INNOMATICS RESEARCH LABS

Deep Learning

pre-requisites

Math for Data Science

- Math's exposure – Yes
  - Some amount of probability
  - Some amount of calculus is required
- PHD or MS in Stats – NO.
- Operation level data scientists – Not much , just basic understanding
- Others involved in designing new algorithms need to have solid foundation in maths and statistics
- Most of the people (90%) are operational level, who need to understand how the algorithms work.

Line concept

- A line is a mathematical object, which can be used for multiple purposes
  - One purpose is to find the shortest distance between two points in a mathematical space
  - It also explains the relationship between two variables, if they influence each other , this is also represented in the form of a line.
- Frame of Reference:
  - The mathematical space or feature space.
- Line practical understanding

Line

Every straight line follow the below eq:

$$o/p^{2} = 3$$

$$y = \boxed{m \quad x} + c \rightarrow \text{intercept}$$

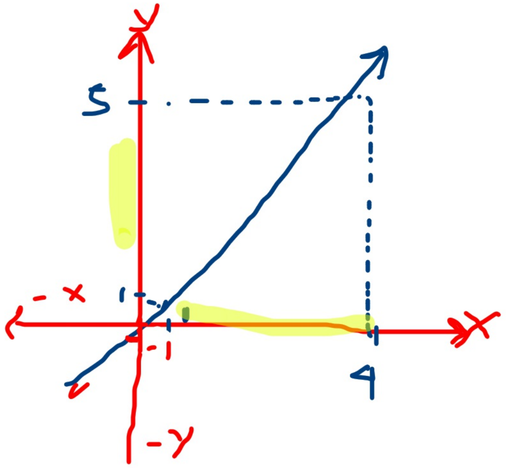
y → Predictable Value

x → i/p value

m → Slope ↳ $\frac{dy}{dx}$ / $\frac{\Delta dy}{\Delta dx}$

$y = c$ if $m/x = 0$

$y = x + c$ if $m = 1$



$A = (1, 1)$ (x, y)

