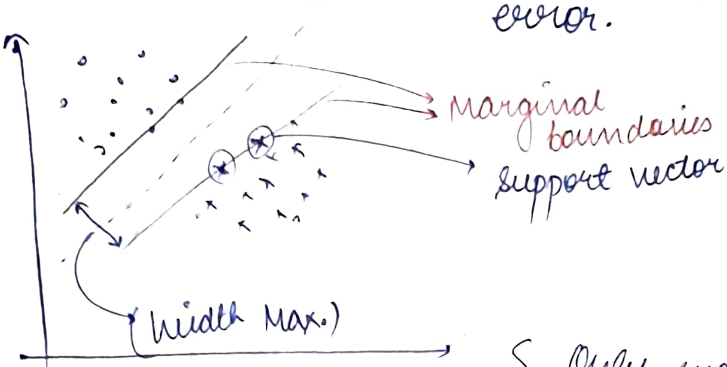
Optimisation → It is required to minimise the error.



Support vector Machine (SVM)

Optimised division line (hyper plane) {b/w the two marginal boundaries (In middle)}.

{Only support vectors are needed to decide the optimal hyper plane.}

SVM is a binary & linear classifier.

Non linear SVM or Kernel SVM for non linear samples.

[Non-linear samples in lower dimensions can be linear separable in higher ~~functions~~ dimensions. (using a mapping func^n).

Multiple SVMs are required for multi class classifier.

$\min_x f(x)$ → objective function   eg $2 < x < 5$ } → constraint

Q. $f(x) = x^4$ ⇒ Find derivative until a non-zero value is obtained.

$f'(x) = 4x^3$ $(x=0)$

$f''(x) = 12x^2$ (Not non zero) at $x=0$

$f'''(x) = 24x$ " "

$f''''(x) = 24$ Order of derivative = 4 (even)
(Non zero)   ∴ $x = 0$ is minima

+ + − + (0, 1, 2)

$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$

$f'(x) = 60x^4 - 180x^3 + 120x^2 + 0 = 60(x^4 - 3x^3 + 2x^2)$

$60x^2(x-1)(x-2)$ $f'(x) = 0 \Rightarrow x = 0, 0, 1, 2$

$f''(x) = 60[4x^3 - 9x^2 + 4x]$ $x=0$ → zero; $x=1$ → Non-zero <0; $x=2$ → Non-zero >0 } (even)

$f'''(x) = 60[12x^2 - 18x + 4]$ $x=0$ → Non-zero   (odd derivative cannot be judged)

$f''''(x) = 60[24x - 18]$

① Mathematical programming techniques.
② Meta heuristic technique
③ Numerical optimization technique.

Optimization

{ gradient descent algorithm }

- Genetic algorithm (GA)
- Particle swarm opt.
- Differential evolution

→ Multidimensional functions :→

$f: \mathbb{R}^d \to R \qquad f(x_1, x_2, \ldots x_d)$

$$\frac{d f(\vec{x})}{d\vec{x}} = \nabla_{\vec{x}} f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix}$$

(gradient) { can be a gradient vector }

eg. $f(x) = x_1^2 x_2 + x_1 x_2^3$

Find gradient of funcn.

$$\nabla_{\vec{x}} f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 x_2 + x_2^3 \\ x_1^2 + 3x_1 x_2^2 \end{bmatrix}$$

rate of change in x_1 direction

$f(x) = w^T x \qquad \vec{w}, \vec{x} \qquad \vec{w} = [w_1, w_2 \ldots w_n]_n \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_n$

$= w_1 x_1 + w_2 x_2 + \ldots w_n x_n$

$\nabla f(x) = w^T$ (a vector)

Gradient Descent Algorithm :-

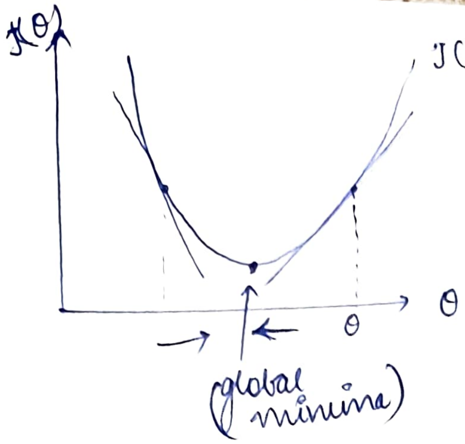
- Initialize $\vec{\theta}$ → (can be vector, scalar) (parameters) θ_1
- while convergence
{
$\theta(t+1) = \theta(t) - \eta \left(\frac{\partial J(\vec{\theta})}{\partial \theta} \right) \to \nabla_{\vec{\theta}} J(\vec{\theta})$
}

