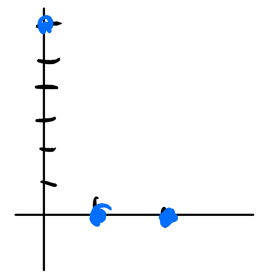
Least Squares or Almost Solving $Ax=b$

Eg: You measure 3 data points
$(0,6)$, $(1,0)$, $(2,0)$
that are supposed to lie on a line
(but they don't due to measurement error)



Q: What line do they want to be on?

Line: $y = Cx + D$

$6 = C\cdot 0 + D$ $\quad$ $(0,6)$ is on the line

$0 = C\cdot 1 + D$ $\quad$ $(1,0)$ is on the line

$0 = C\cdot 2 + D$ $\quad$ $(2,0)$ is on the line

$\rightsquigarrow Ax = b \qquad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \qquad x = \begin{pmatrix} C \\ D \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$

y-values

Inconsistent! What to do?

Key idea: $Ax=b$ has a solution $\iff b \in \mathrm{Col}(A)$

Solve $A\hat{x} = \hat{b}$ where $\hat{b} \in \mathrm{Col}(A)$ is as close to b as possible.

$$\hat{b} = b_V = \text{orthogonal projection onto } V = \mathrm{Col}(A)$$

We know how to compute $\hat{b}$: solve

$A^TA\hat{x} = A^Tb \rightsquigarrow \hat{b} = A\hat{x}$

and we have $\hat{x}$ too.

Upshot: Replace the inconsistent system $Ax=b$ by the normal equation $A^TA\hat{x}=A^Tb$ consistent.

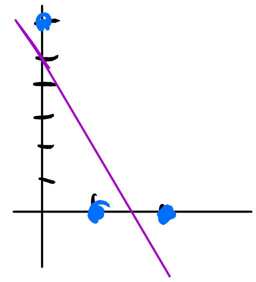
Eg: $A=\begin{pmatrix}0&1\\1&1\\2&1\end{pmatrix}$ $b=\begin{pmatrix}6\\0\\0\end{pmatrix}$ $A^TA=\begin{pmatrix}5&3\\3&3\end{pmatrix}$ $A^Tb=\begin{pmatrix}0\\6\end{pmatrix}$

$A^TA\hat{x}=A^Tb$: $\left(\begin{array}{cc|c}5&3&0\\3&3&6\end{array}\right)\xrightarrow{\text{RREF}}\left(\begin{array}{cc|c}1&0&-3\\0&1&5\end{array}\right)\rightsquigarrow\hat{x}=\begin{pmatrix}-3\\5\end{pmatrix}$

$C=-3 \quad D=5$

best-fit line:

$y=-3x+5$



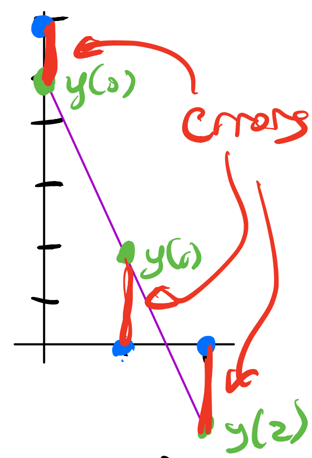
Q: What is being minimized by this line?

$y=-3x+5=Cx+D$

y-values I got

$$\hat{b}=A\hat{x}=\begin{pmatrix}0&1\\1&1\\2&1\end{pmatrix}\begin{pmatrix}C\\D\end{pmatrix}=\begin{pmatrix}0C+D\\C+D\\2C+D\end{pmatrix}=\begin{pmatrix}y(0)\\y(1)\\y(2)\end{pmatrix}=\begin{pmatrix}5\\2\\-1\end{pmatrix}$$

$b=\begin{pmatrix}6\\0\\0\end{pmatrix}=$ y-values I wanted



Minimized: $\|b-\hat{b}\|^2$

$$=\left\|\begin{pmatrix}6-y(0)\\0-y(1)\\0-y(2)\end{pmatrix}\right\|^2=\left\|\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\|^2$$

$\hookrightarrow=(6-y(0))^2+(0-y(1))^2+(0-y(2))^2$

"least squares"

Procedure (least squares): Start with matrix eqn $Ax=b$.

(1) Solve the normal equation $A^TA\hat{x}=A^Tb$

A solution $\hat{x}$ is a least squares solution of $Ax=b$.

It minimizes $\|A\hat{x}-b\|$ over $\hat{x}\in\mathbb{R}^n$.

↳ best approximate solution

→ remember $\hat{b}=A\hat{x}$ is the closest vector to b in $\mathrm{Col}(A)$.

Eg: Find the best-fit parabola thru $(-1,\tfrac{1}{2})$, $(1,-1)$, $(2,-\tfrac{1}{2})$, $(3,2)$.

