DCS-404Artificial Intelligence and Machine Learning
Lecture-01
Foundational Math Review-01
Linear Algebra.

Siman Giri

Learning Outcome:

- Review and revise some fundamental concepts from mathematics we will be using through out the course.

A. Preliminary Concepts

Sets and Functions.

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**:
  - $x \in A$**(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:
    - $X = \{x_1, \ldots, x_n\}$ **or;** $X = \{x: x \text{ is an even integer and } x > 0\}$
- Some of the **important sets** are:

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

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:** $\{4, 5, 8\} \subset \{2, 3, 4, 5, 6, 7, 8, 9\}$
  - 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:
    - $A \cup B = \{x: x \in A \text{ or } x \in B\}$ and $A \cap B = \{x: x \in A \text{ and } x \in B\}$.
    - 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 $A_1, \dots, A_n$.
  - **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 $A \subset U: U \text{ universal set}$. Complement of A is :
    - $A' = \{x: x \in U \text{ and } x \notin \mathrm{A}\}$.
  - **Difference** of two sets A and B is:
    - $A \backslash \mathrm{B} = A \cap B' = \{x: x \in A \text{ and } x \notin \mathrm{B}\}$.

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.
  - $A \times B = \{(a, b): a \in A \text{ and } b \in B\}$

Example:
If $A = \{x, y\}, B = \{1, 2, 3\}, \text{and } C = \emptyset$, 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:
  - $A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n): a_i \in A_i \text{ for } i = 1, \ldots, n\}$.

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**.

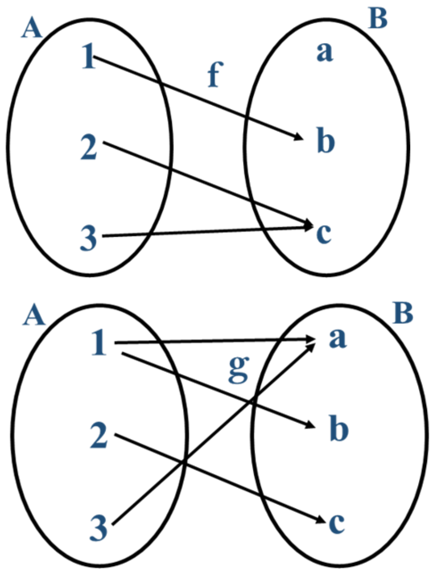


Figure:1 Mappings and relations.

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**.

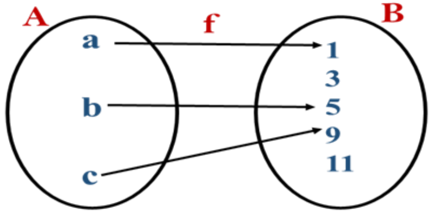


Fig: A well defined function.

- **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.

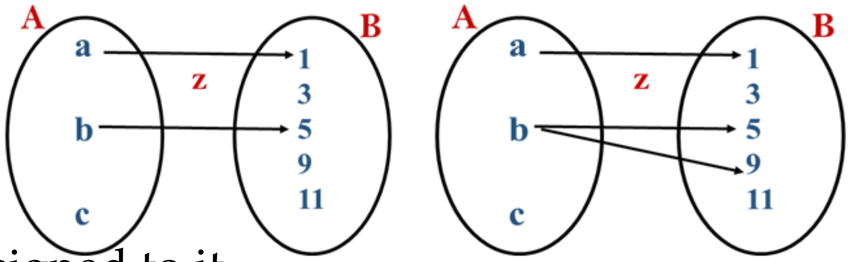


Fig: Not a Functions

A.4 Relations: (Domain and Range).

- Let $f: A \rightarrow B$ 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:
  - $\forall x_1, x_2 \in A: x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$.

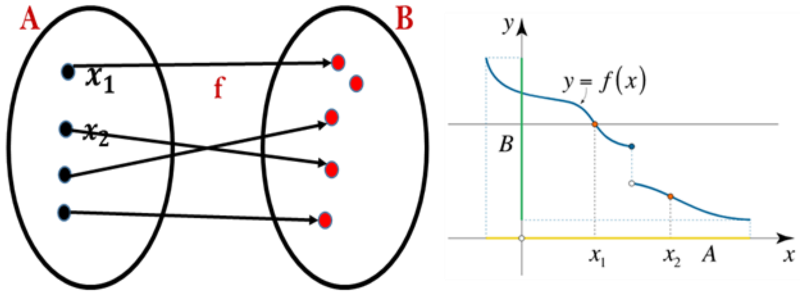


Fig: Injective Function.

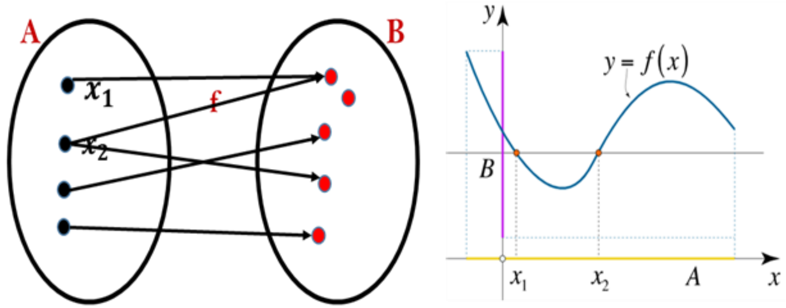


Fig: Non-Injective Function.