$f(\theta)$

$\eta \to$ (the constraint value) $0 < \eta < 1$
Step size / Learning Rate

~~t → st~~
t → step number or iteration number

J → objective function



$$\theta(t+1) = \theta(t) - \eta \frac{\partial J(\theta)}{\partial \theta}$$

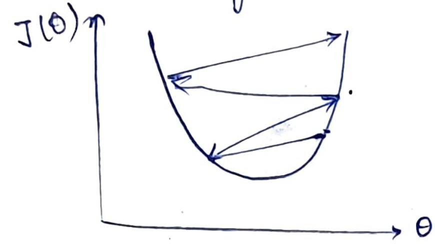
+ve +ve

$\theta(t+1) < \theta(t)$ approaches global minima point

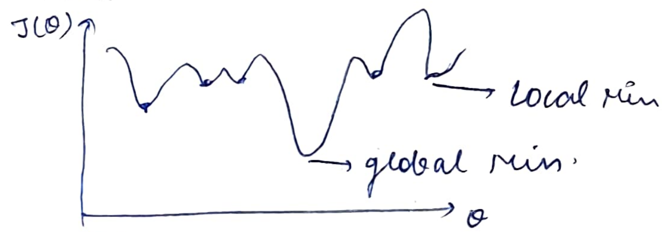
If $\frac{\partial J(\theta)}{\partial \theta} \Rightarrow$ -ve then also

$\theta(t+1)$ approaches global point

If value of η is too large then it will cause overshooting. { $\theta(t+1)$ will not reach global min. point}



Limitation :→



High prob. that we don't end up at global min.

$$\boxed{\text{If } \left(\frac{\partial J(\theta)}{\partial \theta} = 0\right) \text{ then } \theta(t+1) = \theta(t)}$$

This algorithm works well with convex funcⁿ.

$\eta \rightarrow$ step size it has to be chosen very carefully such that we reach minima (global) in proper amount of iterations.

$\boxed{\text{Parameter}}$ → is learned through process.
eg → θ

$\boxed{\text{Hyperparameter}}$ → Is generally fixed by us.
eg. epoch, η

eg. 150 → sample size   15 × 10 ← Batch size ↓ (1 epoch)
iterations

No. of samples = 1 epoch

In batch based G.D.] → iterations = epochs.

Linear Algebra

- 2D matrix → grayscale image
- 3D matrix → coloured image

Tensor) Multi dimensional collection of no.s.

- Scalar $\alpha \in R$
- Vector $x \in \mathbb{R}^n$ (collection of no.s)
- Matrix $w \in \mathbb{R}^{m \times n}$ (2D collection of no.s)

∴ Scalar is a ~~0-D~~ tensor (0th order), vector is ~~1D~~ tensor (1st order), Matrix is 2nd order tensor.

VECTOR

Let v is ~~a~~ a n-dimensional vector, this vector can be represented in n-dim. coordinate system.

we generally not used in M.L. $\vec{V}$ or $V = [v_1, v_2 \ldots v_n]$ or $(v_1, v_2 \ldots v_n)$

For add^n of two vectors their dimensions must be same.

DOT PRODUCT :- $v_1 = [x_1, x_2 \ldots x_n]$

$v_2 = [y_1, y_2 \ldots y_n]$

$$\boxed{v_1 \cdot v_2 = |v_1||v_2| \cos\theta}$$

$$v_1 \cdot v_2 = \sum_{i=1}^{n} x_i y_i = \alpha \in R \text{ (scalar)}$$

Matrix Notation of DOT PRODUCT = $v_1 \times v_2^T$ or $v_1^T \times v_2$

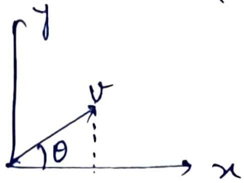
eg. $v_1 = (1, 1, -1)$ $v_2 = (2, 3, 1)$

$v_1 . v_2 = 2 + 3 - 1 = 4$ Ans.

Length / Magnitude of a vector) $\vec{V} = (x_1, x_2 \ldots x_n)$

$$|\vec{V}| = \sqrt{x_1^2 + x_2^2 + \ldots x_n^2}$$

Angle of a vector



$$\tan\theta = \frac{y \text{ component}}{x \text{ component}}$$

Linear Combination of vectors :- Let S is a set of vectors that contains k vectors of same dimensions

$S = \{v_1, v_2 \ldots v_k\}$, then a new vector

$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots \alpha_k v_k = \sum_{i=1}^{k} \alpha_i v_i$ then V is called linear combination of vectors. $\alpha_1, \alpha_2 \ldots \alpha_k$ are scalars.

eg $V_1 = (1, 2, -1)$ $V_2 = (1, 1, 0)$ $V_3 = (0, 1, -1)$

$$\alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 = \alpha_1 (1, 2, -1) + \alpha_2 (1, 1, 0) + \alpha_3 (0, 1, -1)$$
$$= (\alpha_1 + \alpha_2, \; 2\alpha_1 + \alpha_2 + \alpha_3, \; -\alpha_1 - \alpha_3)$$

eg. $V_1 = [1 \; 2 \; 3]$
$V_2 = [2 \; 0 \; 3]$

$$V = V_1 + 2V_2 = [5 \; 2 \; 9]$$

$$V = V_1 + 2V_2 = [V_1 \; V_2] \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Basis vectors:- The vectors whose linear combination will generate all the possible vectors in that vector space.

Linearly dependent vectors:- If we can represent a vector in terms of some other vector.

$$V = \alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 - - - \alpha_n V_n = 0$$
when not all $\alpha_1 - - - \alpha_n = 0$. {atleast one $\alpha_i \neq 0$}.

