UNIT-1 DEEP LEARNING

Machine Learning

It is a branch of artificial Intelligence That develops algorithms by learning The hidden patterns of the datasets used it to make predictions on new Similar type data, without being explicitly programmed for each task.

It is used in many different applications like recom-mmendation Systems, fraud detection, portfolio optimization, automated task, and so on.

Deep Learning

Deep learning is a branch of artificial Intelligence (AI) That teaches Computers to process data in a way That is inspired by the human brain.

Deep learning models Can recognize Complex patterns in pictures (images), text, sounds, and other data to prod-uce accurate insights and predictions.

Applications of Deep Learning :

* Self Driving Cars
* News Aggregation & Fraud News Detection
* Natural language Processing
* Health care
* Automatic Machine Translation
* Automatic Handwriting Generation

* Automatic Game playing
* Automatic facial recognition
* Medical Image Analysis
* Digital Assistants etc.

Scalars

In deep learning, a scalar refers to a single numeric value, as opposed to a vector or a matrix, which are collection of multiple values

=> Scalars have no direction and represent a magnitude only.

=> A Scalar is a single number

=> We write scalars in lower case, non-bold typeface font.

[Scalar is a zero ('0') dimensional tensor]

=> Examples:

$x \in R$

for Example, if you have a grayscale image, each pixel's intensity can be represented as a scalar value, In The Context of neural networks, scalars are commonly used as input, outputs, error values, or intermediate values in the computations performed by the network

Vector:

A vector is an array of numbers or a list of scalar

Values. The single value in the array/list of vector are called the entries or components of the vector.

⇒ The vector variables are usually denoted by lower case letter with a right arrow on top ($\vec{x}, \vec{y}, \vec{z}$) or bold faced lower case letter ($\mathbf{x}, \mathbf{y}, \mathbf{z}$).

⇒ Mathematical notation:-

$\vec{x} \in R^n$, where x is a vector, The expression says that vector $\vec{x}$ has n real value scalars.

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{or} \quad \vec{x} = [x_1, x_2, \ldots x_n]$$

where $x_1, x_2 \ldots x_n$ are the entries/components of the vector $\vec{x}$

Vector is also called 1 Dimensional tensor.

In deep learning, a vector is a one-dimensional array of numerical elements. Each element in the vector is identified by an index or a position

⇒ Vectors are commonly used to represent features, inputs, outputs or parameters in various components of deep learning models.

⇒ Examples * Input vector * output vector. * weight vector

Matrices :

It is a rectangular array of real-valued scalars arranged in m horizontal rows and n vertical columns.

⇒ Each element belongs to the i^{th} row and j^{th} column

⇒ The elements are denoted

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

Matrices are denoted by bold-font upper-case letter

* Matrices is also called 2-dimensional Tensors.

Mathematical notation : $A \in R^{m \times n}$

⇒ Matrices are fundamental to many aspects of deep learning particularly in representing weights, inputs, and transformations between layers in neural networks

⇒ In the context of deep learning

1. weight matrices
2. Input matrices
3. Activation matrices.

Mathematically, matrices play a crucial role in expressing the computations involved in neural networks.

Different operations

* Addition (or) Subtraction

$$(A \pm B)_{ij} = A_{ij} \pm B_{ij}$$

* Scalar multiplication

$$(cA)_{ij} = c \cdot A_{ij}$$

* Matrix Multiplication

$$(AB)_{ij} = A_{i1} + B_{1j} + A_{12} B_{2j} + \dots + A_{in} B_{nj}$$

[Note: Defined only if the number of columns of the left matrix is the same as the no. of rows of the right matrix]

* Transpose : It has the rows and columns exchanged

$$(A^T)_{i,j} = A_{j,i}$$

Example
$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -6 & 7 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 2 & -6 \\ 3 & 7 \end{bmatrix}$$

* Identity matrix (I_n) :

It has ones on the main diagonal, and zeros else where.

Example : $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- → Identity matrix
- → Diagonal Identity matrix
- → Scalar Identity matrix

Diagonal Identity matrix

It has any one same number on the main diagonal, and zeros elsewhere.

Example: $\begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ (or) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ etc

Scalar Identity matrix

It has any numbers on the main diagonal, and zero elsewhere

Example: $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ (or) $\begin{bmatrix} 6 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ etc.

