6CS012 – Artificial Intelligence and Machine Learning.
Lecture – 01
Foundational Math Skills for AI and ML.
A quick revision on Linear Algebra and Derivative.

Siman Giri {Module Leader – 6CS012}

Learning Outcomes!!!

- Review and revise some fundamental concepts from Mathematics – **Linear Algebra and Derivative** we will be using through out the course.
  - *Cautions !!!*
    - We will omit many important topics in Linear algebra and Matrix Calculus, which we believe are not essential for understanding deep learning.
- In Particular we will discuss:
  - Why do we need Linear Algebra for Machine/Deep Learning?
    - {**Almost**} Everything we need to know about vector and matrices for Machine/Deep Learning.
  - A very big picture on Definition of Derivative and Matrix Calculus.

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:
  - $\mathbf{a_1x_1 + \ldots\ldots + a_nx_n = b}$;
  - **linear maps** such as:
    - $\mathbf{(x_1, \ldots\ldots, x_n) \mapsto a_1x_1 + \cdots + a_nx_n}$;
  - 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."**

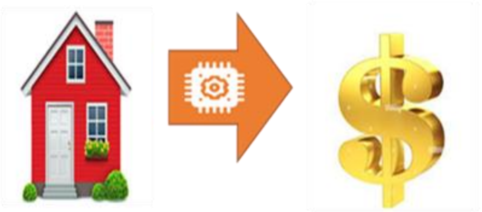
Fig: What is Linear Algebra?

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.



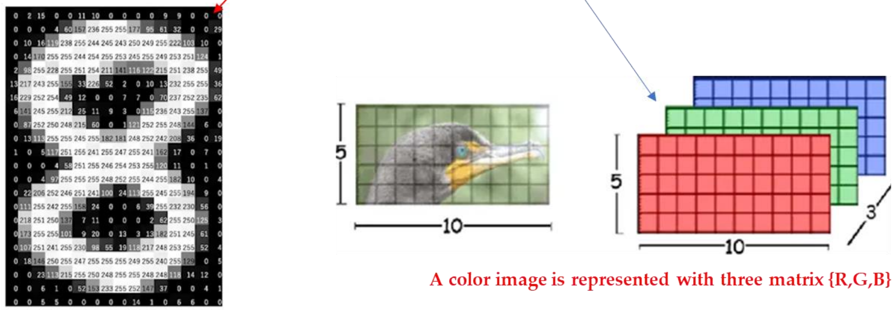
Data: Features/Descriptor of House

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

How would you represent this, for computer?
Matrix.

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

Each pixel value is the shade of Gray 0 – 255.



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**.

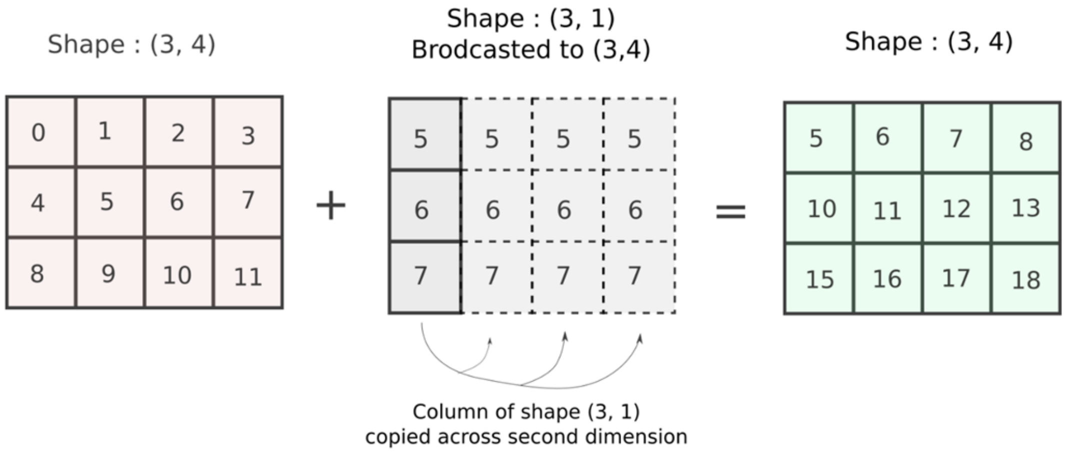


Fig: Idea of Vectorizatons.

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.