Linear independent vecter :- $V = \alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 - - \alpha_n V_n = 0$
when all $\alpha_1 - - \alpha_n = 0$.

eg. $S = \{(1,0), (0,1)\}$ check if they are independent?

$$\alpha_1 V_1 + \alpha_2 V_2 = 0 \Rightarrow \alpha_1 (1,0) + \alpha_2 (0,1) = 0$$
$$(\alpha_1, \alpha_2) = 0$$
$$\alpha_1 = 0, \quad \alpha_2 = 0$$ $\therefore$ linearly independent

Note :→

If zero vector is present in sample vector space then the vectors will always be dependent.

SPAN :→

Span is of a set of vectors is the set of all vectors obtained using a linear combination of original vectors.

eg. $\mathbb{R}^2$ $(1,0)$ $(0,1)$ then the span of these vectors in 2-D vector space will be the whole $\mathbb{R}^2$ as $(1,0)$ and $(0,1)$ are basis vectors.

Basis Vectors (i) are the vectors that span the entire vector space and (ii) are linearly independent.

eg. for $\mathbb{R}^3$ space $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ serve as basis vectors.

Orthogonal vectors:- $V_1 \cdot V_2 = 0$ $\quad V_1 \perp V_2$

A set of vectors $\{V_1, V_2, \ldots V_n\}$ are mutually orthogonal if every pair of vectors is orthogonal i.e.

$$V_i \cdot V_j = 0 \quad \forall\ i \neq j$$

eg. $\{\overset{(V_1)}{(1, 0, -1)}, \overset{(V_2)}{(1, \sqrt{2}, 1)}, \overset{(V_3)}{(1, -\sqrt{2}, 1)}\}$ this set of vectors ~~are~~ is mutually orthogonal $\because$ $V_1 \cdot V_2 = 0$, $V_1 \cdot V_3 = 0$, $V_2 \cdot V_3 = 0$

Orthonormal vectors:-

A set of vectors $\{V_1, V_2 \ldots V_n\}$ is said to be orthonormal if every vector in the set has magnitude 1 and they are mutually orthogonal.

e.g. $S = \{ \overset{V_1}{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)}, \overset{V_2}{\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)} \}$

$|V_1| = 1$, $|V_2| = 1$ and $V_1 \cdot V_2 = 0$

Note:-

A orthogonal set of non-zero vectors is linearly independent but the converse is not true for all the cases.

Gram-Schmidt Algorithm is used to convert linear independent vectors into ~~orthogonal~~ orthonormal vectors.

{Assignment} Learn about it!

Matrix

→ element wise multiplication (just like add^n and subtraction). For element wise multiplication order must be same.
↳ $A \otimes B$ or $A \odot B$.

Broadcasting:- (Only a programming concept) $A_{ij} + b_j = C_{ij}$

eg $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ $B = [1\ 1\ 1]$ $A+B = \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix}$

But this concept is theoritically not correct.

Points to remember!

① If $A \times B = 0$ then its not necessary that $A = 0$ or $B = 0$

② If $A \times B = A \times C$, then its not necessary that $B = C$.

③ Transpose i) $(A+B)^T = A^T + B^T$, (iii) $(AB)^T = B^T A^T$
ii) $(\alpha A)^T = \alpha A^T$, (iv) $(A^T)^T = A$

④ Inverse $AA^{-1} = I = A^{-1}A$, $(AB)^{-1} = B^{-1}A^{-1}$
$(A^{-1})^{-1} = A$ $(A^T)^{-1} = (A^{-1})^T$
$(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$ $\{ |A^{-1}| = \frac{1}{|A|} \}$

Norm

(mapping a object using a single value)
⇓
matrix, vector, tensor etc.

Norm is basically related to finding length.

Norm is mapping ~~vecto~~ vector / matrix / tensor to a scalar. Norms are way of measuring the length of vector, matrices etc.

$f(0)$ is that satisfies,

① $f(x) = 0$ $\xrightarrow{\text{when}}$ $x = 0$ [ ]

② $f(x+y) \leq f(x) + f(y)$ (Triangle inequality)

③ for all $\alpha \in R$, $f(\alpha x) = |\alpha| f(x)$ (Linear property)

Types of Norms

→ Euclidean Norm / 2 Norm / L^2 Norm

$V = [V_1, V_2, \dots V_n]$

$\|V\|_2 = \sqrt{V_1^2 + V_2^2 + \dots V_n^2}$

eg. $V = [-5, 3, 2]$

→ 1-Norm / L^1 Norm

$\|V\|_1 = |V_1| + |V_2| + \dots + |V_n|$

so in general, p-Norm

$\|V\|_p = (|V_1|^p + |V_2|^p + \dots + |V_n|^p)^{1/p}, \ p \geq 1$

↳ If $p = 1$ → 1 Norm
$p = 2$ → 2 Norm

→ ∞-Norm / Max Norm

$\|V\|_\infty = \max(|V_1|, |V_2|, \dots |V_n|)$

Orthogonal Matrix $\boxed{AA^T = I}$ which means $A^T = A^{-1}$

Frobenius Norm :-

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

~~$Tr(AB) = Tr(BA)$~~

also called Euclidean Norm

defined as the square root of the sum of the absolute square of its elements.

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{i=1}^{n} |a_{ij}|^2} = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$$

Frobenius norm is used for matrices, while L2 & euclidean norms are used for vectors. ✓

$\boxed{\text{Assignment}}$ → Frobenius Norm

Trace of a matrix $= \Sigma a_{ii}$ (principal diagonal elements).

can be calculated for non sq matrix also.

