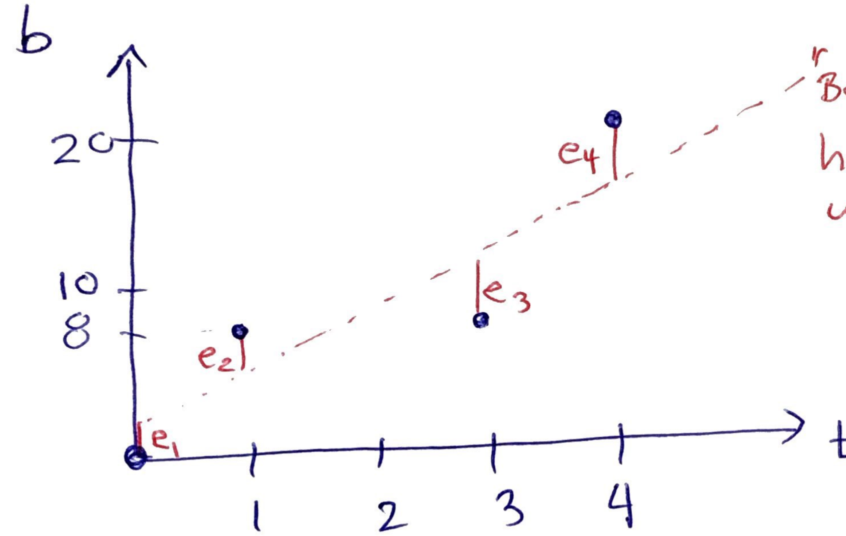
Linear Regression

Data: (time, observation)
t $\quad$ b

$(0,0), \ (1,8), \ (3,8), \ (4,20)$



"Best fit" line: $y = C + Dt$
here, C and D are unknown real numbers.

Ans: $y = 1 + 4t$

We want to fit a line. How best to do it? The line equation is $y = C + Dt$.

Ideally, our best-fit line passes through all the points:

$$\begin{cases} C + 0 \cdot D = 0 \\ C + 1 \cdot D = 8 \\ C + 3D = 8 \\ C + 4D = 20 \end{cases}$$

This is a system of 4 linear eqns in 2 unknowns (C and D). We can write it in matrix form as

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}}_{A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \end{bmatrix}} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}}_{\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}}$$

or $A\vec{x} = \vec{b}$

Bad news: this system is "overdetermined". There is no choice of C and D which yields equality $A\vec{x} = \vec{b}$. After all, our points <u>do</u> <u>not</u> <u>lie</u> <u>on</u> <u>a</u> <u>line</u>.

Instead, what we aim for is

$$A\vec{x} \approx \vec{b}$$

or equivalently that

$$\underbrace{\vec{b} - A\vec{x}}_{\text{This is a vector. We call it } \vec{e}.} \approx 0$$

~~Called the vector of residuals.~~

$$\vec{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \cancel{= \begin{bmatrix} b_1 - (C+Dt_1) \\ b_2 - (C+Dt_2) \\ b_3 - (C+Dt_3) \\ b_4 - (C+Dt_4) \end{bmatrix}}$$

where $e_i = b_i - (C + Dt_i)$

In our case,

$$\vec{e} = \begin{bmatrix} 0 - C \\ 8 - (C+D) \\ 8 - (C+3D) \\ 20 - (C+4D) \end{bmatrix}$$

This is ~~a~~ the vector of vertical distances, (known as residuals).

We want ~~to minimize~~ this vector to be close to 0.

One approach ~~[illegible]~~ is to find values of C and D which <u>minimize the length</u> of the residual vector $\vec{e}$.

The length of a vector V is $\|V\| = \sqrt{V_1^2 + V_2^2 + \cdots V_n}$

Minimizing a length is ~~[illegible]~~ equivalent to minimizing its square, which allows us to ~~[illegible]~~ fend off the pesky square root. So our new goal is:

$$\boxed{\text{find } C, D \text{ which minimize } \|e\|^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2.}$$

In our problem,

$$\|\vec{e}\|^2 = C^2 + (8-C-D)^2 + (8-C-3D)^2 + (20-C-4D)^2$$

Use calculus: Call ~~RHS~~ $f(C,D)$

~~Step 1 Set derivative~~

Step 1 Set $\nabla f = \vec{0}$ (here $\nabla = \begin{bmatrix} \frac{\partial}{\partial C} \\ \frac{\partial}{\partial D} \end{bmatrix}$)

