LINEAR ALGEBRA

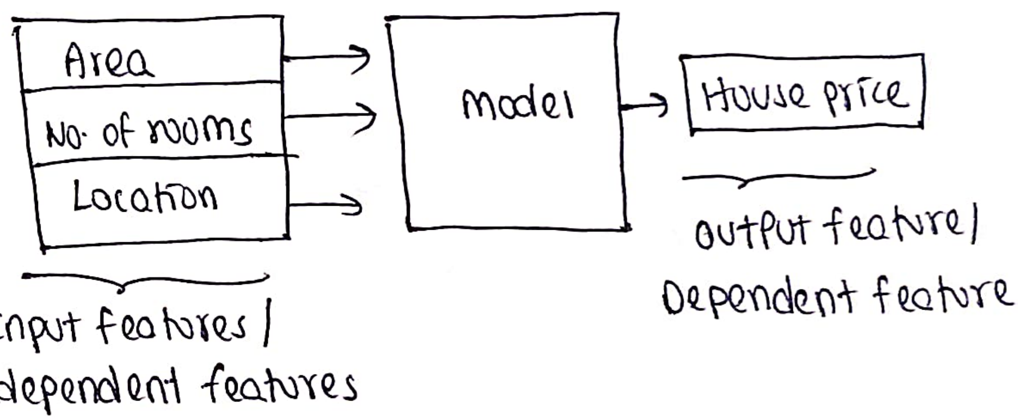
⇒ Linear algebra is a brach of mathematics that focuses on study of vectors, vector spaces, linear transformations and system of linear equations

⇒ Linear algebra provides a framework for understanding the properties of these mathematical objects, which can be represented using matrices and vectors.

Applications of linear algebra:-

① Data representation and manipulation:-

Linear algebra can be used to represent and manipulate the data. For suppose, if we are building the house price prediction model.



The data is represented in the form of vectors.

Area	No. of rooms	Location	Price
1200 sft	3	BLR	45L
1400 sft	5	HYD	50L
1800 sft	4	MAS	51L

This data is represented as [1200, 3, BLR, 45], [1400, 5, HYD, 50] and [1800, 4, MAS, 51] i.e 4-dimensional vectors.

Linear algebra works really well with high dimensional data. we use concepts like dimensionality reductions that help us to do well with high dimensional data.

(ii) Machine learning and Artificial Intelligence:-

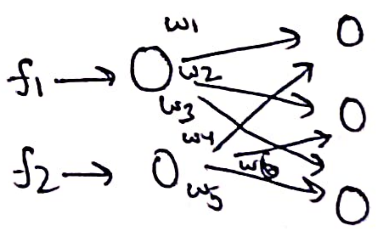
Linear algebra will be used in model training, where we will be performing multiple matrix arithmatic operations and also use linear equations in some algorithms.

(iii) Dimensionality reduction:-

We have algorithms like PCA (Principal component Analysis) that uses concepts like eigen values and eigen vectors to reduce from high dimensional vector to low dimensional vector.

(iv) Neural networks:-

Linear algebra has some applications in forward propagation and backward propogation.



$$\Rightarrow \underbrace{\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \end{bmatrix}}_{\text{matrix multiplication}}$$

we perform arithmatic operations of matrices in neural networks.

Scalars and vectors

→ A scalar is a single numerical value. It represents a magnitude or quantity and has no direction.

Examples:- Speed of a car = 45Kmph
Temperature = 25°c

Applications of scalars in data science:-

Scalar finds application in simple linear regression where we denote the best fit line in the form

$$\boxed{y = mx + c}$$

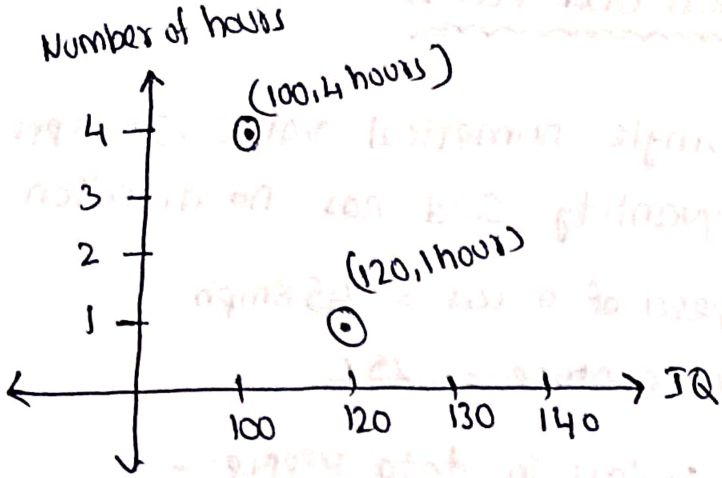
Here ⇒ m = slope of the line
c = y-intercept
} scalars

⇒ A vector is an ordered list of numbers, which can represent a point in space or quantity with both magnitude and direction.

Example:-

Students IQ	Hous of study	Pass/Fail
100	4	Pass
120	1	Fail
130	2	Pass

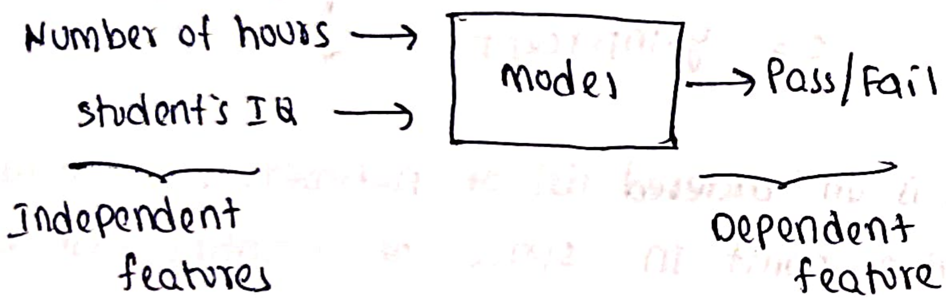
The features like student's IQ and hours of study can be represented as a points in space. Their vectoral denotion will be like (100, 4hours), (120, 1 hour) and (130, 2 hours).

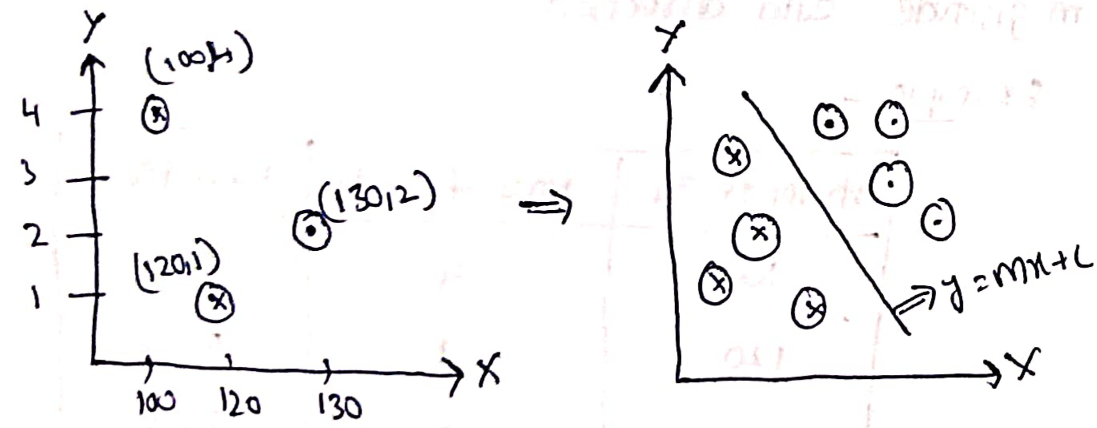


so, with respect to data science, the numericals which can be represented in the space are vectors.