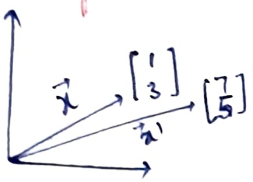
$$Tr(A+B) = Tr(A) + Tr(B)$$

$$Tr(AB) = Tr(BA) \quad , \quad Tr(A) = Tr(A^T)$$

Q: How rank is related to – eigen values? & independent coloumns / vectors?

Rank of matrix = Non-zero eigenvalues

Transformation of a vector :-



$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ $\quad Ax = ? = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}7\\5\end{bmatrix} = \vec{x}'$

When a vector is multiplied with a matrix then its mag. & direction is changed.

$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ $\quad Ax = \begin{bmatrix}3\\3\end{bmatrix} = 3x \Rightarrow \{Ax = 3x\}$

$\therefore$ Generalised eq^n :- $\boxed{Ax = \lambda x}$

(direction is same) This scaling value is called eigen value & this vector is called eigen vector.

So when a matrix A is applied on vector x then it scales x by a factor of / amount of λ.

$A = \begin{bmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}$ find eigen values :→

$$Ax - \lambda x = 0$$

$$|A - \lambda I| = 0$$

$\boxed{\lambda^3 - S_1\lambda^2 + S_2\lambda - S_3 = 0}$

$\boxed{\lambda^3 - 2\lambda^2 + 5\lambda + 4 = 0}$

$$\begin{vmatrix} 2-\lambda & -2 & 3 \\ 1 & 1-\lambda & 1 \\ 1 & 3 & -1-\lambda \end{vmatrix} = 0$$

$\Rightarrow (2-\lambda)[(1-\lambda)(-1-\lambda) - 3] + 2[-1-\lambda-1] + 3[3-1+\lambda] = 0$

$\Rightarrow (2-\lambda)[\lambda^2 - 4] + 2[-2-\lambda] + 3[2+\lambda] = 0$

$\Rightarrow 2\lambda^2 - 8 - \lambda^3 + 4\lambda - 4 - 2\lambda + 6 + 6\lambda = 0$

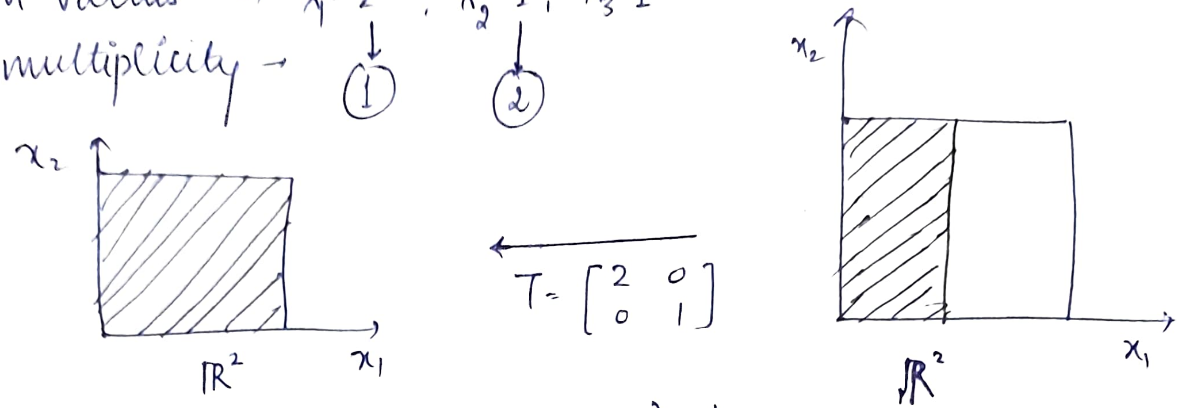
$\Rightarrow$ ~~$-\lambda^3 + 2\lambda^2 + 5\lambda = 0$~~ $\quad \lambda[\lambda^2 - 2\lambda + 5] = 0$

$(+3)(3^2-1)$
$+2(-1-2)$
-4
$2(-1-3)$
$-2-36$ (-4) $+6$
$2+2$
$-1-3$ -4
$-4-5$

For a matrix $A_{3\times3}$, the algebraic multiplicity of an eigen value is the no. of times it is repeated eg.

eigen values $\Rightarrow$ $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 1$

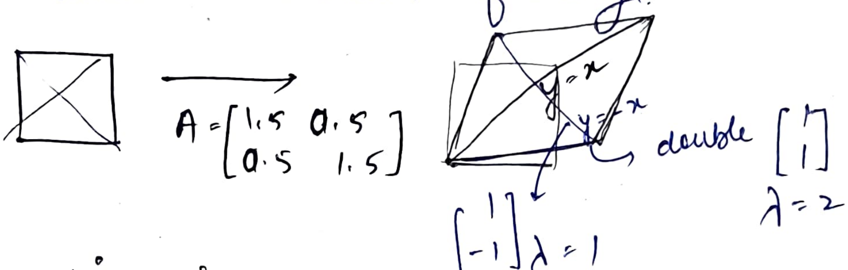
alg. multiplicity $\rightarrow$ 1, 2



$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_2 \end{bmatrix}$ eigen value = $\overset{\lambda_1}{2}, \overset{\lambda_2}{1}$ $\rightarrow$ amount of change

eigen vector $\rightarrow$ direction of change.