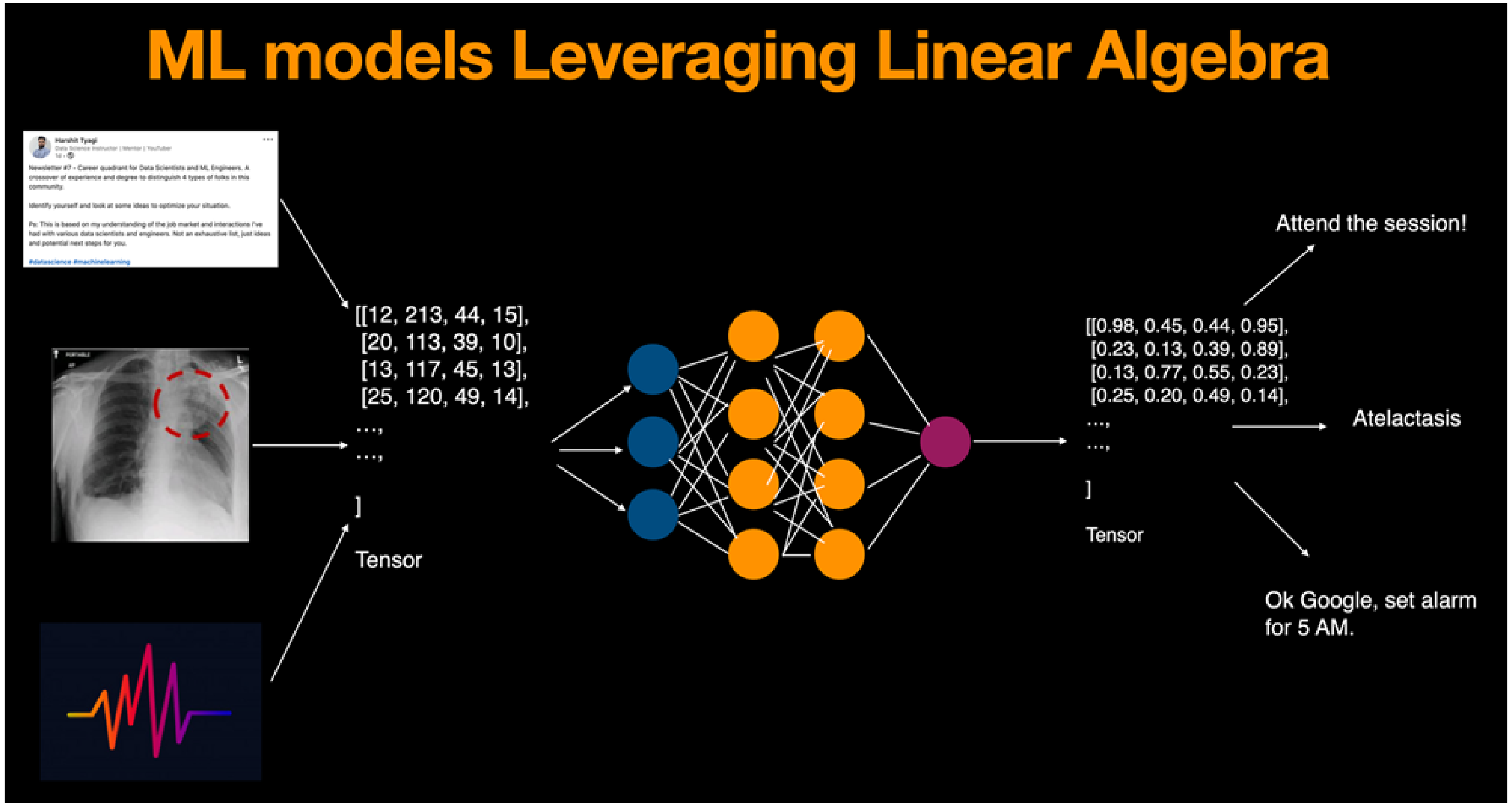


Image By Harshit Tyagi and freeCodeCamp

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**

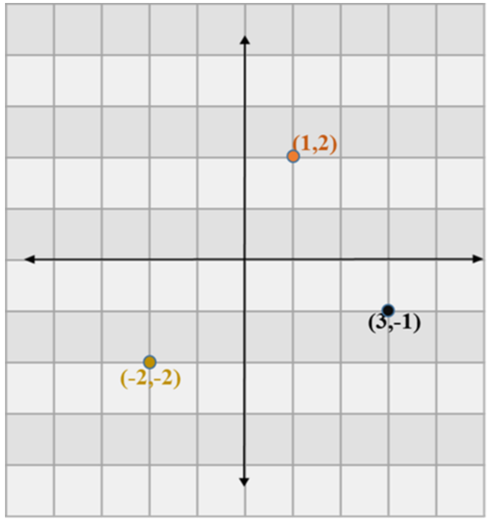


Fig: Vector as a point

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**

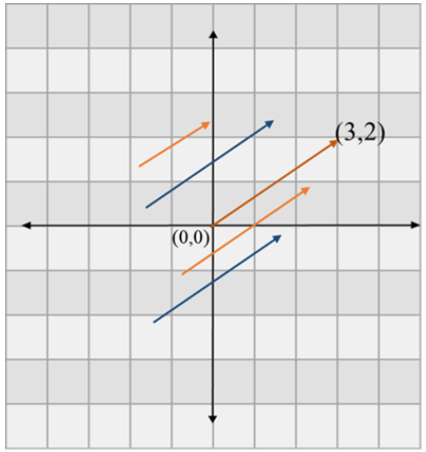


Fig: Vector as Direction

1.2 Vector formal Definition.

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

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

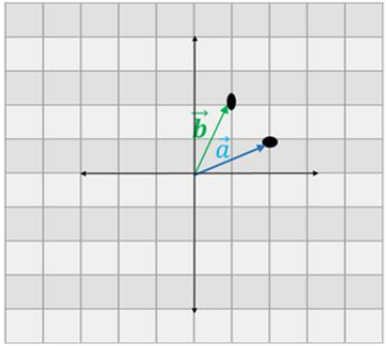


Fig: 2 dimensional vector space

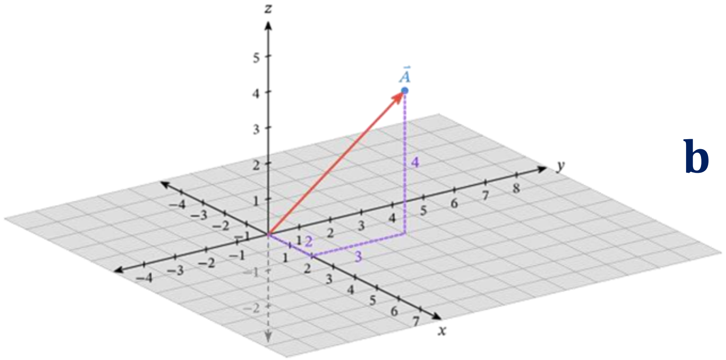


Fig: 3 dimensional vector space

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

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)

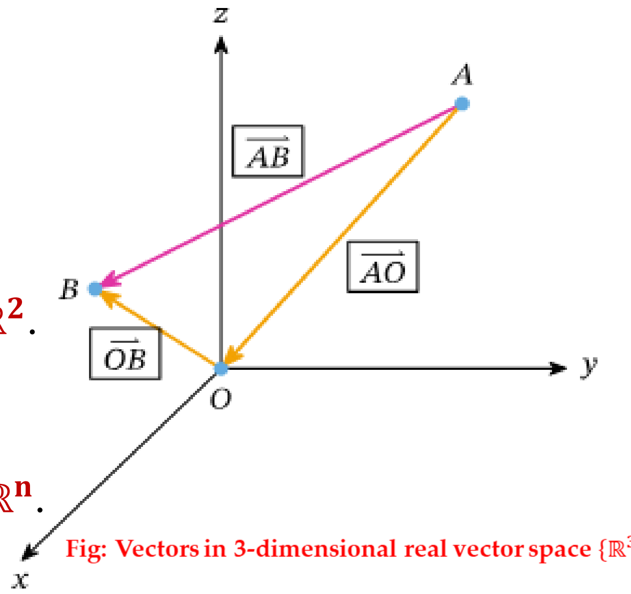


Fig: Vectors in 3-dimensional real vector space $\{\mathbb{R}^3\}$

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 **Arthematic /Algebraic Properties** – We can perform valid mathematical operations.

{details in course note}

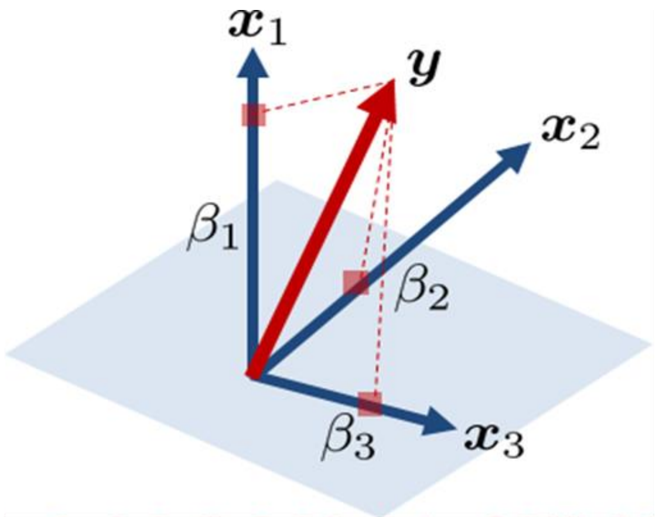


Image from Stanley Chan Book: Introduction to Probability for Data Science.

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 $\mathbf{A}$ has:
    - **m rows** and **n columns**.
    - Each entry/element of $\mathbf{A}$ 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 $\mathbf{C}$.
  - 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 $\mathbf{A}$.
  - For matrix $\mathbf{A} \in \mathbf{C_{m\times n}}$: we let $\mathbf{a_{i*}}$ denote the $\mathbf{i^{th}}$ **row vector of** $\mathbf{A}$.
  - For matrix $\mathbf{A} \in \mathbf{C_{m\times n}}$: we let a_{*i} denote the $\mathbf{j^{th}}$ **column vector** of $\mathbf{A}$.

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:

$$A_{2\times3} \coloneqq \begin{bmatrix} 1 & 2 & 3 \\ 5 & 5 & 4 \end{bmatrix}$$

- 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:

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

- 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:

$$A_{3\times3} \coloneqq \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

- 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:

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

- 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:

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

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

$$A_{3\times3} \coloneqq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

1.6.1 Special Matrices.

- **Symmetric Matrix:**
  - Square matrices are said to be symmetric its equal to its transpose, i.e. $\mathbf{A} = \mathbf{A^T}$. 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 $\mathbf{0_{m\times n}}$. 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
    - $\mathbf{A(a_{ij}) = B(b_{ij})}$.
  - For instance:

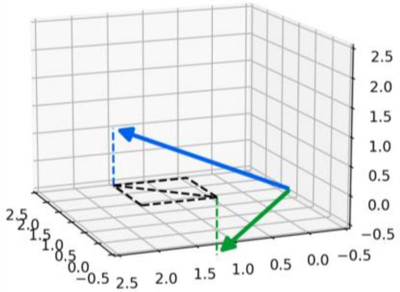
$$B_{2\times2} := \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$$

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

1.7 Interpretation of a Matrix: Collection of Vectors.

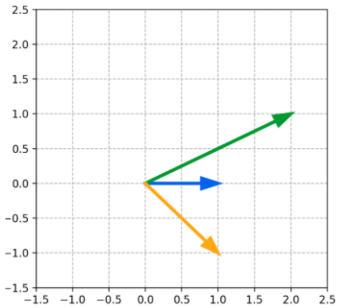
- A matrix can be thought of as a set of vectors.
- For example, for the following matrix:
  - $\mathbf{A} \coloneqq \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & -\mathbf{1} \end{bmatrix}$ can be thought of as
    - a **two three-dimensional** row vectors i.e.
  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

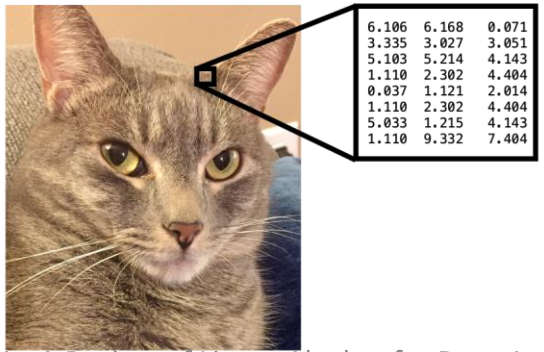
Spreadsheet

	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

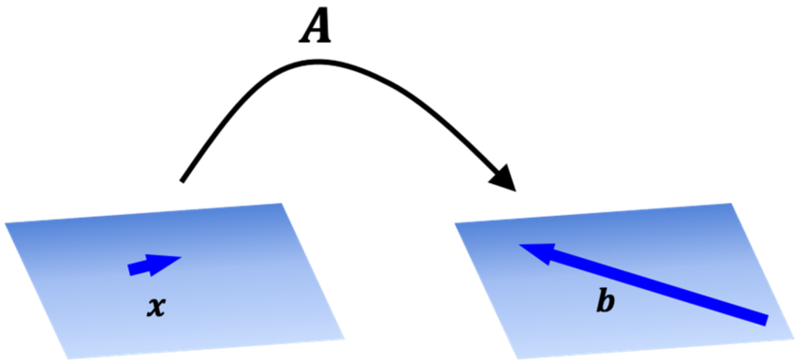
$$\begin{bmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 7.0 & 3.2 & 4.7 & 1.4 \\ 6.5 & 2.8 & 4.6 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \\ 7.1 & 3.0 & 5.9 & 2.1 \end{bmatrix}$$

  - 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**.