A.4 Relations: (Domain and Range).

- Let $f : A \rightarrow B$ 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** $y = f(x)$.

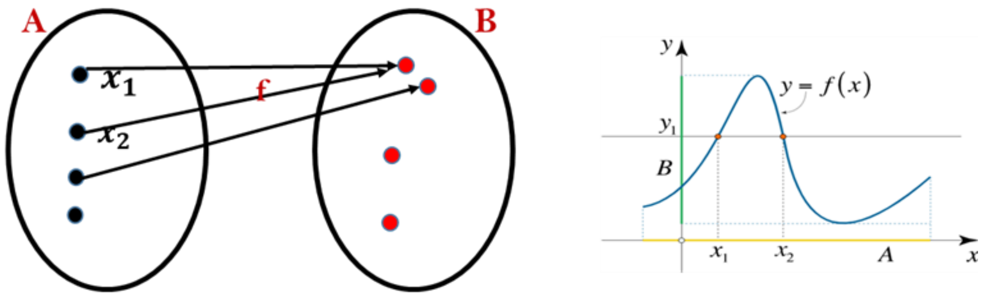


Fig: Surjective Function.

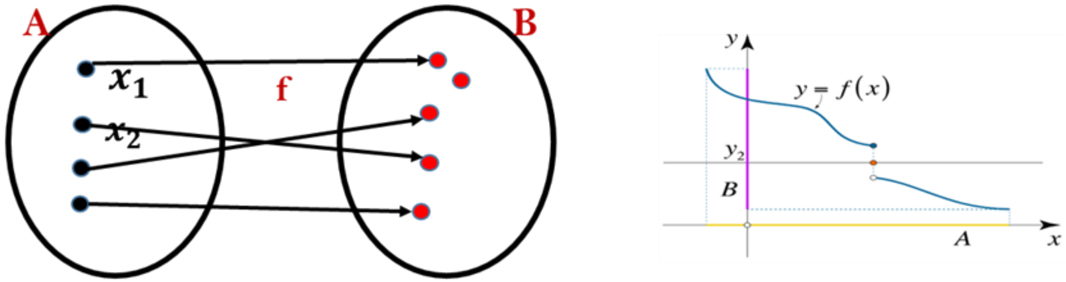


Fig: Non-Surjective Function.

A.4 Relations: (Domain and Range).

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

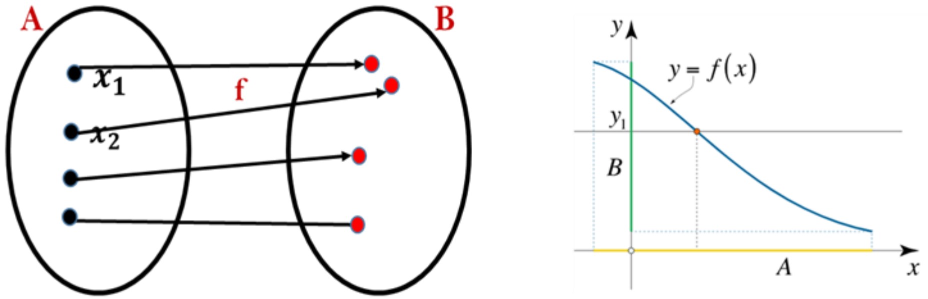


Fig: Bijective Function.

A.5 Summary: Functions and Machine Learning.

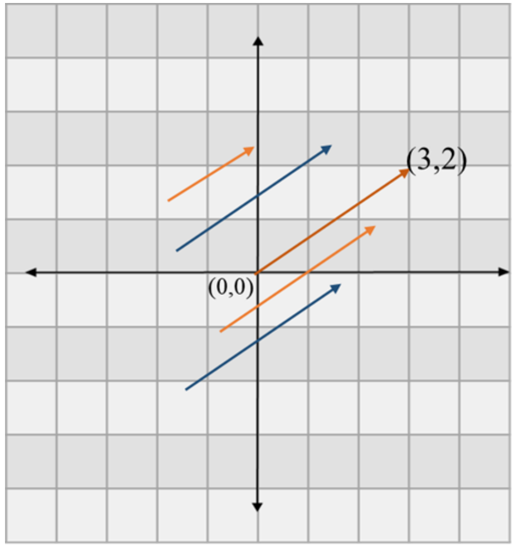
- The ultimate goal of **machine learning** is **learning** a **functions** from **data**, i.e. mappings from domain **(feature vector space(set))** onto the range **(target variable)** of a function.
- The **objective** of **DCS404** is to be able to **understand** all the highlighted **terms** in above **statement**.

Vector; Matrices and Tensors.

