5CS037-Concepts and Technologies of AI
Lecture-01

Linear Algebra for Machine Learning: A Review

Vector and Matrices.

Siman Giri

About the Module!!

What is the module about?

How to pass the module?

What is module about?

- **Learning Objectives**:
  - Understand and Review the basic **Mathematical concepts** required to further understand ML/AI.
  - Build a perspective on **Artificial Intelligence** with it's **social and ethical impacts**.
  - Understand the need for Learning System.
  - Understand the various elements of learning system and it's uses in designing and building **Machine Learning Algorithms**.
  - How to make **better Machine Learning Algorithms**?

How to pass this Module?

- Module Logistics:

S.No	Assessment	Points	Start	Due	Viva
1	Data Wrangling, Data Transformation and Exploratory Data Analysis.	[10]	Week-2	Week-3/4	[Y]
2	Mid-Term Examination	[20]	Week-8	-	[N]
3	Term-Paper[Essay-Ethics of AI]	[20]	Week-7	Week-9	[N]
4	Weekly Workshop Assessment + Class Performance	[10]	Weekly		[Y]
5	Final Portfolio [Regression and Classification]	[40]	Week-10	Week-12	[Y]

- *Plagiarism and Use of AI is completely prohibited, if found will be reported for* **Academic Misconduct**.

How to pass this Module?

- ***Working on Groups***:
  - *We request all of you to form a learning group with 3-5 members with in a section.*
  - *(No cross section group allowed)*
  - *In group you can learn and brain storm together for assignments, weekly workshops, tutorials problem set.*
  - *Individually submit all the assignment requirements code and reports.*
- ***Viva***:
  - *In group you will make presentations for all your assignments and present your work, based on your presentations, respected instructors will conduct a individual viva.*
- ***Our Suggestion***:
  - *Please read your assignment requirements carefully and in detailed.*
  - *Ask lots of question!!!*
  - *Complete all your tasks without the help of chatgpt.*

A. Linear Algebra For Machine Learning.

Preliminary Concepts!

A.1 Sets.

- A **set** is well defined **collection of objects**:
  - Examples: A pack of wolves, A deck of cards, A flock of Pigeons.
- Sets have **elements** or **members**!!
  - Elements/ members: Objects that belongs to set.
  - **Caution!!!** Set it self can be member or elements of other sets.
- Sets are denoted by **capital letters** such as **A or X**:
  - $\boldsymbol{x \in A}$**(read: x is an element of A or x belongs to A).**
  - Set is usually specified by stating the property that determines whether or not an object x belongs to the set or by listing it's elements inside a pair of braces such as:
    - $\boldsymbol{X = \{x_1, \ldots, x_n\}}$ **or;** $\boldsymbol{X = \{x : x \text{ } is \text{ } an \text{ } even \text{ } integer \text{ } and \text{ } x > 0\}}$
- Some of the **important sets** are:

$$\mathbb{N} = \{n : n \text{ } is \text{ } a \text{ } natural \text{ } numbers\} = \{1,2,3 \ldots\}$$
$$\mathbb{Z} = \{n : n \text{ } is \text{ } an \text{ } integer\} = \{-1,0,1, \ldots\}$$
$$\mathbb{R} = \{x : x \text{ } is \text{ } a \text{ } real \text{ } number\}$$
$$\mathbb{C} = \{z : z \text{ } is \text{ } a \text{ } complex \text{ } number\}.$$

A.1 Sets: Relations.

- **Some of the relations**:
  - Set A is subset of B: every element of A is also an element of B.
    - **Notations:** $\boldsymbol{A \subset B}$. **Example:** $\boldsymbol{\{4,5,8\} \subset \{2,3,4,5,6,7,8,9\}}$
  - Set A and Set B are equal: every element of B is in A.
    - **Notations:** $\boldsymbol{A = B}$.
  - Empty sets: Set with no elements are called empty sets and is denoted by ϕ.
- **Some set Operations**:
  - Union(∪) and Intersection(∩) of sets: Union and Intersection of two sets A and B can be defined as:
    - $A \cup B = \{x: x \in A \text{ } or \text{ } x \in B\}$ and $A \cap B = \{x: x \in A \text{ } and \text{ } x \in B\}$.
    - Also written as:
      - $\bigcup_{i=1}^{n} A_i = A_1 \cup \cdots \cup A_n$ and $\bigcap_{i=1}^{n} A_i = A_1 \cap \cdots \cap A_n$. For the union and intersection respectively of the sets $A_1, \ldots, A_n$.
  - **Disjoint** sets: Set A and B are disjoint if they do not have any elements in common.
    - $A \cap B = \emptyset$.
  - **Complement** of A: For all $A \subset U$: U $universal\ set$. Complement of A is :
    - $A' = \{x: x \in U \text{ } and \text{ } x \notin A\}$.
  - **Difference** of two sets A and B is:
    - $A \backslash B = A \cap B' = \{x: x \in A \text{ } and \text{ } x \notin B\}$.

A.1 Sets: Exercise[E.1]-Example.

[E.1] Let $\mathbb{R}$ be the universal set and suppose that:

$A = \{x \in \mathbb{R}: 0 < c \leq 3\}$ and $B = \{x \in \mathbb{R}: 2 \leq x < 4\}$

Then:

$$A \cap B = \{x \in \mathbb{R}: 2 \leq x \leq 3\}$$
$$A \cup B = \{x \in \mathbb{R}: 0 < x < 4\}$$
$$A \backslash B = \{x \in \mathbb{R}: 0 < x < 2\}$$
$$A' = \{x \in \mathbb{R}: x \leq 0 \text{ or } x > 3\}.$$

Proposition: Let A,B and C be sets then;

Proof:

$$i)\ A \cup (B \cup C) = (A \cup B) \cup C.$$
$$ii)\ A \cap (B \cap C) = (A \cap B) \cap C.$$

Solutions[1[i]]:

$$\begin{aligned} A \cup (B \cup C) &= A \cup \{x: x \in B \text{ or } x \in C\} \\ &= \{x: x \in A \text{ or } x \in B, \text{or } x \in C\} \\ &= \{x: x \in A \text{ or } x \in B\} \cup C \\ &= (A \cup B) \cup C. \end{aligned}$$

A.1 Sets: Exercise[P.1]-Practice.

[P.1](De Morgan's Laws). Let A and B be sets. Then proof:

$$1.\ (A \cup B)' = A' \cap B';$$
$$2.\ (A \cap B)' = A' \cup B'.$$

A.2 Sets: Cartesian Products.

- Given sets A and B, we can define a new set $A \times B$ **(called Cartesian product of A and B)** are a set of **ordered pairs** i.e.
  - $\boldsymbol{A \times B = \{(a, b): a \in A \; and \; b \in B\}}$

Example:
If $A = \{x, y\}, B = \{1, 2, 3\}, and \; C = \emptyset$, then
$A \times B = \{(x,1),(x,2),(x,3),(y,1),(y,2),(y,3)\}$.
and,

$$A \times C = \emptyset$$

- Cartesian product of n sets can be defined as:
  - $\boldsymbol{A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n): a_i \in A_i \; for \; i = 1, \ldots, n\}}$.

Trivia!!!
If $[A = A_1 = A_2 = \cdots = A_n]$ then $A \times \cdots \times A$ can be written as A^n.

A.3 Sets-Cartesian Products: Mappings.

- Mappings a.k.a **relations** or **function** are the **subsets** of **Cartesian products** i.e. for Cartesian set $\boldsymbol{A} \times \boldsymbol{B}$ function **f** is
  - $\boldsymbol{f} \subset \boldsymbol{A} \times \boldsymbol{B}$.
  - This represents a special type of relation where $(\boldsymbol{a}, \boldsymbol{b}) \in \boldsymbol{f}$ if every element $\boldsymbol{a} \in \boldsymbol{A}$ there exists a **unique element** $\boldsymbol{b} \in \boldsymbol{B}$. (for every element in A, f assigns a unique element in B).
  - Notations: **for functions**: $\boldsymbol{f}: \boldsymbol{A} \rightarrow \boldsymbol{B}$ and for **ordered pairs**
    - $(\boldsymbol{a}, \boldsymbol{b}) \in \boldsymbol{A} \times \boldsymbol{B}; \boldsymbol{f}(\boldsymbol{a}) = \boldsymbol{b}$ **or** $\boldsymbol{f}: \boldsymbol{a} \rightarrow \boldsymbol{b}$.

A.3 Cartesian Products-Mappings(Domain and Range).

- The set **A** is called the **domain** of f and $\boldsymbol{f(A) = \{f(a): a \in A\} \subset B}$ is called the **range(co-domain)** of f **and its elements are called image under** f.
  - The elements of the function's domain as **input values** and the elements in the function's range as **output values**.