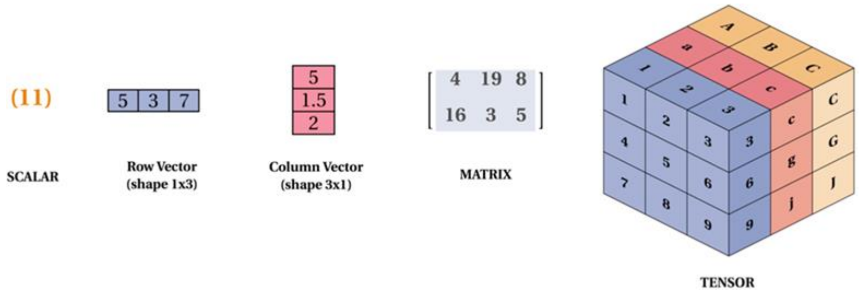


$$A.X \coloneqq \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}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

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:

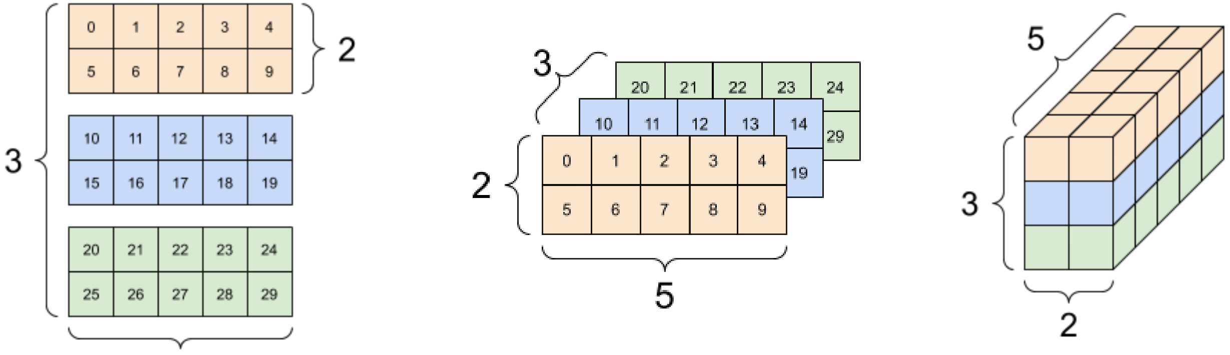
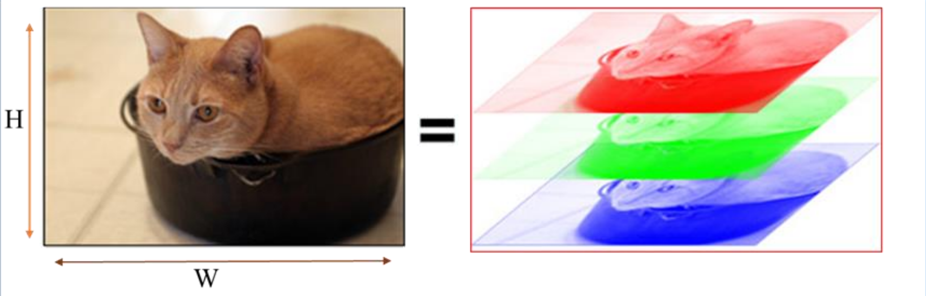


Fig: {Different Form} Visualization of Tensor with 3 axes.

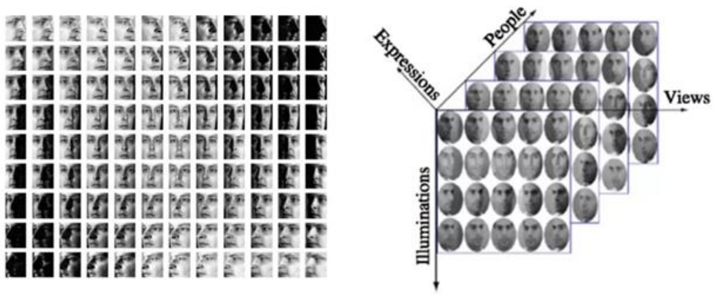
Tensor → Example.

- 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 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:
  - $\mathbf{u}^T\mathbf{v} \in \mathbb{R} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n] \cdot \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \dots \\ \mathbf{v}_n \end{bmatrix} = \sum_{i=1}^{n} \mathbf{u}_i \times \mathbf{v}_i$
- **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.

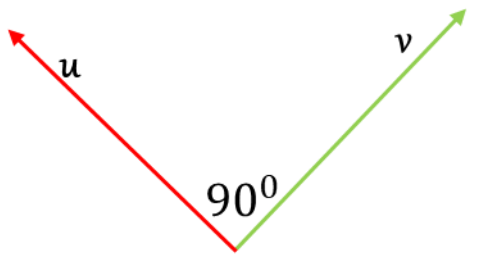
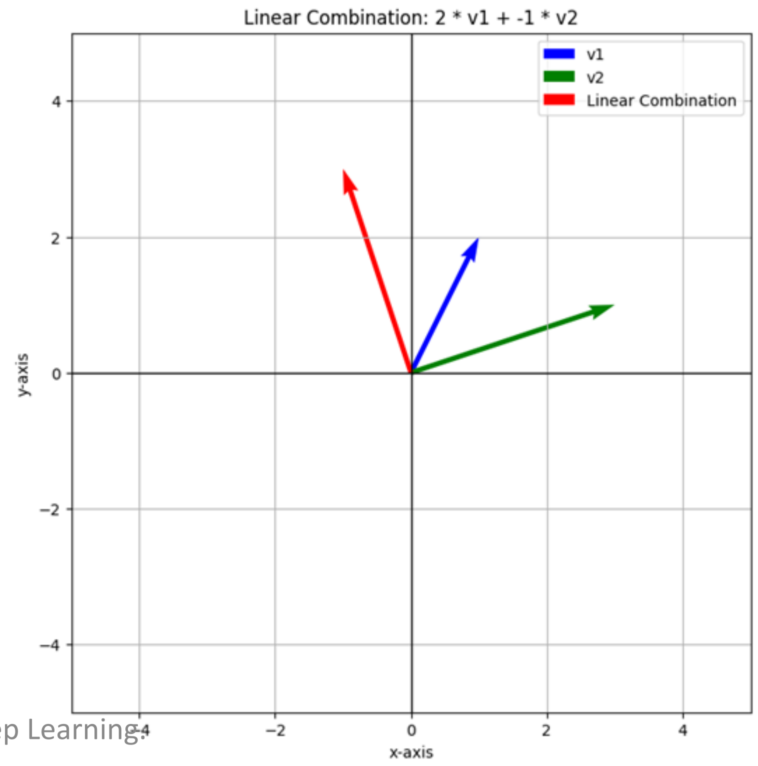


Fig: Orthogonal Vectors

2.2 Linear Combinations of Vectors.

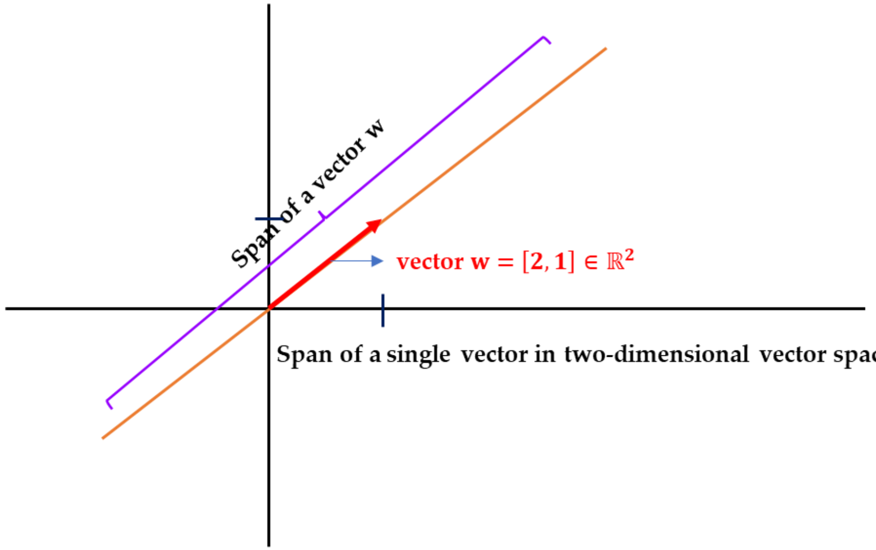
- Idea Combining two or more than vectors to form a new vector.
- Definition:
  - A vector v is a linear combination of a set of vectors $\mathbf{v_1, v_2, \dots, v_n}$, if it can be expressed as:
    - $\mathbf{v} = \mathbf{c_1 v_1} + \mathbf{c_2 v_2} + \cdots + \mathbf{c_n v_n}$
      - where:
   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