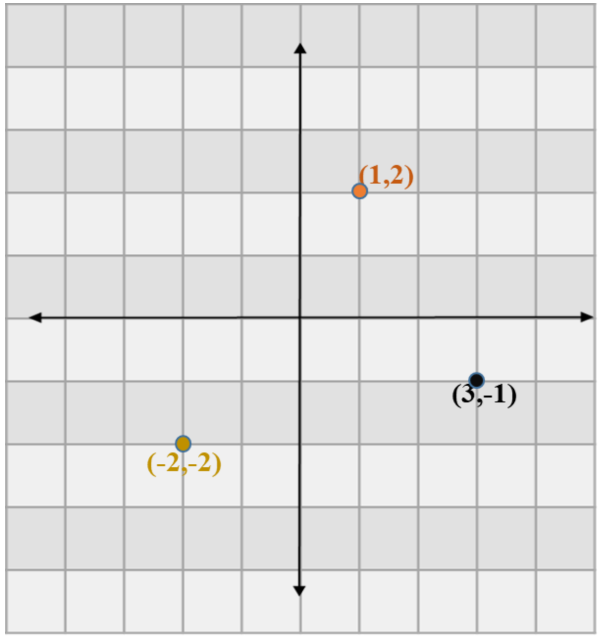
1. Vector

1.1 Vectors: Interpretation.

- **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



- **point in space:**
  - E.g., in 2D we can visualize the data points with respect to a coordinate origin



1.2 Vectors: Definition.

- **In Math/Physics:**
  - Vectors are quantity having both **direction and magnitude** written as $\vec{\mathbf{v}}$ (**arrow** represents the **direction** and **v: magnitude** which is proportional to length)
- **In Computer science:**
  - Vectors are **one dimensional ordered array** of **real value numbers** (**scalars**).
  - Denoted by **bold-font lower case**, can be written in **column** or **row** form.

$$\boldsymbol{v} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{5} \end{bmatrix} \textbf{ or; } \boldsymbol{v} = [\mathbf{1} \quad \mathbf{0} \quad \mathbf{5}]$$

    - length of an array defines the **dimension of vector** i.e. how many **axis** are required to represent the **vector in graph**.
  - To generalize: we can write $\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{n}}$. Here $\boldsymbol{n}$ signifies the dimension.

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \ddots \\ \boldsymbol{v}_n \end{bmatrix}$$

1.3 Vector Operations: Addition.

- **Parallelogram Rule for Vector Addition:**
  - If v and w are vectors in **n-space{$\mathbb{R}$^n}** 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 :

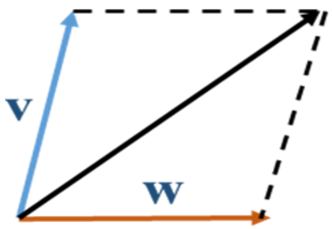


Fig: a

- **Triangle Rule for Vector Addition:**
  - If v and w are vectors in **n-space{$\mathbb{R}$^n}** 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 :

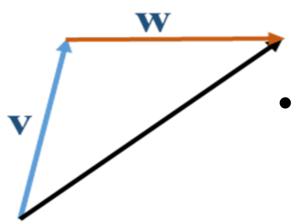


Fig: b

Implementation:

Add **element** by **element**

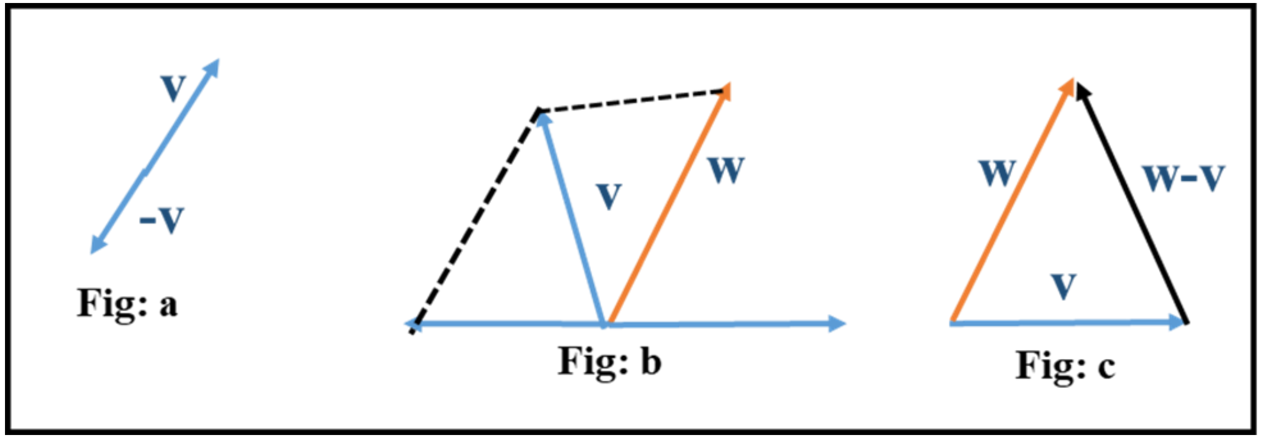
i.e.

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} \in \mathbb{R}^n; v = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} \in \mathbb{R}^n \text{then:}$$

$$w + v = \begin{bmatrix} w_1 + v_1 \\ w_2 + v_2 \\ \dots + \cdots \\ w_n + v_n \end{bmatrix} \in \mathbb{R}^n$$

1.3 Vector Operations: 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).**



Implementation:

Subtract **element** by **element**

i.e.