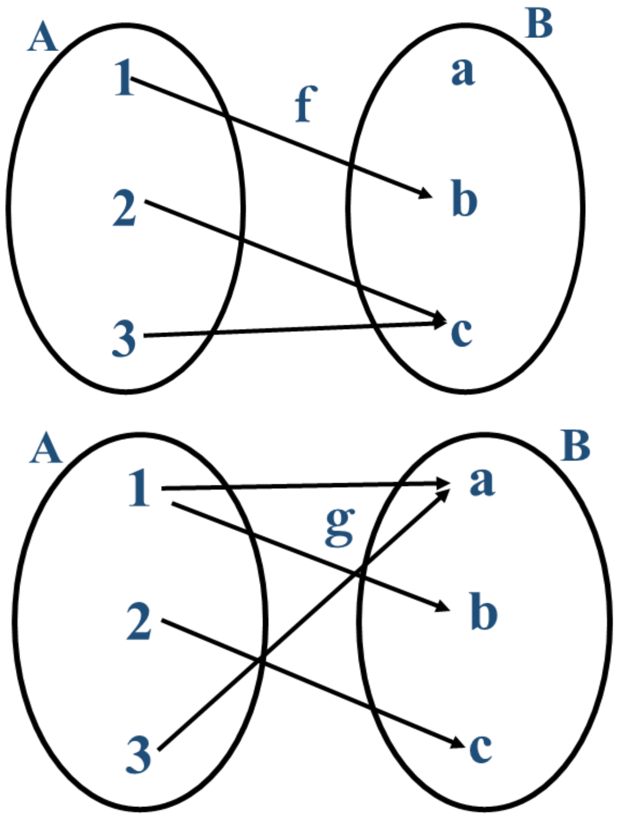


Figure:1 Mappings and relations.

A.4 Relations: (Domain and Range).

- Let $f: A \rightarrow B$ be a function from the domain A to the codomain B.
- **well-defined**: A **relation/function** is well-defined if **each element** in the **domain** is assigned to a **unique element** in the **range**.

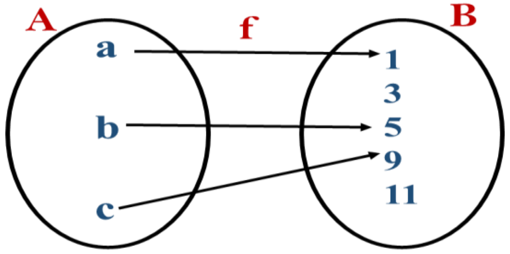


Fig: A well defined function.

- **Not a functions**:

  i) domain has **no image** associated with it.

  ii) one of the elements in the domain has **two images** assigned to it.

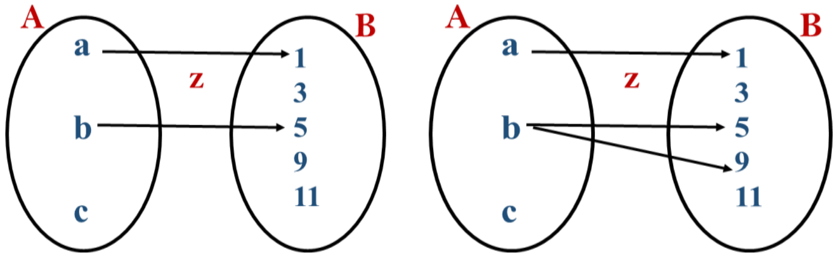


Fig: Not a Functions

A.4 Relations: (Domain and Range).

- Let $f: A \to B$ be a function from the domain A to the codomain B.
- **injective (one to one)**: The function f is called injective (or one-to-one) if it maps distinct **elements** of A to **distinct elements** of B. In other words, for every element y in the co-domain B there exists at most one pre-image in the domain A:
  - $\forall x_1, x_2 \in A: x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$.

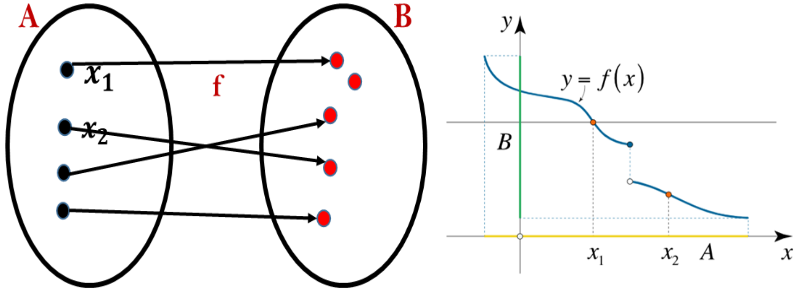


Fig: Injective Function.

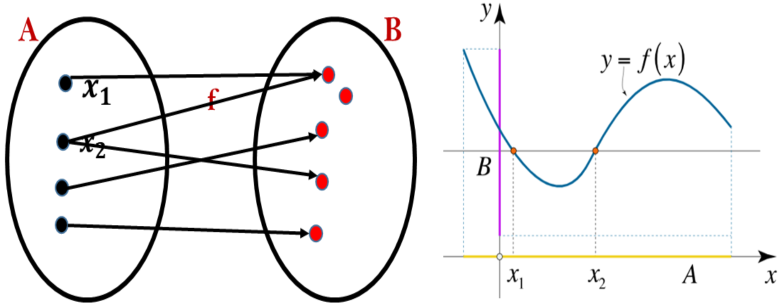


Fig: Non-Injective Function.

A.4 Relations: (Domain and Range).

- Let $f: A \to B$ be a function from the domain A to the codomain B.
- **surjective (onto):** A function f from **A** to **B** is called surjective (**or onto**) if for every y in the co-domain/range **B** there exists at least one x in the domain **A**:
  - $\forall_y \in B$: $\exists_x \in A$ **such that** $y = f(x)$.

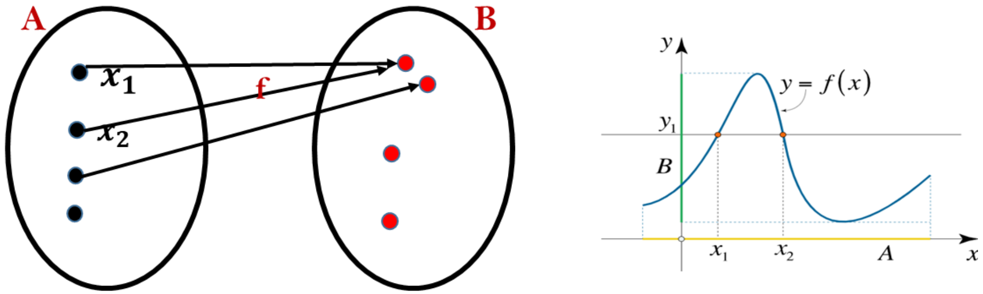


Fig: Surjective Function.

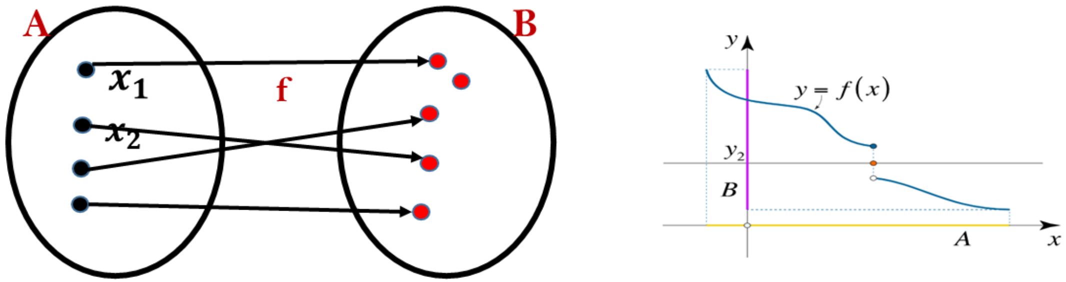


Fig: Non-Surjective Function.

A.4 Relations: (Domain and Range).

- Let $\boldsymbol{f}$: $\boldsymbol{A} \longrightarrow \boldsymbol{B}$ be a function from the domain $\boldsymbol{A}$ to the codomain $\boldsymbol{B}$.
- **bijective (one-to-one and onto):** A function $\boldsymbol{f}$ from **A** to **B** is called bijective(**or one-to-one and onto**) if for every $\boldsymbol{y}$ in the codomain/range **B** there exists exactly one element $\boldsymbol{x}$ in the domain **A**:
  - $\forall_{\boldsymbol{y}} \in \boldsymbol{B}$: $\ni !_{\boldsymbol{x}} \in \boldsymbol{A}$ **such that** $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$.

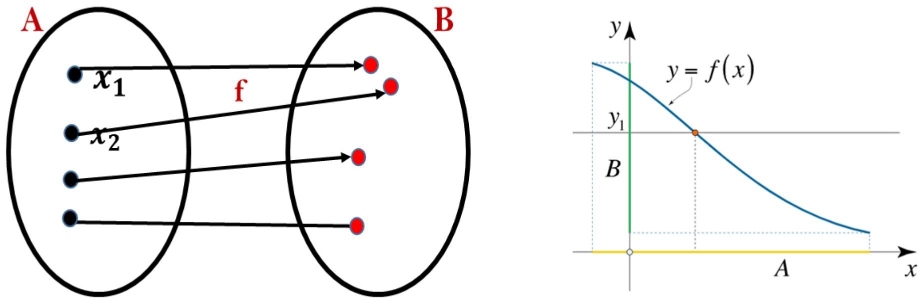


Fig: Bijective Function.

