - **Perpendicular and each has length 1.**



2.4 Linear Combinations of Vectors.

- Idea Combining two or more than two vectors to form a new vector.
- Definition:
  - A vector v is a linear combination of a set of vectors $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}$, if it can be expressed as:
    - $\mathbf{v} = \mathbf{c_1 v_1} + \mathbf{c_2 v_2} + \cdots + \mathbf{c_n v_n}$
      - where:
        - $\mathbf{c_1}, \mathbf{c_2}, \dots, \mathbf{c_n}$ are scalars (**coefficients**).
        - $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}$ are vectors in a vector space.
- Example in $\mathbb{R}^2$:
  - Let $\mathbf{v_1} = [1, 2]$ and $\mathbf{v_2} = [3, 1]$,
  - If we take scalars $\mathbf{c_1} = \mathbf{2}$ and $\mathbf{c_2} = \mathbf{-1}$,
    - then their linear combination will
      - produce a new *vector v* in same **vector space**.
      - $\mathbf{v} = \mathbf{2} \times \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix} + (\mathbf{-1}) \times \begin{bmatrix} \mathbf{3} \\ \mathbf{1} \end{bmatrix}$ ■

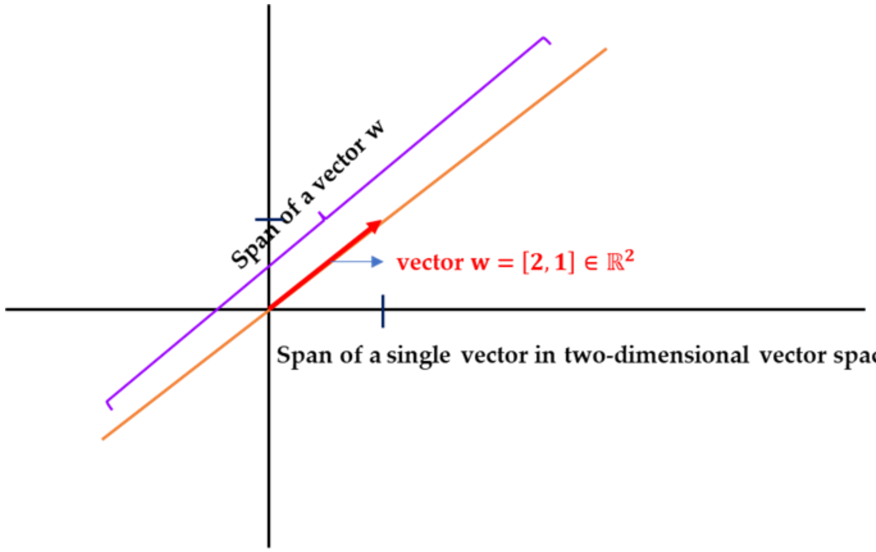


2.5 Span of a Set of vectors.

- **Span** is a consequences of **Linear combination of vectors** and can be thought as a **subset** inside a **vector space** (also known as **vector subspace**).
- A subspace, $\mathbb{S}$ **of real vector space** $\mathbb{R}^{\mathbf{n}}$ is thought of a **flat (having no curvature) surface** with in $\mathbb{R}^{\mathbf{n}}$:
  - is a collection of **all the vectors in** $\mathbb{S}$ which satisfies the following (algebraic) conditions:
    - The ***origin* (0 *vector*)** is contained in $\mathbb{S}$.
    - If vector $\boldsymbol{v}_1$ **and** $\boldsymbol{v}_2$ are in $\mathbb{S}$; then $\boldsymbol{v}_1 + \boldsymbol{v}_2 \in \mathbb{S}$.
    - If $\boldsymbol{v}_1 \in \mathbb{S}$ and $\boldsymbol{\alpha}$ a scalar then $\boldsymbol{\alpha v}_1 \in \mathbb{S}$.
- The **span of a set of vectors** $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n\} \in \mathbb{R}^{\boldsymbol{n}}$ is the **set** of all **possible linear combinations** of **those vectors**. Formally, the span of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n\}$ is:
  - $\boldsymbol{span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n) = \{\boldsymbol{c}_1\boldsymbol{v}_1 + \boldsymbol{c}_2\boldsymbol{v}_2 + \dots + \boldsymbol{c}_n\boldsymbol{v}_n | \boldsymbol{c}_1, \boldsymbol{c}_2, \dots, \boldsymbol{c}_n \in \mathbb{R}\}$
    - where $\boldsymbol{c}_1, \boldsymbol{c}_2, \dots, \boldsymbol{c}_n$ are **scalar coefficients**.

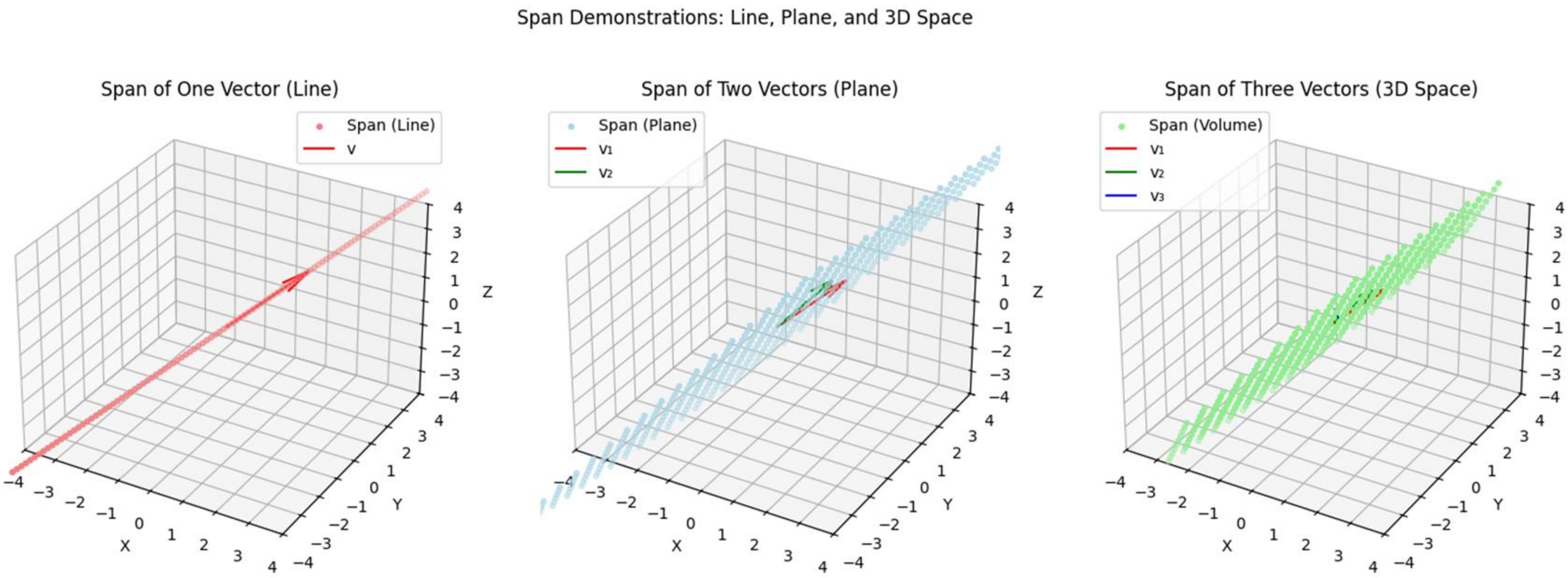
2.5.1 Geometric Interpretation of a Span.

- **The span of one nonzero vector is a line** through the origin in the direction of that vector.
- **The span of two linearly independent vectors is a plane** through the origin.
- **The span of three linearly independent vectors in $\mathbb{R}^3$ is the entire 3D space**, which you can think of as filling a volume (like a cube).



