SYMBOL-TABE-PROJECTION

	Symbol	Read as	what it w
①	x	x	original vector
②	$\hat{x}$	x-hat	estimate/close point
③	$proj_U(x)$	projection of x onto U	explicit projection
④	$\pi(x)$	pi of x	same as ③
⑤	$x \perp y$	x is perpendicular to y	vectors are orthogonal
⑥	$U^\perp$	U-perp	Orthogonal perpendicular.
⑦	$e \in U^\perp$	e belongs to U-perp	residual
⑧	$\hat{x} = \frac{x^T u}{x^T u} u$	projection onto line spanned by u	one line case
⑨	$\hat{x} = (x^T q) q$	same, but q is unit vector	simple line
⑩	$\hat{x} = Px$	projection using Projection matrix	line projection
⑪	$\hat{x} = U(U^T U)^{-1} U^T x$	projection onto subspace spanned by columns of U	general subspace case
⑫	e	residual	leftover
⑬	$e \in U^\perp$	residual lies in orthogonal Complement	error is perpendicular.
⑭	$\hat{x} \in U$	x-hat lies in U	closet point

$Ax \approx b$ - Approximation

$\Downarrow$

SO - IMPORTANT

Why

Because Projection / Approximation is every where

Note :; A said, required miss out

x' ⇒ read as x- prime

x'' ⇒ read as x- double-prime



PROJECTION MATRIX

Let,

$\Rightarrow a v_1 + b v_2$

Suppose we have, two vectors,

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} , v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

All combinations

$a v_1 + b v_2$

look like

$$\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$$

This is whole xy plane in 3D

Suppose, (v_1) (v_2)

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

(This is nothing but Column space for above plane)



Let $y = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$, this point is not on xy plane. So we can find projection

$\boxed{\hat{y} = Py}$ where P is projection matrix

Also

$\hat{y} = Proj_{col(X)}(y)$

$\hat{y} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$

y = reality, $\hat{y}$ = the closest thing model can say

$r = y - \hat{y}$ ⇒ what model explained to explain (X) = feature Matrix / Model

For Line we proved

$$Proj_w(x) = \frac{x^T u}{u^T u} \; u = \hat{x}$$

and we want

$$\boxed{\hat{y} = X\theta} \leftarrow \text{In this format}$$

BEFORE WE CONTINUE

$\Rightarrow (AB)C = A(BC)$

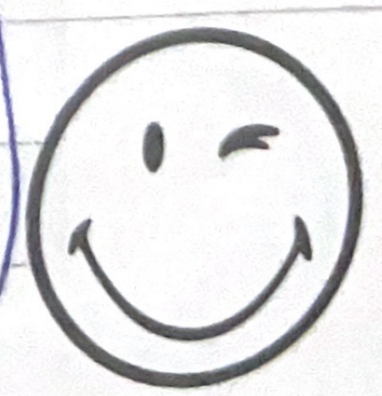
So, $\Rightarrow u(u^T x) = (uu^T)x$

ORDER SHOULD NOT CHANGE

↑

ASSOCIATIVE LAW FOR MATRIX

$$\hat{x} = \frac{u(u^T x)}{(u^T u)} = \frac{(uu^T)x}{(u^T u)} = Px$$



FOR LINE

So $P = \frac{u u^T}{u^T u} \neq 1$

- Denominator $u^T u$ ⇒ Scalar (dot Product)
- Numerator = Matrix

⇓

They are not same. Do not think it will be one.

FOR GENERAL CASE

→ To Study projection as matrix on general case we need to understand matrix inverse first.

(25) What is A^{-1} → Matrix Inverse

⇒ What is meaning of $(X^T X)^{-1}$?

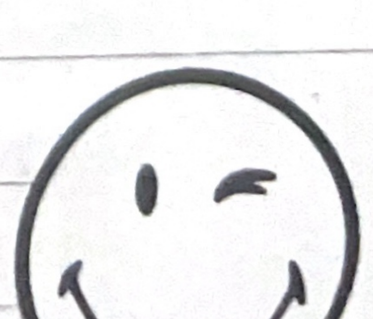
Before we go to inverse of matrix, what is inverse of number

Inverse of number = 3 ⇒ $3^{-1} = \frac{1}{3}$

$3 \times \frac{1}{3} = 1$

Number × Number^{-1} = 1

Same goes with matrix.





Instead numbers like $\Rightarrow 3x = 12$

we have matrix $Ax = b$

$$x = A^{-1}b$$

① ↳ matrix cannot be divided
② ↳ its inverse multiply
③ ↳ gives same effect

Matrix has inverse only if

① ↳ it is square
② ↳ it does not loose information

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad x = \begin{bmatrix} 9 \\ 0 \end{bmatrix}$$

$$Ax = \begin{bmatrix} a \\ 0 \end{bmatrix}$$, b is lost we can never return b , its zero. Inverse dos not make sense here.