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} \in \mathbb{R}^n; v = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} \in \mathbb{R}^n \text{ then:}$$

$$w - v = \begin{bmatrix} w_1 - v_1 \\ w_2 - v_2 \\ \dots - \dots \\ w_n - v_n \end{bmatrix} \in \mathbb{R}^n$$

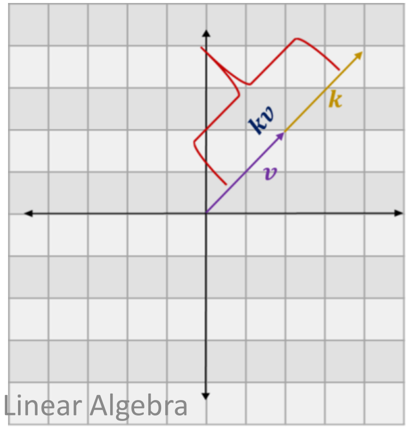
1.3 Vector Operations: Scalar Multiplication.

- **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 number.
- **Scalar Multiplication**:
  - If **v** is a nonzero vector in **n-space{$\mathbb{R}^n$}** , and if **k** is a **nonzero scalar (number)**, 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**

i.e. For $v \in \mathbb{R}^n$, and $k \in \mathbb{R}$ scalar multiplication is:

$$k\,.\,v = \begin{bmatrix} k.v_1 \\ k.v_2 \\ \dots \\ k.v_n \end{bmatrix} \in \mathbb{R}^n$$

  - and **opposite** to that of **v** if **k** is **negative**.
  - If **k = 0** or **v = 0**, then we define **kv to be 0.**



Scalar Multiplication – Properties:
{j,k:= scalar and v,u:= vectors}

- Associativity: **(jk)v = j(kv)**
- Distributive property (Left and Right): **(j+k)v = jv + kv ~ v(j+k) = vj + vk**
- Distributive property vector addition: **j(u + v) = ju + kv**

1.3 Vector Operations: Norm.

Norm: Definition

- For a vector:= $\boldsymbol{v} = [v_1 \quad \cdots \quad v_n] \in \mathbb{R}^n$, then the **norm of v** (also called the **length of v** or the **magnitude of v**) is denoted by $\|\boldsymbol{v}\|$.
- A **norm** can be any function $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}$ that satisfies following properties:
  - For all **v ∈ $\mathbb{R}^n$, f(v) ≥ 0** (**non-negativity**).
  - **f(v) = 0** if and only if **v = 0** (**definiteness**).
  - For all **vector v ∈ $\mathbb{R}^n$** ,and **scalar t ∈ R**, **f(tv) = |t| f(v)** (**homogeneity**).
  - For all **vector v, w ∈ $\mathbb{R}^n$** : **f(v+ w) ≤ f(v) + f(w)** (**triangle inequality**).

Popular Norms:

- **For $p = 2$, we have ℓ_2 norm**

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

  - Also called **Euclidean norm**
  - It is the most often used norm
  - ℓ_2 norm is often denoted just as $\|\mathbf{x}\|$ with the subscript 2 omitted
- **For $p = 1$, we have ℓ_1 norm**

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

  - Uses the absolute values of the elements
  - Discriminate between zero and non-zero elements
- **For $p = \infty$, we have ℓ_∞ norm**

$$\|\mathbf{x}\|_\infty = \max_i |x_i|$$

  - Known as **infinity norm**, or **max norm**
  - Outputs the absolute value of the largest element

1.3 Vector Operations: unit vector.

unit vector: Definition.

- 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 ǀ½ vǀ is a unit vector in the **same direction** as v.
  - More generally, if v is any nonzero vector in $\mathbb{R}^n$, then:
    - $\hat{\boldsymbol{v}} = \frac{1}{\|\boldsymbol{v}\|}\boldsymbol{v}$
- **Normalizing Vector**:
  - The process of **multiplying a nonzero vector** by the reciprocal of its length to obtain a unit vector is called normalizing v.

norm and unit vector: example.

- **Example: Normalizing a vector.**
- Find the unit vector of:

  vector:= $\boldsymbol{v} = [\mathbf{2}, \mathbf{2}, -\mathbf{1}]$.
- Solution:
  - The vector v has length{Norm}:
    - $\|\boldsymbol{v}\| = \sqrt{\mathbf{2}^2 + \mathbf{2}^2 + (-\mathbf{1})^2} = \mathbf{3}.$
  - Thus unit vector:
    - $\hat{\boldsymbol{v}} = \left(\frac{1}{\|\boldsymbol{v}\|}\right)\boldsymbol{v} = \frac{1}{3} \times [2, 2, -1] = \left[\frac{2}{3};\ \ \frac{2}{3};\ \ \left(-\frac{1}{3}\right)\right]$
  - To check:
    - $\|\hat{\boldsymbol{v}}\| = \sqrt{\left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(-\frac{1}{3}\right)^2} = \mathbf{1}.$

1.3 Vector Operations: Dot Product.

dot product: Definition.

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

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

- dot product: properties:
  - If u, v, and w are vectors in $\mathbb{R}^n$ and k a scalar then: **{u, v := vector & k := scalar}**
    - **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)**

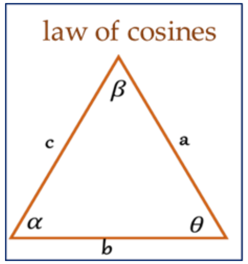
dot product : Example.

- Question: Calculate - $(u - 2v).(3u + 4v)$
- Solution:

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

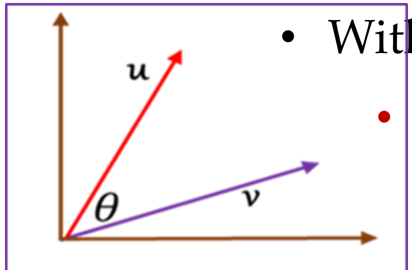
1.4 Angle between two vectors.