Application of vectors in data science / machine learning:-

For example, if we are creating a classification model,



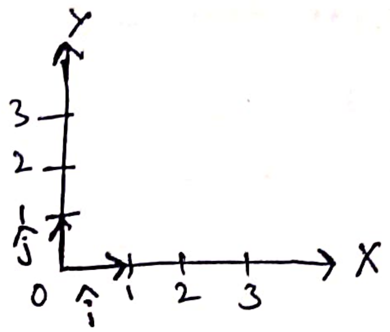


By representing the features in space, we use logistic regression to find a best fit line. The points which lie under the line represent "Fail" and the point which lie above the line represent "Pass".

Representation of vectors in space:-

we use unit vectors to represent points in space.

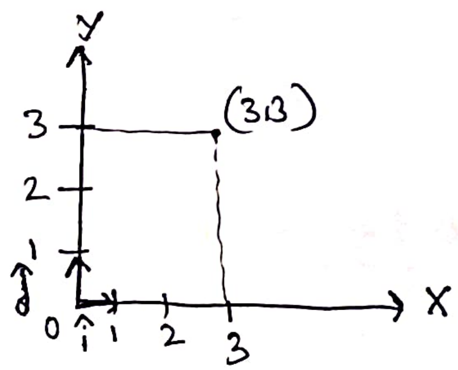
unit vectors have a magnitude of 1.



$\hat{i} \Rightarrow$ magnitude of 1 in X-direction

$\hat{j} \Rightarrow$ magnitude of 1 in Y-direction

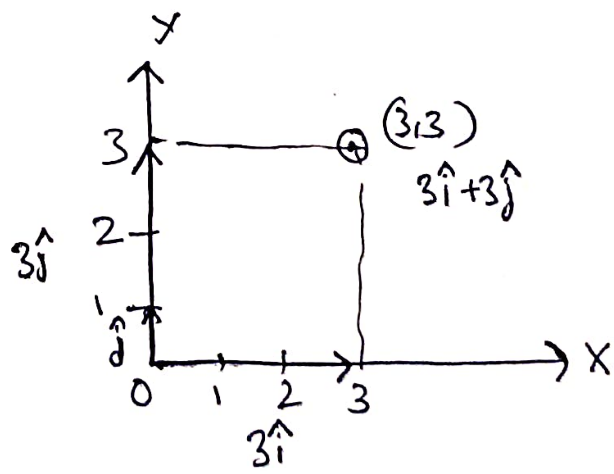
For example, (3,3) is represented as



$x=3 \Rightarrow$ 3 units in the direction of $X = 3\hat{i}$

$y=3 \Rightarrow$ 3 units in the direction of $Y = 3\hat{j}$

(3,3) is represented as $3\hat{i}+3\hat{j}$



we add $3\hat{i}+3\hat{j}$ because we move $3\hat{i}$ (towards right) and then $3\hat{j}$ (towards north).

Addition of two vectors

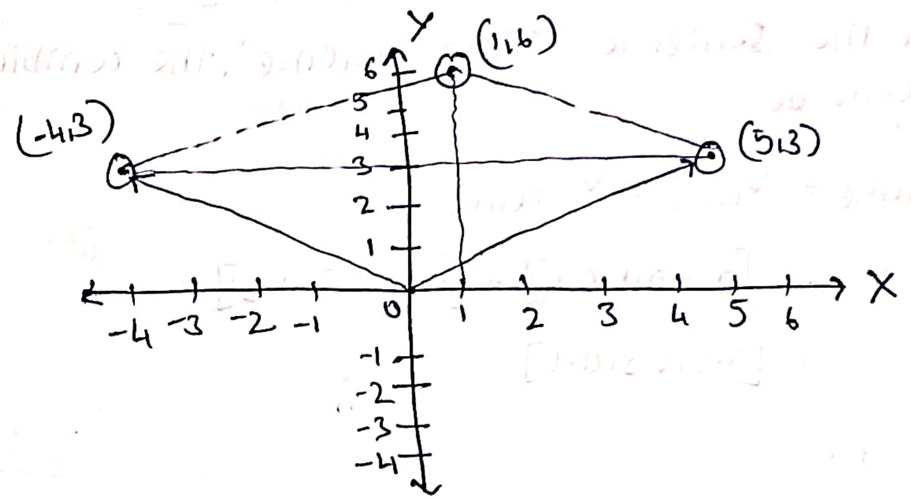
⇒ Suppose we have two vectors which are

$$P_1 = \begin{bmatrix} -4 \\ 3 \end{bmatrix} \text{ and } P_2 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Addition of these two vectors will be:-

$$P_1 + P_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} -4+5 \\ 3+3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

In the coordinate system, it looks like:-



with respect to (-4,3), addition of (5,3) means that the point moves 5 units to the right in X-axis i.e horizontally and 3 units to the top in Y-axis or vertical direction.

with respect to (5,3), addition of (-4,3) means that the points moves 4 units towards the left (due to negative magnitude) and 3 units vertically up in Y-axis.

⇒ similarly for multi-dimensional vectors,

If $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$, then their addition

will be $X + Y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ ---- \\ x_n + y_n \end{bmatrix}$

④ Applications of addition of vectors in Machine Learning:-

(i) Data aggregation

(ii) Feature engineering

(iii) Natural Language Processing

In natural language processing, we convert sentences into multiple vectors called word embeddings.

For example,

If vector of word "Data" is $V_{Data} = [0.2, 0.1, 0.4]$

and vector of word "Science" is $V_{Science} = [0.3, 0.7, 0.2]$

Then for the sentence "Data science", the combined vector will be

$$V_{Data\ science} = V_{Data} + V_{Science}$$

$$= [0.2, 0.1, 0.4] + [0.3, 0.7, 0.2]$$

$$= [0.5, 0.8, 0.6]$$

(iv) Image processing

In image processing, each image is represented of three channels ⇒ Red channel, Blue channel and Green channel.

For example,

Red channel ⇒ $V_R = [255, 128, 0]$

Green channel ⇒ $V_G = [128, 255, 0]$

Blue channel ⇒ $V_B = [64, 64, 255]$

To convert RGB into grey scale,

$$V = \frac{V_R + V_G + V_B}{3} = \frac{1}{3}[V_R + V_G + V_B]$$

$$= \frac{1}{3}[255+128+64, 128+255+64, 255]$$

$$= \frac{1}{3}[447, 447, 255] = [149, 149, 85]$$

Multiplication of vectors

⇒ Threare three types of multiplication of vectors:-

(i) Dot product / Inner product

(ii) Element wise multiplication

(iii) Scalar multiplication

(A) Dot product:-

The dot product of two vectors results in a scalar and is calculated as the sum of products of their corresponding components.

$$A = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

Dot product of vectors A and B will be

$$A \cdot B = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \end{bmatrix} = (2\times4) + (3\times5) = 23$$

Applications of dot product in data science:-

We use dot product in cosine similarity, which is a method to determine how similar two vectors are. It calculates the cosine of the angle between 2 vectors, providing a similarity score that ranges from -1 (dis-similar) to 1 (completely similar).

$$\boxed{\cos\theta = \frac{A \cdot B}{|A| \cdot |B|}}$$

Cosine similarity finds applications in recommendation system where similar recommendations of a movie/product are provided based on what we have already seen or purchased.