$B = (4, 5)$

$$m = \frac{5-1}{4-1} = \frac{4}{3} = 1.3$$

$$C = (0, -1) = -1$$

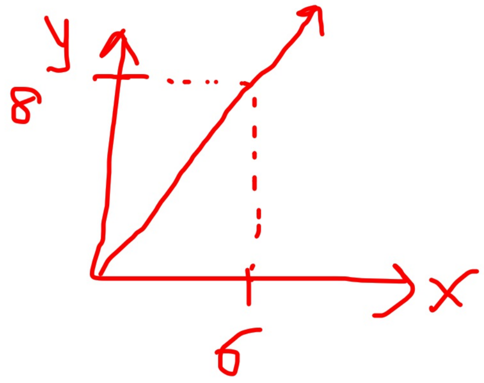
$$y = 1.3x + (-1)$$

Vector

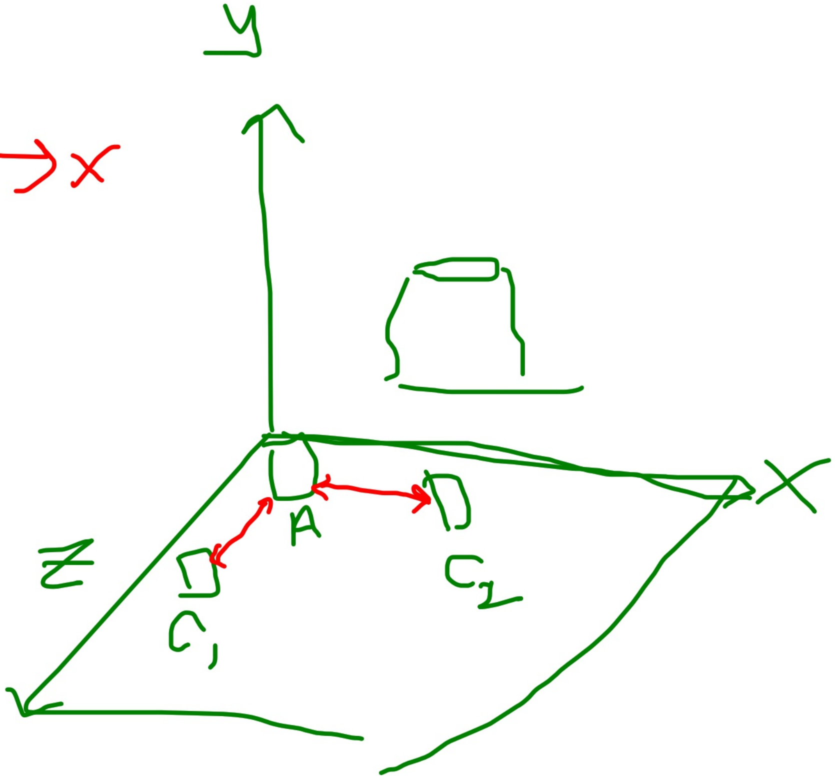
vector/scaler

$$\begin{bmatrix} 6 \\ 8 \end{bmatrix}$$

42

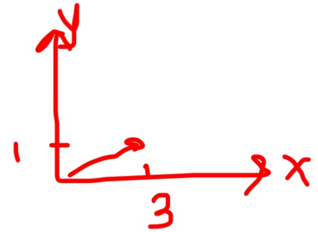


$$A = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$



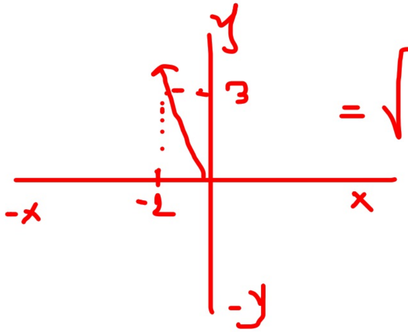
$$C_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \quad C_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$$

$$\vec{V} = \begin{array}{cc} \begin{bmatrix} 3 \\ 1 \end{bmatrix} & \begin{matrix} x \\ y \end{matrix} \end{array}$$



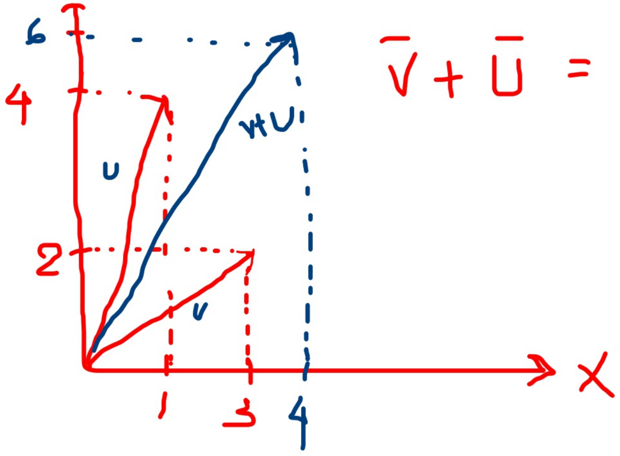
$$= \sqrt{3^2 + 1^2} = \sqrt{10} = 3.2$$

$$\vec{U} = \begin{array}{cc} \begin{bmatrix} -2 \\ 3 \end{bmatrix} & \begin{matrix} x \\ y \end{matrix} \end{array}$$

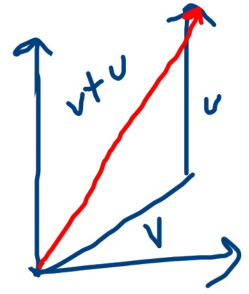


$$= \sqrt{(-2)^2 + (3)^2} = \sqrt{4+9} = \sqrt{13}$$

$$\vec{V} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \vec{U} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