Parabola: $y=Bx^2+Cx+D$

$\tfrac{1}{2}=B(-1)^2+C(-1)+D$ $(-1,\tfrac{1}{2})$ lies on parabola

$-1=B(1)^2+C(1)+D$ $(1,-1)$ lies on parabola

$-\tfrac{1}{2}=B(2)^2+C(2)+D$ $(2,-\tfrac{1}{2})$ lies on parabola

$2=B(3)^2+C(3)+D$ $(3,2)$ lies on parabola

$$\rightsquigarrow \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix}\begin{pmatrix} B \\ C \\ D \end{pmatrix}=\begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix}$$

Least-squares:

$$A^TA=\begin{pmatrix} 99 & 35 & 15 \\ 35 & 15 & 5 \\ 15 & 5 & 4 \end{pmatrix} \qquad A^Tb=\begin{pmatrix} 31/2 \\ 7/2 \\ 1 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 99 & 35 & 15 & 31/2 \\ 35 & 15 & 5 & 7/2 \\ 15 & 5 & 4 & 1 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 53/88 \\ 0 & 1 & 0 & -379/440 \\ 0 & 0 & 1 & -41/44 \end{array}\right)$$

$$\hat{x} = \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix} \rightsquigarrow y = \frac{53}{88}x^2 - \frac{379}{440}x - \frac{41}{44}$$

$$\hat{b} = A\hat{x} = \begin{pmatrix} y(-1) \\ y(1) \\ y(2) \\ y(3) \end{pmatrix} \qquad b = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix}$$

Rem: (1) What if $Ax=b$ has a solution?

$b \in Col(A) \Rightarrow \hat{b} = b \Rightarrow A\hat{x} = \hat{b} = b \Rightarrow x = \hat{x}$

least squares solutions are ordinary solutions.

(2) When is $\hat{x}$ unique?

$Ax = \hat{b}$ has a unique solution

$\Longleftrightarrow$ A has full column rank.

NB $Nul(A^TA) = Nul(A)$

so A^TA has FCR $\Longleftrightarrow$ A has FCR

Orthogonal Column

Def: Nonzero vectors $\{u_1, \dots, u_n\}$ is called

- orthogonal: $u_i \cdot u_j = 0$ when $i \neq j$
- orthonormal: orthogonal & $u_i \cdot u_i = 1$

(orthogonal unit vectors)

Matrix interpretation

$$Q = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{pmatrix} \rightsquigarrow Q^TQ = \begin{pmatrix} u_1 \cdot u_1 & u_1 \cdot u_2 & \cdots & u_1 \cdot u_n \\ u_2 \cdot u_1 & u_2 \cdot u_2 & \cdots & u_2 \cdot u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n \cdot u_1 & u_n \cdot u_2 & \cdots & u_n \cdot u_n \end{pmatrix}$$

- orthogonal: Q^TQ is diagonal (& invertible)
- orthonormal: $Q^TQ = I_n$

$\hookrightarrow$ Q: does this mean $Q^{-1} = Q^T$? Only if square.

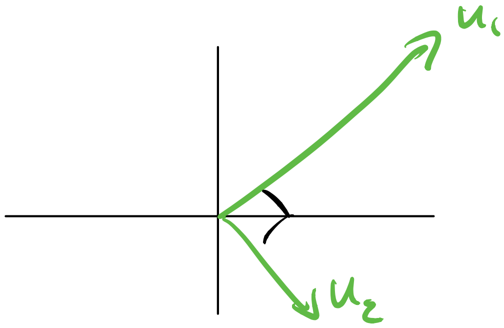
Eg: $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$ is orthogonal: $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = 0$

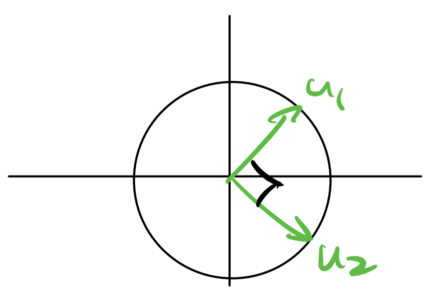
$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 4 \quad \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = 4 \Rightarrow$ not orthonormal.

Divide by lengths: $\left\{ \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$ is orthonormal.

NB: Can always turn an orthogonal set into an orthonormal one by dividing by the lengths.

Picture in $\mathbb{R}^2$:





Properties: Say Q has orthonormal columns.

(1) $Q^TQ = I_n$ ✓

(2) Q has full column rank:

$\rightarrow \mathrm{Nul}(Q) = \mathrm{Nul}(Q^TQ) = \mathrm{Nul}(I_n) = \{0\}$ ✓

(3) $(Qx)\cdot(Qy) = x\cdot y$ "Q· doesn't change angles"

$\rightarrow (Qx)\cdot(Qy) = (Qx)^T(Qy) = x^T \underbrace{Q^T Q}_{} y$

$= x^T I_n y = x^T y = x\cdot y$

(4) $\|Qx\| = \|x\|$, ie "Q· doesn't change lengths"

$\rightarrow \|Qx\|^2 = (Qx)\cdot(Qx) \overset{(3)}{=} x\cdot x = \|x\|^2$

(5) If $V = Col(Q)$ and P_V = projection matrix onto V then

$$\boxed{P_V = QQ^T}$$

$\rightarrow P_V = Q(Q^TQ)^{-1}Q^T = QI_nQ^T = QQ^T$