An example of cosine similarity:-

Avengers movie = [Action, comedy, drama, horror] = [1,2,0,3,1]

Movie B = [2,0,1,1,1]

If a user has watched movie Avengers, then should we recommend movie B to the user can be predicted using cosine similarity.

The cosine similarity is:-

$$\cos\theta = \frac{A.B}{|A||B|} = \frac{[1\ 2\ 0\ 3\ 1][2\ 0\ 1\ 1\ 1]}{|A| \times |B|}$$

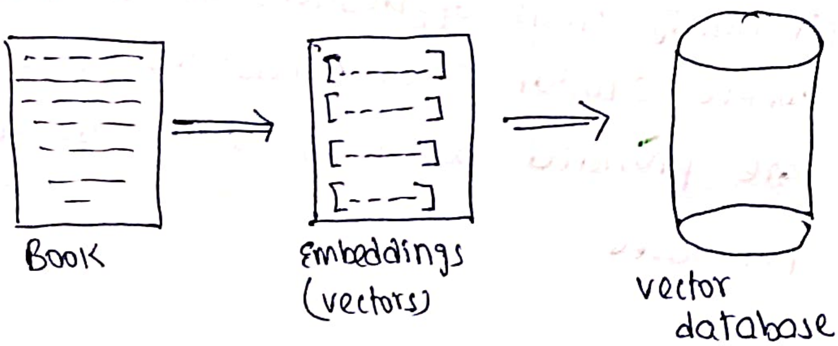
$$= \frac{(1\times2)+(2\times0)+(0\times1)+(3\times1)+(1\times1)}{\sqrt{1^2+2^2+3^2+1^2} \times \sqrt{2^2+1^2+1^2+1^2}}$$

$$= \frac{6}{4\times\sqrt{7}} = \frac{3}{2\sqrt{7}} = 0.586$$

The value of $\cos\theta = 0.586$, is closer to 1, which indicates that genre of movie A has 58.6% similarity to genre of movie B and hence can be recommended.

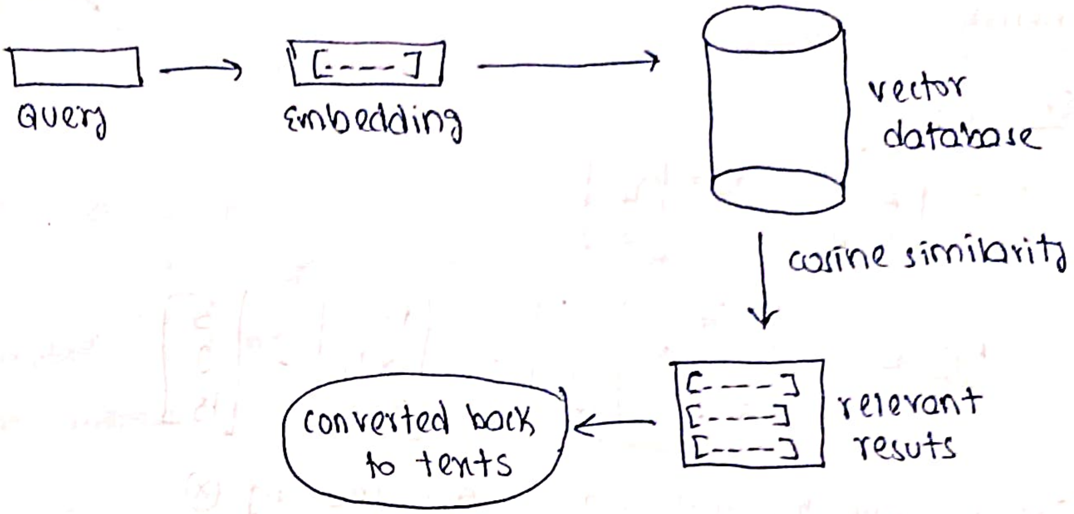
Cosine similarity in RAG:-

In RAG, we store contents of large documents/texts in vector databases. When a query is given, a cosine similarity is applied to obtain similar results.



In the first step, the text is converted into vectors called embeddings which are stored in vector databases.

when user gives a query, that query is converted into a vector and cosine similarity is applied between query vector and vectors present in vector database to obtain relevant results.



Sowmya

Kalyan

Krishna

$1+0=1$
$1+1=2$
$1+2=3$
$1+3=4$
$1+4=5$

5

Sowmya
Kalyan
Krishna

A B C D E F G H I J K L M N
O P Q R S T U V W X Y Z

1+ 2 = 3 4 5 6 7 8 9 10

అ ఆ ఇ ఈ ఉ ఊ
అమ్మ ఇల్లు ఉడత
ఆవు ఈగ ఊయల

A – Apple D – Dog.
B – Ball E – elephant
C – Cat F – Frog

⊛ Element wise multiplication:-

In element wise multiplication, corresponding elements of 2 vectors are multiplied to form a new vector of same dimension.

Example:-

$$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad B = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

$$A \otimes B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \otimes \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 1\times3 \\ 2\times4 \\ 3\times5 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \\ 15 \end{bmatrix}$$

Element wise multiplication is denoted by $\otimes$

⊛ Application of element wise multiplication in data science:-

⇒ Element wise multiplication finds an application in feature engineering.

Example:-

Products	cost	Discount
A	1000	10%
B	500	20%
C	200	15%

To get the discounted price, we can perform element wise multiplication between cost and discount.

Discounted price = cost $\otimes$ Discount

$$= \begin{bmatrix} 1000 \\ 500 \\ 200 \end{bmatrix} \otimes \begin{bmatrix} 10\% \\ 20\% \\ 15\% \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 30 \end{bmatrix}$$

In this way, there are various cases where we can perform multiplication / addition / subtraction element-wise in feature engineering to obtain a new feature.

→ Similarly, in Deep learning, there a various concepts of RNN, LSTM RNN, GRU RNN where there are applications of element-wise multiplication ($\otimes$) and element-wise addition ($\oplus$).

There are concepts like forget gate and input gate where these find applications.

② Scalar multiplication:-

It involves multiplication of vector by a scalar, resulting in a vector where each component is scaled by the vector.