Q.



$\boxed{\text{eigen vectors from a symm. matrix are orthogonal}}$

Properties of eigen values:-

1. $AX = \lambda X \Rightarrow A^{-1}AX = A^{-1}\lambda X \Rightarrow IX = \lambda A^{-1}X$
$\rightarrow \boxed{A^{-1}X = \frac{1}{\lambda}X}$

If λ is eigen value for matrix A then $\frac{1}{\lambda}$ will be eigen value for A^{-1}!

2. If λ is eigen value for matrix A then λ^m will be eigen value for A^m.

3. Sum of eigen values = tr (matrix A) (trace)
Product of eigen values = det (A).

4. Eigen vectors with distinct eigen values are linearly independent.

Quadratic Form :-

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix

$\boxed{q(X) = X^T A X}$ X is a n-D vector eg. $A = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 3 & 1 \\ 2 & 1 & 4 \end{bmatrix}$ $\quad X \to$ 3-D vector $\quad$ ~~$X = (x_1) + x_2 + x_3$~~

quadratic form →

$$q(x) = \underset{1\times3}{(x_1 \;\; x_2 \;\; x_3)} \underset{3\times3}{\begin{pmatrix} 1 & -1 & 2 \\ -1 & 3 & 1 \\ 2 & 1 & 4 \end{pmatrix}} \underset{3\times1}{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}$$

: quadratic expression

$= x_1(x_1 - x_2 + 2x_3)$
$+ x_2(-x_1 + 3x_2 + x_3)$
$+ x_3(2x_1 + x_2 + 4x_3)$ $= x_1^2 - x_1x_2 + 2x_1x_3 - x_1x_2 + 3x_2^2 + x_2x_3$
$+ 2x_1x_3 + x_2x_3 + 4x_3^2$

$= x_1^2 + 3x_2^2 + 4x_3^2 - 2x_1x_2 + 2x_2x_3 + 4x_1x_3$

coefficient of x_1^2 $= 1 = a_{11}$ $\quad a_{22} \quad a_{33} \quad a_{12}+a_{21} \quad a_{23}+a_{32} \quad a_{13}+a_{31}$ } → short trick

from matrix

q. $q(x) = x_1^2 - 2x_2^2 - 4x_1x_3$ $\quad A = ?$

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

$\boxed{a_{31} + a_{13} = -4}$
$a_{11} = 1$
$a_{22} = -2$
$a_{33} = 0$

Positive semi-Definite Matrix: → A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is said to be +ve semi definite if for all $x \in \mathbb{R}^n$, $\boxed{q(x) = X^TAX \geq 0}$.

eg. $A = \begin{bmatrix} 1 & 3 \\ 3 & 10 \end{bmatrix}$ $\quad q(X) = [x_1 \;\; x_2]\begin{bmatrix} 1 & 3 \\ 3 & 10 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= x_1(x_1 + 3x_2) + (3x_1 + 10x_2)x_2$

$= x_1^2 + 3x_1x_2 + 3x_1x_2 + 10x_2^2$

$q(x) = x_1^2 + 10x_2^2 + 6x_1x_2 = (x_1 + 3x_2)^2 + x_2^2 \geq 0$
{sum of two squares}

∴ A is a +ve semi definite matrix

Positive definite matrix: → A symm. matrix $A \in \mathbb{R}^{n\times n}$ is said to be +ve definite matrix if $\forall x \in \mathbb{R}^n$,

$\boxed{q(X) = X^TAX > 0}$

* If all eigen values are non-ve i.e ≥ 0 then the matrix is said to be +ve semi-definite.

If all eigen values are +ve i.e > 0 then the matrix is said to be +ve definite matrix.

---

* Columns of an invertible matrix are linearly independent.

Proof = do yourself.

$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ $\quad \lambda^2 - 2\lambda + (-3) = 0$

$\Rightarrow \lambda^2 - 2\lambda - 3 = 0 \Rightarrow (\lambda - 3)(\lambda + 1) = 0$

$\lambda_1 = 3, \ \lambda_2 = -1$ eigen values

eigen vectors :- $\longrightarrow$ $V_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad V_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$$\underset{A}{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}} \underset{V_1}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \underset{\lambda_1 \ V_1}{3\begin{bmatrix} 1 \\ 1 \end{bmatrix}} \rightarrow (1)$$

$$\underset{A}{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}} \underset{V_2}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}} = \underset{\lambda_2 \ V_2}{-1\begin{bmatrix} -1 \\ 1 \end{bmatrix}} \rightarrow (2)$$

$$\underset{A}{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 3 & -1 \end{bmatrix} \rightarrow \underset{V_1 \ V_2}{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}} \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$$

($V_1 \ V_2$ under the columns of the second matrix; $\lambda_1 V_1 \ \lambda_2 V_2$ under the columns of the result)

(*)

$\boxed{AV = V\Lambda}$ & $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

$\hookrightarrow$ diagonal matrix with eigen values as the diagonal elements.

and V is matrix containing eigen vectors as its columns.