* Inverse Matrix :-

For a Square matrix A with rank, is its inverse matrix if Their product is an identity matrix. I

$$A^{-1}A = AA^{-1} = I.$$

→ Additive Inverse $= (-A)$

→ Multiplicative Inverse. (A^{-1})

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad = \frac{1}{|ad-bc|} \cdot adj(A)$$

Additive Inverse

Example

$$A = \begin{bmatrix} -5 & 6 \\ 8 & 9 \end{bmatrix} \quad -A = \begin{bmatrix} 5 & -6 \\ -8 & -9 \end{bmatrix}$$

Tensors

Tensors are n-dimensional arrays of scalars.

→ In deep learning, tensors are multi-dimensional arrays that generalize the concept of scalars, vectors, and matrices to higher dimensions.

→ Tensors can be considered as a mathematical representation of data that can have any number of dimensions.

→ They are a fundamental data structure used to represent inputs, outputs, and parameters in deep learning models. The computations in neural networks are often expressed in terms of tensor operations.

Linear Combination

Given a set of vectors $\{v_1, v_2, \ldots v_k\}$ in a vector space V, any vector of the form

$$v = a_1\vec{v_1} + a_2\vec{v_2} + \ldots a_k\vec{v_k}$$

for some scalars $a_1, a_2, \ldots a_k$, is called a linear combination of $v_1, v_2, \ldots v_k$.

Example: Let $v_1 = (1,2,3)$, $v_2 = (1,0,2)$

Express $u = (-1,2,-1)$ as a linear combination of v_1 and v_2. find scalars a_1 and a_2.

Ans:

$$u = a_1 v_1 + a_2 v_2$$

$v_1 = (1,2,3)$ $\qquad a_1 + a_2 = -1$

$v_2 = (1,0,2)$ $\qquad 2a_1 + 0a_2 = 2$

$u = (-1,2,-1)$ $\qquad 3a_1 + 2a_2 = -1$

$$\left(\begin{array}{cc|c} 1 & 1 & -1 \\ 2 & 0 & 2 \\ 3 & 2 & -1 \end{array}\right) \qquad \begin{array}{l} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 3R_1 \end{array}$$

$$\left(\begin{array}{cc|c} 1 & 1 & -1 \\ 0 & -2 & 4 \\ 0 & -1 & 2 \end{array}\right)$$

So $a_2 = -2$ and $a_1 = 1$.

Example-2

Express $u = (-1,2,0)$

$v_1 = (1,2,3)$ $\quad v_2 = (1,0,2)$

$$\left(\begin{array}{cc|c} 1 & 1 & -1 \\ 2 & 0 & 2 \\ 3 & 2 & 0 \end{array}\right) \qquad \begin{array}{l} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 3R_1 \end{array}$$

$$\left(\begin{array}{cc|c} 1 & 1 & -1 \\ 0 & -2 & 4 \\ 0 & -1 & 3 \end{array}\right)$$

This System has no solution, so u can't be expressed as a linear combination of v_1 and v_2.

Linear Independence

Definition :- Given a set of Vectors $\{v_1, v_2, \dots v_k\}$, in a vector space V, they are said to be linearly independent if the equation

$$c_1\vec{v_1} + c_2\vec{v_2} + \dots + c_k v_k = 0$$

~~otherwise~~

if $\{v_1, v_2, \dots v_k\}$ are not linearly independent

They are called linearly dependent.

Example 1:

Determine whether $v_1 = (1,2,3)$ and $v_2 = (1,0,2)$ are linearly dependent or independent. Consider the Homogeneous system.

Solution :-

$$c_1(1,2,3) + c_2(1,0,2) = (0,0,0)$$

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 3 & 2 & 0 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

only solution is the trivial solution $a_1 = a_2 = 0$.

so linearly independent.

Example 2: Determine whether $v_1 = (1,1,0)$ and $v_2 = (1,0,1)$ and $v_3 = (3,1,2)$ are linearly dependent. Want to find solutions to the system of equations.

Solution $c_1(1,1,0)+c_2(1,0,1)+c_3(3,1,2)=(0,0,0)$

which is equivalent to solving

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Span

Definition: Given a set of vectors $\{v_1, v_2, \dots v_k\}$ in a vector space V, the set of all vectors which are a linear combination of $v_1, v_2, \dots v_k$ is called the span of $\{v_1, v_2, \dots v_k\}$ i.e

$$\operatorname{Span}\{v_1, v_2, \dots v_k\} = \{v \in V \mid v = a_1v_1 + a_2v_2 + \dots a_kv_k\}$$

* Given a set of vectors $S = \{v_1, v_2, \dots v_k\}$ in a vector space V, S is said to span V if $\operatorname{span}(S)=V$.