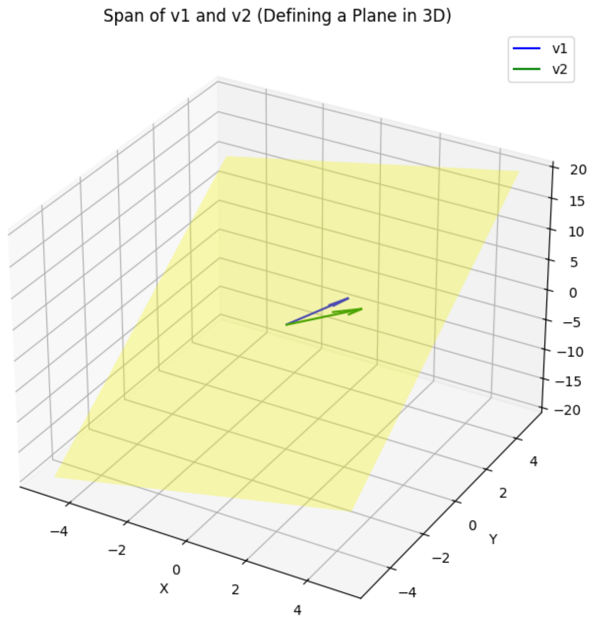
2.3 Span of a Set of vectors.

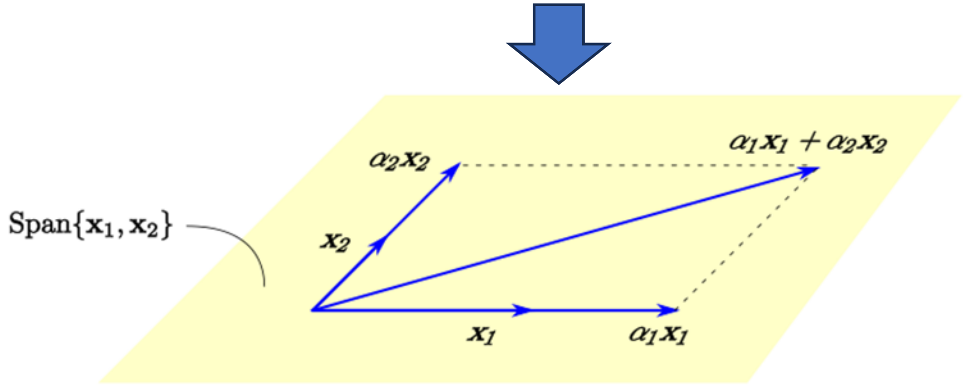
- **Span** is a consequences of **Linear combination of vectors** and can be thought as a **subset** inside a **vector space** (also known as **vector subspace**).
- A subspace, $\mathbb{S}$ **of real vector space** $\mathbb{R}^{\mathbf{n}}$ is thought of a **flat surface (having no curvature) surface** with in $\mathbb{R}^{\mathbf{n}}$:
  - is a collection of **all the vectors in** $\mathbb{S}$ which satisfies the following (algebraic) conditions:
    - The *origin* $(\mathbf{0}$ *vector*$)$ is contained in $\mathbb{S}$.
    - If vector $\boldsymbol{v}_{\mathbf{1}}$ **and** $\boldsymbol{v}_{\mathbf{2}}$ are in $\mathbb{S}$; then $\boldsymbol{v}_{\mathbf{1}} + \boldsymbol{v}_{\mathbf{2}} \in \mathbb{S}$.
    - If $\boldsymbol{v}_{\mathbf{1}} \in \mathbb{S}$ and $\boldsymbol{\alpha}$ a scalar then $\boldsymbol{\alpha v}_{\mathbf{1}} \in \mathbb{S}$.
- The **span of a set of vectors** $\{\boldsymbol{v}_{\mathbf{1}}, \boldsymbol{v}_{\mathbf{2}}, \dots, \boldsymbol{v}_{\boldsymbol{n}}\} \in \mathbb{R}^{\boldsymbol{n}}$ is the **set** of all **possible linear combinations** of **those vectors**. Formally, the span of $\{\boldsymbol{v}_{\mathbf{1}}, \boldsymbol{v}_{\mathbf{2}}, \dots, \boldsymbol{v}_{\boldsymbol{n}}\}$ is:
  - $\boldsymbol{span}(\boldsymbol{v}_{\mathbf{1}}, \boldsymbol{v}_{\mathbf{2}}, \dots, \boldsymbol{v}_{\boldsymbol{n}}) = \{\boldsymbol{c}_{\mathbf{1}}\boldsymbol{v}_{\mathbf{1}} + \boldsymbol{c}_{\mathbf{2}}\boldsymbol{v}_{\mathbf{2}} + \cdots + \boldsymbol{c}_{\boldsymbol{n}}\boldsymbol{v}_{\boldsymbol{n}} | \boldsymbol{c}_{\mathbf{1}}, \boldsymbol{c}_{\mathbf{2}}, \dots, \boldsymbol{c}_{\boldsymbol{n}} \in \mathbb{R}\}$
    - where $\boldsymbol{c}_{\mathbf{1}}, \boldsymbol{c}_{\mathbf{2}}, \dots, \boldsymbol{c}_{\boldsymbol{n}}$ are **scalar coefficients**.

2.3.1 Geometric Interpretation of a Span.



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

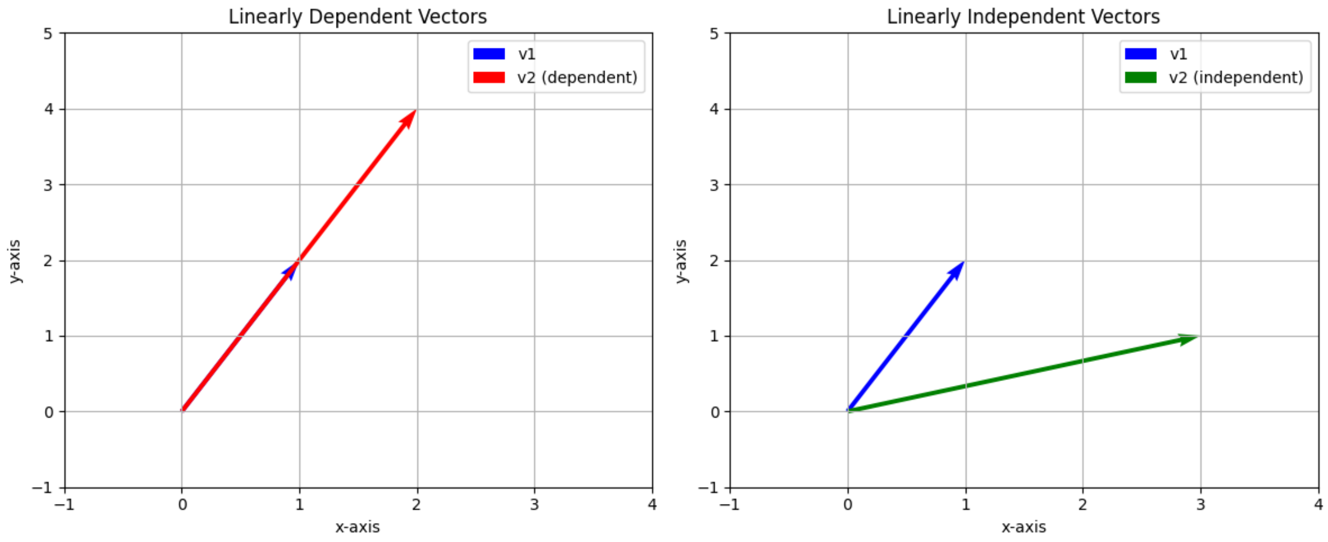




Span of a two vector in three-dimensional space is a plane.

2.4 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.



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

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

  - In the above, $\mathbf{A_{(i,j)}}$ 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.