Eigen Decomposition :$\longrightarrow$ $\quad AV = V\Lambda$

Post multiply by V^{-1} if it exists

{diagonalisation of matrix}

$\{A = PDP^{-1}\}$

$\Rightarrow \boxed{AVV^{-1} = V\Lambda V^{-1}}$

$\Rightarrow AI = \cancel{VV} \ V\Lambda V^{-1} \Rightarrow \boxed{A = V\Lambda V^{-1}}$

$$\underset{n \times n}{A} = \underset{n \times n}{V \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ V_1 & V_2 & \cdots & V_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}} \ \underset{n \times n}{\Lambda \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \ \lambda_n \end{bmatrix}} \ \underset{n \times n}{V^{-1}}$$

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_n$

In case of symmetric matrix :⟶ i.e. A is a symm matrix then eigen vectors are orthogonal so V will be a orthogonal matrix $\hookrightarrow \boxed{V^{-1} = V^T}$

$$\therefore \quad A = V \Lambda V^{-1} \Rightarrow \boxed{A = V \Lambda V^T}$$

Singular Value Decomposition

for non-square matrix

$$\boxed{A = U \Sigma V^T}$$

$m \neq n$

$$\underset{m\times n}{\boxed{}} = \underset{m\times m}{\boxed{U}}\ \underset{m\times n}{\boxed{\Sigma}}\ \underset{n\times n}{\boxed{V^T}}$$

$U \rightarrow m$ no. of orthonormal vectors $= \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{bmatrix}_{m\times m}$

$u_1, u_2, \dots u_m$ are called left singular vectors.

size of $u_i = m \times 1$

$V \rightarrow$ eigen vectors (orthogonal matrix) $\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}_{n\times n}$ size of $v_i = n \times 1$

$v_1, v_2 \dots v_n$ are called right singular vectors.

All $v_1, v_2 \dots v_n$ are also orthonormal.

$\Sigma \rightarrow$ the diagonal matrix $\begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ \vdots & & & \ddots & \sigma_n \\ 0 & 0 & \cdots & & 0 \end{bmatrix}_{m\times n}$

$\sigma_1, \sigma_2, \dots \sigma_n$ are called singular values.

Now,

$$A = U \Sigma V^T$$

$$A^T A = (U \Sigma V^T)^T (U \Sigma V^T) = V \Sigma^T U^T U \Sigma V^T$$

(U is an orthonormal matrix $\therefore U^T U = U^{-1} U = I$)

$$A^T A = V \Sigma^T \Sigma V^T$$

$$\boxed{A^T A \simeq V \Sigma^2 V^T}$$

$$\boxed{A^T A = V \Sigma^T \Sigma V^{-1}}$$

$$\Sigma^T \Sigma = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \\ & & & \sigma_n^2 \end{bmatrix}_{n\times n}$$

$A^T A$ is a symmetric matrix $\because (A^T A)^T = A^T A$

Comparing with ~~A = VN~~ $[A = V \Lambda V^{-1}]$

here V is matrix containing eigen vectors of $A^T A$. These eigen vectors will be orthogonal $\because$ $A^T A$ is a symmetric.

Σ So first calculate $A^T A$ $\rightarrow$ calculate its eigen values (λ) and then its eigen vectors.

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_3} \; \cdots \end{bmatrix}$$

Now for U,

$\cancel{AA}$ $A = U \Sigma V^T$

$AA^T = (U\Sigma V^T)(U \Sigma V^T)^T$

$= U \Sigma \underline{V^T V} \Sigma^T U^T$

$$\boxed{AA^T = U \Sigma \Sigma^T U^T}$$

U is collection of eigen vectors obtained from $\underline{AA^T}$ and are orthogonal $\because$ AA^T is symm.

Eigen values obtained from $A^T A$, AA^T will have same non-zero values.

<u>The eigen vectors ~~are~~ have to made orthonormal.</u>

Find the SVD:- $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}_{3\times 2}$ $\quad A^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}_{2\times 3}$

$$A A^T = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$|A - \lambda I|$

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

eigen values of AA^T:- $\cancel{\lambda^3 - 4\lambda^2 + 4\lambda - 0 = 0}$

$\Rightarrow$ $\cancel{\lambda(\lambda^2 - 4\lambda + 4) = 0}$

$\Rightarrow$ $\cancel{\lambda(\lambda - 2)^2 = 0}$

$\cancel{\lambda = 0, 2, 2}$ $\quad \boxed{\lambda = 0, 1, 3}$

$\lambda = 0$, $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \Rightarrow \cancel{\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & -1/2 & 1/2 \end{bmatrix} [X] = 0}$

$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1/2 & -1/2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$ $\quad \Rightarrow \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{bmatrix}$ $\quad X_1 = k\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

$2x_1 + x_2 + x_3 = 0$
$x_2 - x_3 = 0$ $\rightarrow x_2 = x_3$ $\quad x_1 = -x_2$

$X_1 = k_1\begin{bmatrix} -1 \\ +1 \\ 1 \end{bmatrix}$

$\lambda = 1$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} X_2 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$x_1 + x_2 + x_3 = 0$ $\qquad x_2 = -x_3$

$x_1 = 0$

$X_2 = k_2\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$

$\lambda = 3$, $\begin{bmatrix} -1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}[X_3] = 0 \Rightarrow \begin{bmatrix} -1 & 1 & 1 \\ 1 & -2 & 0 \\ 0 & 1 & -1 \end{bmatrix} X_2 = 0$

$$\Rightarrow \begin{bmatrix} -1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} X_2 = 0$$

$$\begin{bmatrix} -1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$\Rightarrow -x_1 + x_2 + x_3 = 0$ $\quad x_1 = 2x_2$

