Lesson 11

Chapter 4

Projections

Linear regression

Recall $V, V^\perp \le \mathbb{R}^n$

$\mathcal{B} = b_1 \dots b_k$ basis for V ; $A = \begin{bmatrix} b_1 \cdots b_k \end{bmatrix}$

$P_V \quad \mathbb{R}^n \to \mathbb{R}^n$ orthogonal projection on V

$P_V(b) = b_V$ if $b = b_V + e$, b_V in V, e in $V^\perp$

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

1) Since $b = b_V + e$ $\quad b_V \in V \quad e \in V^\perp$

$b_i^T b = b_i^T b_V$ for $i = 1, \dots, k$

2) $A^T b = A^T b_V$ for all b in $\mathbb{R}^n$

3) $b_V = x_1 b_1 + \cdots + x_k b_k = A \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} b_1 \cdots b_k \end{bmatrix} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}}_{\hat{x}}$

4) If $\mathcal{B}$ is orthonormal

$b_i^T b_V = b_i^T \cdot b = x_i$

$A^T b_V = \begin{bmatrix} b_1^T \\ b_2^T \\ b_k^T \end{bmatrix} b_V = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = A^T b$ so

$P_V(b) = A A^T b = \begin{bmatrix} b_1 \cdots b_k \end{bmatrix} \begin{bmatrix} b_1^T \\ \vdots \\ b_k^T \end{bmatrix} b = \begin{bmatrix} b_1 \cdots b_k \end{bmatrix} \begin{bmatrix} b_1^T b \\ b_2^T b \\ \vdots \\ b_k^T b \end{bmatrix}$

$= b_1^T b \cdot b_1 + b_2^T b \, b_2 + \cdots + b_k^T b \cdot b_k$

5) If B is not orthonormal

$A^T b = A^T b_V = A^T A \hat{x}$

$A^T A$ is invertible so

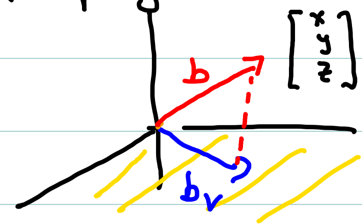
$(A^T A)^{-1} A^T b = \hat{x}$

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

Example

Given $V = xy$ plane in $\mathbb{R}^3$, find P

(matrix of orthogonal projection on V)



1) Find basis for V : $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

Notice that this basis is orthonormal: all vectors in it have length 1 and are pairwise orthogonal

2) Write $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

3) $P = A(A^T A)^{-1} A^T$:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \Bigg(\overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}^{I_2} \Bigg)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Easy because basis was orthonormal

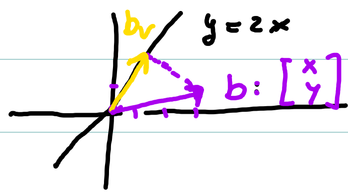
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = P \quad ; \quad P \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

Check that $P^2 = P$ $\quad P = P^T$

Alternatively:

$$P_V \begin{bmatrix} x \\ y \\ z \end{bmatrix} = [x\ y\ z] \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + [x\ y\ z] \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Orthogonal projection on line $y = 2x$:



1) Find a basis for V

$$B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

2) $A = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

3) $$P = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \Big(\underbrace{\underset{1\times 2}{\begin{bmatrix} 1 & 2 \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}}_{1\times 1} \Big)^{-1} \begin{bmatrix} 1 & 2 \end{bmatrix}$$

$$P = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \underset{1\times 2}{\begin{bmatrix} 1 & 2 \end{bmatrix}} = \frac{1}{5} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix};$$ so for example

The orthogonal projection of $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ is :

$$P \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \underset{2\times 1}{\begin{bmatrix} 4 \\ 1 \end{bmatrix}} = \begin{bmatrix} 6/5 \\ 12/5 \end{bmatrix};$$

we could also write

$$P \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad ; \quad 5 = \left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\|^2$$

$$\frac{1}{\|b_1\|^2} b_1^T \begin{bmatrix} x \\ y \end{bmatrix} \cdot b_1$$

$$\left(\frac{b_1}{\|b_1\|} \right)^T \begin{bmatrix} x \\ y \end{bmatrix} \cdot \frac{b_1}{\|b_1\|}$$

Note : $P = P^2$, $P = P^T$

$$\mathrm{Null}(P) = \mathrm{span}\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix} \right) = \mathrm{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)^{\perp}$$

In general the orthogonal projection in $\mathbb{R}^n$ of b on line through origin parallel to b_1 $(V = \text{span}(b_1))$ is:

$$Pb = \frac{1}{\|b_1\|^2} b_1^T b \cdot b_1$$

if $\|b_1\| = 1$ $\quad Pb = b_1^T b \cdot b_1$

Example:

Given $P = \begin{bmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{bmatrix}$.