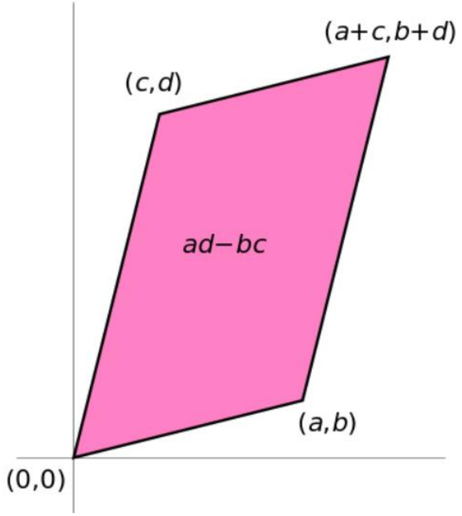
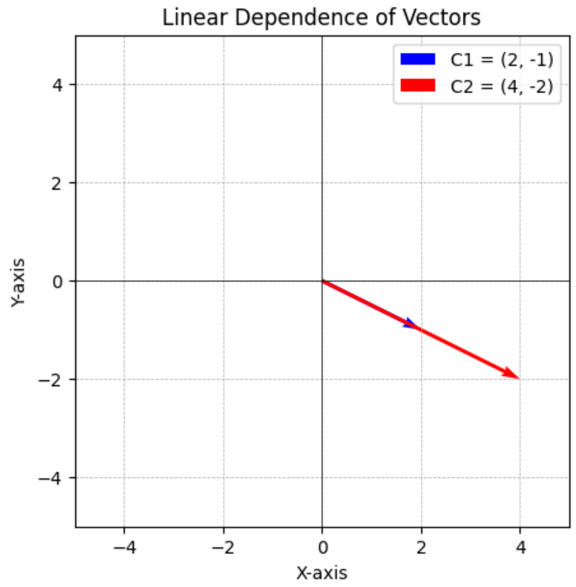
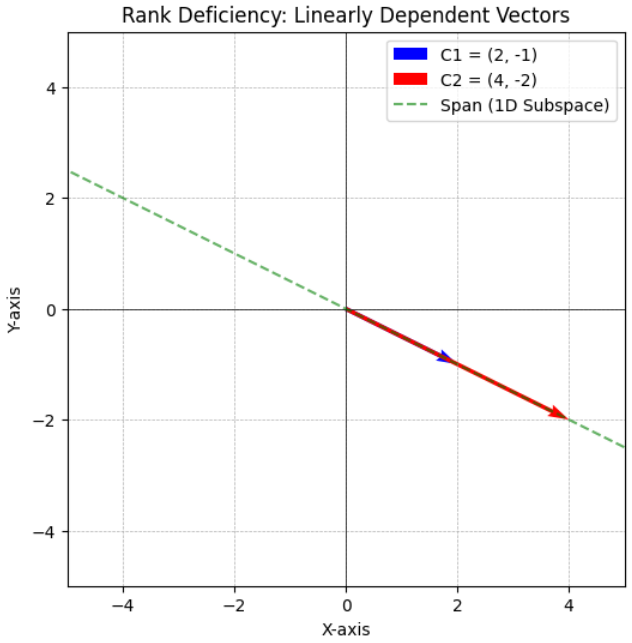


Fig: determinant represents area (or volume) of the parallelogram described by the vectors in the rows of the matrix

3.2 Rank of a Matrix.

- For **$m \times n$ matrix** the rank of the matrix is the largest number of linearly independent row or columns.
- For Example:
    - For Matrix; $\mathbf{B} \coloneqq \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}$ Find the **Rank and Interpret**.
    - **Our Observation:**
        - The second **column c_2** can be written as : $c_2 = 2 \times c_1$
            - 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.
        - Since the **full rank of 2×2 matrix in $\mathbb{R}^2$ vector space is 2**,
            - B is considered **rank – deficient**.





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:**
  - **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.

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:
      - $\begin{bmatrix} \mathbf{a_{11}} & \mathbf{a_{12}} \\ \mathbf{a_{21}} & \mathbf{a_{22}} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} = \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix}$ $\mathbf{i.e.} \rightarrow \ \mathbf{Ax = b}$
    - 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}
        $\mathbf{x = A^{-1}b} \blacksquare$ {you know how to find A^{-1}}

3.5 Matrix – Matrix Multiplication.

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

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

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

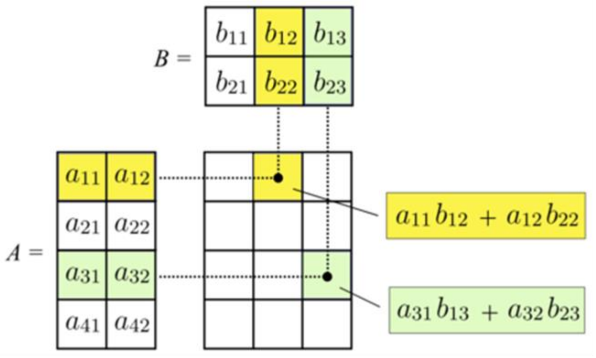


Fig: Schematic representation of Matrix product

Properties of Matrix – Matrix Multiplication:

1. Associativity: $\mathbf{(AB)C = A(BC)}$
2. Associativity with scalar Multiplication: $\mathbf{\alpha(AB) = (\alpha A)B}$
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?

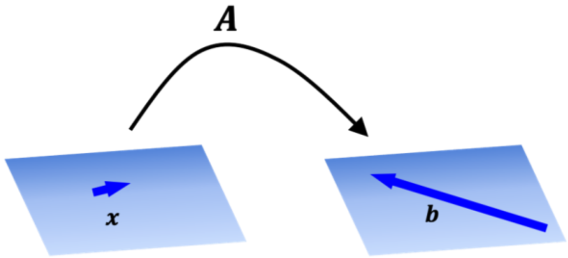


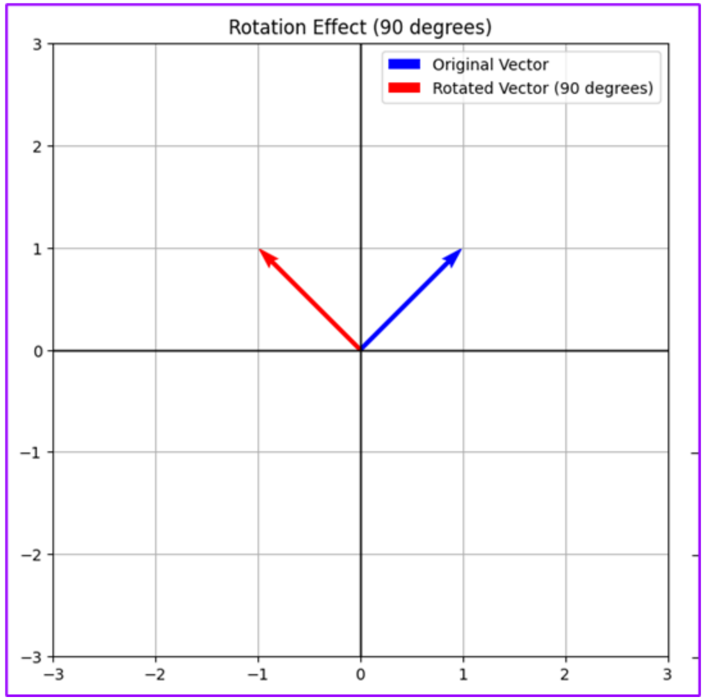
Fig: How my vector will Transformed?

Matrix – vector multiplication can result in:

Change in magnitude or,
Change in direction or,
Both changes depending on the matrix involved.

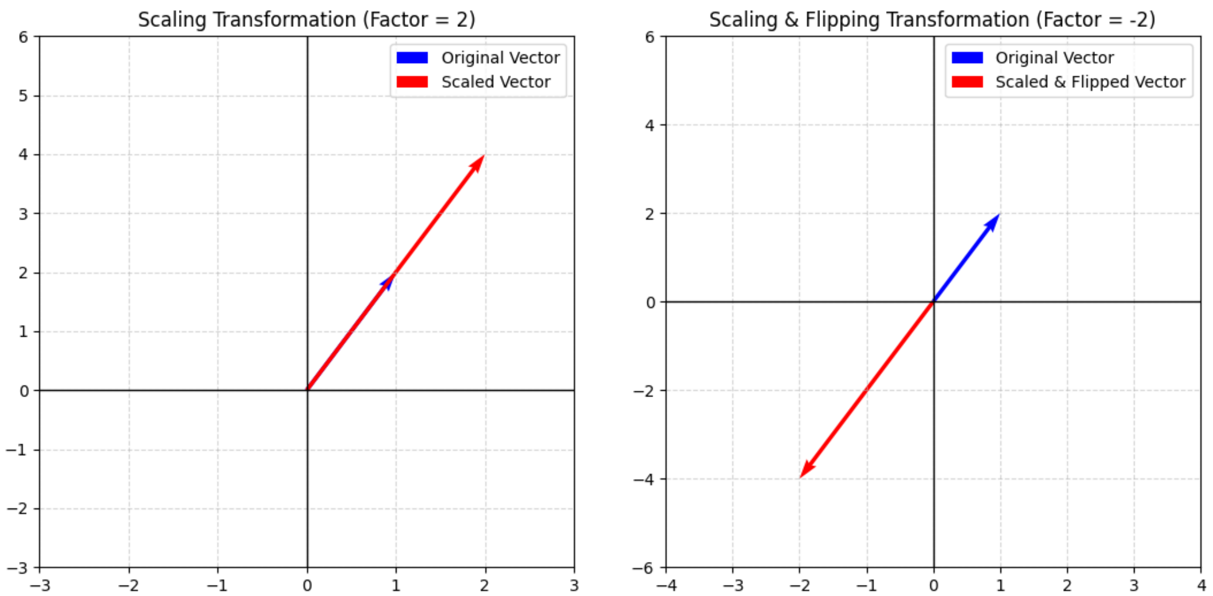
3.6.1 Geometric Interpretation of Matrix – vector Multiplication.