Span of a single vector in two-dimensional vector space is a line.

2.5.1 Geometric Interpretation of a Span.



2.5.2 Span and Learning.

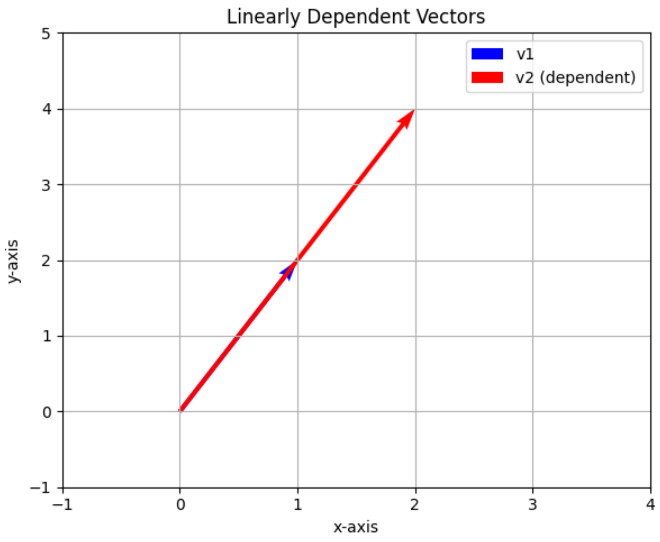
- **Feature Representation and Dimensionality:**
  - Each data point is often represented as a vector in a feature space.
  - The **span of your features** essentially defines the **space where your data lives**.
  - If your features are **linearly dependent** (i.e., don't add new "directions"), your data effectively lies in a **lower-dimensional subspace**.
  - This is why **dimensionality reduction** techniques (like PCA) look for a **smaller spanning set (basis)** that **captures most of the variance**.
- **Model Expressiveness:**
  - Linear models like **linear regression** find **weights** that are **linear combinations of features**.
  - The **predictions** lie in the **span of the feature vectors** (plus bias).
  - Understanding the span helps explain whether your model can represent the target well (e.g., if the target lies in the span of your features).
- **Feature Span and Model Weights:**
  - Your features form a vector space; the model's predictions are linear combinations of these features.
  - When features are **linearly dependent or highly correlated**, the effective **span** of your features is "smaller" than the number of features suggests.
  - This can cause the model weights to become **unstable or non-unique, hurting generalization.**

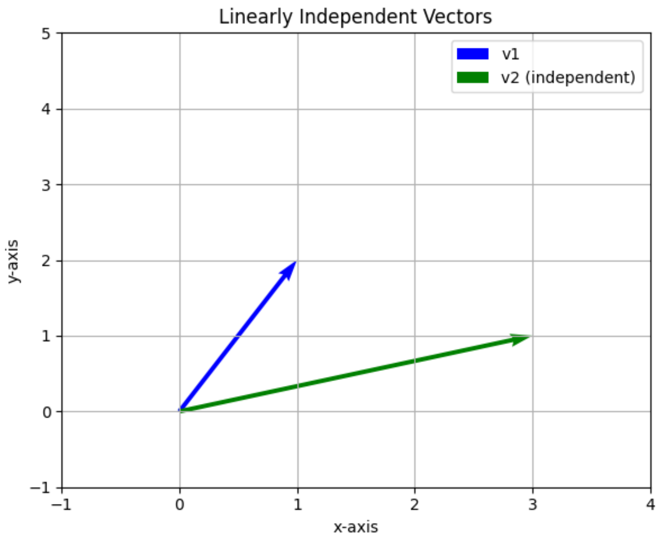
2.5.2 Span and Learning.

- **Regularization and Overfitting:**
  - When features are **highly correlated (dependent),** models **can overfit** by giving **extreme weights** to **redundant features.**
  - Understanding span and linear dependence helps motivate **regularization** techniques like Ridge or Lasso to constrain weights.
  - **Lasso (L1 Regularization):**
    - Pushes some feature coefficients **exactly to zero**, effectively **removing those features** from the model. This reduces the **dimension of the span** of the selected features, creating a smaller subspace that still explains the data well.
  - **Ridge (L2 regularization):**
    - Shrinks the coefficients of correlated features **towards zero but rarely exactly zero**, so it **keeps all features** but reduces their impact.
    - This controls instability caused by overlapping spans without changing the span's dimension.

2.6 Linearly Independent and Dependent Vectors.

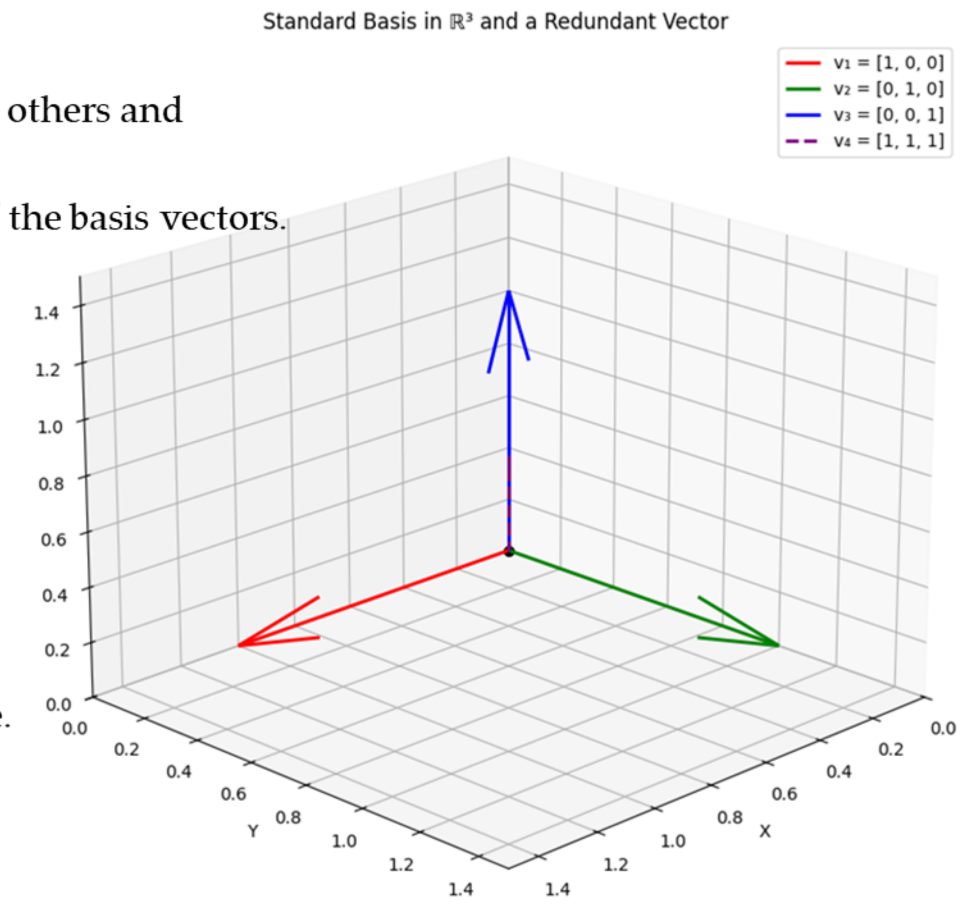
- A set of vectors $v_1, v_2, \ldots, v_n$ in a **vector space** $\mathbb{R}^n$ is:
  - **Linearly dependent** if at least one vector can be written as a linear combination of the others.
    - Mathematically, this means there exists at least one scalars $c_1, c_2, \ldots, c_n$ which is not zero, such that:
      - $\mathbf{c_1 v_1 + c_2 v_2 + \cdots c_n v_n = 0}$; **at least one** $\mathbf{c \neq 0}$.
  - **Linearly Independent** if the only possible solution for above equation is $c_1 = c_2 = \cdots = c_n = 0$, i.e. no vector set can be written as a combination of the others.