CALCULATE INVERSE

① adjugate / cofactor formulae

② row reduction / Gauss–Jordan elimination

$$[A \mid I] \rightarrow [I \mid A^{-1}]$$

Row REDUCTION

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

Steps ⇒ $R_3 \leftarrow R_3 - R_1$

$R_3 \leftarrow R_3 + 2R_2$

$R_3 = \frac{1}{3} R_3$

$R_2 \leftarrow R_2 - R_3$

$R_1 \leftarrow R_1 - 2R_2$



Then,

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/3 & -2/3 & 2/3 \\ 0 & 1 & 0 & 1/3 & 1/3 & -1/3 \\ 0 & 0 & 1 & -1/3 & 2/3 & 1/3 \end{array}\right]$$

This part is A^{-1}

3x3 Matrix with No Inverse

$$B = \left[\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 0 & 1 & 1 \end{array}\right]$$

① Notice row 2 is double row 1
② two rows dependent
③ Means matrix loses information

⇓

{{ Its determinant is Zero and inverse does not exists.

** ⇒ For mat 3×4, no inverse exist

For non squared matrix we have pseudoinverse

(26) $A^+ \Rightarrow$ Pseudoinverse

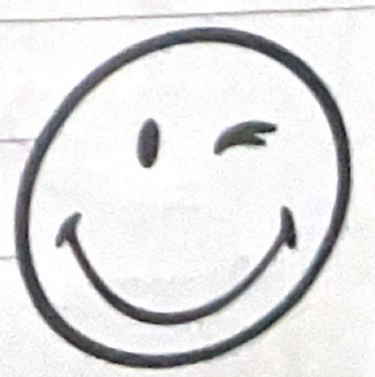
It is not ordinary inverse, but behaves like best possible inverse in least square problem.

This is why, in regression, when matrices are not square, we don't invert X directly

We instead work with

$$(X^T X)^{-1} X^T$$

We will talk about it later





(27) $|A|$ → Determinant

$||A||$ ⇒ Normalization ⇒ A^TA

But

$|A|$ ⇒ Determinant

DO NOT CONFUSE

⇒ It is required to know if inverse is possible.

Thought is interested only in its own survival

This kills Maths : OPEN BRAIN

THINK DIFFERENTLY

All this comes up later.

Determinants / Inverse / Cofactor / Minimum Adjugat

⇒ BACK TO PROJECTION

Remember, $r = y - \hat{y}$

From, $X\theta = y = \hat{y} + r$

Small $\|r\|$ = good fit
Large $\|r\|$ = bad fit

⇐ RULE





(28) $\min_\theta \|X\theta - y\|^2$ Least-Square.

$\Rightarrow$ Choose θ so that the model output $X\theta$ is as close as possible to the target vecto y.

Why SQUARE

- It keeps the same minimizer
- It is algebraically easier
- It punishes big errors more strongly

SHAPE-THINKING

• $X \in R^{N \times D}$	• $y \in R^N$
• $\theta \in R^D$	• Difference in R^N
• $X\theta \in R^N$	• Norm Squared is scalar

Now,

$$X^T X\theta = X^T y$$ → RIGHT

Remember, If $U \subseteq R^n$,

① $\hat{x} = proj_U(x)$

② $error = x - \hat{x}$

③ $z = x - \hat{x} \in U^\perp \Rightarrow x = \hat{x} + z$

④ $\hat{x} \in U$

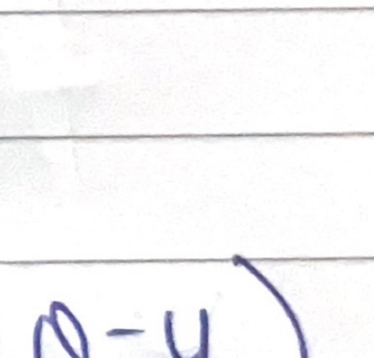
Let, $f(\theta) = \|X\theta - y\|^2$

Write it as, $f(\theta) = (X\theta - y)^T (X\theta - y)$

Expand,

$$f(\theta) = \theta^T X^T X\theta - 2y^T X\theta + y^T y$$

$$\frac{d(f(\theta))}{d\theta} = 2X^T X\theta - 2X^T y = 0$$



$$X^T X\theta = X^T y$$

Again

Transpose Rules

① $(A-B)^T = A^T - B^T$

So $(X\theta - y)^T = \theta^T X^T - y^T$

② $(AB)^T = B^T A^T$

So $(X\theta)^T = \theta^T X^T$

Hence

$$(X\theta - y)^T (X\theta - y) = (\theta^T X^T - y^T)(X\theta - y)$$

Expand

$$= \theta^T X^T X\theta - \theta^T X^T y - y^T X\theta + y^T y$$

Middle two terms are Scalar,
How? Thats your home wok ⇒ U-do.

Now,

$$r = y - X\theta$$

For best θ, the residual must be orthogonal to every column of X

That means

$$X^T r = 0$$

Substitute $r = y - X\theta$

$$X^T(y - X\theta) = 0$$

Distribute X^T

$$X^T y - X^T X\theta = 0$$

So $\boxed{X^T y = X^T X\theta}$ Again





$$\hat{\theta} = (X^T X)^{-1} X^T y$$

$$P = X (X^T X)^{-1} X^T$$

WE WILL STUDY MORE ON PROJECTIONS IN

IT'S NOT YET OVER