- **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:
    - $\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{-1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$
  - **Effect**: This matrix rotates the vector without changing its length.
- **Example Calculation**:
  - Given the vector $\mathbf{v} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \rightarrow \mathbf{Rv} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$
  - 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 **k** is **the scaling factor**:
    - If $\mathbf{k > 1}$, the vector is stretched.
    - If $\mathbf{0 < k < 1}$, the vector is compressed.
    - If $\mathbf{k < 0}$, 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** $\boldsymbol{v}$:
    - $\mathbf{i)}\ \begin{bmatrix} \mathbf{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{2} \\ \mathbf{4} \end{bmatrix}$
    - $\mathbf{ii)}\ \begin{bmatrix} \mathbf{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{-2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{-2} \\ \mathbf{-4} \end{bmatrix}$



4. Eigen Value Problem.

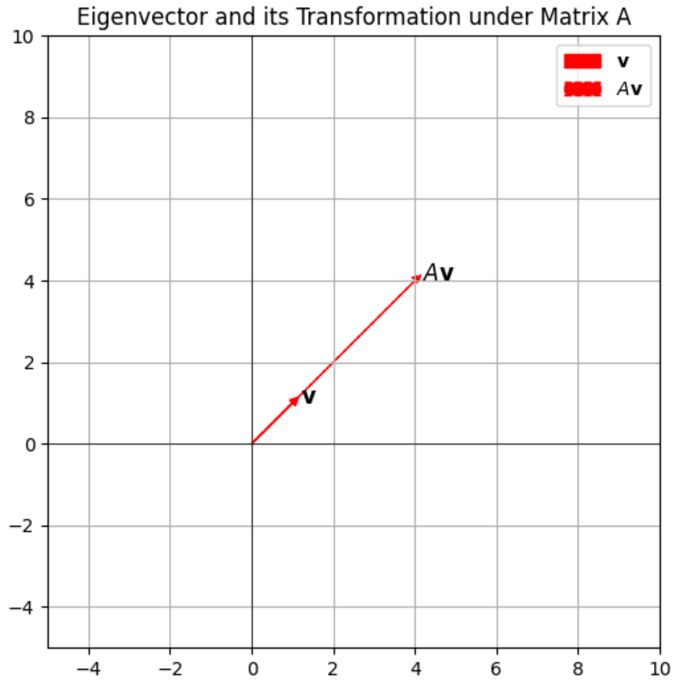
{aka eigen Value Decomposition.}

4.1 Eigen Vector and Eigen Value.

- An **eigenvector** of a **square matrix A** is a **non-zero vector v** that, when **multiplied by A,** results in a scalar multiple of itself.
  - In other words, it is a vector that does not change direction when the linear transformation represented by A is applied to it.
    - Instead, it only gets scaled by a certain factor, called the **eigenvalue**.
- Mathematically, for any **Matrix – vector** pair if following holds:
  - $\boldsymbol{Av = \lambda v}$
    - then the **vector** $\boldsymbol{v}$ is called eigen vector and the **scaling factor** $\boldsymbol{\lambda}$ is called eigen value.
- **Key points about eigen vectors:**
  - **Non-zero**: Eigenvectors are always non-zero vectors, $\mathbf{i.e.\, v \neq 0}$.
  - **Scaling:** The transformation A simply scales the eigenvector by the **eigenvalue** $\boldsymbol{\lambda}$;
    - it does not change the **vector's direction**.
  - **Multiple eigenvectors**: For each eigenvalue, there can be infinitely many eigenvectors, all scalar multiples of each other. They form a subspace (called the eigenspace) corresponding to that eigenvalue.

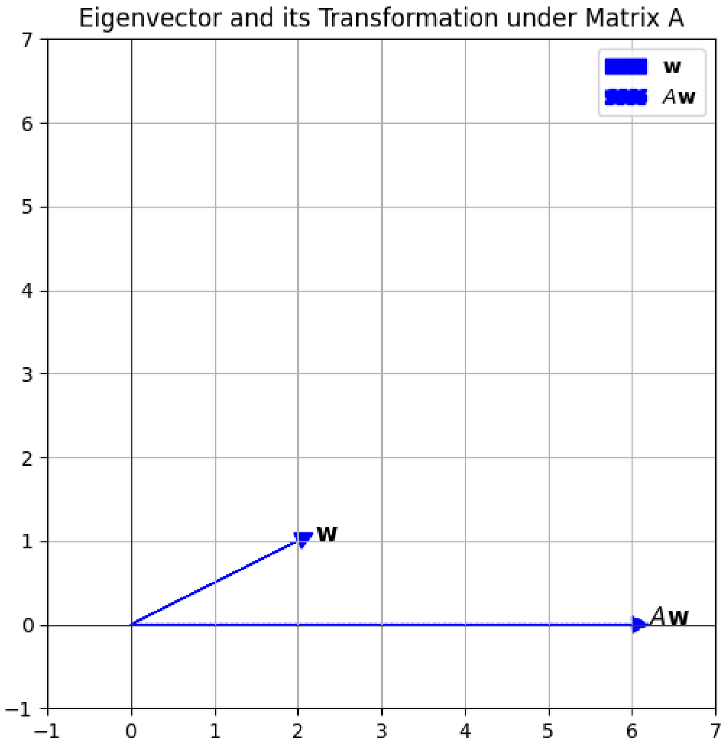
4.1.1 Identify the Eigen Vector.

- **Consider a matrix**
  - $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ -4 & 8 \end{bmatrix}$ and
    - vectors $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$; $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.
- **Which are Eigen vectors?**
  - For $\mathbf{v}$ → **we check if v is an eigenvector by calculating Av**:
    - $\mathbf{Av} = \begin{bmatrix} 2 & 2 \\ -4 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow 4\mathbf{v}.$
    - So, $\mathbf{v}$ is an **eigen vector** with **eigenvalue** $\lambda = 4$.
- What about w?



4.1.1 Identify the Eigen Vector.

- **Consider a matrix**
  - $\mathbf{A} = \begin{bmatrix} \mathbf{2} & \mathbf{2} \\ \mathbf{-4} & \mathbf{8} \end{bmatrix}$ and
    - vectors $\mathbf{v} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}$; $\mathbf{w} = \begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix}$.
- **Which are Eigen vectors?**
  - For w → **we check if *w* is an eigenvector by calculating Aw**:
    - $\mathbf{Aw} = \begin{bmatrix} \mathbf{2} & \mathbf{2} \\ \mathbf{-4} & \mathbf{8} \end{bmatrix} \begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{6} \\ \mathbf{0} \end{bmatrix} \neq \boldsymbol{\lambda w}.$
    - So, **w** is not an **eigen vector there does not exist a**
      - **scalar $\boldsymbol{\lambda}$** under which $\mathbf{Aw} = \boldsymbol{\lambda}\mathbf{w}$ holds true.



4.2 Eigen Value Problem.

- The **eigenvalue problem** is a fundamental concept in linear algebra and plays a critical role in various fields such as machine learning, physics, and computer science.
- It involves **finding scalar values** (called **eigenvalues**) and corresponding **non-zero vectors** (called **eigenvectors**) for a given **square matrix**.
  - Mathematically, Given a **square matrix A**, the eigenvalue problem is to find **scalars λ** and **eigen vector v** that satisfy the following equation.
    - $\mathbf{Av} = \lambda\mathbf{v}$.

4.3 Steps to solve the Eigenvalue Problem.

- Write the characteristic equation:
  - To find the eigenvalues, we rewrite the equation as:
    - $(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I})\mathbf{v} = \mathbf{0}$ {called characterstic equation}
    - Where:
      - **I** is the **identity matrix** of the **same size as** A,
      - $\boldsymbol{\lambda} \rightarrow$ **eigen values**.
      - $\mathbf{v} \rightarrow$ **eigen vector**.
      - *Cautions: the matrix $A - \lambda I$ must be singular i.e. $det(A - \lambda I) = 0$.*
- Compute the characteristic polynomial:
  - Solve
    - $\mathbf{det(A - \lambda I) = 0}$,
  - which gives a **polynomial equation in $\boldsymbol{\lambda}$** which is called characteristic polynomial.
- Solve the characteristic polynomial:
  - Solve the polynomial equation to find the eigenvalues $\boldsymbol{\lambda_1, \lambda_2, \ldots \lambda_n}$.
- Find the eigen vectors:
  - For each **eigen value $\boldsymbol{\lambda_i}$**,
    - substitute it back into the equation $(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I})\mathbf{v} = \mathbf{0}$ and solve for the **eigenvector $\boldsymbol{v}$.**

4.3.1 Example Problem.

Eigenvalues of Matrix A

Consider a matrix A:

$$A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$$

The characteristic equation is:

$$\det(A - \lambda I) = \det \begin{pmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix} = 0$$