Example Find span $\{v_1, v_2\}$, where $v_1=(1,2,3)$ and $v_2=(1,0,2)$.

Solution Span $\{v_1, v_2\}$ is the set of all vectors $(x,y,z)\in\mathbb{R}^3$ such that $(x,y,z)=a_1(1,2,3)+a_2(1,0,2)$.

$$\left(\begin{array}{cc|c} 1 & 1 & x \\ 2 & 0 & y \\ 3 & 2 & z \end{array}\right) \qquad \begin{array}{l} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 3R_1 \end{array}$$

$$\left(\begin{array}{cc|c} 1 & 1 & x \\ 0 & -2 & y-2x \\ 0 & -1 & z-3x \end{array}\right) \qquad R_2 \to \frac{-1}{2} R_2$$

$$\left(\begin{array}{cc|c} 1 & 1 & x \\ 0 & 1 & x-\frac{1}{2}y \\ 0 & -1 & z-3x \end{array}\right) \qquad R_3 \to R_3 + R_2$$

$$\left(\begin{array}{cc|c} 1 & 1 & x \\ 0 & 1 & x-\frac{1}{2}y \\ 0 & 0 & z-2x-\frac{1}{2}y \end{array}\right)$$

So solutions when $4x+y-2z=0$. Thus span $\{v_1, v_2\}$ is the plane $4x+y-2z=0$.

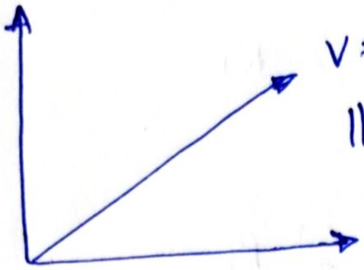
Norms: The norm of a Vector can be any function That maps a vector to a positive Value. Different functions can be used, and we will see a few examples. These functions can be called norms if they are characterized by the following properties: Norms are non-negative values

Two reasons to use norms

1. To estimate how "big" a Vector/tensor is
2. To estimate "how close" one tensor is to another

Example

1)

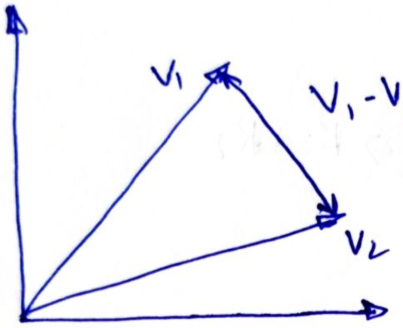


$v = (3,4)$

$\|v\| = \sqrt{3^2+4^2} = 5$

"how-big" Vector/~~Tensor~~

2)



"how close" one tensor is to another.

$\|\Delta v\| = \|V_1 - V_2\|$

$\Downarrow$

Ex: How close is one image to another?

Properties of Norms

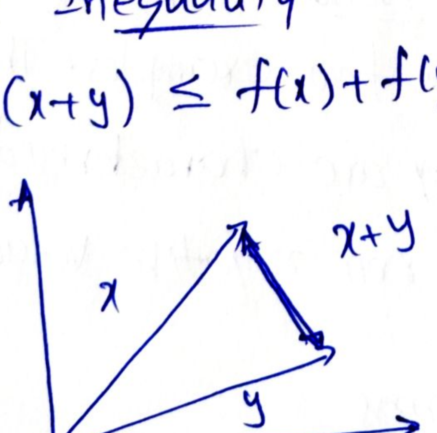
Mathematically, a norm is any function f that satisfies

* $f(x) = 0 \Rightarrow x = 0$

  Any norm should satisfy that is if the vector has length 0, then it must be the 0 vector.

* Triangle Inequality

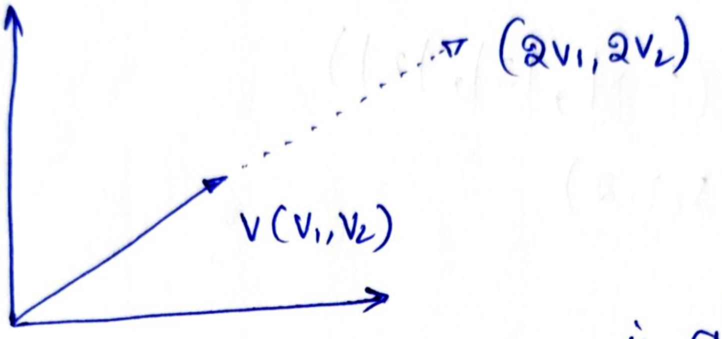
  $$f(x+y) \le f(x) + f(y)$$



* Linearity

  $\forall \alpha \in R$, ~~f(αx)~~ $f(\alpha x) = |\alpha| f(x)$

If I take a vector and simply scale it up, take a string extend it by two times each of the coordinates. will increase by a factor of 2, so let us say if I increase it by a factor of alpha



These Three properties is also called.