A.5 Summary: Functions and Machine Learning.

- The ultimate goal of **machine learning** is **learning** a **functions** from **data**, i.e. mappings from domain **(feature vector space(set))** onto the range **(target variable)** of a function.
- The **objective** of **5CS037** is to be able to **understand** all the highlighted **terms** in above **statement**.

B. Introduction to Linear Algebra.

Why to study Linear Algebra for Machine Learning?

B.1 What is Linear Algebra?

- Linear Algebra is the branch of **mathematics** concerning **linear equations** such as: $a_1x_1 + \ldots\ldots + a_nx_n = b$; **linear maps** such as:$(x_1, \ldots.., x_n) \mapsto a_1x_1 + \cdots + a_nx_n$; and **there representations** in **vector spaces and through matrices**.

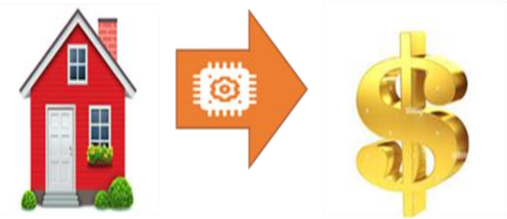
--Wikipedia.

"When in doubts, assume outputs are linear function of inputs."

B.2 Linear Algebra for Machine Learning: Why and Where?

- Data Representations:

Task: House Price Prediction.



Data: Features/Descriptor of House

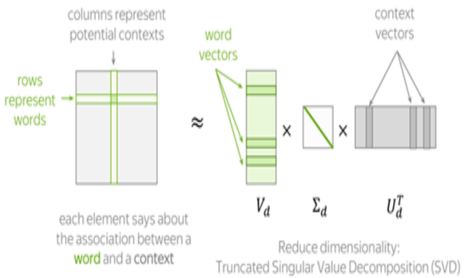
Area	Rooms	Price
1080	8	1,00,000.00
1200	10	1,50,000.00

How would you represent this, for computer?

Matrix.

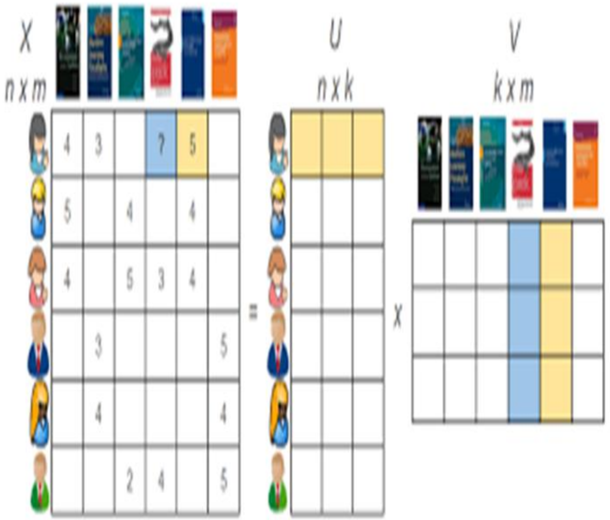
$$\begin{bmatrix} 1080 & 8 \\ 1200 & 10 \end{bmatrix} \begin{bmatrix} 1,00,000 \\ 1,50,000 \end{bmatrix}$$

- Dimension Reductions:



Matrix Decompositions

- Building Machine Learning Models:



Recommendation System

& many more!!!

B.3 Linear Algebra: Final Thoughts!

"linear algebra is the **mathematics of data**. **Matrices** and **vectors** are the **language of data**.."

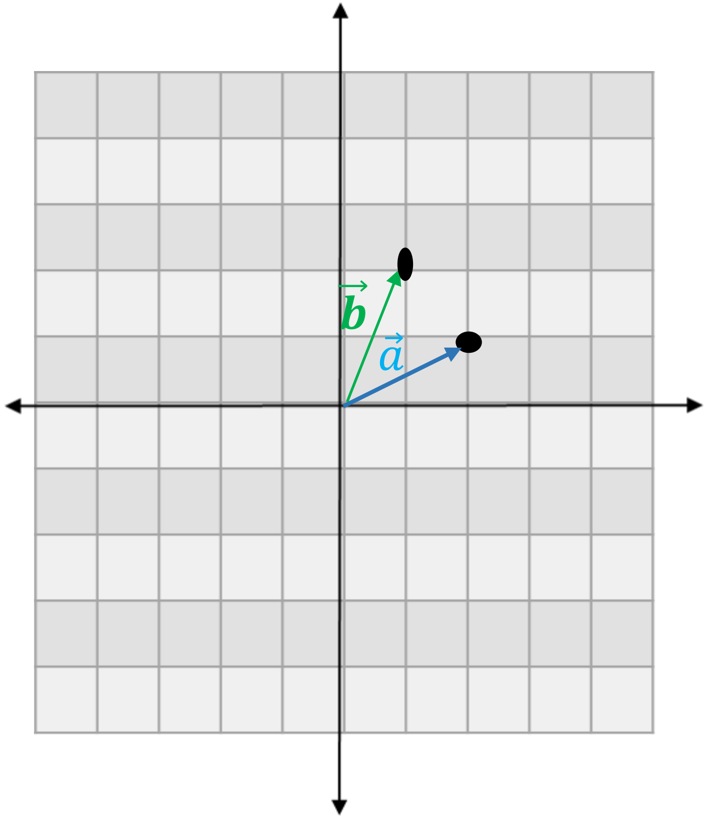
—Jason Brownlee

1. Vectors.

What is a Vector?

1.1 Vector: Introduction!!

- What is a vector?
  - For Computer Science:
    - One-dimensional array of numbers that represents a particular point in **space**.
    - Example: $\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \vec{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
    - Is $\vec{a}$ == $\vec{b}$?
    - One-dimensional ordered array of numbers that represents a particular point in **space**.
- Dimension Space: length of the array i.e.(How many value we require to represent a particular vector?)
    - $\vec{a}$ ***is in*** $\mathbb{R}^2$ ***dimensional space.***
    - $\vec{c} = [1, 2, 3]$ is in which dimensional space?
      - $\mathbb{R}^3$
- In upcoming slides:
  - we will review the basic properties of vectors in two and three dimensions with the goal of extending these properties to vectors in $\mathbb{R}^n$.



1.2 Vector - Algebraic operations: Addition.

- *Parallelogram Rule for Vector Addition:*
  - *If v and w are vectors in 2-space or 3-space that are positioned so their initial points coincide, then the two vectors form adjacent sides of a parallelogram, and the sum v + w is the vector represented by the arrow from the common initial point of v and w to the opposite vertex of the parallelogram : Figure (a).*
- *Triangle Rule for Vector Addition:*
  - *If v and w are vectors in 2-space or 3-space that are positioned so the initial point of w is at the terminal point of v, then the sum v + w is represented by the arrow from the initial point of v to the terminal point of w :Figure (b).*

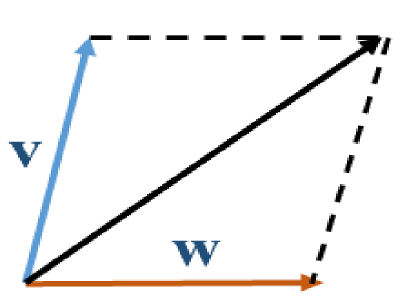


Fig: a

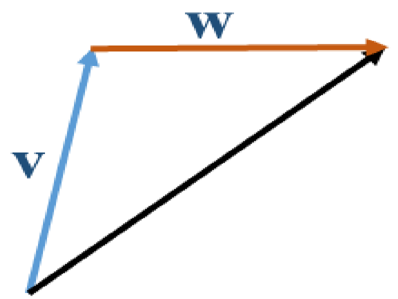


Fig: b

1.2 Vector - Algebraic operations: Addition.

- Implementation:
  - Add **element-by-element** i.e.
    - $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$; $b = \begin{bmatrix} b_1 \\ \vdots \\ b_2 \end{bmatrix}$ then
    - $a + b = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}$.
  - Example:
    - $\boldsymbol{a} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$; $\boldsymbol{b} = \begin{bmatrix} -4 \\ -1 \end{bmatrix}$
    - $\boldsymbol{a} + \boldsymbol{b} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