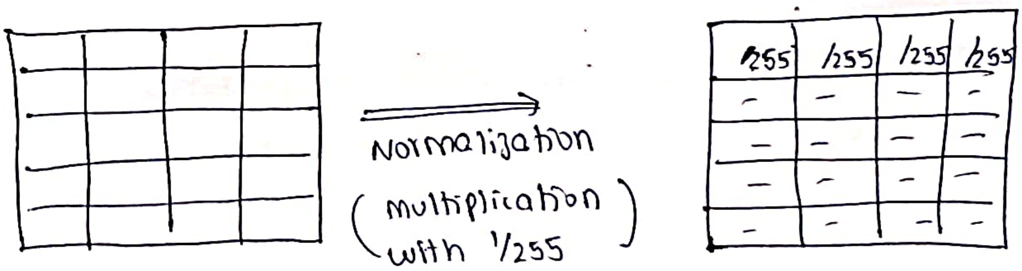
$$A = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} \qquad c = 4$$

$$cA = 4 \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 4\times3 \\ 4\times5 \\ 4\times7 \end{bmatrix} = \begin{bmatrix} 12 \\ 20 \\ 28 \end{bmatrix}$$

Applications of Scalar Multiplication:-

① Normalisation in image processing

An image is represented by pixel values in grey scale and their values will be in range 0 to 255. when the values of pixels are huge, for simpler computation, we perform normalization. where each pixel value is divided by 255.



Neural networks work well when input values are in similar scale, the raw pixel values (0-255) can cause large gradients. so, normalization reduces this issue.

(ii) Data conversion:

For example, if there is a vector of heights (in meters), we can convert it into centimeters by performing scalar multiplication with 100.

$$\text{Height (in m)} = \begin{bmatrix} 1.60 \\ 1.70 \\ 1.83 \end{bmatrix}$$

$$\text{Height (in cm)} = 100 \times \text{Height (in m)} = 100 \begin{bmatrix} 1.60 \\ 1.70 \\ 1.83 \end{bmatrix}$$

$$= \begin{bmatrix} 160 \\ 170 \\ 183 \end{bmatrix}$$

So in data conversions / transformations, we can perform scalar multiplication.

$1 \times 1 = 1$

$1 \times 2 = 2$

$1 \times 3 = 3$

Matrices

⇒ A matrix is a rectangular array of numbers / symbols / expressions assigned / arranged in rows and columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

a_{ij} ⇒ Element of i^{th} row and j^{th} column

Applications of matrices in data science:-

⇒ Matrices are used in data representation

	Math score	Physics score	Biology score
student 1 →	[55	65	75]
student 2 →	[65	60	55]
student 3 →	[70	45	80]

The same can be represented in matrix as

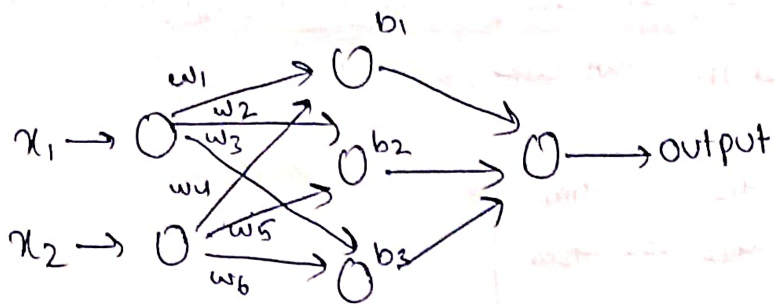
$$\begin{array}{cl} \begin{bmatrix} 55 & 65 & 75 \\ 65 & 60 & 55 \\ 70 & 45 & 80 \end{bmatrix} & \begin{array}{l} \leftarrow \text{student 1} \\ \leftarrow \text{student 2} \\ \leftarrow \text{student 3} \end{array} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \text{Math} & \text{Physics} & \text{Biology} \end{array} & \end{array}$$

⇒ confusion matrix in machine learning
It is used to calculate accuracy of the model.

$$\text{confusion matrix} = \begin{bmatrix} \text{True positive} & \text{False negative} \\ \text{False positive} & \text{True negative} \end{bmatrix}$$

Confusion matrix is a 2x2 matrix that helps us find metric scores such as accuracy, precision and recall to understand performance of model.

⇒ In neural networks, there is a technique call forward propogation where matrix operations are performed



$$z = w^T x + b$$

where $w = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \end{bmatrix}$

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

$$\boxed{z = \begin{bmatrix} w_1 & w_4 \\ w_2 & w_5 \\ w_3 & w_6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}$$

Involves matrix multiplication and addition operations

Matrix operations:-

Matrix operations are used to manipulate and analyze multidimensional data efficiently.

→ matrix addition and subtraction
→ scalar matrix multiplication
→ matrix multiplication

Matrix addition and subtraction:-

Addition or subtraction of corresponding elements of 2 matrices can be done, only if they have same dimensions.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}_{3\times3} \qquad B = \begin{bmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \\ 1 & 2 & 3 \end{bmatrix}_{3\times3}$$

Since both A and B have same dimensions, they can be added to result in a matrix with same dimesions

$$A+B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}_{3\times3} + \begin{bmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \\ 1 & 2 & 3 \end{bmatrix}_{3\times3}$$

$$= \begin{bmatrix} 1+4 & 2+5 & 3+6 \\ 4+7 & 5+8 & 6+9 \\ 7+1 & 8+2 & 9+3 \end{bmatrix} = \begin{bmatrix} 5 & 7 & 9 \\ 11 & 13 & 15 \\ 8 & 10 & 12 \end{bmatrix}_{3\times3}$$

Scalar multiplication:-

Scalar multiplication involves multiplying every element of a matrix by a scalar value

If $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$ and $c = 4$, then

$$cA = 4 \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 2\times4 & 3\times4 \\ 4\times4 & 5\times4 \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 16 & 20 \end{bmatrix}$$

Suppose we have a matrix representing product prices in dollars and we want to adjust it to the inflation by a factor of 1.05,

$$\text{Original prices} = \begin{bmatrix} 10 & 20 & 30 \\ 15 & 25 & 35 \\ 20 & 30 & 40 \end{bmatrix}$$

Factor = 1.05

Adjusted prices = Factor x original prices

$$= 1.05 \begin{bmatrix} 10 & 20 & 30 \\ 15 & 25 & 35 \\ 20 & 30 & 40 \end{bmatrix} = \begin{bmatrix} 10.5 & 21.00 & 31.50 \\ 15.75 & 26.25 & 36.75 \\ 21 & 31.5 & 42 \end{bmatrix}$$

$$\text{Adjusted prices} = \begin{bmatrix} 10.5 & 21 & 31.50 \\ 15.75 & 26.25 & 36.75 \\ 21 & 31.5 & 42 \end{bmatrix}$$

In this way, if we want to scale matrix by a factor, we can perform scalar multiplication to get adjusted results.

④ Matrix multiplication:-

It involves the dot product of rows of the first matrix with the columns of the second matrix.

For two matrices $A(m \times n)$ and $B(n \times p)$, the result of matrix multiplication is $C(m \times p)$

Example:-

$$A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}_{1\times 3} \quad B = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}_{3\times 1}$$

$$A \times B = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = (1\times 2) + (2\times 3) + (3\times 4)$$

$$= 2+6+12 = 20$$

Another example:-

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{2\times3} \qquad B = \begin{bmatrix} 7 & 9 & 11 \\ 8 & 10 & 12 \end{bmatrix}_{2\times3}$$

A and B cannot be multiplied because number of columns in A is not equal to number of rows in B.