Expanding the determinant:

$$(4-\lambda)(3-\lambda) - 2 \times 1 = 0$$

Simplifying:

$$\lambda^2 - 7\lambda + 10 = 0$$

Solving this quadratic equation:

$$\lambda_1 = 5, \quad \lambda_2 = 2$$

4.3.1 Example Problem.

Eigenvectors of Matrix A

Next, we find the eigenvectors:
For $\lambda_1 = 5$, solve $(A - 5I)v = 0$:

$$(A - 5I) = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}$$

Solving the system:

$$\begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

This gives the eigenvector:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

For $\lambda_2 = 2$, solve $(A - 2I)v = 0$:

$$(A - 2I) = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}$$

Solving the system:

$$\begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

This gives the eigenvector:

$$v_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Conclusion: For the matrix A, the eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = 2$, with corresponding eigenvectors $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

4.4 Eigenvalue Decomposition.

- **Eigenvalue Decomposition** is a process where a **square matrix is factorized** into
  - its eigenvalues and eigenvectors.
  - Specifically, for a matrix A, if it can be decomposed into a product of three matrices:
    - $\mathbf{A} = \mathbf{V\Lambda V^{-1}}$
    - where:
      - $\mathbf{A}$ is the original matrix.
      - $\mathbf{V}$ is the matrix whose columns are the eigenvectors of A.
      - $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal entries are the eigenvalues of A.
      - $\mathbf{V^{-1}}$ is the inverse of the matrix V.
- One of the application of Eigenvalue decomposition is **Principal Component Analysis** used for dimensionality reduction purposes.
    - {This workshop we will implement PCA with eigen value decomposition and try to compress the image.}

5. Matrix and Derivative.

{Finding the Slope for Univariate Function.}

5.1 What is Derivative?

- The **derivative of a function** measures **how** the **output value** of the function changes as we make **small adjustments** to its **input**.
- Notations:
  - The derivative of a function $\mathbf{f(x)}$ is represented by $\frac{\mathbf{d}}{\mathbf{d(x)}}\left(\mathbf{f(x)}\right)$ or $\frac{\mathbf{df(x)}}{\mathbf{d(x)}}$ or $\mathbf{f'(x)}$ and is defined as:
- If we have a function $\mathbf{f(x)}$, the derivative $\mathbf{f'(x)}$ at a point $\mathbf{x}$ tell us the rate of change of function $\mathbf{f}$ at that point.
- This rate of change is crucial for optimization techniques,
  - such as finding maxima or minima, which are frequently used in training machine learning models.

5.2 Derivatives: Scalar Function.

- **Most popular:**
  - **Derivative of a Scalar function i.e. Scalar derivatives $\mathbf{f}: \mathbb{R} \rightarrow \mathbb{R}$**
    - A scalar function is a function that maps a **real number $\mathbf{x}$** to another **real number $\mathbf{f(x)}$.**
      - $\mathbf{f(x) = x^2}$
      - Here x: a real number and $f(x)$: also a real number.
    - We are interested in the **rate** at which $\mathbf{f(x)}$ **changes** as $\mathbf{x}$ **changes**.
    - The derivative is the heart of calculus, buried inside this definition:
      - $\mathbf{f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}}$ **when the limit exists.**
      - popularly known as the "**limit definition of the derivative**" or "**derivative by using the first principle**"
    - But what does it mean?

5.2.1 Derivative First Principle: Interpretation.

- **Derivative of a function is a measure of local slope.**
  - **1st Example: For Linear Function $y = f(x) = 2x$.**

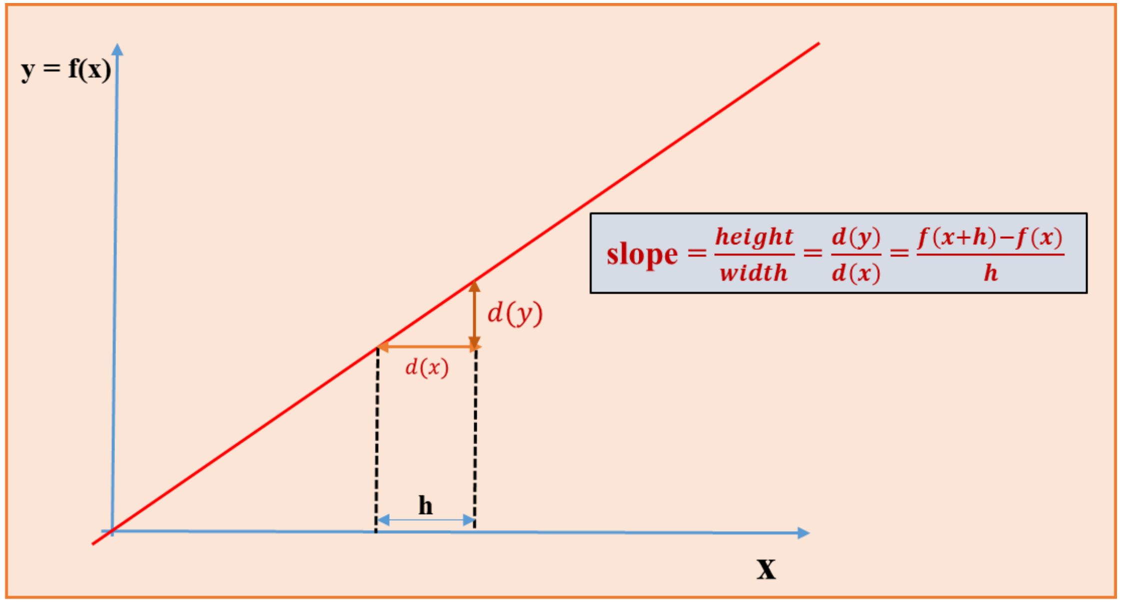
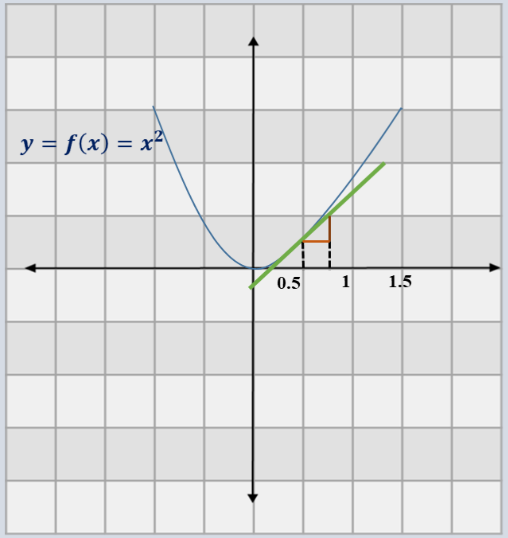


Fig: Derivative → Interpretation.

What for non linear function?

5.2.2 Derivative First Principle: Interpretation.

- **2nd Example: For Non - Linear Function $y = f(x) = x^2$.**



I want to find the slope at point 0.5. What does this mean?
You draw a tangent at x = 0.5 and find the slope of that tangent line.

Slope $= \frac{height}{width} = \frac{d(y)}{d(x)}\Big|_{x=0.5} = 1$

- The derivative of a function at a **point is the slope of the tangent drawn to that curve** at that point.
  - (slope) derivative of a linear function (straight line) is constant at all the point not for the non-linear function.
- It also represents the **instantaneous rate of change** at a point on the function.

5.3 Some Common Rules for determining Derivative.

Rule	Function	Derivative
Sum – Difference Rule	$f(x) \pm g(x)$	$f'(x) \pm g'(x)$
Multiplication by Constant	$c.f(x)$	$c.f'(x)$
Product Rule	$f(x).g(x)$	$f'(x).g(x) + f(x).g'(x)$
Quotient Rule	$f(x)/g(x)$	$\frac{f'(x).g(x) - f(x).g'(x)}{(g(x))^2}$
Chain Rule	$f(g(x))$	$f'(g(x)).g'(x)$

!!! Hands on practice in Tutorial.

5.4 Derivative of some common function.

Function - Type	Function - Notation	Derivative
Constant function	$f(x) = c$; where c is real constant.	$f'(x) = (c)' = 0$.
Identity function	$f(x) = x$	$f'(x) = (x)' = 1$.
Linear function	$f(x) = mx$	$f'(x) = (mx)' = m$.
Function of the form	$f(x) = x^n$	$f'(x) = (x^n)' = nx^{n-1}$.
Exponential function of the form	$f(x) = a^x$; where $a > 0$	$f'(x) = (a^x)' = a^x \ln(a)$.
Exponential function	$f(x) = e^x$	$f'(x) = (e^x)' = e^x$.
Logarithmic function	$f(x) = \ln(x)$	$f'(x) = (\ln(x))' = \frac{1}{x}$.
Sinusoidal function	$f(x) = \sin(x)$	$f'(x) = (\sin(x))' = \cos(x)$.
Cosine function	$f(x) = \cos(x)$	$f'(x) = (\cos(x))' = -\sin(x)$.
Tangent function	$f(x) = \tan(x)$	$f'(x) = (\tan(x))' = \sec^2(x)$.