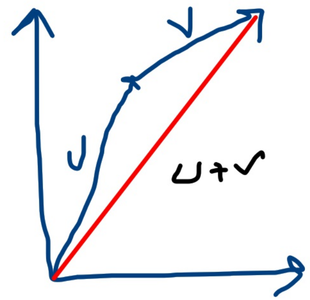


$$\bar{V} + \bar{U} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$



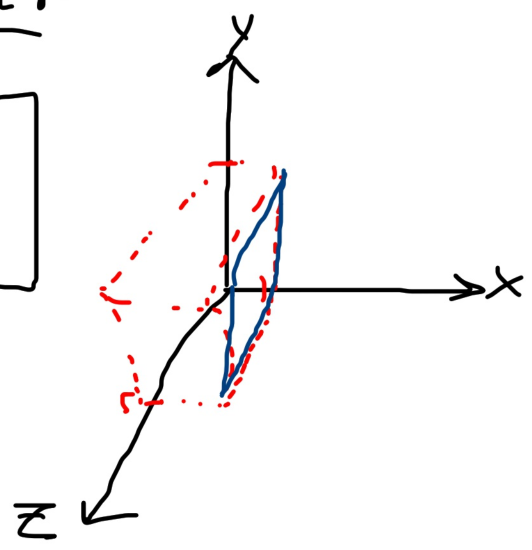
$$\sqrt{4^2 + 6^2} = \sqrt{16 + 36}$$

$$= 7.52$$

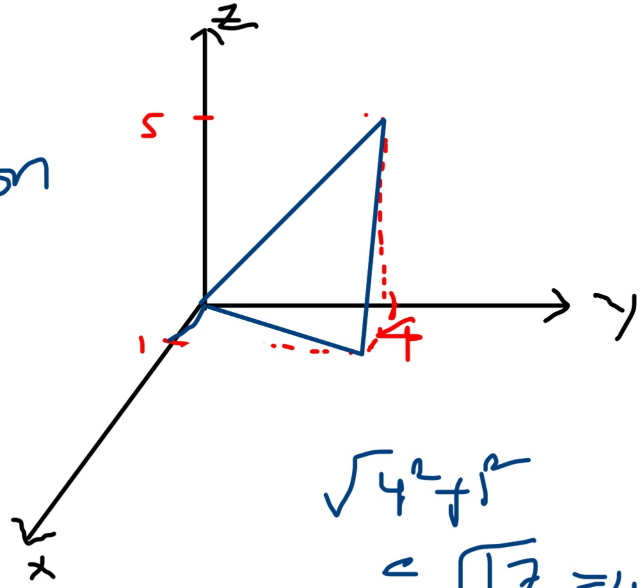


3D Vector

$$v = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$$



rotation $\rightarrow$



$$\sqrt{4^2+1^2} = \sqrt{17} = 4.12$$

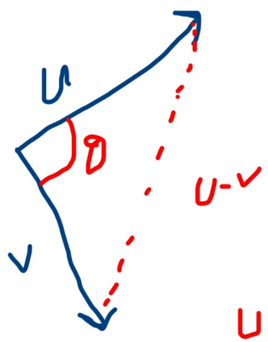
$$\sqrt{1^2+4^2+5^2} = 6.4$$

$$\bar{U} \cdot \bar{V} = U_1 \cdot V_1 + U_2 \cdot V_2 + U_3 \cdot V_3$$

$$U = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} \qquad V = \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix} \qquad U \cdot V = 1 \cdot 4 + (-2) \cdot 3 + 5 \cdot 1$$

$$= 4 - 6 + 5 = 3$$

$$\vec{U} \cdot \vec{V} = \|\vec{U}\| \cdot \|\vec{V}\| \cdot \cos\theta$$



$$\cos\theta = \frac{\vec{u}\cdot\vec{v}}{|u|\cdot|v|}$$

$$u = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \quad v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \frac{2\cdot 1 + 1\cdot 1 + (-2)\cdot 1}{\sqrt{2^2+1^2+(-2)^2}\cdot\sqrt{1^2+1^2+1^2}} = \frac{1}{3\cdot\sqrt{3}}$$

$$\cos\theta = \frac{1}{3\cdot\sqrt{3}}$$

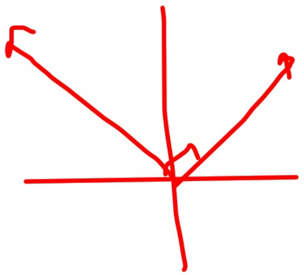
$$\theta = \arccos\left(\frac{1}{3\cdot\sqrt{3}}\right)$$

$$\downarrow$$

$$\text{degree} = \text{radian} \times \frac{180}{\pi}$$

$\bar{u}\cdot\bar{v} = 0$ orthogonal vector

$$u = \begin{bmatrix} -3 \\ 4 \end{bmatrix} \quad v = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$