Transposing B, $B^T = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}_{3\times2}$

And now, A and B^T can be multiplied becouse number of columns in matrix A (3) and number of rows in matrix B (3) are equal.

$$A \times B^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{2\times3} \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}_{3\times2}$$

$$= \begin{bmatrix} (1\times7)+(2\times9)+(3\times11) & (1\times8)+(2\times10)+(3\times12) \\ (4\times7)+(5\times9)+(6\times11) & (4\times8)+(5\times10)+(6\times12) \end{bmatrix}$$

$$= \begin{bmatrix} 7+18+33 & 8+20\times36 \\ 28+45+66 & 32+50+72 \end{bmatrix} = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$$

Matrix multiplscation is one of the most important matrix operations used in deep learning, where we deal with multiple ~~rows and~~ weights and biases.

Functions

⇒ A function is a mathematical relationship that uniquely associates elements of one set (called domain) with elements of another set (called co-domain). In simpler terms, a function maps input to outputs in a specific way.

Notation of a function:-

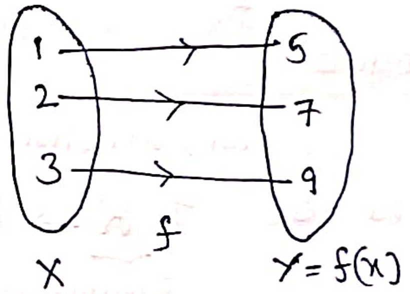
A function mapping elements from set X (domain) to set Y (codomain) is denoted by $f: X \to Y$.

If x is an element of set X, then $f(x)$ is the corresponding element of set Y.

Example of function:-

$f(x) = 2x+3$ ⇒ This function maps real number x to a real number $f(x) = 2x+3$

$$\left.\begin{aligned} f(1) &= 2(1)+3 = 5 \\ f(2) &= 2(2)+3 = 7 \\ f(3) &= 2(3)+3 = 9 \end{aligned}\right\}$$



Functions find applications in dimensionality reduction, where transformation functions are used.

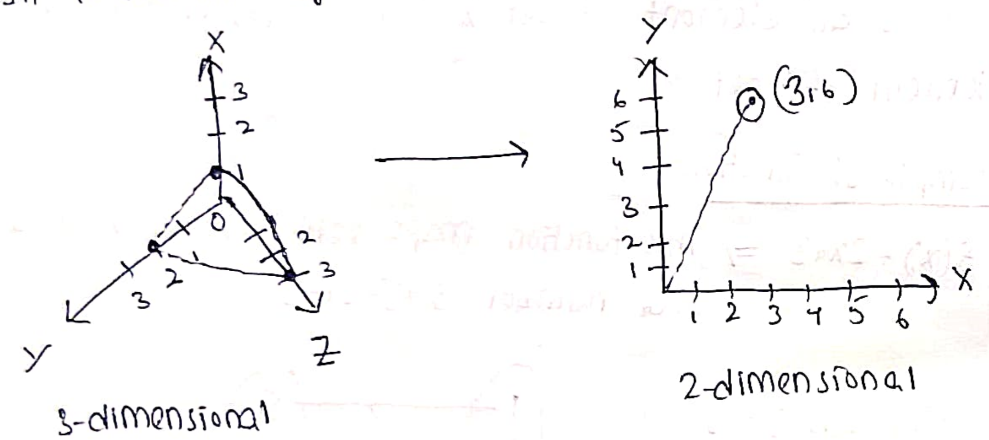
vector transformation

⇒ These transformations are generally used in dimensionality reduction. For example, let us convert a 3-dimensional vector into a 2-dimensional vector,

$$f(x,y,z) = (x+y, 2z)$$

$$f\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

In co-ordinate system,



Since this is a transformation function, it can also be represented as $\underbrace{T: R^3}_{\text{3-dimensions}} \rightarrow \underbrace{R^2}_{\text{2-dimensions}}$

⇒ Vector transformations refer to operations that map vectors from one space to another, often changing their magnitude, direction or both. These transformations are typically described using matrices and are fundamental in many fields like engineering.

⊛ Example of transformations:-

① **Scaling:-**

Scaling is a transformation that changes the magnitude of vector while keeping their direction same.

$$v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$v' = 2v = 2\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}2\times 2\\3\times 2\end{bmatrix} = \begin{bmatrix}4\\6\end{bmatrix}$$

Scaling transformations are generally used in:-

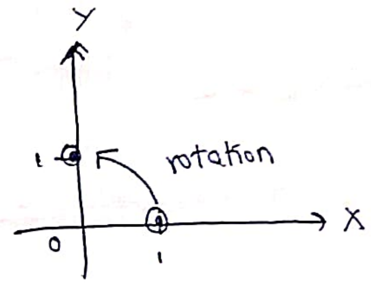
(i) Data normalization

(ii) Computer graphics to re-size objects.

(ii) Rotation:-

Rotation is a transformation function that turns vectors around the origin.

$$v = \begin{bmatrix}1\\0\end{bmatrix}$$



Then the rotation,

$$v' = \begin{bmatrix}0\\1\end{bmatrix}$$

Rotation transformations are used in:-

(i) Image processing to rotate the image.

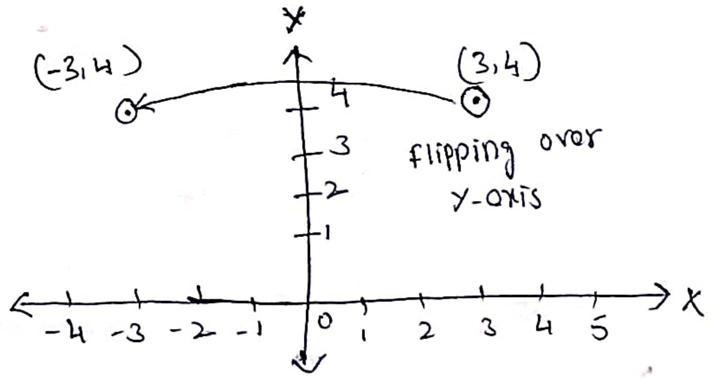
(ii) Robotics to adjust robot orientation

(iii) 3-d graphics to rotate objects.

(iii) Reflection:-

It is a transformation that flips vectors over a specified axis or plane.

$$v = \begin{bmatrix}3\\4\end{bmatrix}$$



Reflection is used in image processing to create mirror images.

Linear transformation

⇒ A linear transformation is a function between two vector spaces that preserve the operations of vector addition and scalar multiplication. This means that T is a linear transformation from a vector space V to vector space W, it is denoted by,

$$\boxed{T: V \rightarrow W}$$

Theare are two important properties of linear transformation,

(i) Additivity. If there are two vector spaces U and V, then

$$\boxed{T(U+V) = T(U) + T(V)}$$

(ii) Homogenity. For a vector space V and a scalar c,

$$\boxed{T(cV) = cT(V)}$$

⇒ Checking if reflection is a linear transformation or not.

If $X = \begin{bmatrix} x \\ y \end{bmatrix}$ and T is a reflection transformation across the y-axis, then $T(x) = \begin{bmatrix} -x \\ y \end{bmatrix}$

$$T(X) = \begin{bmatrix} -x \\ y \end{bmatrix} = [x \quad y] \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = XA$$

so, we can represent T(X) as $T(X) = AX$, where $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

checking the first property of linear transformation.