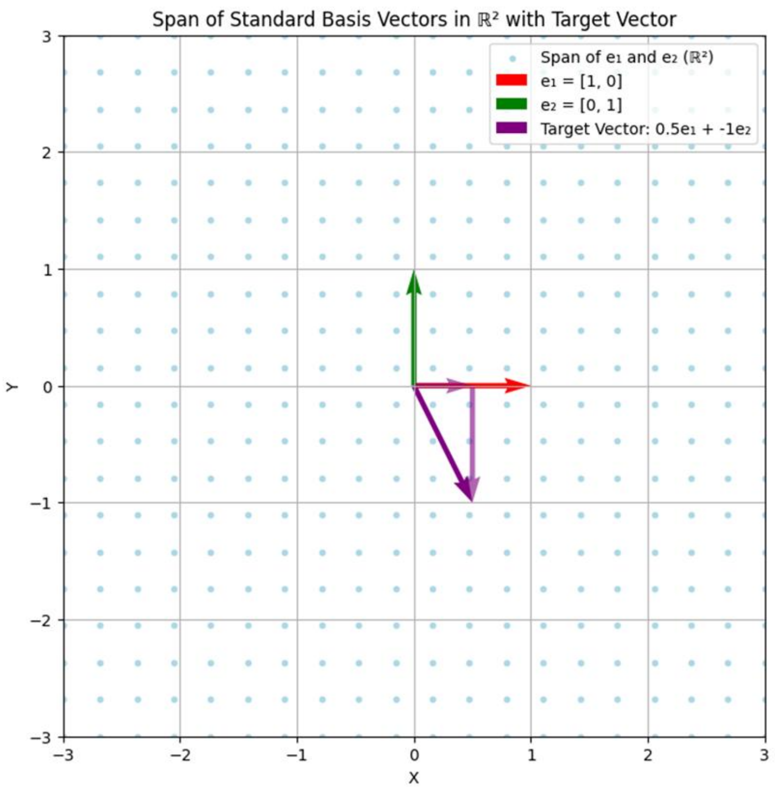
2.7 Basis of a Vector Space.

- Definition:
  - A basis of a **vector space V** is a **set of vectors:** $\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_k}\}$ such that:
    - The vectors **are linearly independent**,
      - i.e. No vector in a basis can be written as a combination of the others and
    - They **span the space V,**
      - i.e. every vector in V can be written as a linear combination of the basis vectors.
  - Basis is the minimal set needed to span the entire vector space.
- **Example in $\mathbb{R}^3$:**
  - Following three vectors forms the basis in $\mathbb{R}^3$ vector space
    - $\mathbf{B} \coloneqq \{\mathbf{v_1} = [1, 0, 0], \mathbf{v_2} = [0, 1, 0], \mathbf{v_3} = [0, 0, 1]\}$ as they are:
    - Linearly independent and
    - There span is all of $\mathbb{R}^3$
  - If you add a fourth vector (e.g. $v_4 = [1,1,1]$),
    - $\mathbf{B} \coloneqq \{\mathbf{v_1} = [1, 0, 0], \mathbf{v_2} = [0, 1, 0], \mathbf{v_3} = [0, 0, 1], \boldsymbol{v_4} = [1, 1, 1]\}$
    - you still span $\mathbb{R}^3$, but now it not minimal, Thus not a basis anymore.



2.7.1 Basis of a Vector Space: Intuition.

- **Intuition:**
  - A basis is like a **"minimal" coordinate system for a space.**
  - It gives you the smallest set of building blocks from which you can construct any vector in the space.
  - Think of:
    - $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$ as the **standard basis** for $\mathbb{R}^2$.
  - Any $\mathbf{x} \in \mathbb{R}^2$ can be written as:
    - $\mathbf{x} = \mathbf{a}\begin{bmatrix}1\\0\end{bmatrix} + \mathbf{b}\begin{bmatrix}0\\1\end{bmatrix}$; **here a, and b are any scalar in** $\mathbb{R}$.