- 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:



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

- For angel between two vectors **u and v**:
  - We can replace above expression with vector lengths:
    - $\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{u}\|\|\boldsymbol{v}\|\boldsymbol{cos\theta}.$
  - With a bit of algebraic manipulation:
    - $\boldsymbol{cos\theta} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle / \|\boldsymbol{u}\|\|\boldsymbol{v}\|$ **then;** $\boldsymbol{\theta} = \mathbf{arccos}\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right)$



angle between vectors

- **Vector 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 $\mathbb{R}^{\boldsymbol{n}}$; this is equal to pair of vector forming a $\mathbf{90^0}$ angle.

Fig: Orthogonal Vectors

Vector; Matrices and Tensors.

2. Matrices

2.1 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 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 ≠ to the number of columns, i.e. $A^{m\times n}$ with $m \neq n$. For instance:

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

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

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

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

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

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

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

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

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

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

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

4.2 Special Matrices.

- **Symmetric Matrix:**
  - Square matrices are said to be symmetric its equal to its transpose, i.e. $A = A^T$. 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. $D = \alpha I$. 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 $A(a_{ij}) = B(b_{ij})$. For instance:

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

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

- **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:

$$X := \begin{bmatrix} h_1 & w_1 \\ h_2 & w_2 \\ h_3 & w_3 \\ h_4 & w_4 \\ h_5 & w_5 \end{bmatrix}$$

4.3 Matrix Operation: Arithmetic.

Matrix-Matrix : ($\pm$)

- 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:

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

Matrix-Scalar Multiplication

- Matrix-scalar multiplication is an element-wise operation.
- Each element of the matrix A is multiplied by the scalar α is defined as:

$$\boldsymbol{a_{ij}} \times \boldsymbol{\alpha}, \text{ such that } (\boldsymbol{\alpha A})_{ij} = \boldsymbol{\alpha}(\boldsymbol{A})_{ij}.$$

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

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

4.3 Matrix Operation: Arithmetic.

Matrix-Vector Multiplication.

- 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:

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

- **Example:**

Given Matrix A and column vector x ; Compute $A \times x$:

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 3 & -1 \\ 1 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{3\times3} \text{ and } x = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix} \in \mathbb{R}^{3\times1}$$

$$Ax = 3 \times \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (-1) \times \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} + 4 \times \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \times 1 + 3 \times 0 + 3 \times 1 \\ -1 \times 0 - 1 \times 3 - 1 \times 2 \\ 4 \times -2 + 4 \times -1 + 4 \times 1 \end{bmatrix}$$

$$= \begin{bmatrix} -5 \\ -7 \\ 5 \end{bmatrix} \in \mathbb{R}^{3\times1}$$

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:**

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

Where:

$$a_{mn} \times b_{mn} := c_{mn}$$

- **Example:**

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

4.3 Matrix Operation: Arithmetic.

Trace

- *Trace* of a matrix is the sum of all diagonal elements

$$\mathrm{Tr}(\mathbf{A}) = \sum_i a_{ii}$$

- A matrix for which $\mathbf{A} = \mathbf{A}^T$ is called a ***symmetric matrix.***

Determinant

- *Determinant* of a matrix, denoted by $\det(\mathbf{A})$ or $|\mathbf{A}|$, 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

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

  - In the above, $\mathbf{A}_{(i,j)}$ is a minor of the matrix obtained by removing the row and column associated with the indices i and j

4.4 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:

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

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

- Matrix – Matrix Multiplication Properties:
  - Associativity:

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

Associativity with scalar multiplication:

$$\alpha(\boldsymbol{AB}) = (\alpha\boldsymbol{A})\boldsymbol{B}$$

  - Distributive with addition:

$$\boldsymbol{A}(\boldsymbol{B} \pm \boldsymbol{C}) = \boldsymbol{AB} \pm \boldsymbol{AC}$$

  - Caution! In matrix-matrix multiplication orders matter, it is not commutative i.e.

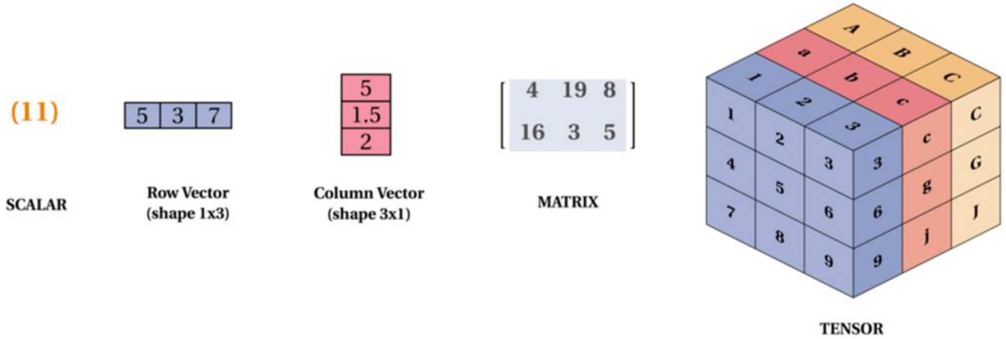
$$\boldsymbol{AB} \neq \boldsymbol{BA}.$$

Vector; Matrices and Tensors.

