Lecture 13:

Least Square Approximation:

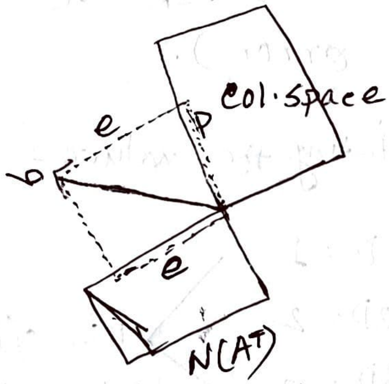
①

Recap of the previous lecture →

If b is in the column space, then $Pb = b$

If $b \perp$ column space, then $Pb = 0$

Do you remember in which space the error vector, e is?



Pb

$(I-P)b$

$p + e = b$

Why do we need projection?

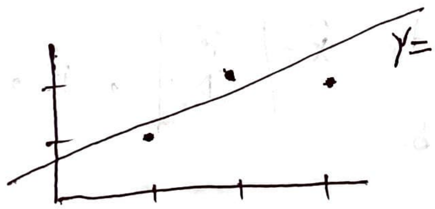
→ It often happens that $Ax=b$ has no solution. The usual reason is → too many equations. When error, $e = b - Ax$ down to zero, then x is an exact solution to $Ax=b$.

But if 'b' isn't in the columnspace of A, we try to get $\hat{x}$ by multiplying ~~AI~~ A^T in the both side of $Ax=b$.

So, when $Ax=b$ has no solution, multiply A^T and solve $A^TA\hat{x} = A^Tb$

Here's a good application of projection, "Fitting a straight Line into m points."

Suppose you have three points, (1,1), (2,2), (3,2)

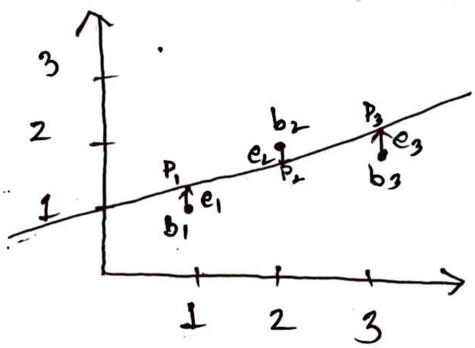


y = how far up
x = how far along
m = slope/gradient
b = the y intercept

This line is also called least squares Regression Line.

→ If your data shows a linear relationship between the X and Y variables, you will want to find the line that best fits this linear relationship. That line is called a regression line. This line makes the vertical distance from the data points as small as possible, It's called least squares because the best line of fit is one that minimizes ~~the~~ the variance. (the sum of the squares of the errors).

NOW,



Substituting the values –

$$C + D = 1$$
$$C + 2D = 2$$
$$C + 3D = 2$$

this system doesn't have a solution. Because?

But it does have a best solution.

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

What do you need to find? $\hat{x}$.
Sounds familiar?

so, $\|e\|^2 = \|Ax - b\|^2$ → we have to minimize this.

$\Rightarrow e_1^2 + e_2^2 + e_3^2$

If there was a fourth data point above the vertical 3, it will be called an outlier.

Suppose, the points on the line from the data points are the projection points P_1, P_2, P_3.

so, we have to find, $\hat{x} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$, P, P?

so, we have $A^T A \hat{x} = A^T b$

Now the variance can be minimized by 3 ways –

1. geometry → Finding the nearest point
2. algebra
3. calculus.

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}; \quad A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

so,
$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

$$\Rightarrow \quad 3\hat{C} + 6\hat{D} = 5$$
$$6\hat{C} + 14\hat{D} = 11$$

$$2D = 1 \Rightarrow \boxed{D = 1/2, \quad C = 2/3}$$

Best line, $2/3 + 1/2\, t = y$

How to do it by calculus? we know $e_1^2 + e_2^2 + e_3^2 = (C+D-1)^2 + (C+2D-2)^2 + (C+3D-2)^2$

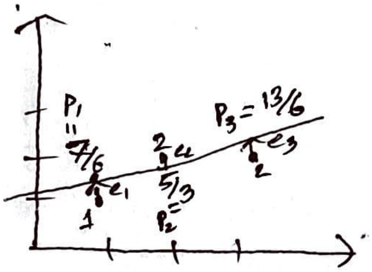
How do you minimize this? You take the partial derivative with respect to C and D.

so, $dE/dC = 2(C + D.D - \cancel{1}) + 2(C. \ldots\ldots$

$dE/dD = \ldots\ldots\ldots$

~~You'll~~ You'll get the same two equations.

So, going back to the previous picture,



$$y = 2/3 + 1/2\, t$$
$t \to x$

$$e_1 = b_1 - p_1 = = 1/6 \; ; \quad e_2 = -2/6 \; ; \quad e_3 = 1/6$$

$2 - \frac{5}{3}$

$\frac{6-5}{3} = -1/3$

So, as, $b = p + e$

$$\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 7/6 \\ 5/3 \\ 13/6 \end{bmatrix} + \begin{bmatrix} -1/6 \\ 2/6 \\ -1/6 \end{bmatrix}$$

$e \perp p \rightarrow$ can you prove?

Also, $e \perp A \rightarrow$ Can you prove?

A little question for you $\rightarrow$ I wanna prove that A^TA is invertible.

If 'A' has independent columns, then A^TA is invertible.

<u>Proof:</u> If ~~A~~ A^TA isn't invertible, then

~~suppose,~~ $A^TAx = 0$

$\Rightarrow x^T A^T A x = 0$

$\Rightarrow (Ax)^T (Ax) = 0$

$\Rightarrow Y^T Y = 0$

$\hookrightarrow Y = 0$ [Because vectors are getting multiplied here]

$\therefore Ax = 0.$

As, A has independent columns, so $x = 0$