$-x_2 + x_3 = 0$

$x_2 = x_3$

$X_3 = k_3\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$

X_1, X_2, X_3 are mutually orthogonal.

$$U = \begin{bmatrix} -1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}_{3\times 3}$$ ✓

Now,

$$A^TA = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

eigen vectors of A^TA :-

$\lambda^2 - 3\lambda + 2 = 0$

$\Rightarrow \cancel{\lambda^2 - \lambda - 2\lambda + 2 = 0}$

$\lambda(\lambda-1) - 2(\lambda-1) = 0$

$\lambda = 1, 3$

$\lambda = 1$, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}[X_1] = 0 \Rightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} X_2 = 0$

$x_1 + x_2 = 0$ $\qquad x_1 = -x_2$

$X_1 = k_1\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\lambda = 3$, $\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$

$-x_1 + x_2 = 0.$

$X_2 = k_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$V = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}_{2\times 2} \qquad \Sigma = \begin{bmatrix} \sqrt{1} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix}$$

$$\boxed{A = U \Sigma V^T}$$

ass Qn $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$

Reduced SVD :-

$$A = \left[\begin{array}{cccc|c} u_1 & u_2 \cdots & u_{\gamma} & \cdots & u_m \end{array}\right] \left[\begin{array}{cccc|c} \sigma_1 & & & & 0 \\ 0 & \sigma_2 \cdots & & \sigma_r & \\ \hline 0 & & & & 0 \end{array}\right] \left[\begin{array}{ccc|c} v_1 & v_2 - & v_{\gamma} & -v_n \end{array}\right]^T$$

$$= [U_L \mid U_R] \left[\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array}\right] [V_L \mid V_R]^T$$

$$= (U_L \Sigma_{TL} \mid 0) \begin{pmatrix} V_L^T \\ V_R^T \end{pmatrix}$$

If $m > n$: $A_{m\times n} = \underset{m\times n}{U} \; \underset{(n\times n)}{\Sigma} \; \underset{n\times n}{V^T}$

$m < n$:- $A_{m\times n} = \underset{m\times m}{U} \; \underset{m\times m}{\Sigma} \; \underset{m\times n}{V^T}$ $\quad \{V = m\times m\}$ $\quad V^T = m\times n.$

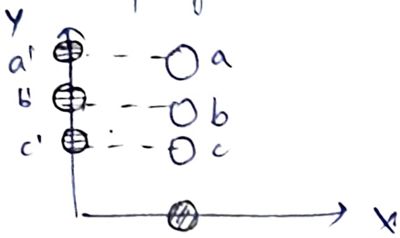
Principal Component Analysis (PCA) :-

- To reduce the dimensions of feature vector.
  {Model complexity is reduced, time taken for processing reduces}
- The original info / pattern of data / relationship b/w data should remain intact.

Applications of K-means clustering :- Image segmentation, facebook friend requests

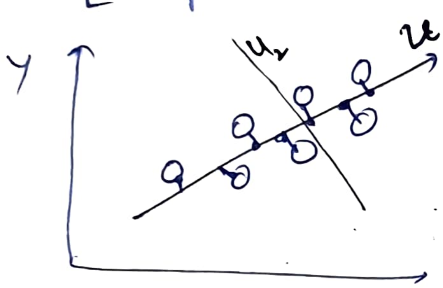
PCA are projection problems (projection of data)

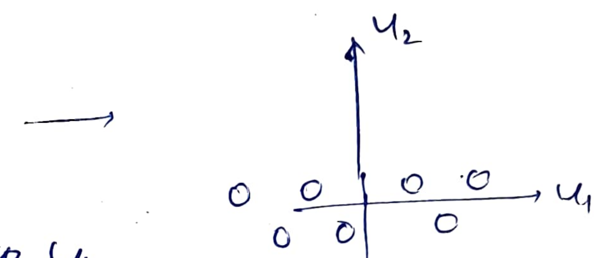
Suppose there are three 2D points & we take their projection on x & y axis.



The projection of y axis is better ∵ info is not lost, which is lost in projection of x-axis.

✱ [variance & dispersion of data should be max after projection on the axis.]





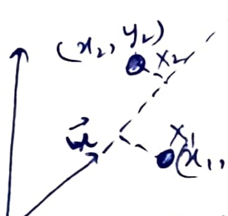
u_2 is orthogonal to u_1.
Projection of u_1 is better (variance is more).

Now we want to project the matrix X below on a vector x such that (i) the variance of the projected data is maximum
(ii) x is a unit vector.

$$X = \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \begin{array}{c} f_1 \quad f_2 \; \cdots \; f_d \\ \left[\begin{array}{c} \\ \\ \\ \\ \end{array} \right]_{n \times d} \end{array} \rightarrow n \times d' \qquad d' <<< d$$

But lets start with a simpler problem.

$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$



$$p = proj_{\vec{x}} X_1 = \frac{X_1 \cdot x}{|x|}$$

$$= X_1 \cdot \hat{x}$$

(x_1, y_1) $\quad$ (r, s)

linear combination of original points.

$$= r x_1 + s y_1$$

PCA is a linear transformation.