- ⇒ Idea of '0' (zero)
- ⇒ Idea of Triangle Inequality
- ⇒ Idea of linearity

__Types__ of __Norms__ (Vector Norms)

1) $\|v\|_1$ (or) Manhatten distance Norm $= |v_1| + |v_2| + \cdots |v_n|$

2) $\|v\|_2$ (or) Euclidean Norm $= (v_1^2 + v_2^2 + v_3^2 + \cdots + v_n^2)^{1/2}$

3) $\|v\|_p$ (or) P-Norm $= (|v_1|^p + |v_2|^p + \cdots + |v_n|^p)^{1/p}$ if $p \geq 1$

4) $\|v\|_\infty$ (or) Infinite Norm (or) max-norm

$$\|v\|_\infty = \max(|v_1|, |v_2|, \ldots, |v_n|)$$

__Example__ $v = (2, 3, -5)$

$\|v\|_1 \Rightarrow$

$|2| + |3| + |-5| = 2+3+5 = 10$

$V_1 = (-5, 3, 2)$

$$\|V\|_2 = \sqrt{(-5)^2 + 3^2 + 2^2}$$

$$= \cong 6.16$$

∞-Norm $= \|v\|_\infty$

$$\max(|-5|, |3|, |2|)$$

$$\max(5, 3, 2)$$

$$= 3$$

Trace of Matrix

* The trace of a matrix is given by the sum of its diagonal elements

$$Tr(A) = \sum_i A_{ii}$$

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 7 & 2 & 6 \\ 8 & 9 & 3 \end{bmatrix} \qquad Tr(A) = 1 + 2 + 3 = 6$$

Properties of Trace

1) Linearity 2) Cyclic Property.

Linearity: If Scalar c, A & B are two same size matrices.

$$tr(c \cdot A) = c \cdot tr(A)$$

$$tr(A+B) = tr(A) + Tr(B)$$

Example

$C = 5 \qquad A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}$

$tr(A) = 7 \qquad tr(B) = 7$

$tr(C.A) = 35 \qquad C.tr(A) = 35$

2. Cyclic Property.

$tr(ABC) = tr(BCA) = tr(CAB)$

Special Matrices and Vectors

* Diagonal Matrix -
* Symmetric matrix
* Unit Vector
* Orthogonal Vectors
* Orthogonal Matrices.

Diagonal Matrix

only diagonal entries are non-zero

$D_{ij} = 0$ if $i \neq j$

Example: $D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Symmetrix matrix

A matrix is equal to its transpose

$A = A^T$

* **Unit Vector**

Vector with unit "length"

$$\|V\|_2 = 1$$

means $\sqrt{v_1^2 + v_2^2 + \dots v_n^2}$

* **Orthogonal Vectors** - Mutually perpendicular.

$$\vec{x} \cdot \vec{y} = 0 \Rightarrow x^T y = 0$$

Example $\begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}$ (or) $\begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}$

* **Orthogonal Matrix**

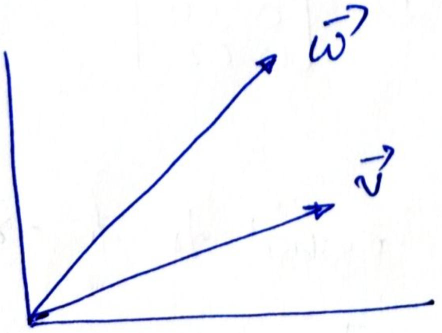
Transpose is equal to inverse.

$$A^T = A^{-1}$$

$$A^T A = A A^T = I$$

Eigen Decomposition

* Extremely useful for square symmetric matrices.
* Every real matrix can be thought of as a combination of rotation and stretching



$$\underset{\substack{n \times n \\ \downarrow \\ \text{matrix}}}{A} \; \underset{\substack{n \times 1 \\ \downarrow \\ \text{vector}}}{\vec{v}} = \underset{\substack{n \times 1 \\ \downarrow \\ \text{vector}}}{\vec{w}}$$

* Eigen vectors for a matrix are those special vectors That only stretch under the action of the matrix.

