Week-2 - LINEAR ALGEBRA

$Ax = b$

$A(m \times n)$, $x(n \times 1)$, $b(m \times 1)$

if $m = n$

↳ no. of equations are variables are same.

$m > n$

↳ More eq^n than var

- Usually no solution.

$m < n$

↳ less eq^n than var

- usually multiple solutions.

Full Row Rank

- when all the rows of the matrix are LI
- data sampling does not present a linear relationship - samples are independent.

Full Column Rank

- when all the columns of the matrix are LI
- Attributes are LI

☆ $\boxed{\text{Row Rank} = \text{Column Rank}}$

Case 1 : $m = n$

- $m = n$
  - A is full rank and $|A| \neq 0$ → $x = A^{-1}b$ Unique soln
  - A is not full rank
    - Consistent (∞ soln)
    - Inconsistent (No soln)

Case 2 : $m > n$

↳ case of not enough variables or attributes.
↳ no-solution case.

Optimaization perospective

① Minimize $(Ax - b)$

$\Downarrow$

Denote $(Ax - b) = e_{(m \times 1)}$ error term

• one could minimize all the errors collectively by minimizing $\boxed{\sum_{i=1}^{m} e_i^2}$

$\Downarrow$ same.

Minimize $(Ax - b)^T (Ax - b)$

$\Downarrow$

$\min \left[\left(x^T A^T A x - 2b^T A x + b^T b \right) \frac{1}{2} = f(x) \right]$

the soln to this optimization problem is obtained by differentiating $f(x)$ WRT x and setting the differential to zero

$$\boxed{\nabla f(x) = 0}$$

$$\boxed{f(x) = x^T A^T A x - 2b^T A x + b^T b}$$

$$\therefore \nabla f(x) = 2(A^T A)x - 2A^T x = 0$$

$$(A^T A)x = A^T b$$

$$(A^T A)x = A^T b$$

$$\boxed{x = (A^T A)^{-1} A^T b}$$

eg. $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$

$x = (A^T A)^{-1} (A^T b)$

$= \begin{bmatrix} 0.2 & -0.6 \\ -0.6 & 2.8 \end{bmatrix} \begin{bmatrix} 20 \\ 5 \end{bmatrix}$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

R code

```
A = matrix (c (1,2,3,0,0,1), ncol = 2, byrow = F)
B = matrix (c (1,2,5), ncol = 1, byrow = F)
x = inv (t(A) %*% A) %*% t(A) %*% b
```

Case 3: $m < n$

- more variables than eqn.
- ∞ soln case.

Optimization

- $\min\left(\frac{1}{2}x^T x\right)$ s.t $Ax = b$.

- Define lagrangian function $f(x, \lambda)$

$$\min\left[f(x,\lambda) = \frac{1}{2}x^T x + \lambda^T(Ax - b)\right]$$

- diff the langrangian WRT x, and setting to zero

$$\boxed{x + A^T\lambda = 0}$$

$x = -A^T \cancel{A} \lambda$

$\Downarrow$

$Ax = b = -AA^T\lambda$

thus we obtain $\lambda = -(AA^T)^{-1} b \Rightarrow$ all rows are LI

$x = -A^T\lambda$

☆ $\boxed{x = A^T (AA^T)^{-1} b}$

eg $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ $B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

~~A = ma~~

$x = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 14 & 3 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$$x = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -0.2 \\ 1.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.2 \\ -0.4 \\ 1 \end{bmatrix}$$

R code

```
A = matrix (c(1,0, 2,0, 3,1), ncol =3 ), byrow=F)
b = c(2,1)
library( MASS)
x = t(A) %*% inv (A %*% t(A)) %*% b
```

→ Generalization

The concept we used to generalize the soln is for all the 3 cases is called MOORE - PENROSE pseudo - inverse of matrix

↳ the pseudo inverse is used as follows

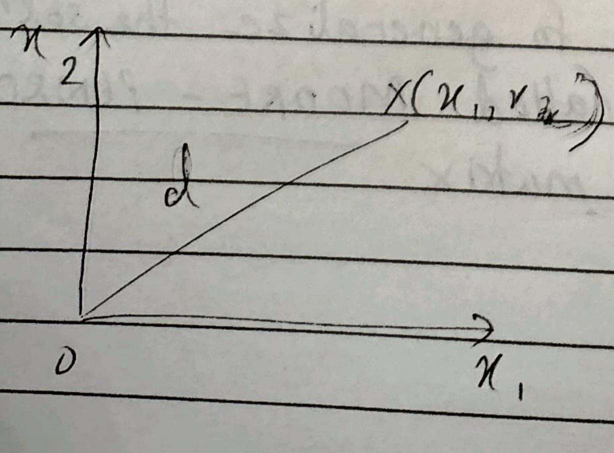
$$Ax = b$$
$$\Downarrow$$
$$\boxed{x = A^{+}b}$$