(i) Additivity

Let $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$ and $V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$ be two vector spaces

Then as per property of linear transformation, vector addition of two vector spaces must be preserved

$$\boxed{T(U+V) = T(U) + T(V)}$$

checking LHS,

$$U+V = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} U_1+V_1 \\ U_2+V_2 \end{bmatrix}$$

$$T(U+V) = A(U+V) = A\begin{bmatrix} U_1+V_1 \\ U_2+V_2 \end{bmatrix}$$

checking RHS,

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \Rightarrow T(U) = A \cdot U = A\begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \Rightarrow T(V) = A \cdot V = A\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

$$T(U)+T(V) = A\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + A\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

$$= A\left(\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}\right) = A\begin{bmatrix} U_1+V_1 \\ U_2+V_2 \end{bmatrix}$$

Since both LHS and RHS are same, reflection satisfies the vector addition property.

(ii) checking homogenity

Let $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$ and scalar $C = 4$

As per homogenity property of linear transformation,

$$T(CU) = CT(U)$$

checking LHS,

$$CU = 4\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 4U_1 \\ 4U_2 \end{bmatrix}$$

$$T(CU) = A \cdot CU = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 4U_1 \\ 4U_2 \end{bmatrix} = \begin{bmatrix} -4U_1 \\ 4U_2 \end{bmatrix}$$

checking RHS.

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \Rightarrow T(U) = A \cdot U = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} -U_1 \\ U_2 \end{bmatrix}$$

$$CT(U) = C \cdot \begin{bmatrix} -U_1 \\ U_2 \end{bmatrix} = 4\begin{bmatrix} -U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} -4U_1 \\ U_2 \end{bmatrix}$$

Since both LHS and RHS are equal, we can say that the reflection transformation satisfies property of homogenity too.

Reflection transformation satisfies both properties of linear transformation, so it can be considered as a linear transformation.

⇒ An example that doesn't follow linear transformation

$$T(x) = x + b = x + \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

This is a transformation, $T: R^2 \to R^2$

① Checking addition of vectors property.

As per linear transformation properties, addition of vectors should be preserved.

$$T(U) + T(V) = T(U+V)$$

Checking LHS,

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \Rightarrow T(U) = U + b = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} U_1+1 \\ U_2+1 \end{bmatrix}$$

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \Rightarrow T(V) = V + b = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} V_1+1 \\ V_2+1 \end{bmatrix}$$

$$T(U) + T(V) = \begin{bmatrix} U_1+1 \\ U_2+1 \end{bmatrix} + \begin{bmatrix} V_1+1 \\ V_2+1 \end{bmatrix} = \begin{bmatrix} U_1+V_1+2 \\ U_2+V_2+2 \end{bmatrix}$$

Checking RHS,

$$U + V = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} U_1+V_1 \\ U_2+V_2 \end{bmatrix}$$

$$T(U+V) = (U+V) + b = \begin{bmatrix} U_1+V_1 \\ U_2+V_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} U_1+V_1+1 \\ U_2+V_2+1 \end{bmatrix}$$

Since LHS and RHS are not equal

This doesn't satisfy vector addition property required to be a linear transformation. So, $T(x) = x + b$ is not a linear transformation.

(ii) checking homogenity.

As per property of homogenity of linear transformation.

$$cT(u) = T(cu)$$

checking LHS.

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \Rightarrow T(u) = u+b = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} u_1+1 \\ u_2+1 \end{bmatrix}$$

$$cT(u) = c\begin{bmatrix} u_1+1 \\ u_2+1 \end{bmatrix} = \begin{bmatrix} cu_1+c \\ cu_2+c \end{bmatrix}$$

checking RHS,

$$cu = c\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}$$

$$T(cu) = cu+b = \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} cu_1+1 \\ cu_2+1 \end{bmatrix}$$

LHS = RHS only when $c=1$ but doesn't hold when c takes any other value. So, this doesn't satisfy homogenity property too.

Since $T(x) = x+b$ doesn't satisfy properties of linear transformation, T is not a linear transformation.

Applications of linear transformation

Linear transformations are fundamental in data science for several reasons. They provide mathematical framework for manipulating and analyzing data, which is crucial for various data processing tasks, model building and interpretation.

(A) Dimensionality reduction:-

→ Principal component analysis is widely used for reducing the dimensionality of datasets while retaining as much variance as possible.

→ PCA involves finding a set of orthogonal axes (principal components) and projecting the data onto these axes.

The transformation of data points in original space to the new space is defined by the principal components is a linear transformation which helps in:-

(i) Reducing computational cost.

(ii) Mitigating the curse of dimensionality

(iii) Visualizing high-dimensional data.

(B) Feature engineering:-

Linear transformations can be used to create new features from the existing ones. Interactions b/w features can be captured through linear combinations, which can then be used in machine learning models to improve predictive performance. Techniques like linear regression, ridge regression and linear discriminant analysis (LDA), all rely on linear transformations to find meaningful feature representations.

(C) Data preprocessing:-

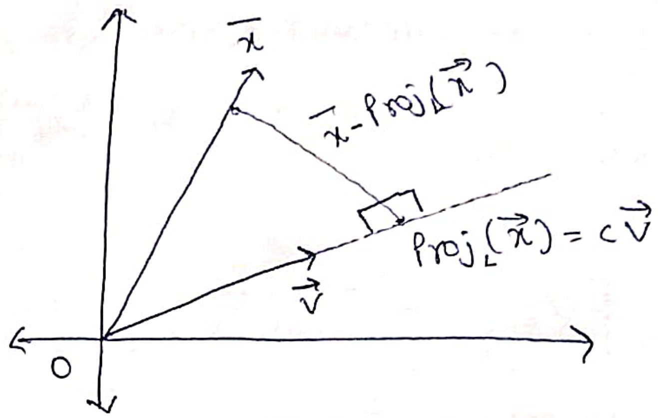
Linear transformations can be used to scale data, making it suitable for machine learning models.

standardization transforms data to have a mean of zero and a standard deviation of one. while normalization scales data to a specific range. These transformations are essential for ensuring that all features contribute equally to the model, especially in algorithms like gradient descent.

(4) Neural networks

In neural networks, especially deep learning models, the layers consist of linear transformations followed by non-linear activation functions. The weights in neural networks can be seen as a series of linear transformations that map input data to intermediate layers and eventually to the output layer. This linear capability is important for the network's ability to learn complex patterns in data.

Introduction to projections



$\bar{x} - Proj_L(\vec{x})$ and $Proj_L(\vec{x})$ are perpendiculer to each other. when two vectors are prependicular, their dot product is 0.

$$(\vec{x} - Proj_L(\vec{x})) \cdot Proj_L(\vec{x}) = 0$$

$$\Rightarrow (\vec{x} - c\vec{v}) \cdot c\vec{v} = 0$$

$$\Rightarrow \vec{x} \cdot \vec{v} - c\vec{v} \cdot \vec{v} = 0$$

$$\Rightarrow c\vec{v} \cdot \vec{v} = \vec{x} \cdot \vec{v}$$

$$\boxed{c = \frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}}$$

so, the projection of vector $\vec{x}$ on extension of $\vec{v}$ is

$$Proj_L(\vec{x}) = c\vec{v} = \frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v}$$

$$\boxed{Proj_L(\vec{x}) = \frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v}}$$

Inverse function

→ An inverse of a function is a function that reverses the effect of the original function.