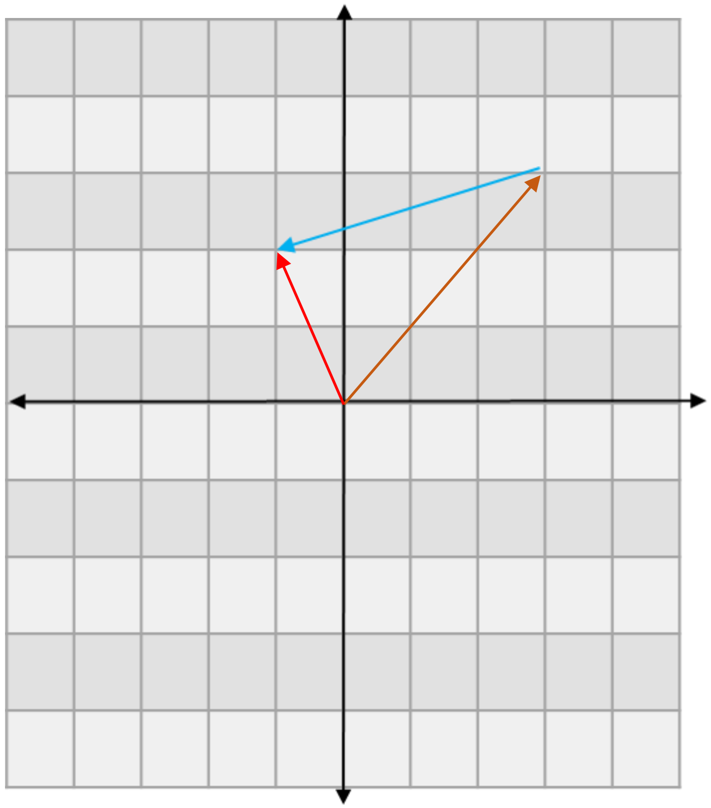
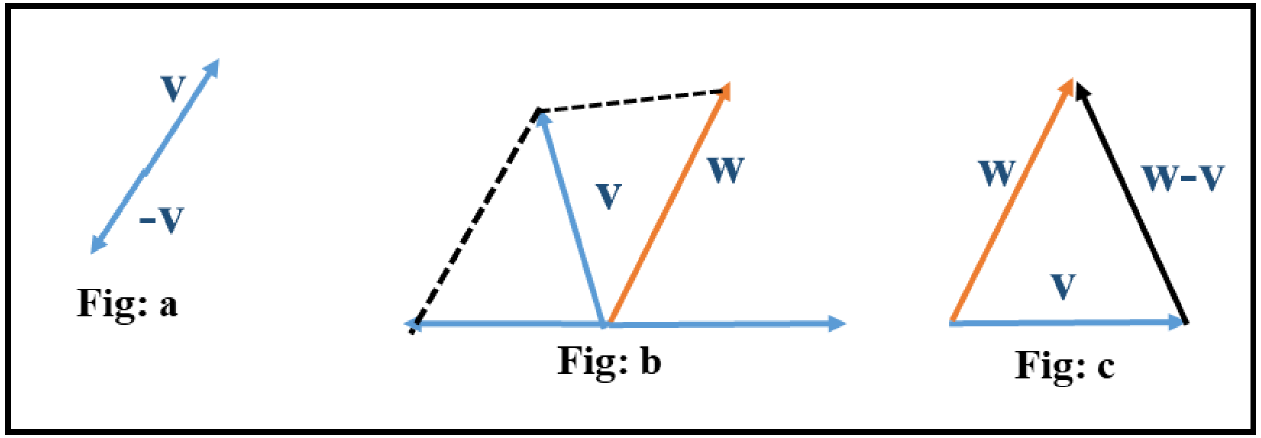


Fig: Geometric Interpretation of Vector addition.

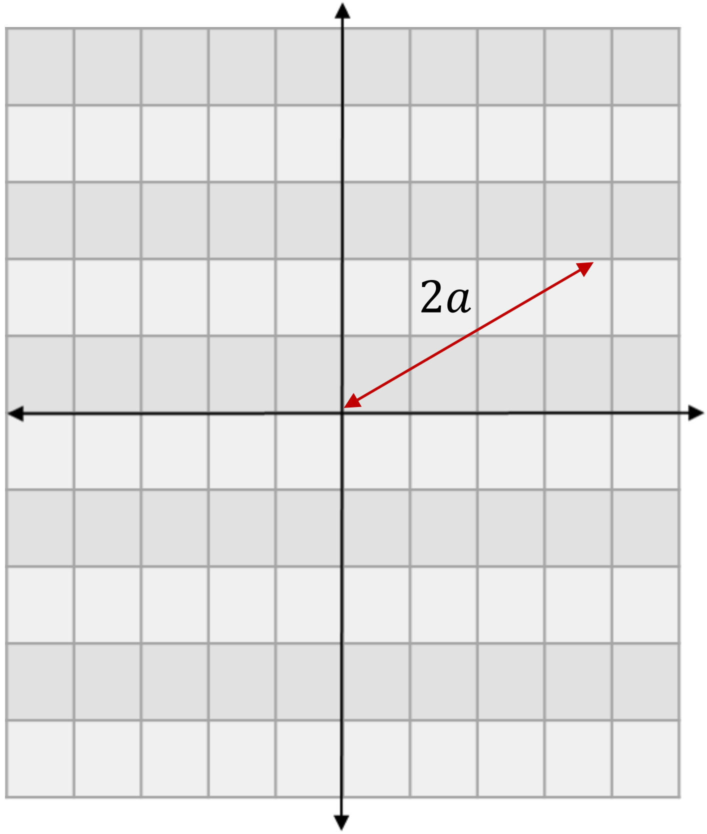
1.2 Vector - Algebraic operations: Subtraction.

- *Vector Subtraction*:
  - *The negative of a vector v, denoted by –v, is the vector that has the same length as v but is oppositely directed (Fig: a), and the difference of v from w, denoted by w – v, is taken to be the sum w – v = w + (–v).*



1.2 Vector - Algebraic operations: Multiplication.

- **Scalar**:
  - It is a number real or complex.
  - Vectors of interest are real then the set of scalars are also real.
  - Why the name scalars?
    - It scales the vector by given numbers.
- **Scalar Multiplication**:
  - If **v** is a nonzero vector in **2-space or 3-space**, and if **k** is a **nonzero scalar**, then we define the **scalar product(multiplication)** of **v by k** to be the **vector** whose **length is |k|** times the **length of v** and
    - **whose direction** is the **same** as that of **v** if **k** is **positive**
    - and **opposite** to that of **v** if **k** is **negative**.
    - If $k = 0$ or $v = 0$, then we define kv to be 0.



1.2 Vector - Algebraic operations: Multiplication.

- Scalar Multiplication - Example:
  - Vector –scalar multiplication is an element-wise operation. It's defined as:
  - For vector $\boldsymbol{u} = [\boldsymbol{u_1}, \boldsymbol{u_2} \dots \boldsymbol{u_n}]$ and scalar $\boldsymbol{\alpha}$ Scalar multiplication is :

$$\boldsymbol{\alpha u} = \begin{bmatrix} \boldsymbol{\alpha u_1} \\ \vdots \\ \boldsymbol{\alpha u_n} \end{bmatrix}$$

- Scalar Multiplication-Properties:
  - Vector-scalar multiplication satisfies a series of important properties:
    - Associativity: $(\boldsymbol{\alpha\beta})\boldsymbol{u} = \boldsymbol{\alpha}(\boldsymbol{\beta u})$
    - Left-distributive property: $(\boldsymbol{\alpha} + \boldsymbol{\beta})\boldsymbol{u} = \boldsymbol{\alpha u} + \boldsymbol{\beta u}$
    - Right-distributive property: $\boldsymbol{u}(\boldsymbol{\alpha} + \boldsymbol{\beta}) = \boldsymbol{u\alpha} + \boldsymbol{u\beta}$
    - Right-distributive property for vector addition: $\boldsymbol{\alpha}(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{\alpha u} + \boldsymbol{\alpha v}$

Basic Vector Notations!!!

- In computer science: we represent vector as: **a**.
- two vectors in same dimensional space are equal if $\boldsymbol{a_i} = \boldsymbol{b_i}$ for all $\boldsymbol{i}$.
- **row vector**: $\boldsymbol{a} = [\boldsymbol{a_i}, \dots, \boldsymbol{a_n}] \in \mathbb{R}^{\boldsymbol{n}}$
- **column vector**: $\mathbf{a} = \begin{bmatrix} \boldsymbol{a_i} \\ \dots \\ \boldsymbol{a_n} \end{bmatrix}$
- **transpose of Vector**: $\boldsymbol{Coloumn\ Vector} \leftrightarrow \boldsymbol{Row\ Vector}.$
- $$\vec{\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{v_1} \\ \vdots \\ \boldsymbol{v_n} \end{bmatrix} = [\boldsymbol{v_1} \quad .. \quad \boldsymbol{v_n}]^{\boldsymbol{T}}$$

1.3 Vector: Definition.

- If n is a positive integer, then an ordered n-tuple is a sequence of n real numbers [v1, v2,...,vn]. The set of all ordered n-tuples is called n-space and is denoted by $\mathbb{R}^n$.
  - **Vectors in $\mathbb{R}^n$:**
    - Let $\mathbb{R}^n = \{(x_1, \dots, x_n): x_j \in \mathbb{R} \text{ for } j = 1, \dots, n\}$. Then,
      - $\vec{x} = [x_1, \dots, x_n]$ is called a **vector**.
    - The number x_j are called the components of $\vec{x}$.
- Operations on vectors in $\mathbb{R}^n$, will all be natural extensions of the familiar operations on vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.
  - If v = [v1, v2,...,vn] and w = [w1, w2,...,wn] are vectors in $\mathbb{R}^n$, and if **k** is any scalar, then we define
    - v + w = [v1 + w1, v2 + w2,...,vn + wn]
    - kv = (kv1, kv2,... , kvn)
    - –v = (–v1, –v2,..., –vn)
    - w – v = w + (–v) = (w1 – v1, w2 – v2,...,wn – vn)
- How could you tell whether an array is **vector** or **not**?
- Why did we define vector in $\mathbb{R}^n$? Can't we define in terms of $\mathbb{C}^n$?