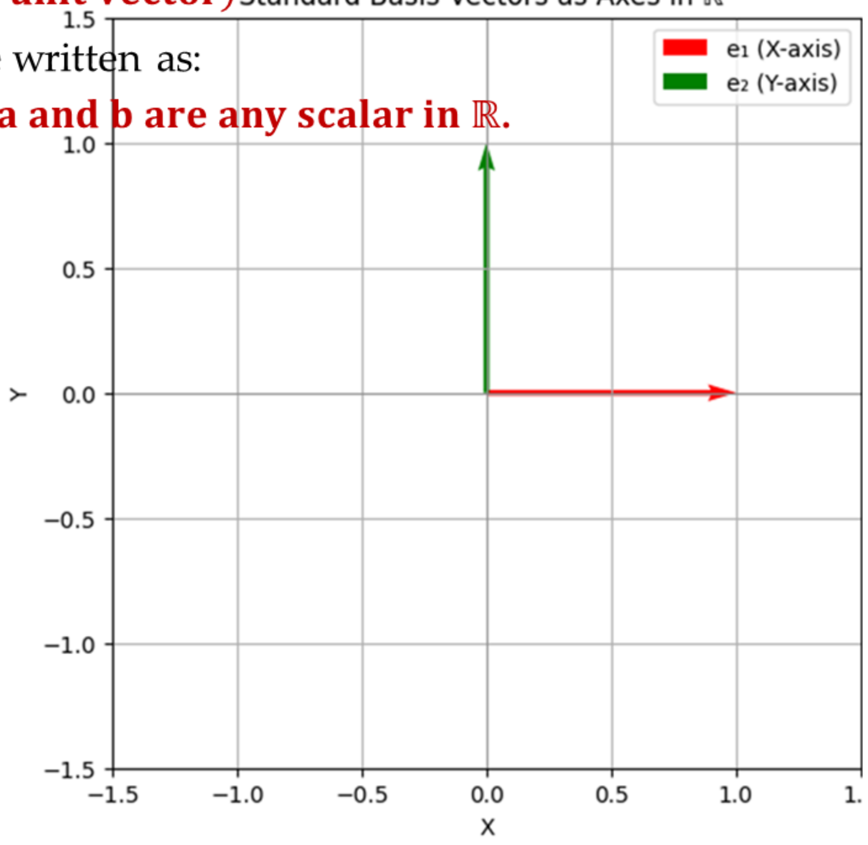


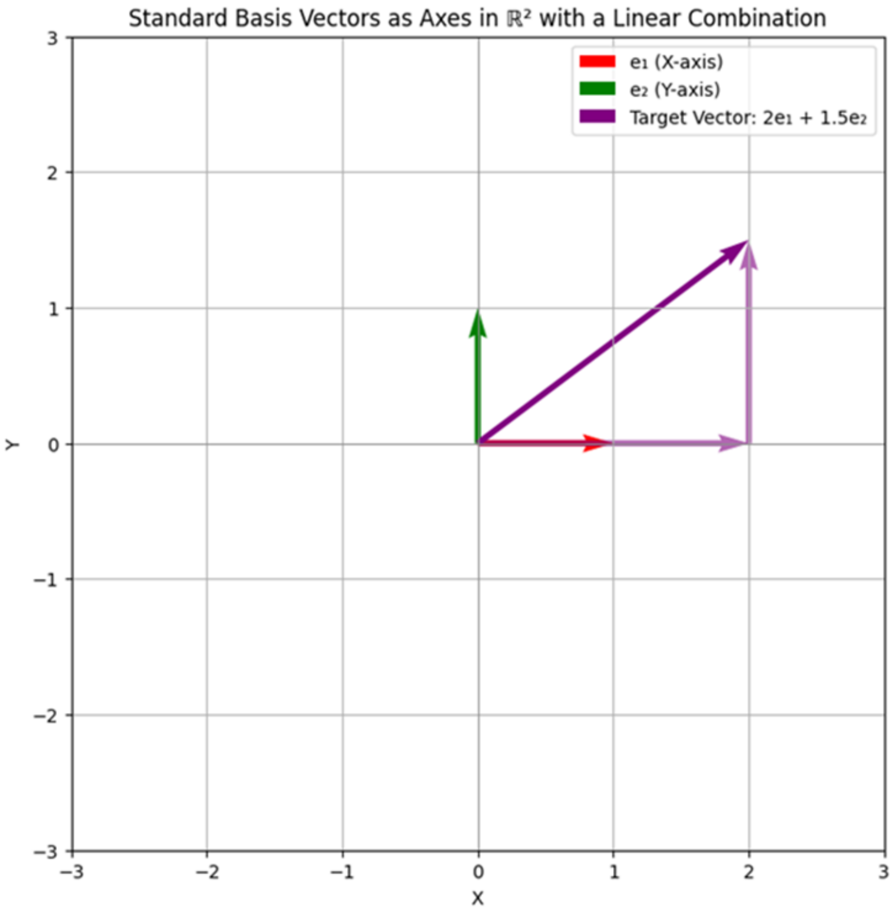
2.7.2 Standard Basis.

- **Standard Basis:**
  - For $\mathbb{R}^n$, the standard basis is: $\{\mathbf{e_1}, \mathbf{e_2}, \dots, \mathbf{e_n}\}$
    - Where:
      - $\mathbf{e_i} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{1\ in\ i^{th} position} \\ \vdots \\ \mathbf{0} \end{bmatrix} \Rightarrow$ **a vector with 1 in position i, 0 elsewhere**
- **Connection to Unit Vectors:**
  - Each standard basis vector e_i is also a unit vector:
    - $\|\mathbf{e_i}\| = \sqrt{\mathbf{1^2 + 0^2 + \cdots + 0^2}} = \mathbf{1}$
    - And they point **along the coordinate axes**.
- So, we say: The standard basis vectors are unit vectors along the coordinate directions of $\mathbb{R}^n$.

2.7.2.1 Standard Basis: Example

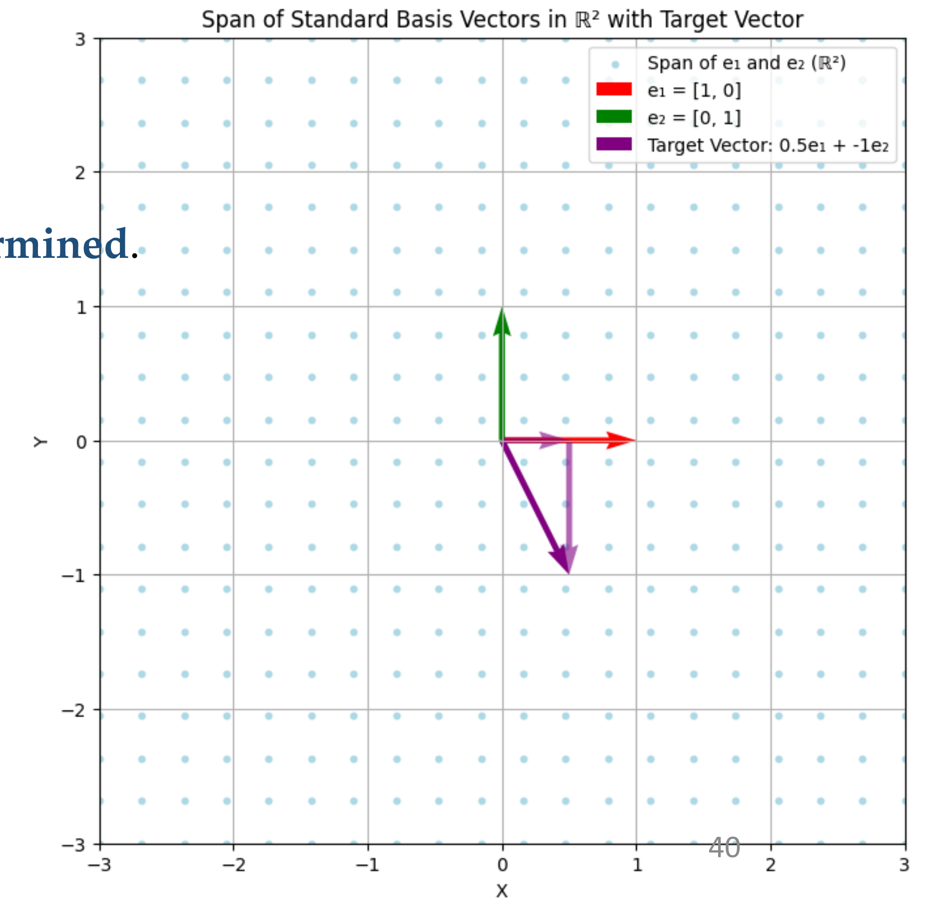
- Standard Basis in $\mathbb{R}^2$:
  - $\mathbf{e_1} = [\mathbf{1}, \mathbf{0}]$ $(\mathbf{x}$ $-$ **axis unit vector**$)$
    $\mathbf{e_2} = [\mathbf{0}, \mathbf{1}]$ $(\mathbf{y}$ $-$ **axis unit vector**$)$
- Any vector $\mathbf{x} \in \mathbb{R}^{\mathbf{2}}$ can be written as:
  - $\mathbf{x} = \mathbf{a}\mathbf{e_1} + \mathbf{b}\mathbf{e_2}$**, here a and b are any scalar in** $\mathbb{R}$**.**





2.7.3 Properties of Basis.

1. **Unique Representation:**
    - Every vector in a vector space can be uniquely represented as a linear combination of the basis vectors.
    - What it means:
        - If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of **vector space V**,
        - Then any $\mathbf{x} \in \mathbf{V}$ can be written as:
            - $\mathbf{x} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n$
                - And the **coefficients $\alpha_1, \alpha_2, \dots, \alpha_n$ are uniquely determined**.
    - Why?
        - Because the **basis vectors are linearly independent**
            - there is **only one way to combine** them to
                - **reach a given vector in the space.**



2.7.3.1 Properties of Basis.

2. **Same number of elements in any Basis:**
   - All bases of a **finite dimensional vector space have the same number of vectors**.
   - What it means:
     - If one **basis of V has n vectors**, then **every basis of V has exactly n vectors**.
     - You cannot have a basis of 2 vectors and another basis of 3 vectors for the same space.
   - Why?
     - Because the number of vectors in a basis corresponds to the dimension of the space.
3. **Dimension of the Vector space:**
   - The **number of vectors** in a basis is called the **dimension of the vector space**.
   - Examples:
     - $$\mathbb{R}^2: \text{dimension} = 2 \ \rightarrow \text{basis has 2 vectors}$$
       $$\mathbb{R}^3: \text{dimension} = 3 \ \rightarrow \text{basis has 3 vectors}$$
       $$\text{A plane in } \mathbb{R}^3: \text{dimension} = 2 \ \rightarrow \text{any basis of that plane will have 2 vectors.}$$
     - We can have a 2-dimensional subspace of $\mathbb{R}^n$ that has a basis of 2 linearly independent vectors.
     - **Cautions:** A basis of a vector space must have **exactly as many vectors as the dimension of the space**.