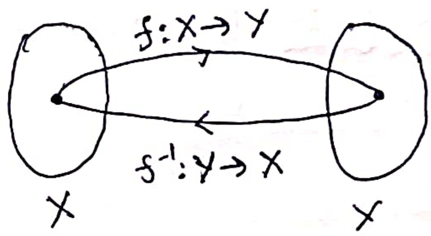
If you have a function f that maps an element x from set X to an element y in a set Y, the inverse function f^{-1} maps y back to x

$$\boxed{f : X \to Y, \text{ then } f^{-1} : Y \to X}$$

⇒ The inverse functions satisfy the following conditions

(i) For all $x \in X$, $f(f^{-1}(y)) = y$

(ii) For all $y \in Y$, $f^{-1}(f(x)) = x$



If we get some value after applying the function, then applying inverse function on that value will bring back the original value.

Ⓐ Identity function:-

Identity function maps every element x in X to itself.

$$\boxed{I : X \to X}$$

Properties of identity function:-

(i) Preservation of elements

(ii) Linearity i.e identity function is a linear transformation.

(iii) Identity matrix

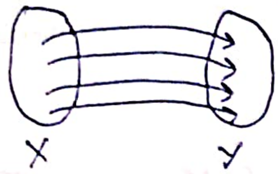
(iv) Having same inverse

⇒ A function only has an inverse if it is bijective.

Bijective follows two properties:-

① Injective (one to one)

Different elements in domain map to different elements in co-domain



(ii) Surjective (onto)

Every element in the co-domain is the image of atleast one element in the domain

② Example of finding inverse of a function:-

Suppose, $f(x) = 2x+3$

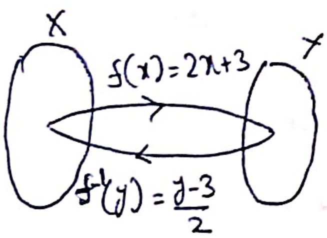
Let $f(x) = y \Rightarrow y = 2x+3$

$\Rightarrow y-3 = 2x$

$$\boxed{x = \frac{y-3}{2}}$$

$f(x) = y \Rightarrow f^{-1}(y) = x = \frac{y-3}{2}$

$$\boxed{f^{-1}(y) = \frac{y-3}{2}}$$



$f(f^{-1}(y)) = f\left(\frac{y-3}{2}\right) = 2\left(\frac{y-3}{2}\right)+3 = y$

$f^{-1}(f(x)) = f^{-1}(2x+3) = \frac{(2x+3)-3}{2} = x$

Both properties of inverse functions are satisfied

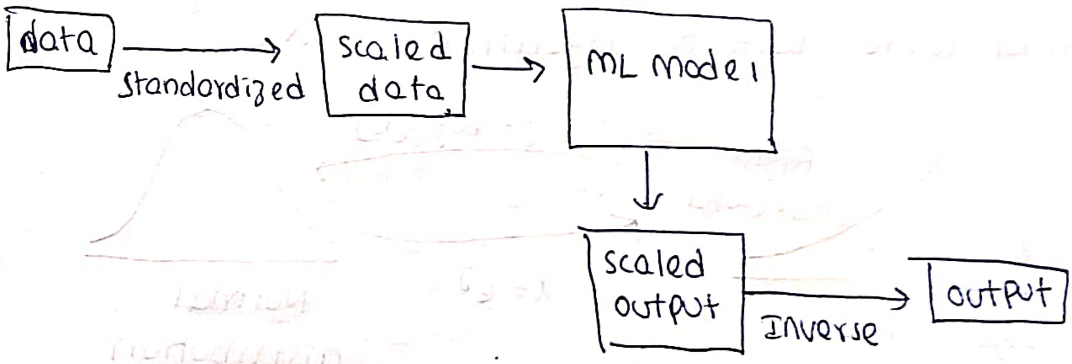
(A) Application of inverse functions:-

In standardization, we convert data to get it in a range for efficient computation before giving it to model. once model provides the response, it must be inversed to get the actual output.

For example,

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \xrightarrow{f(x) = \frac{x-\mu}{\sigma}} \begin{bmatrix} -1.5/2 \\ -0.5/2 \\ +0.5/2 \\ +1.5/2 \end{bmatrix}$$

$\Rightarrow$ This data will be sent to the model

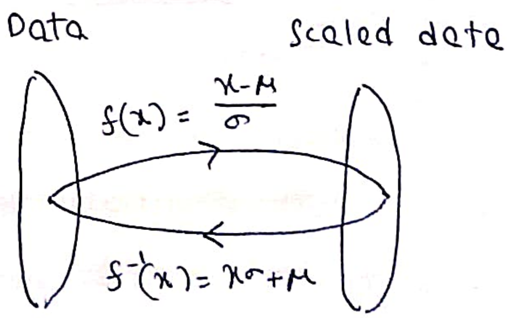


$$f(x) = y = \frac{x-\mu}{\sigma}$$

$$\Rightarrow x = y\sigma + \mu$$

$$\boxed{f^{-1}(y) = \sigma y + \mu}$$

This inverse function helps us to get actual data from scaled data



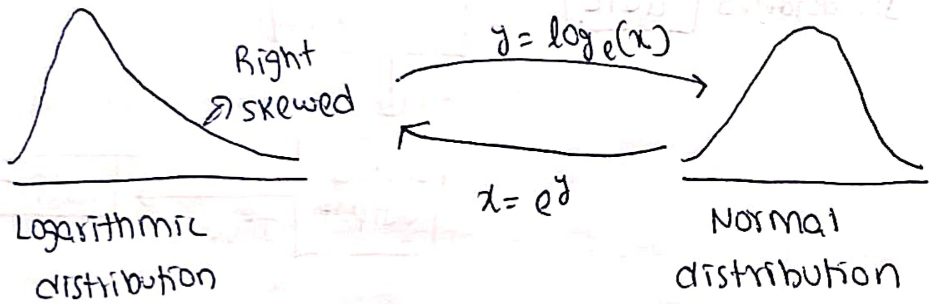
After training a machine learning model, on standardized data, the predictions are often re-scaled back to original scale to interpret the results in a meaningful way.

⇒ Similarly, when we perform feature scaling with min max normalization,

Original ⇒ $y = \dfrac{x - \min(x)}{\max(x) - \min(x)}$

Inverse ⇒ $x = y(\max(x) - \min(x)) + \min(x)$

⇒ And sometimes, we convert logarithmic curve to normal curve for analysis. Inverse function helps us to convert normal curve back to logarithmic curve



In financial data analysis, income or sales data often exhibit skewness. Applying a log transformation can stabilize the variance and make patterns more visible. After model prediction, the inverse log transformation is applied to interpret results at the original scale.