1.3 Vector Space: Definition.

- 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.
- In other words:
  - For addition of two vectors:
    - takes two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$, and it produces the third vector $\boldsymbol{u} + \boldsymbol{v} \in \mathbb{R}^2$.
    - (addition of vectors – gives another vector in the set)
  - For scalar Multiplication:
    - Takes a scalar $\boldsymbol{c} \in \boldsymbol{F}$ and a vector $\boldsymbol{v} \in \mathbb{R}^n$ produces a new vector $\boldsymbol{cv} \in \mathbb{R}^n$.
    - (multiplying a vector by a scalar – gives another vector in the set)

1.4 Vector Space: Axioms 1.

- If **u, v**, and **w** are **vectors** in $\mathbb{R}^n$, and if **k** and **m** are **scalars**, then:
  - **Commutative properties**:
    - $u + v = v + u$
    - $ku = uk$
  - **Associative Properties**:
    - $(u + v) + w = u + (v + w)$
    - $k(mu) = (km)u$
  - **Distributive Properties**:
    - $k(u + v) = ku + kv$
    - $(k + m)u = ku + mu$
  - **Scalar unity and Scalar zero**:
    - $1u = u$
    - $u + 0 = 0 + u = u$
    - $u + (-u) = 0$
    - $0u = 0$ (zero vector).

1.4 Vector Space - Axioms 1: Proof.

- Example (proof b):

$$
\begin{aligned}
(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w} &= \big((\boldsymbol{u}_1+\cdots+\boldsymbol{u}_n)+(\boldsymbol{v}_1+\cdots+\boldsymbol{v}_n)\big)+(\boldsymbol{w}_1+\cdots+\boldsymbol{w}_n) \\
&= (\boldsymbol{u}_1+\boldsymbol{v}_1+\cdots+\boldsymbol{u}_n+\boldsymbol{v}_n)+(\boldsymbol{w}_1+\cdots+\boldsymbol{w}_n) \\
&= \big((\boldsymbol{u}_1+\boldsymbol{v}_1)+\boldsymbol{w}_1,+\cdots+(\boldsymbol{u}_n+\boldsymbol{v}_n)+\boldsymbol{w}_n\big) \\
&= \big(\boldsymbol{u}_1+(\boldsymbol{v}_1+\boldsymbol{w}_1),+\cdots+,\boldsymbol{u}_n+(\boldsymbol{v}_n+\boldsymbol{w}_n)\big) \\
&= (\boldsymbol{u}_1+\cdots+\boldsymbol{u}_n)+(\boldsymbol{v}_1+\boldsymbol{w}_1+\cdots+\boldsymbol{v}_n+\boldsymbol{w}_n) \\
&= \boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})
\end{aligned}
$$

1.4 Vector Space: Axioms 2.

- If v is a vector in $\mathbb{R}^n$ and $\boldsymbol{k}$ is a scalar, then:
  - $0v = 0$
  - $k0 = 0$
  - $(-1)v = -v$

1.5 Linear Combinations of a vector.

- Addition, subtraction, and scalar multiplication are frequently used in combination to form **new vectors**.
- For example, if v_1, v_2, and v_3 are vectors in $\mathbb{R}^n$, then:
  - the vectors $u = 2v_1 + 3v_2 + v_3$ and $w = 7v_1 - 6v_2 + 8v_3$ are formed in this way.
- In general, we make the following definition:
- If w is a vector in $\mathbb{R}^n$, then w is said to be a linear combination of the vectors $[v_1, v_2, \ldots, v_r]$ in $\mathbb{R}^n$ if it can be expressed in the form
  - $w = k_1v_1 + k_2v_2 + \cdots + k_r v_r$ (a)
    - where $k_1, k_2, \ldots, k_r$ are **scalars**. These scalars are called the coefficients of the *linear combination*.
  - In the case where $r = 1$, Eqn. (a) becomes $w = k_1v_1$, so that a *linear combination* of a single vector is just a scalar multiple of that vector.

1.6 Vector: Linear Dependence and Independence.

- Linearly Dependent:
  - A set of vectors is linearly dependent if at least one vector can be obtained as linear combination of other vectors in the set. As you can see in the figure, we can combine vectors v and u to obtain vector w.
  - Mathematically:
    - Consider a set of vectors $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k]$ and scalars $\boldsymbol{\beta} \in \mathbb{R}$. If there is a way to get $\sum_{i=1}^{k} \boldsymbol{\beta}_i \boldsymbol{x}_i = \mathbf{0}$ with at least one $\boldsymbol{\beta} \neq \mathbf{0}$, we have linearly dependent vectors.
    - In other words, if we can get the zero vector as a linear combination of the vectors in the set with weights that are not all zero, we have a linearly dependent vector.

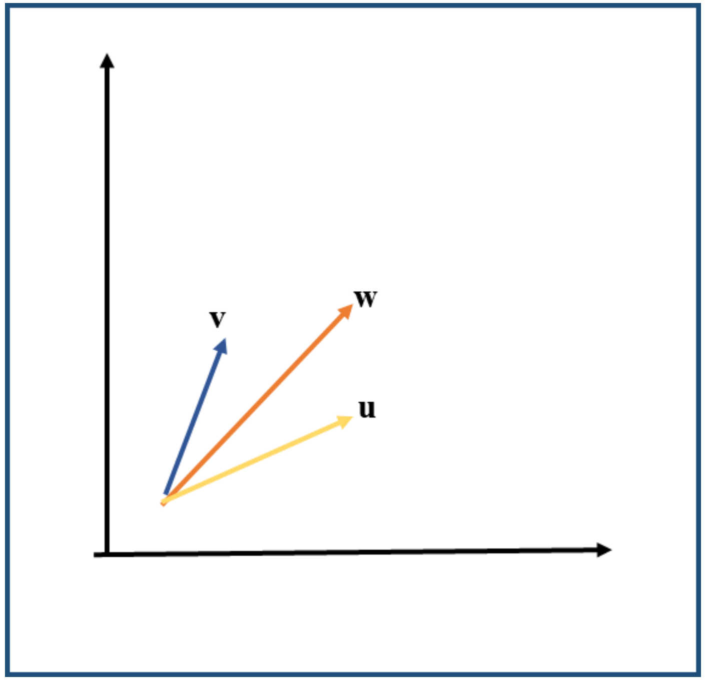


Fig: Linearly dependent Vectors (u+v=w)

1.6 Vector: Linear Dependence and Independence.

- **Linearly Independent:**
  - A **set of vectors** is **linearly independent** if none **vector** can be obtained as **linear combination** of **other vectors** in the set. As you can see in the figure, there is no way we can combine vectors v and u to obtain vector w.
  - **Mathematically:**
    - Consider a set of vectors $[\boldsymbol{x}_1, \dots, \boldsymbol{x}_k]$ and scalars $\boldsymbol{\beta} \in \mathbb{R}$. If only way to get $\sum_{i=1}^{k} \boldsymbol{\beta}_i \boldsymbol{x}_i = \boldsymbol{0}$ is $\boldsymbol{\beta}_1 \dots \boldsymbol{\beta}_k = \boldsymbol{0}$, we have linearly independent vectors.
    - In other words, the only way to get the ***zero vector is by multiplying each vector in the set by zero.***

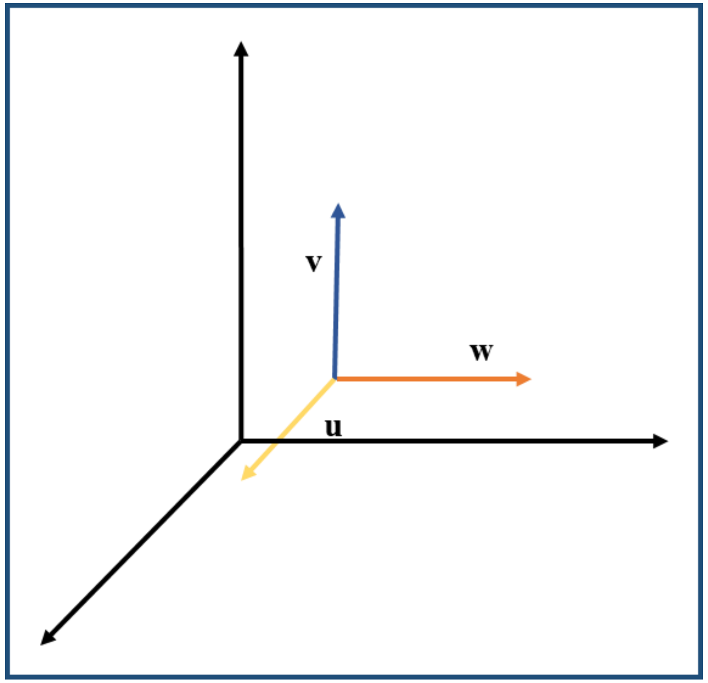


Fig: Linearly independent Vectors.