2.7.3.2 Properties of Basis.

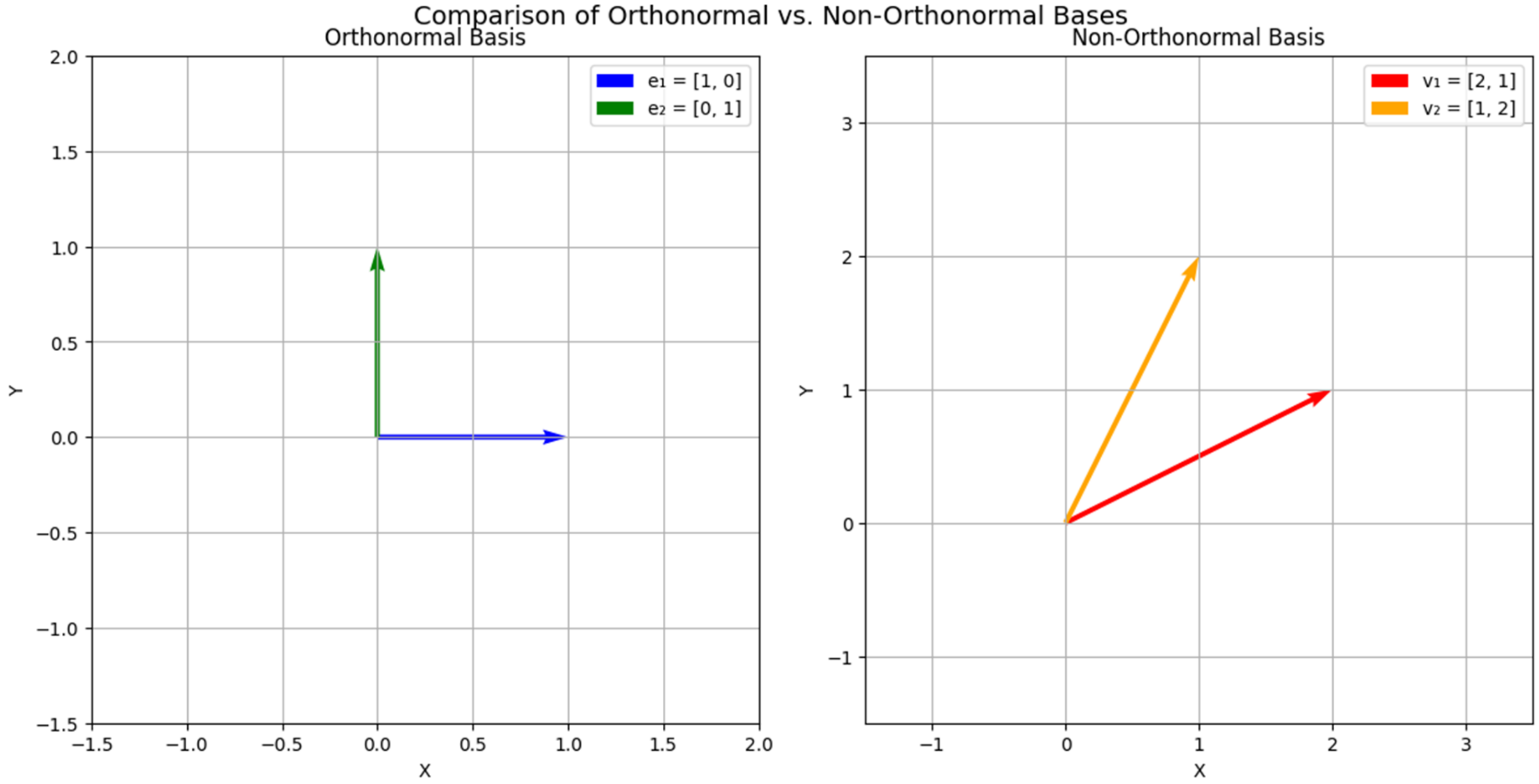
4. **We can have infinitely many bases for the same vector space but all of them must:**
    - Be linearly independent.
    - Span the space.
    - Contain the same number of vectors (equal to the dimension of the space).

- **Example: Multiple Bases in $\mathbb{R}^2$:**
    - **Standard Basis:**
        - $\boldsymbol{\mathcal{B}_1 = \left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}}$
    - Another valid Basis:
        - $\boldsymbol{\mathcal{B}_2 = \left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}}$
    - Both sets have 2 linearly independent vectors.
    - Both span $\mathbb{R}^2 \rightarrow$ every vector in $\mathbb{R}^2$ can be written as unique linear combination of them.
    - **Thus, both are valid basis.**

2.7.4 Orthogonal Basis.

- An orthogonal basis for a vector space is a basis in which all vectors are mutually perpendicular (orthogonal) to each other.
  - If in addition, all the vectors are of unit length, it becomes an orthonormal basis.
- **Formal Definition:**
  - Let $\mathbf{V}$ be an $\mathbf{n-dimensional}$ **inner product space** (like $\mathbb{R}^n$ with dot product). A set of vectors
    - $\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}\} \in \mathbf{V}$ is an **orthogonal basis if :**
      - $\mathbf{v_i} \cdot \mathbf{v_j} = \mathbf{0 \text{ for all } i \neq j}$.
    - **That is each pair of distinct vectors is orthogonal (dot product = 0).**
- **Example in $\mathbb{R}^3$:**
  - The standard basis in $\mathbb{R}^3$: $\mathbf{e_1} = [\mathbf{1}, \mathbf{0}, \mathbf{0}], \mathbf{e_2} = [\mathbf{0}, \mathbf{1}, \mathbf{0}], \mathbf{e_3} = [\mathbf{0}, \mathbf{0}, \mathbf{1}]$
    - is an orthogonal basis and also orthonormal since:
      - **They are mutually perpendicular:** $\mathbf{e_i} \cdot \mathbf{e_j} = \mathbf{0}$ **and**
      - $\|\mathbf{e_i}\| = \mathbf{1}$ **for all** $\mathbf{i}$

Fig: Orthogonal vs Non – Orthogonal basis.