Eigen vectors → Special vectors.

* Eigen values are the factor by which eigenvectors stretch.

mathematically

$$A v = \underbrace{\lambda v}_{w \to \text{ new vector.}}$$

* $\underset{n \times n}{A}$ has n linearly independent eigenvectors.

$$\{ v^{(1)}, v^{(2)} \dots v^{(n)} \}$$

* Concatinate all the vectors (as columns) and make a eigenvector matrix V.

$$V = [v^{(1)}, v^{(2)} \dots v^{(n)}]$$

* If we concatinate the corresponding eigenvalues into a diagonal matrix

$$\Lambda = \text{diag} (\lambda_1, \lambda_2, \dots \lambda_n)$$

Then the factorization / Eigen value decomposition as

$$A = V \Lambda V^{-1}$$

All matrices can be thought of as rotating and streching vectors.

Eigenvectors have pure stretch.

Questions

Q: Decompose the following matrix using eigen value decomposition.

Ans $A = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$ $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\lambda . I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

$A - \lambda I$

$\Rightarrow \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \Rightarrow \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$

find the determinant of the resultant matrix

$\det \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$

$= (7-\lambda)(-1-\lambda) - 3(3)$

$= -7 - 7\lambda + \lambda + \lambda^2 - 9$

$= \lambda^2 - 6\lambda - 16$

Eigen Values: $= \lambda - 8\lambda + 2\lambda - 16$

$\lambda(\lambda - 8) + 2(\lambda - 8)$

$\Rightarrow (\lambda-8)(\lambda+2) \Rightarrow \lambda = 8$ and $\lambda = -2$

for $\lambda = 8$ we get

$B = \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix} \Rightarrow \begin{bmatrix} 7-8 & 3 \\ 3 & -1-8 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix}$

$B . X = 0$

$\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$-x_1 + 3x_2 = 0$ — ①

$3x_1 - 9x_2 = 0$ — ②

$\boxed{x_1 = 3 \quad x_2 = 1}$

for $\lambda = -2$ we get

$C \begin{bmatrix} 7+2 & 3 \\ 3 & -1+2 \end{bmatrix} = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}$

$C . X = 0$

$\begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\left. \begin{array}{l} 9x_1 + 3x_2 = 0 \\ 3x_1 + x_2 = 0 \end{array} \right\}$ $x_1 = 1 \; x_2 = -3$

finall get eigenvector $\boxed{\begin{array}{l} x_1 = 1 \\ x_2 = -3 \end{array}}$

Questions

* List and Explain Various types of matrices and its operations in linear algebra.
* Differentiate between Machine learning & Deep learning.
* Define the terms Matrix and tensor. Explain how these components are used in Deep learning.
* Define The term norms and give its mathematical intentions. (types). Explain the properties of norms.
* Explain following Valid operations on a matrix
  - -> Broad casting
  - -> Hadamard Product
  - -> Trace

Broad casting

$A = [1\ 2\ 3]$ ~~B = [4 5 6 / 7 8 9 / 10 11 12]~~ B = [[4 5 6] [7 8 9] [10 11 12]]

~~A+B =~~ Result (A+B) : [[5 7 9] [8 10 12] [11 13 15]]

Hadamard Product

$A = [1\ 2\ 3]$ $B = [4\ 5\ 6]$

$(A * B) = [4\ 10\ 18]$

* What is Taxicab distance and its applications.

Sol: It is a practical measure for distance in grid-based systems and is widely used in various applications

* The calculate the taxicab distance between two points in grid based System.

Example A (x_1, y_1) B (x_2, y_2)

$$\text{Taxicab distance} = |x_2 - x_1| + |y_2 - y_1|$$

$$(2, 3) \quad (5, 7)$$

$$= |5-2| + |7-3|$$

$$= |3| + |4|$$

$$= 7$$

Applications are

* Robotics & Autonomous Vehicles.
* Image processing & Computer Vision
* Network Routing
* Game Development
* clustering Algorithm
* Telecommunications etc.

* What are the Special matrices useful for data representation Give an example for each

* Explain the following Concepts with Example
  - → linear Combinations
  - → Span
  - → linear dependent / Independent

* Given a Vector A = [9, -3, 7] find the following norms.
  - → Euclidean norm
  - → 1-norm
  - → Infinity norm.