Is it the matrix of a projection ? in hw "projector" ? $P = P^2$ yes

Is it orthogonal ? $P = P^T$? No

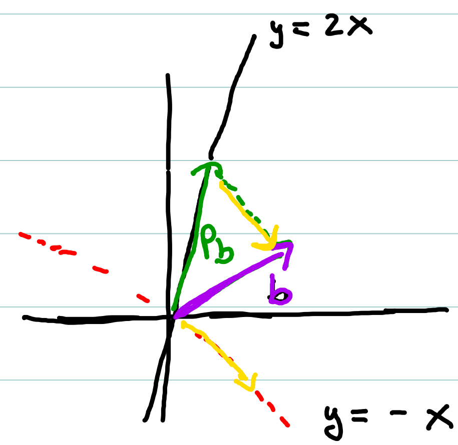
Where are we projecting ? $V = \text{col}(P) = \text{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)$

How are we projecting ? In the direction of $\text{Null}(P) = \text{span}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$

Note $\text{NULL}(P) \not\perp \text{Col}(P)$, but $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is a basis for $\mathbb{R}^2$

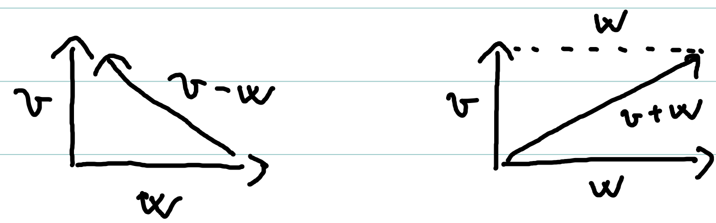
Therefore any b in $\mathbb{R}^2$ can be written as :

$$b = c_1 \underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{\in \text{Col}(P)} + c_2 \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\in \text{NULL}(P)}$$



Pythagorean th: if $v, w \in \mathbb{R}^n$ and $v \perp w$ then

$$\| v - w \|^2 = \| v \|^2 + \| w \|^2$$
$$\| v + w \|^2 = \| v \|^2 + \| w \|^2$$



Proof: $\| v - \underline{w} \|^2 =$

$$= (v-w)^T (v-w) = (v^T - w^T)(v-w) = v^T v - \underbrace{v^T w}_{0} - \underbrace{w^T v}_{0} + w^T w$$

$$= \| v \|^2 + \| w \|^2$$

$\| v + w \|^2 =$

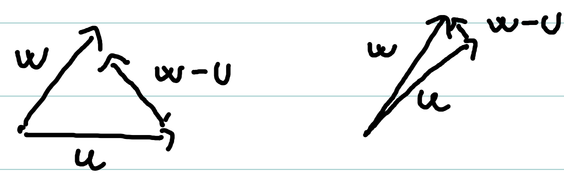
$$= (v+w)^T (v+w) = (v^T + w^T)(v+w) = v^T v + \underbrace{v^T w}_{0} + \underbrace{w^T v}_{0} + w^T w$$

$$= \| v \|^2 + \| w \|^2$$

Th: Let V be a subspace of $\mathbb{R}^n$, and let b_V be the orthogonal projection of b on V then for all $b \in \mathbb{R}^n$ we have

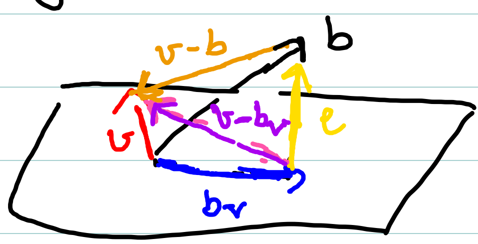
$$\| b_V - b \| \leq \| v - b \|$$

for all v in V. That is, the vector of V that is closest to b is b_V.



Length of $w-u$ tells us how close w and u are.

Proof: $b = b_V + e$ $\qquad e \in V^\perp$



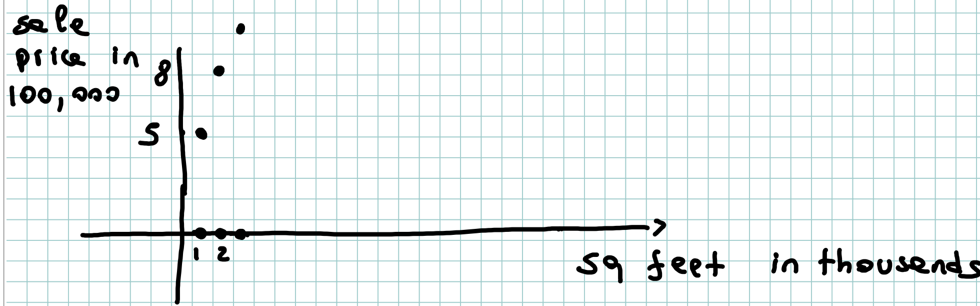
1) $\| b_V - b \|^2 = \| -e \|^2 = \| e \|^2$