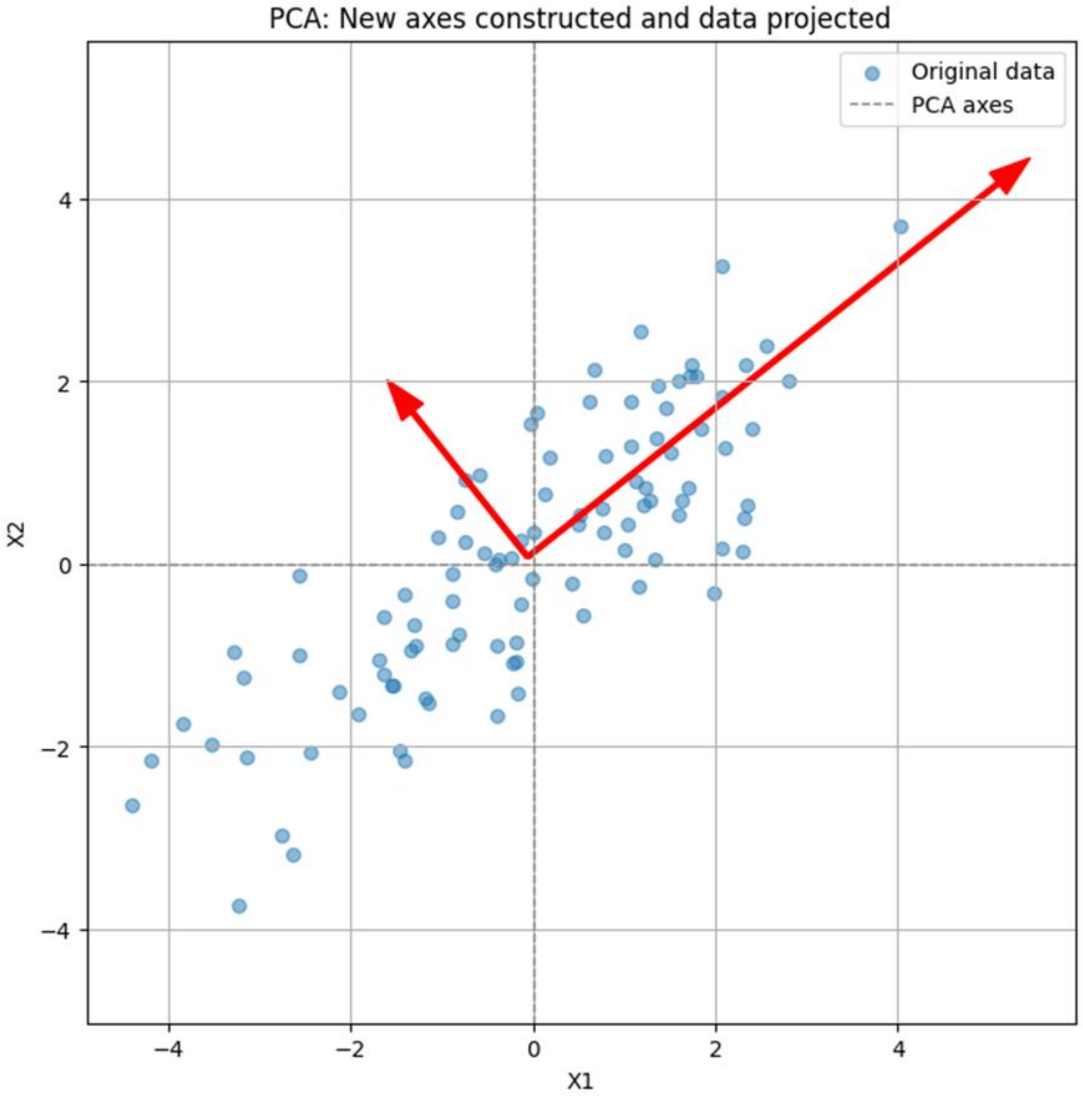


2.8 Linear Models, Linear Combinations and Basis.

- In most classical linear machine learning models, prediction takes the form:
  - $\hat{y} = \mathbf{w}^T\mathbf{x} = \sum_{i=1}^{n} w_i x_i$
    - Each feature $\mathbf{x_i}$ **is weighted by learned parameter** $\mathbf{w_i}$.
  - Implicitly this assumes:
    - **The features are orthogonal i.e. independent.**
    - **Each feature contributes unique information.**
    - **But this assumption does not always hold in real – world data.**
- Correlated Features – Redundant Basis:
  - **When features are correlated:**
    - Two or more features lie in the same direction in the feature space.
    - So, the **feature matrix X has** linearly dependent **columns.**
    - The basis i.e. the feature directions is redundant.
    - The geometry of the model becomes ill – conditioned i.e. numerically unstable.

2.8.1 Linear Models, Linear Combinations and Basis.

- **Solution: change to better basis:**
  - To fix this issue: we can transform the data into a new basis where:
    - Features are uncorrelated (orthogonal)
    - Each axis represents distinct variance.
    - The basis is better aligned with data geometry.
- **How we do that?**
  - We apply **feature extraction techniques** that learn or construct a **new basis** from the data itself.
  - Example: Principal Component Analysis
    - Projects the data onto orthogonal axes (principal components) that maximizes variance.



3. Matrix Algebra.

{Important Matrix Operations.}

3.1 Matrix Determinant.

- **Determinant** of a matrix, denoted by **det(A) or |A|**, is a **real-valued scalar** encoding certain properties of the matrix
  - E.g., for a matrix of size 2×2:

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

  - For larger-size matrices the determinant of a matrix id calculated as

    $\mathbf{det(A)} = \sum_j \mathbf{a_{ij}} (-\mathbf{1})^{\mathbf{i+j}} \mathbf{det}\left(\mathbf{A}_{(\mathbf{i,j})}\right)$

  - In the above, $\mathbf{A}_{(\mathbf{i,j})}$ is a minor of the matrix.
- Properties:
  - $\det(AB) = \det(BA)$
  - $\det(A^{-1}) = \frac{1}{\det(A)}$
  - $\det(A) = 0 \rightarrow A$ is singular i. e. non square matrix.

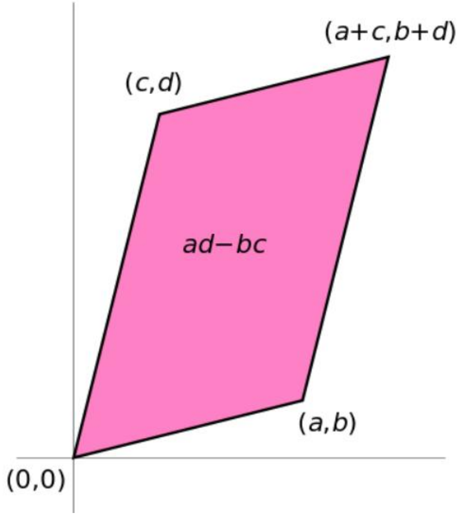
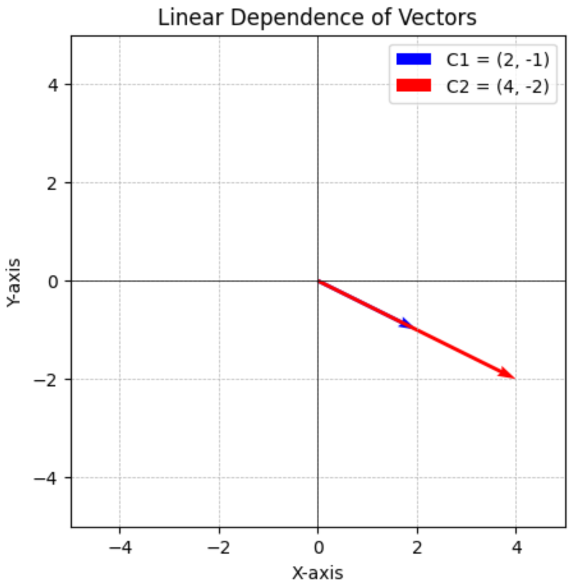
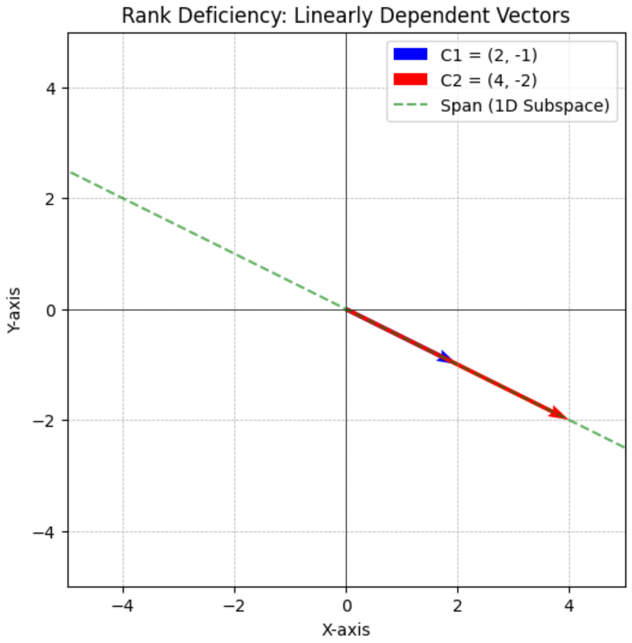