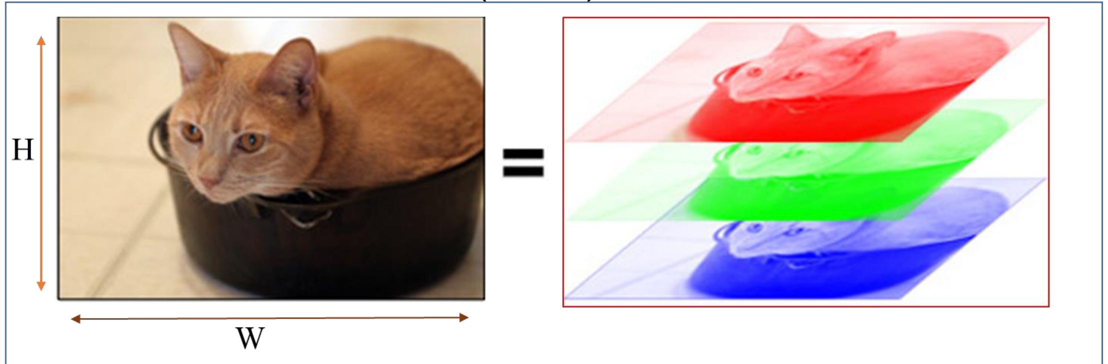
3. Tensors

3.1 Tensors: Definition.

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



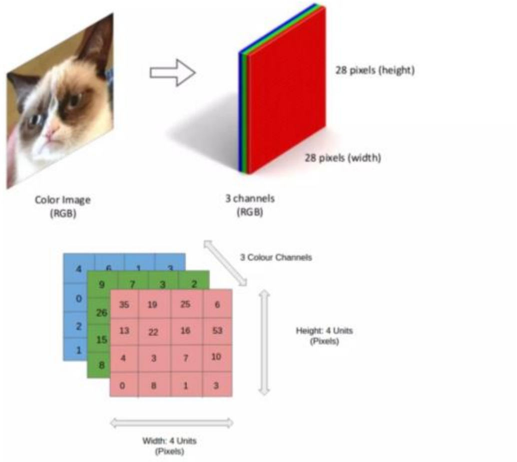
- Tensors in DL are Used to represent an image.
  - image_shape := Height × Width × Color Channel (RGB)



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

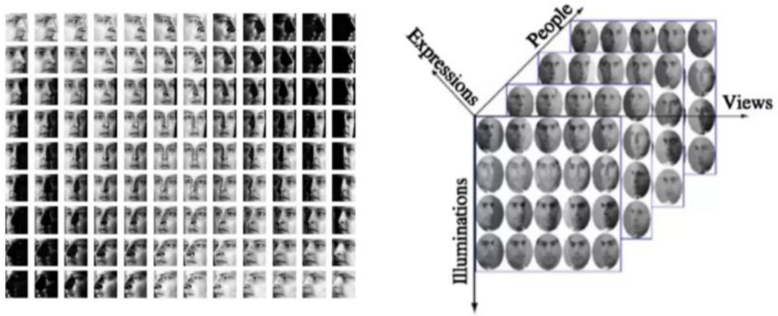
3.2 Tensors: Examples

color image is 3rd-order tensor



color video is 4th-order tensor

facial images database is 6th-order tensor



Calculus.

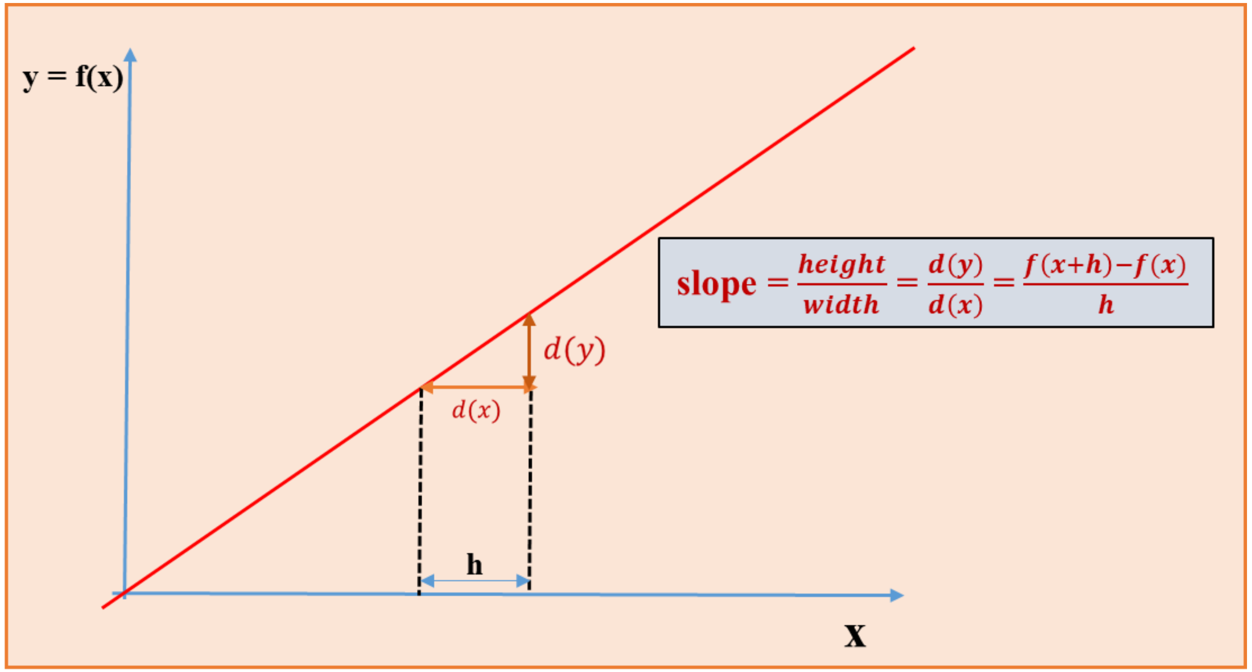
4.Derivative of a Function.