↳ SVD can be used to calculate the pseudo inverse or the generalized inverse (A^{+})

Rcode

```
A = matrix(c(1,2,3,0,0,1), ncol = 2, byrow = F)
b = matrix(c(1,2,5), ncol = 1, byrow = F)
library(MASS)
x = ginv(A) %*% b
```

VECTORS & LENGTH



$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boxed{d = \sqrt{x_1^2 + x_2^2}}$$

$$l = | x^2 - x^1 |_2$$

$$l = \sqrt{(x_1^2 - x_2^1)^2 + (x_2^2 - x_1^2)^2}$$

$$\boxed{l = \sqrt{(x_2 - x_1)^T (x_2 - x_1)}}$$

→ Unit Vector

$$\hat{a} = \frac{A}{|A|}$$

→ orthogonal vectors:

$$\boxed{A \cdot B = \sum_{i=1}^{n} a_i b_i = A^T B = 0}$$

→ orthonormal vectors

↳ orthogonal vectors with unit magnitude.

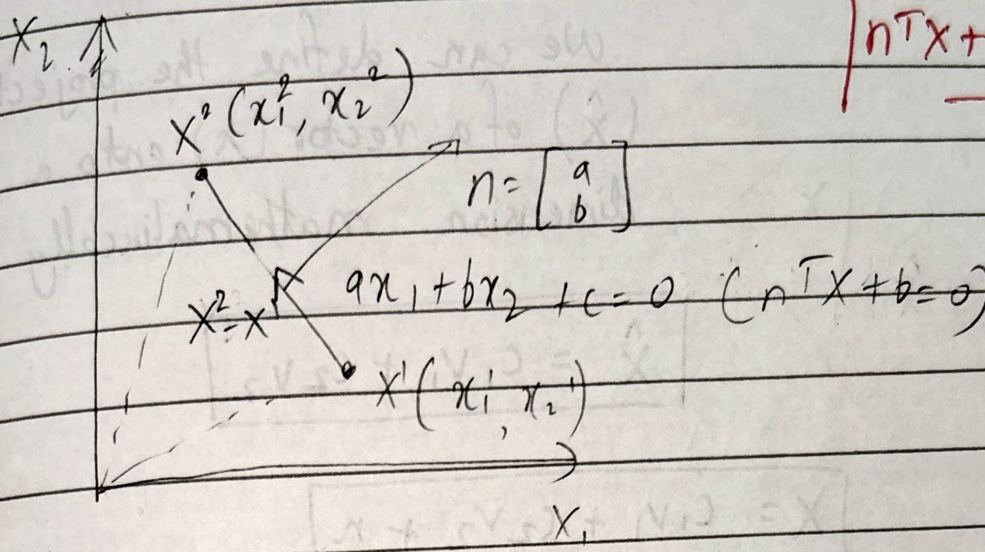
→ Basis vectors:

↳ are set of vectors that are independent and span the space.

↳ basis vectors are not unique.

REPRESENTATION OF LINE & PLANE:

$$\boxed{n^T X + b = 0}$$



Point X^1 ~~lie~~ lies on the line $\Rightarrow n^T X^1 + b = 0 \rightarrow$ (1)
Point X^2 lies on the line $\Rightarrow n^T X^2 + b = 0 \rightarrow$ (2)

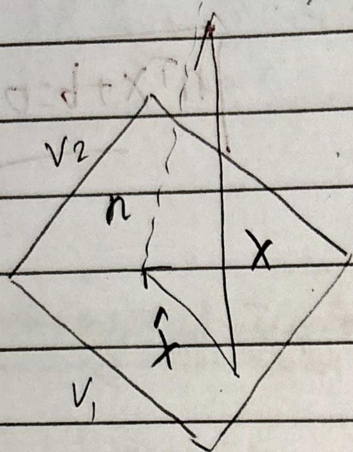
Sub (2) from (1)

$$n^T (X^2 - X^1) = 0$$

$\therefore$ n is $\perp^r$ to $(X^2 - X^1)$

* A single eq^n in 3D space $\rightarrow$ plane
* 2 eq^n in 3D space $\rightarrow$ 3 lines
* 3 eq^n in 3D space $\rightarrow$ point.

⇒ Projections:



We can define the projection ($\hat{x}$) of a vector (x) onto a lower dimension mathematically as

$$\boxed{\hat{x} = c_1 v_1 + c_2 v_2}$$

$$\boxed{x = c_1 v_1 + c_2 v_2 + n}$$

$$\boxed{x = \hat{x} + n}$$

$n^T v_1 = 0$

$v_1^T n = 0$

$n^T v_2 = v_2^T n = 0$

Now, $v_1^T n = 0$

$$v_1^T (x - c_1 v_1 - c_2 v_2) = 0$$

taking $v_1^T v_2 = 0 = v_2^T v_1$

$$v_1^T x - c_1 v_1^T v_1 = 0$$

$$\boxed{c_1 = \frac{v_1^T x}{v_1^T v_1}}$$

$V_2^T n = 0$

$V_2^T (X - C_1 V_1 - C_2 V_2) = 0$

$V_2^T X - C_2 V_2^T V_2 = 0$

$$\boxed{C_2 = \frac{V_2^T X}{V_2^T V_2}}$$

$$\therefore \quad \boxed{\hat{X} = \left(\frac{V_1^T X}{V_1^T V_1}\right) V_1 + \left(\frac{V_2^T X}{V_2^T V_2}\right) V_2}$$