Matrix

$$\begin{array}{cccc} c_1 & c_2 & c_3 & \\ \uparrow & \uparrow & \uparrow & \\ \end{array}$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{array}{l} \rightarrow R_1 \\ \rightarrow R_2 \\ \rightarrow R_3 \end{array}$$

Diagonal

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Scalar

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad d = \text{Same}$$

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

$A \qquad\qquad A^T$

Identity

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad d = 1$$

Symmetric Matrix $\Rightarrow$ $A = A^T \mid A^T = A$

skew symm matrix $\Rightarrow$ $A^T = -A$

$$\begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 4 \\ -2 & -4 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -4 \\ 2 & 4 & 0 \end{bmatrix}$$

Matrix Mul

$$A = [3, 1, 4] \qquad B = \begin{bmatrix} 4 & 3 \\ 2 & 5 \\ 6 & 8 \end{bmatrix}$$

$$\begin{array}{ccc} A & & B \\ 1 \times \boxed{3} & \boxed{3} \times & 2 \end{array}$$

$\hookrightarrow 1 \times 2$

$$[3 \times 4 + 1 \times 2 + 4 \times 6, \quad 3 \times 3 + 1 \times 5 + 4 \times 8]$$

$$\rightarrow [38, 46]$$

Add/sub

$$A = \begin{bmatrix} 3 & -4 \\ 8 & 5 \end{bmatrix} \qquad \begin{bmatrix} 9 & -5 \\ 2 & -6 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 3+9 & -4+(-5) \\ 8+2 & 5+(-6) \end{bmatrix} = \begin{bmatrix} 12 & -9 \\ 10 & -1 \end{bmatrix}$$

Divison

$$\begin{bmatrix} 6 & -10 \\ 1 & -2 \end{bmatrix} = 6 \times (-2) - (-10 \times 1)$$

$$= -12 - (-10)$$

$$= -2$$

$$B^{-1} = \frac{1}{|B|} \cdot \begin{bmatrix} -2 & 1 \\ -10 & 6 \end{bmatrix} = \frac{1}{-2} \cdot \begin{bmatrix} -2 & 1 \\ -10 & 6 \end{bmatrix}$$

Line and plane

- Line with intercept
- WX=C
- Plane:
  - If we have more than one variable upon which y depends, in this case the mathematical space becomes three dimensional
- Hyperplane:
  - If we have more than 2 variables, its called hyperplane and we cant draw it. ☺
  - But, all concepts of planes will be applicable to hyperplane

Summary of line

- Line is a mathematical object, it represents something real it could be distance b/w two entities or explains how two variables interact with each other
- If lines are parallel then they have same slope
- If lines are perpendicular, the slope of line 1 –m , slope of line 2 = -1/m

Vector algebra , magnitude and direction

- Vectors has magnitude and direction
  - This can be expressed mathematically
- How to calculate magnitude
- How to determine the direction vector
- Dimensionality of a vector

Vector operations

- How do we manipulate the vectors?
  - This is to represent the interaction between various forces that come into play in real world
- Addition / Subtraction of a vector
- DOT Product

Matrix algebra

- This is used very frequently in SVM and Neural Networks
- Very efficient for computations purposes , its compact and efficient
- What is a matrix?
- Addition
- Subtraction
- Multiplication

Matrix Multiplication

- If A is an m x n matrix and B is an n x p matrix, then the product AB is an m x p matrix

Transpose

- If A is m x n, the transpose of A is n x m.
- The rows become columns and the columns become rows.
- This is sometimes needed to put things into a form that is compatible for multiplication.