Finding inverse of a matrix

⇒ Inverse of a 2-dimensional matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then inverse of A will be

$$\boxed{A^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \frac{1}{Det(A)}}$$

Finding determinant of a 2-dimensional matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$, Then

Determinant of $A = Det(A) = ad - bc$

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Ⓐ **Determinant:-**

→ The determinant is a scalar value that can be computed from a square matrix. It provides important information about the matrix such as whether the matrix is invertible (has an inverse), and also has the geometrical interpretations, such as describing the scaling factor of linear transformations represented by matrix.

→ For example, if determinant of a matrix is zero, it means that the matrix is not invertible.

so finding determinant of a matrix is a crucial step in finding inverse of a matrix.

⇒ An example of finding inverse of a matrix

$$A = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}$$

Finding determinant of A

$$Det(A) = 4(6) - 2(7) = 24 - 14 = 10$$

$$A^{-1} = \frac{1}{Det(A)} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 0.60 & -0.70 \\ -0.20 & 0.40 \end{bmatrix}$$

verifying whether A^{-1} is right or not,

Let us take matrix $X = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$y = AX = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4+7 \\ 2+6 \end{bmatrix} = \begin{bmatrix} 11 \\ 8 \end{bmatrix}$$

Recovering back X using A^{-1},

$$X = A^{-1}y = \cancel{\frac{1}{10}} \begin{bmatrix} 0.60 & -0.70 \\ -0.20 & 0.40 \end{bmatrix} \begin{bmatrix} 11 \\ 8 \end{bmatrix}$$

$$= \cancel{1} \begin{bmatrix} (0.60)(11) - 0.70(8) \\ (-0.20)(11) + (0.40)(8) \end{bmatrix} = \begin{bmatrix} 6.6 - 5.6 \\ -0.2 + 3.2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

So, we have successfully recovered back the X. Hence, the A^{-1} we have found out is correct.

Eigen values and Eigen vectors

⇒ Eigen values and eigen vectors are fundamental concepts of linear algebra that have numerous applications in various fields such as physics, computer science and data science. They provide insights into properties of linear transformations represented by matrices.

⇒ Eigen value:-

Eigen value is a scalar, that indicates how much an eigen vector is stretched or compressed during linear transformation.

Eigen vector:-

A non-zero vector that only changes in scale and not in direction when a linear transformation is applied.

$$\boxed{Av = \lambda v}$$

For a square matrix A, an eigen vector and its corresponding eigen value λ satisfy the above equation.

⇒ Example:-

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

(i) Finding eigen values

Determinant $(A - \lambda I) = 0$

$$\Rightarrow \left| \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

$$\Rightarrow \text{Det}\left(\begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} \right) = 0$$

$$\Rightarrow (4-\lambda)(3-\lambda) - 2 = 0$$

$$\Rightarrow 12 - 7\lambda + \lambda^2 - 2 = 0$$

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

$$\boxed{\lambda = 2 \text{ or } \lambda = 5}$$

The eigen values of matrix A are $\lambda_1 = 5$ and $\lambda_2 = 2$

(ii) Find the eigen vectors

$$\boxed{(A - \lambda I)v = 0}$$

For $\lambda = 5$.

$$A - 5I = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

Let $v = \begin{bmatrix} x \\ y \end{bmatrix}$, then

$$(A - \lambda I)v = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$$\Rightarrow -x + y + 2x - 2y = 0$$

$$\boxed{x = y}$$

Similarly, for $\lambda = 2$

$$A - 2I = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

$$(A - \lambda I)v = 0 \Rightarrow \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$$\Rightarrow 2x + y + 2x + y = 0$$

$$\boxed{y = -2x}$$

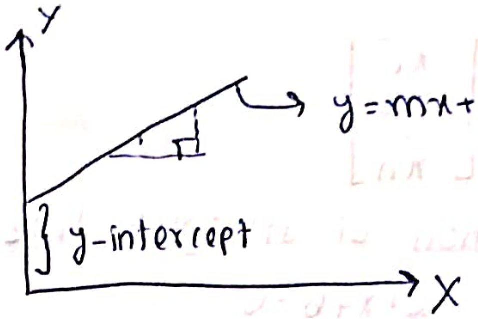
For $\lambda_1 = 5 \Rightarrow v_1 = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

For $\lambda_2 = 2 \Rightarrow v_2 = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

These eigenvalues and eigenvectors describe how the matrix A scales and rotates vectors in its transformation. Eigenvalues indicate factor by which eigenvectors are stretched or compressed and eigenvectors provide the directions in which this stretching or compression occurs

Equation of line, plane and Hyperplane

→ Equation of line in 2-dimensional plane



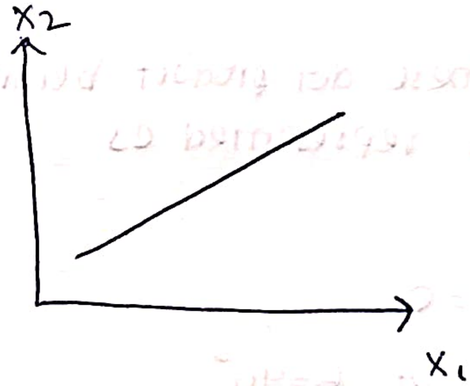
$y = \beta_0 + \beta_1 x$
$y = c + mx$
$ax + by + c = 0$
} Different equations of straight line

$ax + by + c = 0$

$\Rightarrow y = \left(\frac{-ax - c}{b}\right)$

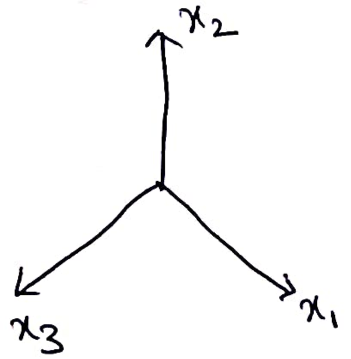
$\Rightarrow y = \left(\frac{-a}{b}\right)x - \frac{c}{b}$

So, when we represent straight line as $ax+by+c=0$, the slope will be $m = \left(\frac{-a}{b}\right)$ and intercept will be $\frac{-c}{b}$ when equated to $y = mx + c$



This line can be represented as $w_1x_1 + w_2x_2 + b = 0$

⇒ Similarly, for a line in 3-dimensional plane



The line can be represented as

$w_1x_1 + w_2x_2 + w_3x_3 + b = 0$

$\Rightarrow w^T x + b = 0$

where, $w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

(⇒ For n-dimensional plane,

$$w_1x_1 + w_2x_2 + \ldots + w_nx_n + b = 0$$

$$w^T x + b = 0$$

where $w = \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix}$

So, the standard representation of straight line will always be in format $w^T x + b = 0$.

⇒ when a line passes through origin, the origin point has all zeroes, so

$$w_1x_1 + w_2x_2 + \ldots w_nx_n + b = 0$$

$$\Rightarrow w_1(0) + w_2(0) + \ldots + w_n(0) + b = 0$$

$$\Rightarrow b = 0$$

So, if line passes through orgin, the equation becomes $\boxed{w^T x = 0}$

If we look at cases where dot product becomes zero, a dot product is represented as

$$A \cdot B = |A||B| \cos\theta$$

$$w^T x = |w||x| \cos\theta = 0$$

This will happen only at $\theta = 90°$

So, if the line passes through orgin, it is always perpendicular to the plane.