2. Norm, Dot Product and Distance in $\mathbb{R}^n$.

How do you estimate or compare two vector?

2. Norm, Product and Distance in $\mathbb{R}^n$.

How do you estimate or compare two vector?

By finding its magnitude i.e. representing vector by a scalar quantity.

2.1 Vector: Norms.

- If v = [v1, v2,...,vn] is a vector in $\mathbb{R}^n$, then the norm of v (also called the **length of v** or the **magnitude of v**) is denoted by **v** , and is defined by the formula:
  - $||v|| = \sqrt{v_1^2 + \cdots + v_n^2} == \sqrt{\sum_{i=1}^{n} v_i^2}.$
- **Formal Definition**:
  - A **norm** can be any function $f: \mathbb{R}^n \to \mathbb{R}$ that satisfies following properties:
    1. For all $v \in \mathbb{R}^n$, $f(v) \geq 0$ (**non-negativity**).
    2. $f(v) = 0$ if and only if $v = 0$ (**definiteness**).
    3. For all $v \in \mathbb{R}^n$, $t \in R$, $f(tv) = |t|f(v)$ (**homogeneity**).
    4. For all $v, w \in \mathbb{R}^n$, $f(v+w) \leq f(v) + f(w)$ (**triangle inequality**).

2.1 Vector – Norms : Examples.

- Examples: $\boldsymbol{l_p}$ family of norms – parameterized by a real number $\boldsymbol{p \geq 1}$, and defined as:
    - $\|\boldsymbol{v}\|_{\boldsymbol{p}} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$

For p = 1: [Euclidean Norm]

$$\|v\|_1 = \sum_{i=1}^{n} |v_i| \text{ -- } l_1 \text{ norm.}$$

For p=2: [Manhattan Norm]

$$\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2} \text{ -- } l_2 \text{ norm.}$$

For p = inf: [max Norm]

$$\|v\|_\infty = \max_i |x_i| \text{ --} l_\infty \text{ norm.}$$

- **Disclaimer**:
    - For practical purposes: length are used as equivalent to norm. Length is a concept from geometry i.e. geometric vectors have length and vectors in $\mathbb{R}^n$ have norm.
        - In practice and in machine learning these concepts are used interchangeably.
    - Many other norms exist, but they are beyond the scope of this review.

2.2 unit Vector!!!

- A vector of **norm 1** is called a **unit vector**.
  - Such vectors are useful for specifying a direction when length is not relevant to the problem at hand.
  - You can obtain a unit vector in a desired direction by choosing any nonzero vector v in that direction and multiplying v by the reciprocal of its length.
  - For example, if v is a vector of **length 2** in $\mathbb{R}^2$, then |½ v| is a unit vector in the **same direction** as v.
  - More generally, if v is any nonzero vector in $\mathbb{R}^n$, then:
    - $\boldsymbol{u} = \frac{\mathbf{1}}{\|\boldsymbol{v}\|}\boldsymbol{v}$
- **Normalizing Vector**:
  - The process of multiplying a nonzero vector by the reciprocal of its length to obtain a unit vector is called normalizing v.
- **Example: Normalizing Vector.**
  - Find the unit vector u that has the same direction as v = (2, 2, –1).
  - Solution:

2.2 unit Vector!!!

- **Example: Normalizing Vector.**
  - Find the unit vector u that has the same direction as v = (2, 2, −1).
  - Solution:

The vector v has length:

$$||v|| = \sqrt{2^2 + 2^2 + (-1)^2} = 3$$

Thus:

$$\hat{v} = \frac{1}{3} \times [2, 2, -1] = \left[\frac{2}{3}, \frac{2}{3}, -\frac{1}{3}\right]$$

To check:

$$||\hat{v}|| = \sqrt{\left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(-\frac{1}{3}\right)^2} = 1.$$

2.2 Vector – Vector Product.

- Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, the quantity $\boldsymbol{u}^T \boldsymbol{y}$, sometimes called the **inner product** or **dot product** of the vectors, is a real number given by:

$$\boldsymbol{u}^T \boldsymbol{v} \in \mathbb{R} = [\boldsymbol{u_1}\ \boldsymbol{u_2} \dots \boldsymbol{u_n}] \begin{bmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \\ \dots \\ \boldsymbol{v_n} \end{bmatrix} = \sum_{i=1}^{n} \boldsymbol{u_i v_i}$$

Disclaimer!!!

- For practical purposes, inner product are used as equivalent to dot product. **Inner products** are a more general concept than dot products (every dot product is an inner product, but not every inner product is a dot product).
- In $\mathbb{R}^n$ the inner product is a dot product defined as:

$$\left\langle \begin{bmatrix} \boldsymbol{u_1} \\ \dots \\ \boldsymbol{u_n} \end{bmatrix}, \begin{bmatrix} \boldsymbol{v_1} \\ \dots \\ \boldsymbol{v_n} \end{bmatrix} \right\rangle := \boldsymbol{u}.\boldsymbol{v} = \sum_{i=1}^{n} \boldsymbol{u_i v_i}$$

- **Outer Product**:
  - Given vectors $\boldsymbol{u} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^n$ (not necessarily of the same size), $\boldsymbol{u}\boldsymbol{v}^T \in \mathbb{R}^{m \times n}$ is called the outer product of the vectors. It is a matrix whose entries are given by $\left(\boldsymbol{u}\boldsymbol{v}^T\right)_{ij} = \boldsymbol{u_i v_j}$, i.e.

$$\boldsymbol{u}\boldsymbol{v}^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} \boldsymbol{u_1} \\ \dots \\ \boldsymbol{u_m} \end{bmatrix} [\boldsymbol{v_1}\ \boldsymbol{v_2} \dots \boldsymbol{v_n}] = \begin{bmatrix} \boldsymbol{u_1 v_1} & \boldsymbol{u_1 v_2} & \dots & \boldsymbol{u_1 v_n} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{u_m v_1} & \boldsymbol{u_m v_2} & \dots & \boldsymbol{u_m v_n} \end{bmatrix}$$

2.3 Dot Product – Properties.

- If u, v, and w are vectors in $\mathbb{R}^n$ and k a scalar then:
  - 0.v = v.0 = 0 [0 vector]
  - (u+v).w = u.w + v.w
  - u.(v-w) = u.v – u.w
  - (u-v).w = u.w – v.w
  - k(u.v) = u.(kv)
- Example: Calculate -
  - $(\boldsymbol{u} - 2\boldsymbol{v}).(3\boldsymbol{u} + 4\boldsymbol{v})$

2.3 Dot Product – Properties.

- Example: Calculate -
  - $(\boldsymbol{u} - 2\boldsymbol{v}).(3\boldsymbol{u} + 4\boldsymbol{v})$

  Solution:

$$\begin{aligned}(\boldsymbol{u} - 2\boldsymbol{v}).(3\boldsymbol{u} + 4\boldsymbol{v}) &= \boldsymbol{u}.(3\boldsymbol{u} + 4\boldsymbol{v}) - 2\boldsymbol{v}.(3\boldsymbol{u} + 4\boldsymbol{v}) \\ &= 3(\boldsymbol{u}.\boldsymbol{u}) + 4(\boldsymbol{u}.\boldsymbol{v}) - 6(\boldsymbol{v}.\boldsymbol{u}) \\ &= 3\|\boldsymbol{u}\|^2 - 2(\boldsymbol{u}.\boldsymbol{v}) - 8\|\boldsymbol{v}\|^2\end{aligned}$$

2.4 Vector – Distance.

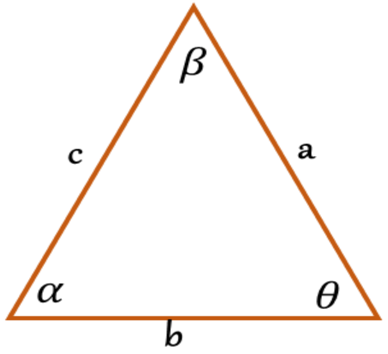
- If $\boldsymbol{u} = [\boldsymbol{u_1}\ \boldsymbol{u_2}\ \dots \boldsymbol{u_n}]$ and $\boldsymbol{v} = [\boldsymbol{v_1}\ \boldsymbol{v_2}\ \dots \boldsymbol{v_n}]$ are points in $\mathbb{R}^{\boldsymbol{n}}$, then we denote the distance between **u** and **v** by **d(u, v)** and define it to be:
- $\boldsymbol{d(u, v) = \ \|u - v\| = \sqrt{(u_1 - v_1)^2 + \cdots + (u_n - v_n)^2}}$

3. Vector Angles and Orthogonality.

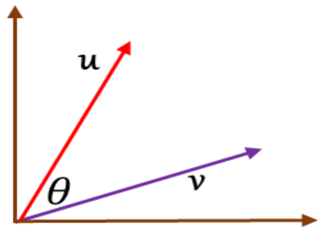
3.1 Vector: Angles.