Now let's expand this concept to matrix X.

$$X = \begin{array}{c} \\ x_1 \\ x_n \end{array}\begin{array}{c} f_1 \; f_2 \cdots f_d \\ \left[\quad \right]_{n \times d} \end{array} \qquad u = [\;]_{d \times 1} \rightarrow x'_{d \times 1}$$

Multivariate data

Projection of X on u = $(X)(u)$

Calculate covariance & maximize it.

$$\Sigma = \sum_{i=1}^{n} (x_i - \mu)^T (x_i - \mu)$$

$\cancel{d \times d \rightarrow \text{symmetric}}$

$\arg\max_{u} (var(X \cdot u))$ → value of u for which variance of projected data is maximum.

↳ such that $||u|| = 1$. u is a unit vector.

$var(X.u) = u^T \Sigma u$

soln :- $\boxed{\Sigma u = \lambda u}$

square matrix.

λ → eigen values of covariance matrix

(u → eigen vector of Σ which is symm.)

λ value more ↑ ⇒ variance more. → corresponding matrix is most powerful.

Now to find projection -

$$X'_{n \times d'} = X_{c\, n \times d} \cdot U_{d \times d'}$$

$U = \begin{bmatrix} u_1 & u_2 \\ \vdots & \vdots \end{bmatrix}$ → (vectors) Principal components

Σ eigen values in descending order.

{Centered data is taken in PCA. data analysis is easy}

How to select principal component!

$$\text{Variance captured by } \lambda_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots \lambda_d} \times 100\%$$

~~Mode~~ Threshold value of variance captured = >90%.

Inputs for PCA

- data
- dimension reduced

$$PC_k = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{i=1}^{d} \lambda_i} > \theta \text{ (Threshold)}$$

What to do → Matrix → covariance → eigen values
[Input = d' (dimens.) eigen vectors ← sort in descending order
select d' columns → $X'_{n\times d'} = X_{c\,n\times d} \cdot U_{d\times d'}$

[PCA can be used for data visualization purpose.].

Solving Linear System of Equations:-

$$\underset{m\times n}{A}\ \underset{n\times 1}{X} = \underset{m\times 1}{B} \Rightarrow X = A^{-1}B$$

A is invertible matrix
{ A is square & non-singular i.e. $|A| \neq 0$ }

Now if A is not a square matrix (then exact soln can not be calculated, we can find approximate soln.)

$A_{m\times n}$

i) $m >> n$ → overdetermined system
ii) $m << n$ → underdetermined system
→ Minimum Norm soln based approximation

Least Square Approach for approximation {for overdetermined}
→ used in Linear regression & many other machine learning problems.

$(m >> n)$

$$AX = b$$

$A \in \mathbb{R}^{m\times n}$

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{21}x_2 = b_2$$
$$a_{31}x_1 + a_{32}x_2 = b_3$$

Obj funcn:- $\arg\min_x \|Ax - b\|_2^2$

$$AX = b \Rightarrow A^T AX = A^T B \Rightarrow X = (A^T A)^{-1} A^T B$$

Least square approximation soln $\Rightarrow \boxed{X = A^+ B}$

($A^+ \Rightarrow$ Left pseudoinverse)

Minimizing error

Eg

$$x_1 = 6 \qquad x_1 + x_2 = 0 \qquad x_1 + 2x_2 = 0$$

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad B = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \qquad (A^TA)^{-1} = \frac{1}{6}\begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}$$

$$(A^TA)^{-1}A^T = \frac{1}{6}\begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}$$

$$(A^TA)^{-1}A^TB = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 30 \\ -18 \end{bmatrix} = \cancel{\frac{1}{6}}\cancel{6}\begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix} \quad \text{Ans.}$$

Minimum Norm solution :→ $m << n$

Obj funcn:- $\arg_x \|AX-B\|_2^2 + \|X\|_2^2$

Minimizing error as well as norm of x.

$AX = B$

$$\Rightarrow \boxed{X = \underbrace{A^T(AA^T)^{-1}}_{\text{Right pseudo inverse}}B} \Rightarrow X = A^+B$$

Q. $x_1 + x_2 + x_3 = 1$, $-x_1 - x_2 + x_3 = 0$

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix}_{2\times 3} \qquad A^T = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}_{3\times 2}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad AA^T = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

$$(AA^T)^{-1} = \frac{1}{8}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

$$A^T(AA^T)^{-1} = \frac{1}{8}\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 2 & -2 \\ 2 & -2 \\ 4 & 4 \end{bmatrix}$$

$$A^T(AA^T)^{-1}B = \frac{1}{8}\begin{bmatrix} 2 & -2 \\ 2 & -2 \\ 4 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/2 \end{bmatrix}$$

sol^n $x_1 = 1/4$, $x_2 = 1/4$, $x_3 = 1/2$ Norm

$$\sqrt{x_1^2 + x_2^2 + x_3^2}$$

$$= \sqrt{\frac{3}{8}} = 0.61$$

Minimum

other sol^ns

$x_1 = 0$

$x_2 = 1/2$

$x_3 = -1/2$

$$\sqrt{0^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2}$$

$$= \sqrt{\frac{1}{2}} = 0.707$$