$f_C = 0$

$$0 = 2C + 2(8-C-D)(-1) + 2(8-C-3D)(-1) + 2(20-C-4D)(-1)$$

$$= 2C - 16 + 2C + 2D - 16 + 2C + 6D - 40 + 2C + 8D$$

$$= -72 + 8C + 16D$$

$$0 \overset{\text{set}}{=} \frac{\partial f}{\partial D} = -2(8-C-D) - 6(8-C-3D) - 8(20-C-4D)$$

$$= -16 + 2C + 2D - 48 + 6C + 18D - 160 + 8C + 32D$$

$$= -224 + 16C + 52D$$

We have a system of equations

$$\begin{cases} 8C + 16D = 72 \\ 16C + 52D = 224 \end{cases}$$

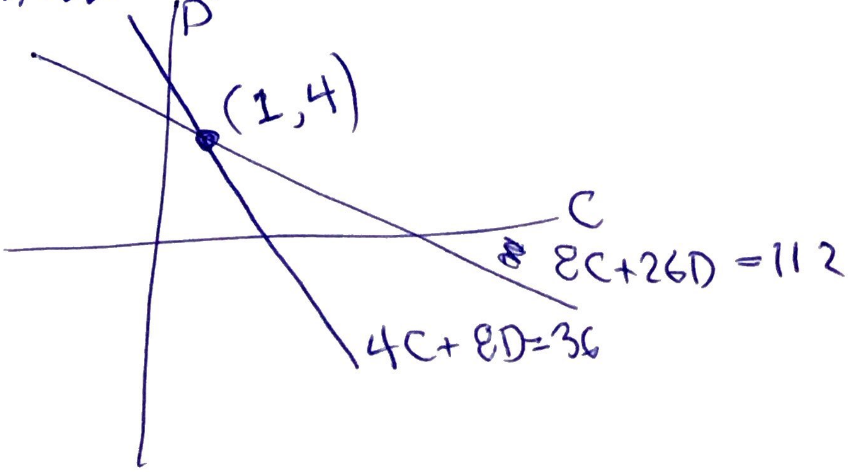
Dividing both equations by 2 (to simplify) gives

$$\left\{\begin{array}{l} 4C + 8D = 36 \\ 8C + 26D = 112 \end{array}\right\}$$

★ Normal Eqns

Solving this system gives $(C,D)=(1,4)$

$\boxed{Q:}$ How do we know this critical point is a minimum?



Recall $A x \approx b$, with $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$,

and unknown $\vec{x} = \begin{bmatrix} C \\ D \end{bmatrix}$.

<u>Transpose:</u>

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix}.$$

If $A\vec{x} \approx \vec{b}$ then $A^TA\vec{x} \approx A^T\vec{b}$

- $$A^TA = \overset{(2\times 4)}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix}} \overset{(4\times 2)}{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}} = \overset{(2\times 2)}{\begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}}$$

$\cancel{A^TA\vec{x} \approx A^T\vec{b}}$

- $$A^T\vec{b} = \overset{(2\times 4)}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix}} \overset{(4\times 1)}{\begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}} = \overset{(2\times 1)}{\begin{bmatrix} 36 \\ 112 \end{bmatrix}}$$

So $A^TA\vec{x} = A^Tb$ is the equation

$$\begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \end{bmatrix} \quad \text{or} \quad \begin{cases} 4C + 8D = 36 \\ 8C + 26D = 112 \end{cases}$$

The normal equations!

If A has L.I. columns

<u>Theorem</u>: The vector $\vec{x}$ that minimizes $\|A\vec{x}-\vec{b}\|^2$ is the solution to the <u>normal eqns</u>

$$A^TA\vec{x}=A^T\vec{b}.$$

The vector $\vec{x}=(A^TA)^{-1}A^T\vec{b}$ is the <u>least squares solution</u> to $A\vec{x}=\vec{b}$.

Remember, $\vec{x}=\begin{bmatrix} C \\ D \end{bmatrix}$ so finding $\vec{x}$ gives us the best-fit line.

<u>Easy to summarize</u>:

If $A\bar{x}=\bar{b}$ has no solution, multiply by A^T and solve

$$A^TA\vec{x}=A^T\bar{b}$$