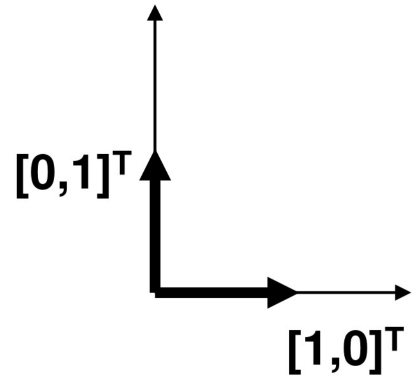
Properties

- The identity matrix has 1s on the main diagonal and 0s elsewhere.
  - Multiplication by the identity matrix yields the original matrix. i.e. AI = IA = A
  - The size of the identity matrix is made to be compatible for the operation intended.
- The zero matrix has 0 in every position.
- If A, B, C are of appropriate sizes, then
  - A(BC) = (AB)C
  - A(B+C) = AB +AC
  - (A+B)C = AC + BC

Matrix inverse

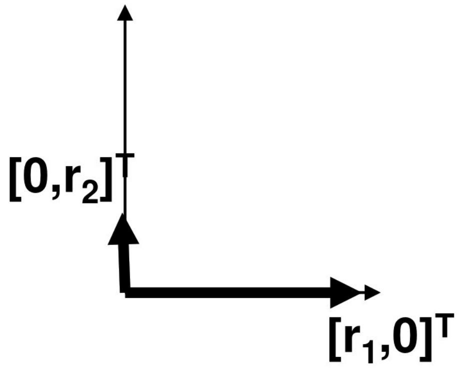
- The inverse (A^{-1}) is defined such at A A^{-1} is I.
- Not every matrix has an inverse. If no inverse exists, then the matrix is called singular (non invertible)
    - If A is nonsingular, so is A^{-1}
    - If A, B are nonsingular, then AB is also non singular and $(AB)^{-1} = B^{-1}A^{-1}$ (Note reversed order.)
    - If A is nonsingular, then so is its transpose and
        - $(A^T)^{-1} = (A^{-1})^T$