6. Matrix and Derivative.

{Finding the Slope for Multi – Variate Function.}

6.1 Derivative of a Multivariate Function.

- (**scalar derivative of**)Multivariate function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ are in the form $f(x, y) = x^2 y$.
- **Partial Derivative:**
  - In mathematics, a **partial derivative** of a function of several variables is its derivative with respect to one of those variables, with the others held constant (as opposed to the total derivative, in which all variables are allowed to vary).
  - This swirly-d symbol,∂, often called "del", is used to distinguish partial derivatives from ordinary single-variable (regular) derivatives.
    - For Example: $f(x, y) = x^2 y$.

$$\frac{\partial}{\partial x} = \underbrace{\frac{\partial}{\partial x} x^2 y} = y \frac{\partial}{\partial x} x^2 = 2xy$$

Treat y as a constant, then take regular derivative.

$$\frac{\partial}{\partial y} = \underbrace{\frac{\partial}{\partial y} x^2 y} = x^2 \frac{\partial}{\partial y} y = x^2 . 1$$

Treat x as a constant, then take regular derivative.

Derivative of $f(x, y) = x^2 y$ are $2xy; x^2$

- Partial derivatives are used in vector calculus and differential geometry.

6.2 {some popular} Nomenclature of Derivative.

- Derivative of a vector/matrix a.k.a Matrix/Vector Calculus is an extension of ordinary scalar derivative to higher dimensional settings.
- Overview of some extended derivative style:

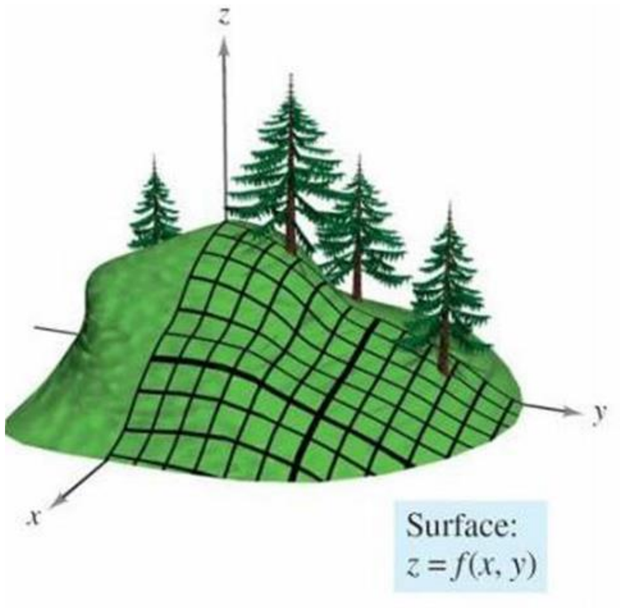
Setting	Derivative	Notation
$\boldsymbol{f}: \mathbb{R} \to \mathbb{R}$	Scalar Derivative	$\boldsymbol{f}'(\boldsymbol{x})$
$\boldsymbol{f}: \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$	Gradient	$\nabla \boldsymbol{f}(\boldsymbol{x})$
$\boldsymbol{f}: \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{m}} \to \mathbb{R}$	Gradient	$\nabla \boldsymbol{f}(\boldsymbol{x})$
$\boldsymbol{f}: \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$	Jacobian	$\boldsymbol{J}_{\boldsymbol{f}}$

6.3 Gradient.

- Gradient:
  - The gradient of a function of multiple variables is the vector of partial derivatives of the function with respect to each variable.
  - Scalar-by-vector $\{f: \mathbb{R} \rightarrow \mathbb{R}^n\}$:
    - The derivative of a scalar function y with respect to a vector $x = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$ is written as:
      - gradients of y: $\nabla y = \frac{\partial y}{\partial x} = \left[\frac{\partial y}{\partial x_1} \; \frac{\partial y}{\partial x_2} \; \cdots \; \frac{\partial y}{\partial x_n}\right]^T$. $\rightarrow$ gradients.
        - **{Stack the partial derivative against all the element of vector *x*}**
  - Scalar-by-Matrix $\{f: \mathbb{R} \rightarrow \mathbb{R}^{n \times m}\}$:
    - The derivative of a scalar function y with respect to a $n \times m$ matrix X is written as:
      - gradients of y: $\nabla y = \frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{n1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1m}} & \cdots & \frac{\partial y}{\partial x_{nm}} \end{bmatrix}$
      - **{Stack the partial derivative against all the element of Matrix X.}**

gradient is also the direction of steepest ascent, What does that mean?

6.4 Gradient: Geometric Interpretation.



He want's to scale the hill:

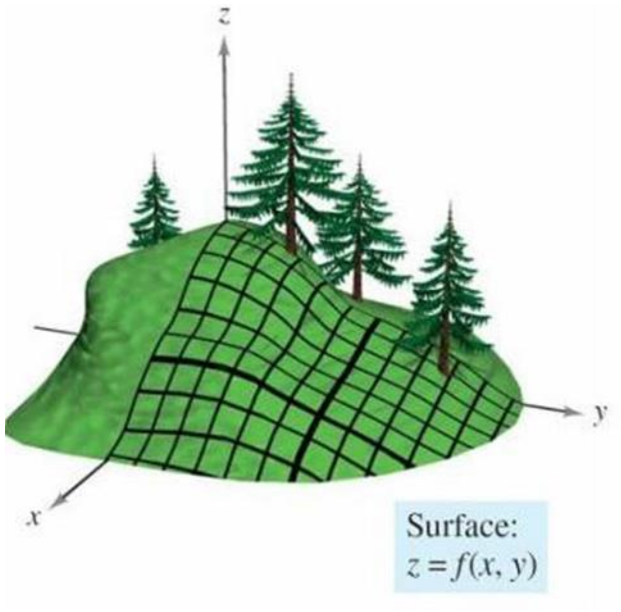
Let's assume he can take two routes:

One through x→ co-ordinate(direction)

One through y → co-ordinate(direction)

Which route would be fastest?

6.4.1 Gradient: Geometric Interpretation.



He want's to scale the hill:
Let's assume he can take two routes:
One through x→ co-ordinate(direction)
One through y → co-ordinate(direction)
Which route would be fastest?

Whichever direction has highest slope(gradient) i.e.
Find the gradient of the surface:

$$z = f(x, y)$$

gradient is a partial derivative of z against x and y stack in the vector.

This is read as: "grad. of z" or "grad z" ← $\nabla z = [\frac{\partial f(x,y)}{\partial x} \quad \frac{\partial f(x,y)}{\partial y}]$

6.5 Gradient: Example 1.

- **$z = f(x, y) = 3x^2y$ find the gradient of z at $[1, 1]$.**
- We know gradient of z is :
  - $\nabla z = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} & \frac{\partial f(x,y)}{\partial y} \end{bmatrix}$
- Finding: $\frac{\partial f(x,y)}{\partial x}$ i.e. y is constant.

$$\boxed{\frac{\partial f(x,y)}{\partial x} = \frac{\partial 3yx^2}{\partial x} = \frac{3y\partial x^2}{\partial x} = 3y2x = 6yx}$$

- Finding $\frac{\partial f(x,y)}{\partial y}$ i.e. x is constant.

$$\boxed{\frac{\partial f(x,y)}{\partial y} = \frac{\partial 3yx^2}{\partial y} = \frac{3x^2\partial y}{\partial y} = 3x^2 \times 1 = 3x^2}$$

- ∇z is:

$$\boxed{\nabla z = [6yx \quad 3x^2]}$$

- ∇z at $[1 \quad 1]$:

$$\boxed{\nabla z = [6 \times 1 \times 1 \quad 3 \times 1^2] = [6 \quad 3]}$$

6.6 Gradient of a Vector – Valued Function : Jacobian.

- **vector-by-vector $\{\boldsymbol{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m\}$:**
  - The derivative of a vector function : $\boldsymbol{y} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n}]^T \in \mathbb{R}^n$ with respect to an input vector $\boldsymbol{x} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_m}]^T \in \mathbb{R}^m$ is written as:

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \cdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix} = J_y$$

  - **J_y: called Jacobian matrix** is a matrix **which contains all the partial derivatives** of each output component with respect to each input variable, providing a full picture how the vector-valued function changes as each input variable changes.

6.7 Derivative: Key Point

- The **derivative** of a **univariate function** is a **scalar**,
  - When the **derivative** of a **multivariate function** is organized and stored in a **vector**, the so-called **gradient**.
    - we denote the **derivative of a multivariable function** **f** using the **gradient symbol** ***Δ*** **{read "del" or "nabla"}**
      - $$\Delta f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \\ \vdots \end{bmatrix}$$
- The **gradient is simply a vector listing** the **derivatives of a function** with respect to **each argument of a function**.

Plan

- Tutorial – Some Hands-on Exercise on Vector, Matrices and Gradient.
- Workshop – Implement PCA with eigenvalue decomposition for Image Compression Application.

The – End.