4.1 Derivatives: Definition.

- **Most popular: Derivative of a function i.e. Scalar derivatives $\boldsymbol{f}: \mathbb{R} \to \mathbb{R}$**
  - The derivative of a function $f(x)$ is represented by $\frac{d}{d(x)}\big(f(x)\big)$ or $\frac{df(x)}{d(x)}$ or $f'(x)$ and is defined as:
  - The derivative is the heart of calculus, buried inside this definition:
    - $\boldsymbol{f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}}$ **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?

4.2 Derivative: Intuition.

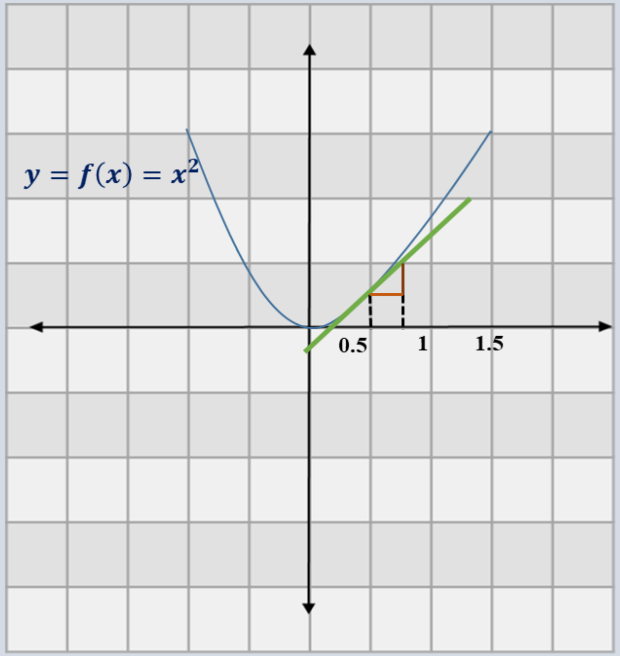
- Derivative of a function is a measure of local slope.
- 1st Example: For Linear Function $y = f(x) = 2x$.



What for non linear function?

4.2 Derivative: Intuition.

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



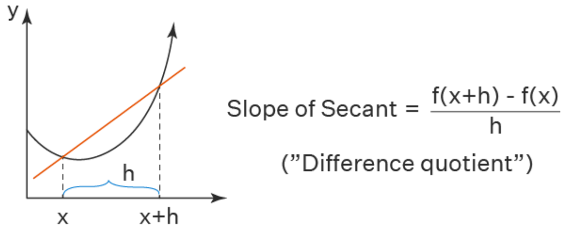
I want to find the slope at point 0.5. What does this mean?
You draw a tangent at x = 0.5 and find the slope of that tangent line.

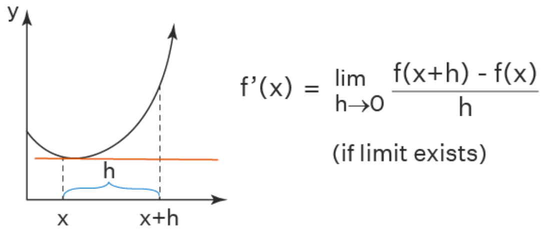
$$\text{Slope} = \frac{height}{width} = \frac{d(y)}{d(x)}\Big|_{x=0.5} = 1$$

- 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.

4.1 Derivatives: Definition.

- **Most popular: Derivative of a function i.e. Scalar derivatives $f: \mathbb{R} \to \mathbb{R}$**
  - The derivative of a function $f(x)$ is represented by $\frac{d}{d(x)}\big(f(x)\big)$ or $\frac{df(x)}{d(x)}$ or $f'(x)$ and is defined as:
  - The derivative is the heart of calculus, buried inside this definition:
    - $\boldsymbol{f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}}$ **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?

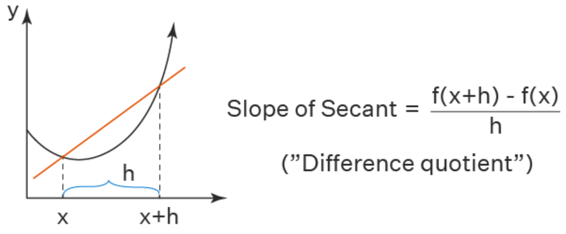


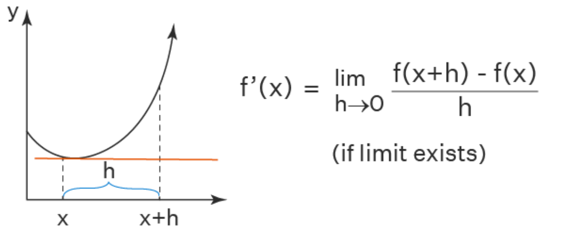


Does that mean every time we plot the function in graph to determine the derivative?