Generalization

$$\hat{X} = \sum_{j=1}^{K} c_j v_j$$

$$\hat{X} = [v_1 \ \cdots \ v_k] \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

$$\boxed{\hat{X} = Vc}$$

Using orthogonality idea,

$$n = X - \hat{X}$$

$$V^T n = 0$$

$$V^T (X - \hat{X}) = 0$$

$$V^T (X - Vc) = 0$$

$$V^T X - V^T V c = 0$$

$$c = (V^T V)^{-1} V^T X$$

$$\boxed{\hat{X} = \left(\frac{V^T X}{V^T V}\right) V}$$

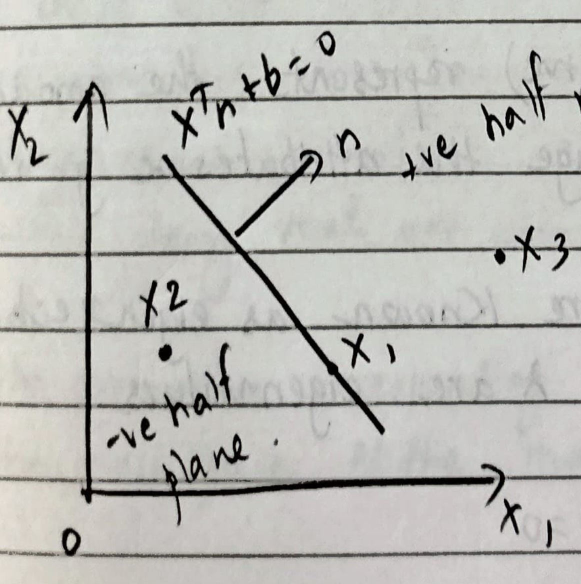
HYPERPLANES

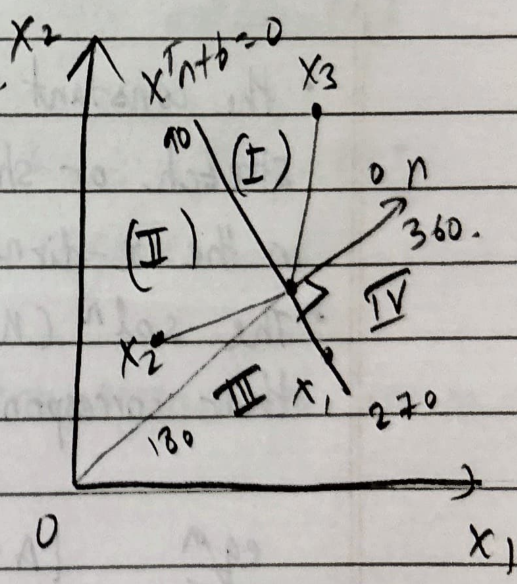
↳ Geometrically, hyperplane is a geometric entity whose dimension is one less than that of its ambient space.

eq^n $\boxed{X^T n + b = 0}$

☆ $X^T n = 0 \Rightarrow$ hyperplane becomes subspace.

⇒ Half Space :





$x^T n + b = 0 \ \forall x \in$ line

$x^T n + b > 0 \ \forall x \in$ subspace in the $\hat{n}$ direction (x_3)

$x^T n + b < 0 \ \forall x \in$ subspace in the $-n$ dir^n (x_2)

Eigen Values and Eigen Vectors

$$\boxed{Ax = \lambda x}$$

- The constant λ (+ve) represents the amount of stretch or shrinkage the attributes x go through in the x-dir^n
- The sol^n (x) are known as eigenvectors and their corresponding λ are eigenvalues

eq^n $|A - \lambda I| = 0$

R code

```
A = matrix(c(8,7,2,3), 2, 2, byrow = TRUE)
ev = eigen(A)
vectors <- ev$vectors
```

- eigenvalues can be complex numbers even for real matrices
- when eigen values become complex, eigenvectors also become complex.
- if matrix is symmetric, then the eigenvalues are always real.
- eigenvectors of symmetric matrices are also real.
- there will always be n linearly independent eigenvectors for symmetric matrices.
- Eigen values of matrices of the form A^TA or AA^T while being real are also non-negative.

- the eigenvectors corresponding to zero eigenvalues are in the null space of the matrix.

- let us assume that there are r eigenvectors corresponding to zero eigenvalue. this means that the null space dimension is r.

↳ that is n-r independent vectors are enough to represent all the vectors in the columns of matrix (column space).