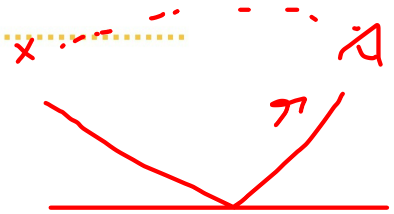
Matrices: Scaling, Rotation, Identity

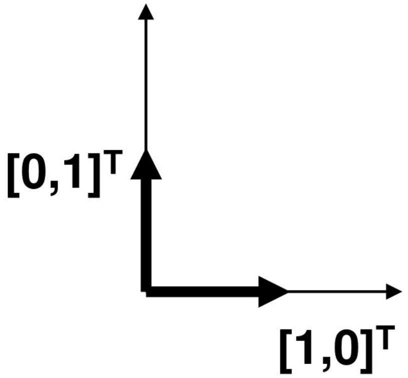


$$\xrightarrow{\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}}$$

scaling



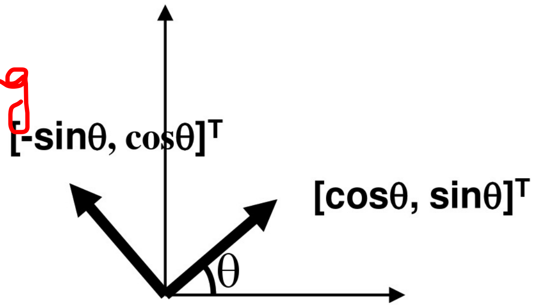




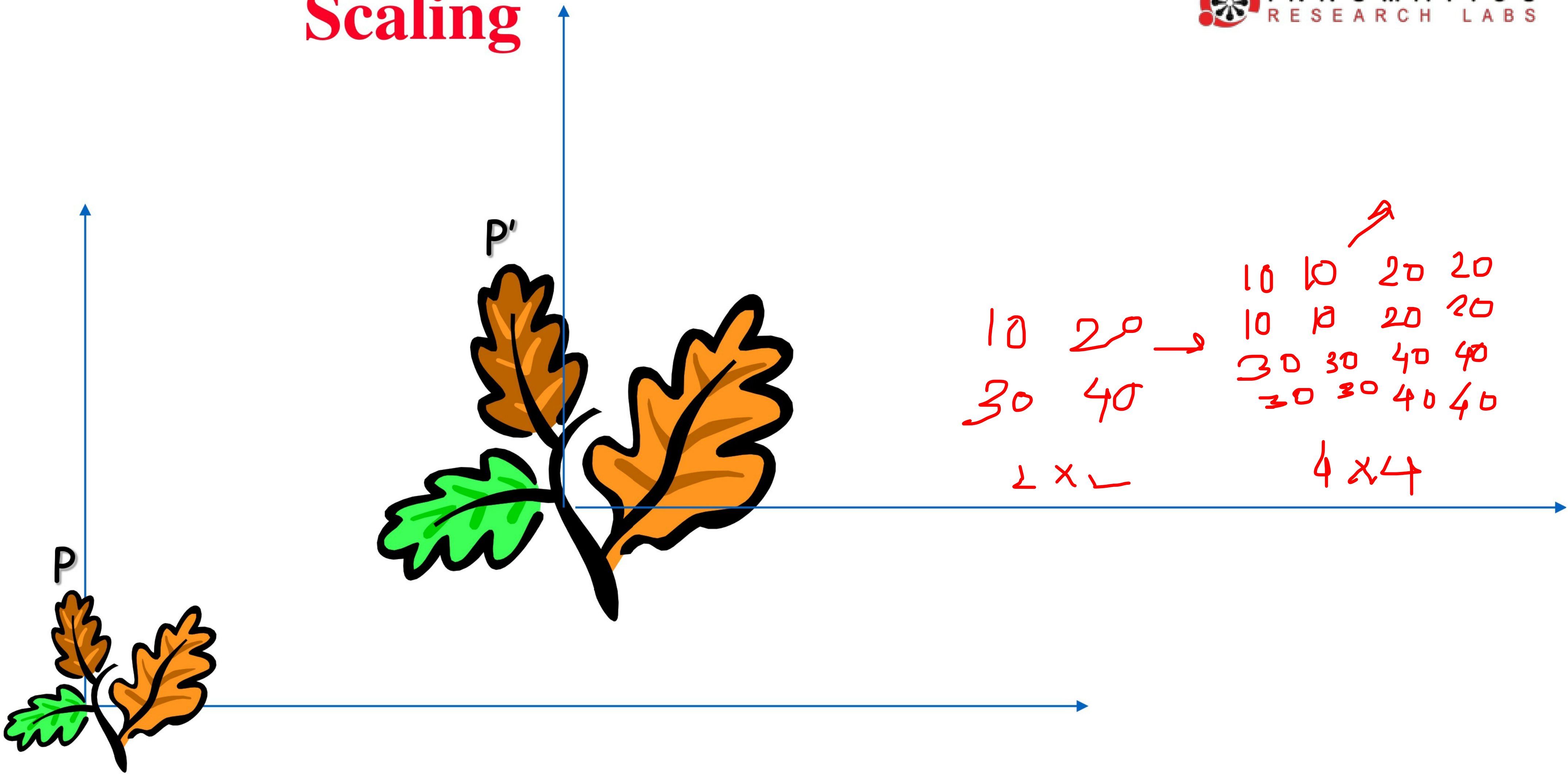
$$\xrightarrow{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}$$