Fig: determinant represents area (or volume) of the parallelogram described by the vectors in the rows of the matrix

3.2 Rank of a Matrix.

- For **m × n matrix** the rank of the matrix is the largest number of linearly independent row or columns.
- For Example:
  - For Matrix; **B** $\coloneqq \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}$ Find the **Rank and Interpret**.
  - **Our Observation:**
    - The second **column** $\boldsymbol{c_2}$ can be written as : $\boldsymbol{c_2 = 2 \times c_1}$
      - Since one **column can be expressed as a multiple of the other**, there is only **one independent column**.
  - Thus, the **rank of B is 1**, meaning it can span only **a 1 D space in** $\mathbb{R}^2$ vector space.
    - All columns of B lie along the same line in $\mathbb{R}^2$
    - Since the **full rank of 2 × 2 matrix in** $\mathbb{R}^2$ **vector space is 2**,
      - B is considered **rank – deficient**.
- **Why it matters?**
  - The matrix cannot invertedly transform $\mathbb{R}^2$.
  - It collapses the 2D space onto a 1D line.
  - This is often a sign of Singularity (determinant = 0), which makes the matrix non – invertible.





3.3 Inverse of a Matrix.

- The inverse of a square matrix A, denoted as A^{-1}, is a matrix that satisfies:
  - $\mathbf{AA^{-1} = A^{-1}A = I}$
    - here **I** is the identity matrix.
- **Conditions for Invertibility:**
  - A **matrix $\mathbf{A_{m\times n}}$** has an **inverse** if and only if :
    - It is a **square matrix $(\mathbf{n \times n})$**.
    - Its **determinant** is **nonzero** i.e. $\mathbf{det(A) \neq 0}$.
    - Its **rank is full**, meaning $\mathbf{rank(A) = n}$.
  - If any of these conditions fail, **the matrix is singular and does not have an inverse**.

3.3.1 Finding inverse of a Matrix.

- If Inverse Exist, we can find the inverse of a Matrix by:
  - **Using Row Reduction:**
    - Row reduction is a method of transforming a matrix into a simpler form (row echelon form – REF) usually the identity matrix for finding the inverse.
    - **REF** can be reached via following valid row operations:
      - Swap two rows
      - Multiply a row by a non zero scalar
      - Add or subtract multiples one to/from another row.
    - It can be done using:
      - Gaussian Elimination: **Transform** the matrix A into REF and then use back – substitution to solve for the inverse.
      - Gauss – Jordan Elimination: **Transform** the matrix A into the **identity matrix** directly, with no need for back substitution.
  - **Using Adjoint (Cofactor) Formula:**
    - Find the inverse of A using the **adjoint** (also called adjugate) **of the matrix**.
      - $\mathbf{A^{-1} = \frac{1}{det(A)} \times adj(A)}$,
      - For 2×2 matrix:
        - $\mathbf{A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}}$;
          - $\mathbf{A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}}$

3.4 System of Linear Equation.

- A **system of linear equation** is a **collection of one or more linear equations** that share a common **set of variables**. For example:
  - $2x + y = 5$
    $3x + 4y = 6$
- **Types of Systems as per solution behavior:**
  - **Consistent System**: A system that has at least one solution.
    1. **Unique Solution**: Occurs when the system has a single solution.
    2. **Infinite Solutions**: Occurs when the system has many solutions.
  - **Inconsistent System**: A system that has no solution.
- **Types of Systems as per Equation behavior:**
  - **Determined system:** system has exactly as many equations as unknowns (i.e. variables).
    - Usually, exact solution and has **Square Shape**.
  - **Underdetermined system:** System has fewer equations than unknowns.
    - $x + y + z = 1 \rightarrow$ **one equation three varaibles**.
    - Usually, **infinite solution** and has **Wide matrix shape i.e. m < n**.
  - **Overdetermined system:** More equations than unknowns: **m > n.**
    - Usually **no exact solution**, but you can find a **best approximation** using **least squares**.

3.4.1 Solving System of Linear Equations.

- There are different techniques , our interest is **Matrix Method (aka Matrix Inversion Method).**
  - Any system of Linear Equation:
    - $\mathbf{a_{11}x_1 + a_{12}x_2 = b_1}$
      $\mathbf{a_{21}x_1 + a_{22}x_2 = b_2}$
  - can be represented in the form:
    - $\begin{bmatrix} \mathbf{a_{11}} & \mathbf{a_{12}} \\ \mathbf{a_{21}} & \mathbf{a_{22}} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} = \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix}$ i.e. $\rightarrow$ $\mathbf{Ax = b}$
  - here:
    - $\mathbf{A} \rightarrow$ is a matrix of coefficients with size $m \times n$, m is the number of equations and n is the number of variable.
    - $\mathbf{x} \rightarrow$ **is** a column vector representing the unknown variables with size $\mathbf{n \times 1}$.
    - $\mathbf{b} \rightarrow$ is a column vector representing the constants with size $\mathbf{m \times 1}$.
- The equation can be modified:
  - $\mathbf{A^{-1}Ax = A^{-1}b}$ {Multiplying both side by A^{-1}}
    $\mathbf{Ix = A^{-1}b}$ {I is the identity matrix}
    $\mathbf{x = A^{-1}b}$ ∎ {you know how to find A^{-1}}

3.5 Matrix – Matrix Multiplication.

- Matrix multiplication between $\mathbf{A} \in \mathbb{R}^{\mathbf{n}\times\mathbf{p}}$ and $\mathbf{B} \in \mathbb{R}^{\mathbf{n}\times\mathbf{p}}$ with resultant matrix $\mathbf{C} \in \mathbb{R}^{\mathbf{m}\times\mathbf{p}}$ can be defined as:

$$\boxed{A.B := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{np} \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mp} \end{bmatrix}}$$

$$\boxed{c_{ij} := \sum_{l=1}^{n} a_{il} b_{lj}; \text{ with i=1,...m; and j=1,...,p}}$$

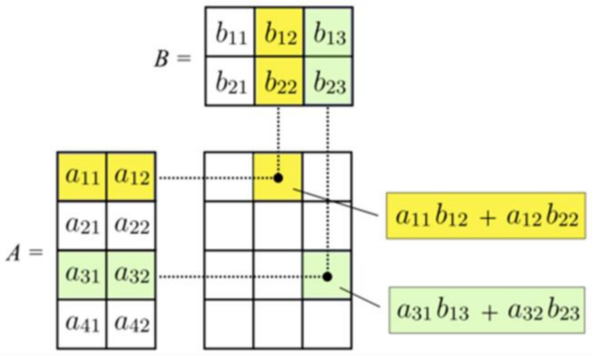


Fig: Schematic representation of Matrix product

Properties of Matrix – Matrix Multiplication:

1. Associativity: $\mathbf{(AB)C = A(BC)}$
2. Associativity with scalar Multiplication: $\boldsymbol{\alpha}\mathbf{(AB) = (}\boldsymbol{\alpha}\mathbf{A)B}$
3. Distributive with sum: $\mathbf{A(B \mp C) = AB \mp BC}$
4. Cautions!! In matrix – matrix multiplication orders matter, it is not commutative i.e. $\mathbf{AB \neq BA}$.

3.6 Matrix – Vector Multiplication.

- Matrix-vector multiplication is an operation between a matrix and a vector that produces a new vector.
- Matrix-vector multiplication equals to taking the dot product of each column n of matrix-A with each element of vector-x resulting in vector y and is defined as:

$$A.X := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_{mn} \end{bmatrix}$$

- Matrix – vector multiplication can be interpreted as taking a **linear combination** of the columns of a matrix A **weighted** by elements of **vector x**.
  - What can be the consequences of such operation?