- In machine learning, the **angle** between a pair of **vectors** is used as a **measure of vector similarity**.
- Inner products are used to define angles and orthogonality (guided by Cauchy–Schwarz inequality).
  - Cauchy-Schwarz inequality –
    - $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$
    - ***{In words: the absolute value of the inner product of a pair of vectors is less than or equal to the products of their length.}***
- The definition of the angle between vectors can be thought as a generalization of the **law of cosines** in trigonometry, which defines for a triangle with sides a, b, and c, and angle ϑ are related as:
  - $\boldsymbol{c^2 = a^2 + b^2 - 2abcos\theta}$.

- *Law of Cosines and Angle between vectors.*



$c^2 = a^2 + b^2 - 2abcos\theta$



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

We can replace this expression with vector lengths:

$$\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2(\|\boldsymbol{x}\| \|\boldsymbol{v}\|)cos\boldsymbol{\theta}$$

With a bit of algebraic manipulation, we can clear the previous equation to:

$$cos\boldsymbol{\theta} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle / \|\boldsymbol{u}\| \|\boldsymbol{v}\|$$

And there we have a **definition for (cos) angle** θ. Further from Cauchy-Schwartz inequality we know that **(cos) angle** θ must be:

$$-1 \leq \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|} \leq 1$$

3.2Vector: Orthogonality.

- A pair of vectors **u and v** are **orthogonal** if their inner product is zero i.e. $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{0}$.
- Notation for a pair of orthogonal vectors is $\boldsymbol{u} \perp \boldsymbol{v}$.
- In the $\mathbb{R}^n$; this is equal to pair of vector forming a $\mathbf{90^0}$ angle.

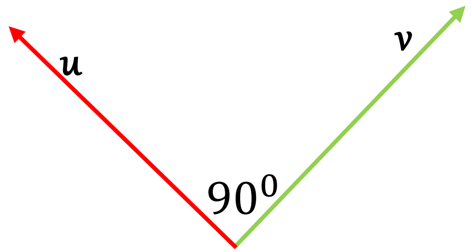


Fig: Orthogonal Vectors

4.Matrices.

Definition and Basic Matric Operation.

4.1 Matrix: 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 a *italicized* upper-case letter like A.
  - For two dimensions: we say the matrix A has m rows and n columns. Each entry of A is defined as a_{ij}.
  - Thus a matrix $A^{m \times n}$ is define as:

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

4.2 Special Matrices.

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

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

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

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

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

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

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

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

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

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

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

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

4.2 Special Matrices.

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

$$\boxed{\boldsymbol{A}_{3\times3} := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 6 \\ 3 & 6 & 1 \end{bmatrix}}$$

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

$$\boxed{\boldsymbol{A}_{3\times3} := \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}}$$

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

$$\boxed{\boldsymbol{A}_{3\times3} := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}$$

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

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

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

- **Design Matrix:**
  - A design matrix is a matrix containing **data** about multiple characteristics of **several individuals or objects**. Each row corresponds to an individual and each column to a characteristic. For instances:
    - If we measure the height and weight of five individuals, we can collect the measurements in a design matrix having five rows and two columns.
    - Each row corresponds to one of the ten individuals, the first column contains the height measurements and the second one reports the weights:

$$\boxed{\boldsymbol{X} := \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{w}_1 \\ \boldsymbol{h}_2 & \boldsymbol{w}_2 \\ \boldsymbol{h}_3 & \boldsymbol{w}_3 \\ \boldsymbol{h}_4 & \boldsymbol{w}_4 \\ \boldsymbol{h}_5 & \boldsymbol{w}_5 \end{bmatrix}}$$

4.3 Matrix Operation: Arithmetic.

- **Matrix-matrix addition/subtraction:**
  - Matrices are added or subtracted in a element wise fashion. The sum (+ or-) of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ and is defined as:

$$\boxed{\boldsymbol{A} \pm \boldsymbol{B} := \begin{bmatrix} a_{11} \pm b_{11} & \cdots & a_{1n} \pm b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & \cdots & a_{mn} \pm b_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}}$$

- **Matrix-Scalar Multiplication:**
  - Matrix-scalar multiplication is an element-wise operation. Each element of the matrix $\boldsymbol{A}$ is multiplied by the scalar $\boldsymbol{\alpha}$ is defined as: $\boldsymbol{a_{ij}} \times \boldsymbol{\alpha}$, such that $(\boldsymbol{\alpha A})_{ij} = \boldsymbol{\alpha}(\boldsymbol{A})_{ij}$.

$$\boxed{\begin{gathered} \alpha = 2 \text{ and } A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\ \alpha A = 2 \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \end{gathered}}$$

4.3 Matrix Operation: Arithmetic.

- **Matrix-vector Multiplication(dot product):**
  - Matrix-vector multiplication equals to taking the dot product of each column $\boldsymbol{n}$ of matrix-$\boldsymbol{A}$ with each element of vector-$\boldsymbol{x}$ resulting in vector $\boldsymbol{y}$ and defined as:

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

- **Hadamard Product:**
  - It is tempting to think in matrix-matrix multiplication as an element-wise operation, as multiplying each overlapping element of A and B. Such operation is called **Hadamard product ; defined as:**
    - $\boldsymbol{a_{ij}.b_{ij} \coloneqq c_{ij}}$

- $$\boldsymbol{A.B} = \begin{bmatrix} 0 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 \times 1 & 2 \times 3 \\ 1 \times 2 & 4 \times 1 \end{bmatrix} = \begin{bmatrix} 0 & 6 \\ 2 & 4 \end{bmatrix}$$

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

:

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

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

Matrix-matrix multiplication has a series of important properties:

- Associativity:
  $(AB)C = A(BC)$

Associativity with scalar multiplication:
  $\alpha(AB) = (\alpha A)B$

- Distributive with addition:
  $A(B \pm C) = AB \pm AC$
- Caution! In matrix-matrix multiplication orders matter, it is not commutative i.e.
  $AB \neq BA$.

4.4 Matrix Operation Summary: Arithmetic Properties.

- Properties of Matrix Arithmetic:
  - Assuming that the sizes of the matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid:
    - $A+B = B+A$ [Commutative law for matrix addition]
    - $A+(B+C) = (A+B)+c$ [Associative law for matrix addition]
    - $A(BC) = (AB)C$ [Associative law for matrix multiplication]
    - $A(B+C) = AB + AC$ | $A(B-C) = AB-AC$ [Left distributive law]
    - $(B+C)A = BA + CA$ | $(B-C)A = BA-CA$ [Right distributive law]
    - $a(B + C) = aB + aC$ | $a(B-C) = aB - aC$
    - $(a+b) C = aC + aC$ | $(a-b)C = aC - bC$
    - $a(bC) = (ab)C$
    - $a(BC) = (aB)C = B(aC)$

4.4 Matrix operations: Algebraic.

- **Identity Matrix:**
  - An identity matrix is a square matrix with ones on the diagonal from the upper left to the bottom right, and zeros everywhere else. We denote identity matrix as $\boldsymbol{I_n}$. We define $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ as:

$$\boldsymbol{I_n} := \begin{bmatrix} \mathbf{1_{11}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1_{nn}} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

4.5 Matrix operations Algebraic: Inverse.

- Consider the square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$. We define $\boldsymbol{A}^{-1}$ as a matrix with property:
  - $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n = \boldsymbol{A}\boldsymbol{A}^{-1}$
- Inverse of a matrix can be used to solve system of linear equation i.e.
  - For a equation:
    - $\boldsymbol{Ax} = \boldsymbol{y}$
  - If A has an inverse:
    - $\boldsymbol{A}^{-1}\boldsymbol{Ax} = \boldsymbol{A}^{-1}\boldsymbol{y}$

$$\boldsymbol{Ix} = \boldsymbol{A}^{-1}\boldsymbol{y}$$

  - Thus :
    - $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$
- **Singular/Non Singular matrix:**
  - If A is a square matrix, and if a matrix B of the same size can be found such that AB = BA = I, then A is said to be invertible (or non singular) and B is called an inverse of A. If no such matrix B can be found then A is said to be singular.