4.1 Derivatives: Definition.

- **Most popular: Derivative of a function i.e. Scalar derivatives $\boldsymbol{f}: \mathbb{R} \rightarrow \mathbb{R}$**
  - The derivative of a function $f(x)$ is represented by $\frac{d}{d(x)}\left(f(x)\right)$ or $\frac{df(x)}{d(x)}$ or $f'(x)$ and is defined as:
  - The derivative is the heart of calculus, buried inside this definition:
    - $\boldsymbol{f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}}$ **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?





Does that mean every time we plot the function in graph to determine the derivative?

No we have some common rules and formula for that.

4.3 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	$f(x) = \tan(x)$	$f'(x) = (\tan(x))' = \sec^2(x)$.

4.4 Derivation Rules.

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.

Calculus.

5.Derivative of a Function with multiple variable.

5.1 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.

5.2 Derivative of a Vector/Matrix.

- 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:

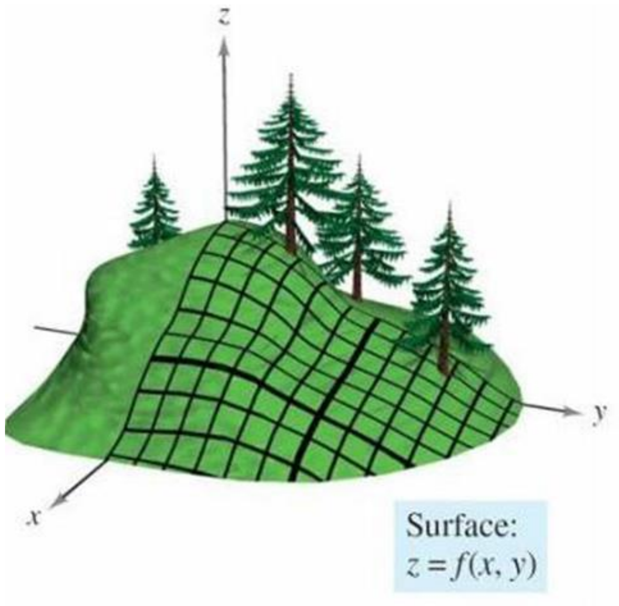
Setting	Derivative	Notation
$f: \mathbb{R} \to \mathbb{R}$	Scalar Derivative	$f'(x)$
$f: \mathbb{R}^n \to \mathbb{R}$	Gradient	$\nabla f(x)$
$f: \mathbb{R}^{n\times m} \to \mathbb{R}$	Gradient	$\nabla f(x)$
$f: \mathbb{R}^n \to \mathbb{R}^m$	Jacobian	J_f

5.3 Derivative of a Vector: 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 $\{f: \mathbb{R}^n \rightarrow \mathbb{R}\}$:
    - 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: $\nabla y = \frac{\partial y}{\partial x} = \left[\frac{\partial y}{\partial x_1} \; \frac{\partial y}{\partial x_2} \; \cdots \; \frac{\partial y}{\partial x_n}\right]^T.$ $\longrightarrow$ gradients.
        - **{Stack the partial derivative against all the element of vector *x*}**
  - Scalar-by-Matrix $\{f: \mathbb{R}^{n\times m} \rightarrow \mathbb{R}\}$:
    - 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.}**

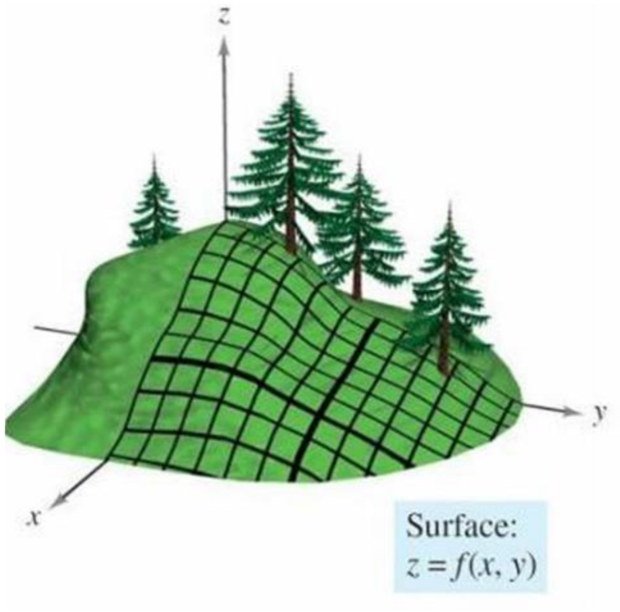
gradient is also the direction of steepest ascent, What does that mean?

5.4 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?

5.4 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?

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" ← $\nabla z = [\frac{\partial f(x,y)}{\partial x} \quad \frac{\partial f(x,y)}{\partial y}]$

5.5 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 $[1 \quad 1]$: $\boxed{\nabla z = [6 \times 1 \times 1 \quad 3 \times 1^2] = [6 \quad 3]}$

5.6 Derivative of a Vector: Jacobian.

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

Thank You and Questions!!!

p.neupane.276@Westcliff.edu
9861434787