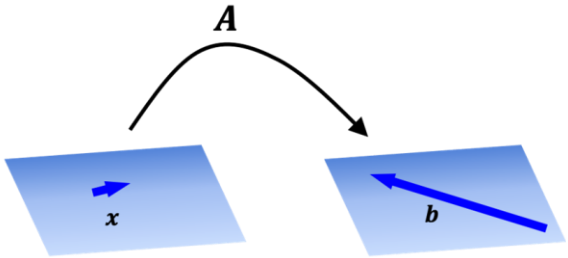
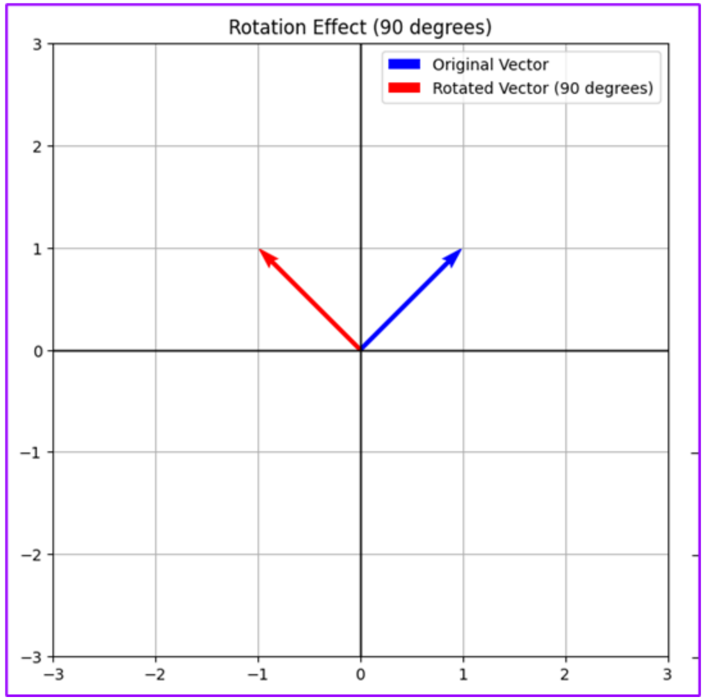


Fig: How my vector will Transformed?

Matrix – vector multiplication can result in:
Change in magnitude **or**,
Change in direction or,
Both changes depending on the matrix involved.

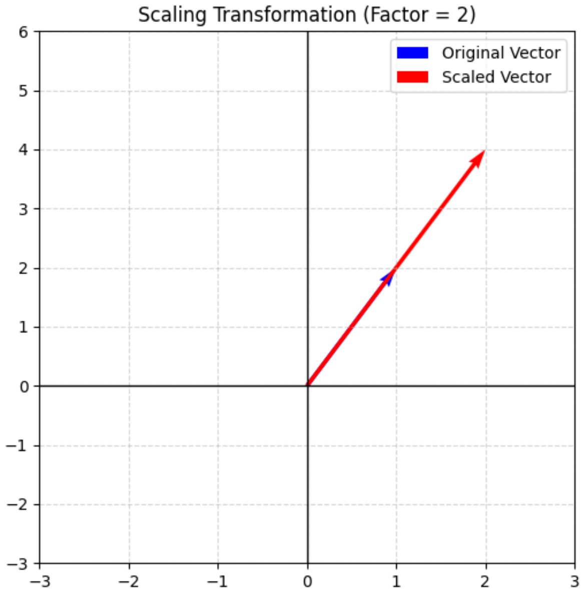
3.6.1 Geometric Interpretation of Matrix – vector Multiplication.

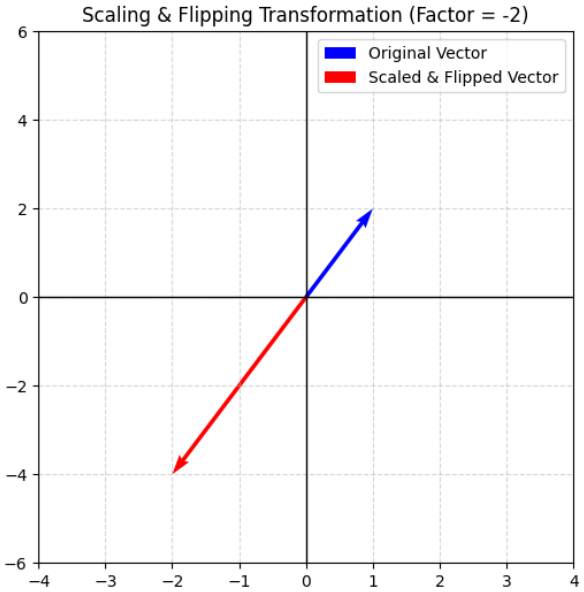
- **Rotation Matrix:**
  - A rotation matrix rotates a vector by a specified angle while preserving its magnitude.
  - **Example**: A 2D rotation matrix that rotates a vector by 90 degrees counterclockwise:
    - $\mathbf{R} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
  - **Effect**: This matrix rotates the vector without changing its length.
- **Example Calculation**:
  - Given the vector $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rightarrow \mathbf{Rv} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
  - The **magnitude remains 1**, but the **direction changes** from the **x-axis to the y-axis**.



3.6.2 Geometric Interpretation of Matrix – vector Multiplication.

- **Scaling Matrix:**
  - A scaling matrix increases or decreases the magnitude of the vector without changing their direction.
    - $\mathbf{S} = \begin{bmatrix} \mathbf{k} & \mathbf{0} \\ \mathbf{0} & \mathbf{k} \end{bmatrix}$
  - where $\mathbf{k}$ is **the scaling factor**:
    - If $\mathbf{k} > \mathbf{1}$, the vector is stretched.
    - If $\mathbf{0} < \mathbf{k} < \mathbf{1}$, the vector is compressed.
    - If $\mathbf{k} < \mathbf{0}$, the vector is **flipped** and scaled.
- Example:
  - Given a vector: $\boldsymbol{v} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix}$ and;
  - a scaling matrices
    - $\boldsymbol{i})\ \begin{bmatrix} \mathbf{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{2} \end{bmatrix}$ and $\boldsymbol{ii})\ \begin{bmatrix} \mathbf{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{-2} \end{bmatrix}$;
  - Applying $\boldsymbol{S}$ **to** $\boldsymbol{v}$:
    - $\mathbf{i})\ \begin{bmatrix} \mathbf{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{2} \\ \mathbf{4} \end{bmatrix}$
    - $\mathbf{ii})\ \begin{bmatrix} \mathbf{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{-2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{-2} \\ \mathbf{-4} \end{bmatrix}$





3.7 Formalizing Matrix – vector Multiplication.

- **Linear Transformation: An overview.**
  - A **linear transformation** is a **function T** that maps vectors from one vector space to another (possibly the same space), preserving two key operations:
  - **Additivity:**
    - $\mathbf{T(u+v) = T(u) + T(v)}$
  - **Homogeneity (Scalar Multiplication):**
    - $\mathbf{T(cv) = cT(v)}$
      - for all vectors u, v, and scalars c.
- Every **linear Transformation T** from $\mathbb{R}^{\mathbf{n}}$ **to** $\mathbb{R}^{\mathbf{m}}$ can be represented by an $\mathbf{m \times n}$ **matrix A**.
  - Applying T to a vector x is the same as multiplying by the matrix:
    - $\mathbf{T(x) = Ax}$
      - Each column of A shows how the transformation acts on a **basis vector** of the input space.
      - Multiplying **A by x** combines these transformed basis vectors weighted by the **coordinates in x.**

Thank You