4.5 Matrix operations Algebraic: Inverse.

- Method for computing the inverse of an **invertible matrix** of any size:
  - **Determinant of a matrix**:
    - The ***determinant*** of a matrix is a **single numerical value** which is used when calculating the inverse or when solving systems of linear equations.
    - The determinant of a matrix **A** is denoted **|A|**, or sometimes **det(A)**. The determinant is only defined for square matrices.
    - A matrix is said to be ***singular*** if its **determinant** is ***zero***.
  - **Determinant of 2X2 matrix:**

$$\boxed{\begin{gathered} For\ A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}: \\ \det(A) = ad - bc. \end{gathered}}$$

  - **Determinant of 3X3 matrix:**

$$\boxed{\begin{gathered} For\ A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}: \\ \det(A) = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} \end{gathered}}$$

4.5 Matrix operations Algebraic: Inverse.

- Method for computing the inverse of an invertible matrix of any size:
  - The matrix:
      - $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
    - is invertible if and only if $ad\text{-}bc \neq 0$, in which case the inverse is given by the formula:
      - $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

5.System of Linear Equations.

5.1 What is Linear Equations?

- In mathematics, a linear equation is an equation in the form:
  - $\boldsymbol{\alpha_1 x_1 + .. + \alpha_n x_n + b = 0}$
    - Where:
      - $\boldsymbol{x_1, .., x_n}$ ***are the variables(unknown)***
      - $\boldsymbol{b, \alpha_1, .., \alpha_n}$ ***are coefficients which are often real numbers.***
        - -Wikipedia
- Geometrically;
  - If you plot a linear equation on a graph it represents the straight line.

5.2 What is system of Linear Equations?

- A *system of linear equations* consists of *two or more linear equations* with same variable.
- For Example:
  - $$\begin{cases} x + 2y - z = 1 \\ 2x + 2y + z = -2 \\ -x + y - z = 0 \end{cases}$$
- Solving System of Linear Equations:
  - Graphing method
  - Substitution method
  - Elimination method
  - Matrix method

5.3 The Racing Problem.

- You are playing catching up with your dog. From previous experiment we know that you can run at the speed of **0.2 km/minute**, your dog can run with the speed of **0.5 km/minute**. You had a 5 minute head start.
  - **How far you can go before your dog catches you?**
  - **You and dog had same speed of 0.5km/minute and you still had a 5 minute head start.**
  - **You and dog had same speed of 0.5km/minute and you do not have a head start.**
  - **(Hint: Plot in Graph.)**

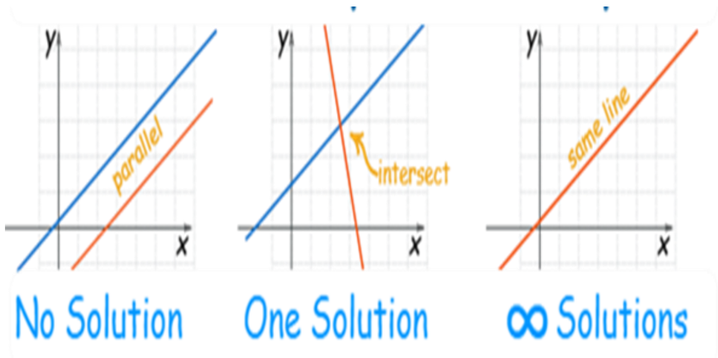
5.3 The Racing Problem-Solution.

- Step:1
  - Create all possible linear equations.
    - You run at 0.2 km/min i.e.
      - $\boldsymbol{d = 0.2t \quad ---(i)}$
        - Here; $\boldsymbol{d = distance\ travelled\ and\ t = time\ taken.}$
    - Dog runs at 0.5 km/min i.e.
      - $\boldsymbol{d = 0.5(t - 5) \quad ---(ii)}$
- Step:2
  - Graph both the equations.

5.3 The Racing Problem-Solution.

- Step:1
  - Create all possible linear equations.
    - You run at 0.2 km/min i.e.
      - $\boldsymbol{d = 0.2t \ ---(i)}$
        - Here; $\boldsymbol{d = distance\ travelled\ and\ t = time\ taken.}$
    - Dog runs at 0.5 km/min i.e.
      - $\boldsymbol{d = 0.5(t-5) \ ---(ii)}$
- Step:2
  - Graph both the equations.

Solutions of System of Linear Equations.



A system of equations is called *inconsistent* if it has **no solutions**. It is called *consistent* otherwise.

5.4 System of Linear Equations: Matrix.

- *For following system of linear equations: what is the ideal way to represent it:*
  - $$\begin{cases} \boldsymbol{x+2y-z=1} \\ \boldsymbol{2x+2y+z=-2} \\ \boldsymbol{-x+y-z=\ \ 0} \end{cases}$$

5.4 System of Linear Equations: Matrix.

- For following system of linear equations: what is the ideal way to represent it:
  - $$\begin{cases} x + 2y - z = 1 \\ 2x + 2y + z = -2 \\ -x + y - z = 0 \end{cases}$$
- Matrices are ideal to represent systems of linear equations. Consider the matrix **M** and vectors **w** and **y** in $\mathbb{R}^3$, we can set up a system of linear equations as $\boldsymbol{Mw} = \boldsymbol{y}$ as:
  - $$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$; which is equivalent to:

$$\boxed{\begin{aligned} m_{11}w_1 + m_{12}w_2 + m_{13}w_3 &= y_1 \\ m_{21}w_1 + m_{22}w_2 + m_{23}w_3 &= y_2 \\ m_{31}w_1 + m_{32}w_2 + m_{33}w_3 &= y_3 \end{aligned}}$$

6.Solving system of linear equations: Matrices.

6.1 Gauss Elimination Method.

- Gauss Elimination is a robust algorithm to solve system of linear equations.
- It works by *eliminating* terms from a system of equations, such that it is simplified to the point where we obtain the **row echelon form** of the matrix.
  - A matrix is in **row echelon form** when all rows contain zeros at the bottom left of the matrix. For instance:
    - $\begin{bmatrix} p_1 & a & b \\ 0 & p_2 & c \\ 0 & 0 & p_3 \end{bmatrix}$
      - The **p values** along the diagonal are the **pivots** also known as basic variables of the matrix.
- There are three ***elementary transformations*** in **Gaussian Elimination** that when combined, allow simplifying any system to its row echelon form:
  - Addition and subtraction of two equations (rows)
  - Multiplication of an equation (rows) by a number
  - Switching equations (rows)

6.2 Gauss Elimination Method: Example.

- For the following system $Aw = y$:

$$\begin{bmatrix} 1 & 3 & 5 \\ -1 & 2 & -1 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

6.2 Gauss Elimination Method: Solution.

Writing in augmented form:

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 2 & 2 & -1 & 1 \\ 1 & 3 & 2 & 2 \end{array}\right]$$

Perform: $R_2 : R_2 - 2R_1$

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 0 & -4 & -11 & 3 \\ 1 & 3 & 2 & 2 \end{array}\right]$$

Perform: $R_3 : R_3 - R_1$

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 0 & -4 & -11 & 3 \\ 0 & 0 & -3 & 3 \end{array}\right]$$

This is a row echelon form.

Writing in the original form :

$$\begin{bmatrix} 1 & 3 & 5 \\ 0 & -4 & -11 \\ 0 & 0 & -3 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 3 \end{bmatrix}$$

If we write in the equation form:

$$-3w_3 = 3 \Rightarrow w_3 = -1.$$

By back substitution:

$$-4w_2 + (-11)(-1) = 3$$
$$-4w_2 = -8 \Rightarrow w_2 = 2$$

Final equation:

$$w_1 + 3(2) + 5(-1) = -1$$
$$w_1 + 6 - 5 = -1$$
$$w_1 = -2.$$

6.3 Gauss -Jordan Method:

- The only difference between **Gaussian Elimination** and **Gauss-Jordan Elimination,** is that in Gauss-Jordon elimination we obtain ***reduced row echelon form***:
- Reduced row echelon form: for instance:
  - $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
    - Pivots must be 1 and
    - Entries above the pivots must be 0.

6.3 Gauss -Jordan Method: Example.

Writing in augmented form:

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 2 & 2 & -1 & 1 \\ 1 & 3 & 2 & 2 \end{array}\right]$$

Perform: $R_2 : R_2 - 2R_1$

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 0 & -4 & -11 & 3 \\ 1 & 3 & 2 & 2 \end{array}\right]$$

Perform: $R_3 : R_3 - R_1$

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 0 & -4 & -11 & 3 \\ 0 & 0 & -3 & 3 \end{array}\right]$$

This is a row echelon form.

Perform: $R_3 : \frac{R_3}{3}$ and $\frac{R_2}{-4}$:

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 0 & 1 & 2.75 & -0.75 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

Perform: $R_1 : R_1 - 3R_2$

$$\left[\begin{array}{ccc|c} 1 & 0 & -3.25 & 1.25 \\ 0 & 1 & 2.75 & -0.75 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

Perform: $R_1 : R_1 + 3.25R_3$ and $R_2 : R_2 - 2.75R_3$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

This is a reduced row echelon form.

Thank You!!!

Questions????