rotation

x'ing



Scaling

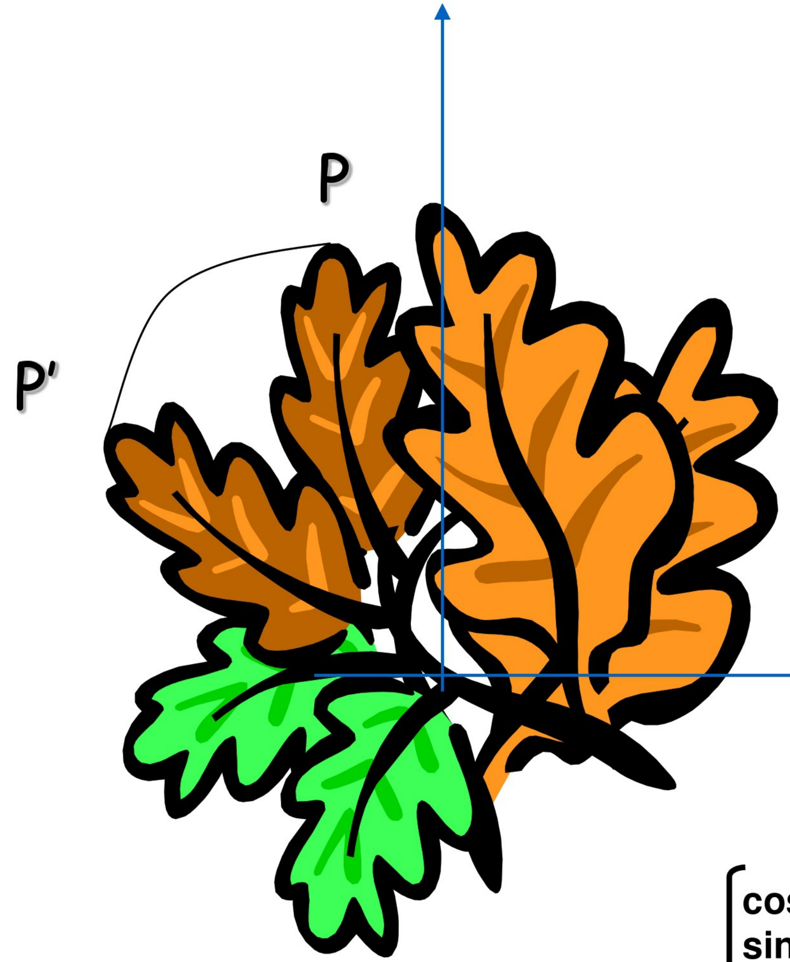


$$\begin{bmatrix} 10 & 20 \\ 30 & 40 \end{bmatrix} \rightarrow \begin{bmatrix} 10 & 10 & 20 & 20 \\ 10 & 10 & 20 & 20 \\ 30 & 30 & 40 & 40 \\ 30 & 30 & 40 & 40 \end{bmatrix}$$

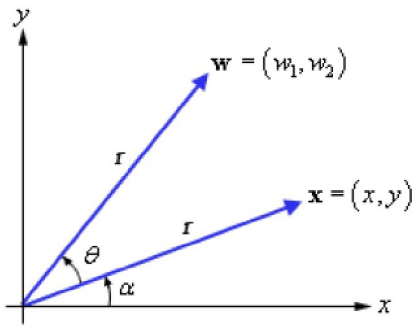
2 x 2 4 x 4

$$\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$$

a.k.a: dilation (r >1), contraction (r <1)



$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



Inverse of a Matrix

- Identity matrix:
  AI = A
- Inverse exists only for square matrices that are non-singular
  - Maps N-d space to another N-d space bijectively
- Some matrices have an inverse, such that:
  $AA^{-1} = I$
- Inversion is tricky:
  $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

  Derived from non-commutativity property

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

Determinant of a Matrix

- Used for inversion
- If det(A) = 0, then A has no inverse

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \det(A) = ad - bc$$

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

Introduction to functions

- When we say building models, we are essentially creating a surface in mathematical space, the surface could be a line, plane or hyperplane , these are used to do classification or to predict some number. The surface is called Function and represented as "f"
  - Y = f (   )
- The surfaces are called functions are inturn called models

Differential of a function

- Derivatives or differentiations
- Partial derivatives
- Differential of a function at a particular point:
    - How much change in y, if x is changed by a small margin and this is also called slope

Maxima and minima of a function

- TO find the minima or maxima we need the tangent at all the points to that function and tangents are differentials
- If you equate your derivate to zero , either you have your x as your max or x as your minima
- Double differential? It helps to identify the local minima or local maxima

Chain rule

- Lets understand this with an example

Maxima and minima applications in ML and DL

- Example : Price of the car increases as power of car increases
- Partial derivates is used across many of the machine learning / DL algorithms to reduce the error and arrive at appropriate solution


Q
&
A