2) $\| v - b \|^2 = \| \underbrace{v - b_V}_{\text{in } V} \underbrace{- e}_{\text{in } V^\perp} \|^2 = \| v - b_V \|^2 + \| e \|^2$ Pythagorean th

$\| v - b_V \| > 0$ if $v \neq b_V$

therefore $\| b_V - b \| \leq \| v - b \|$

Linear regression



(1, 5) (2, 8) (3, 10)

I would like to find a line $C + Dx = y$ going through my 3 points

$C + D = 5$ point 1
$C + 2D = 8$ point 2
$C + 3D = 10$ point 3

I want to solve:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 10 \end{bmatrix}$$

$A \quad x \quad = \quad b$ No solutions

b not in col (A)

Which vector in col (A) is closest to b?

Idea: $V = \text{col}(A)$

Replace b with b_V the projection of b on V

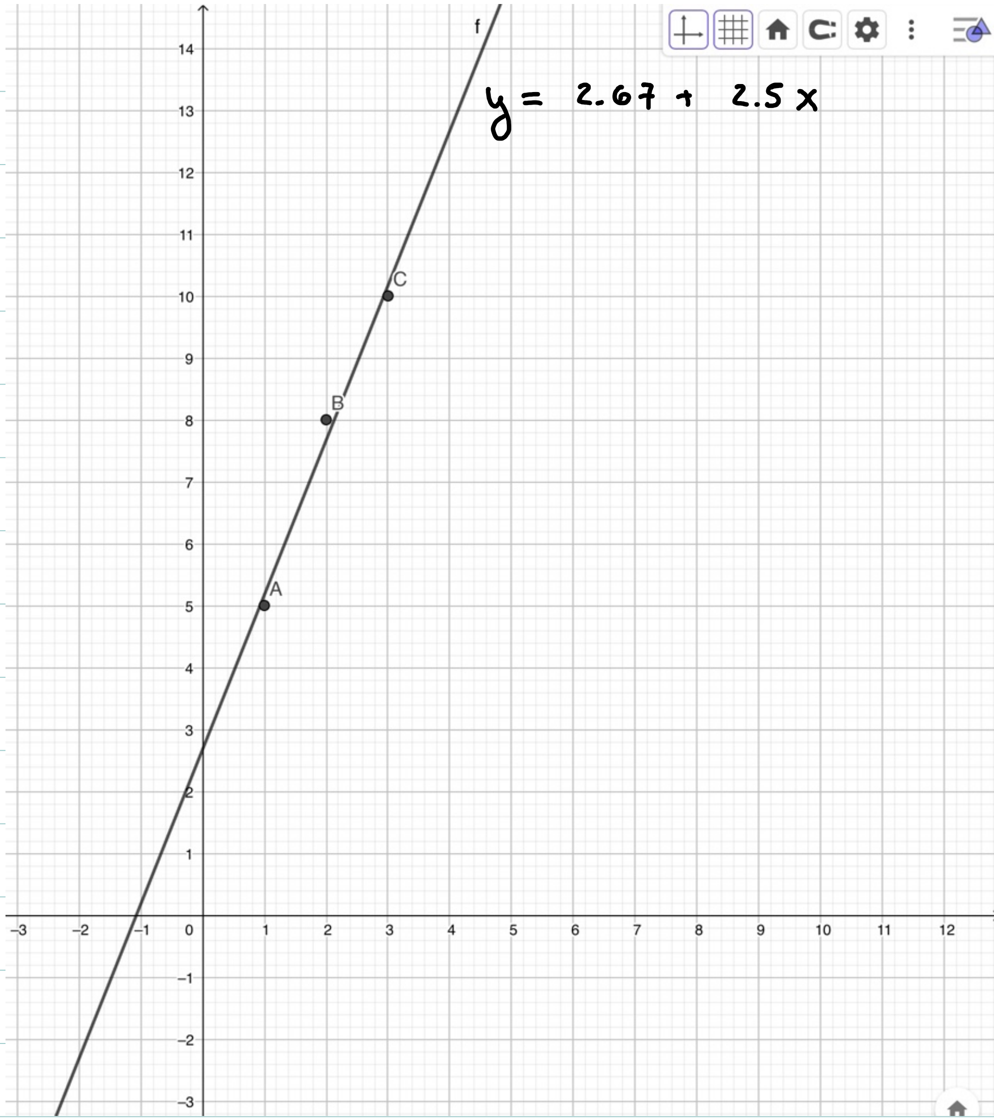
and solve $Ax = b_V$ instead of $Ax = b$

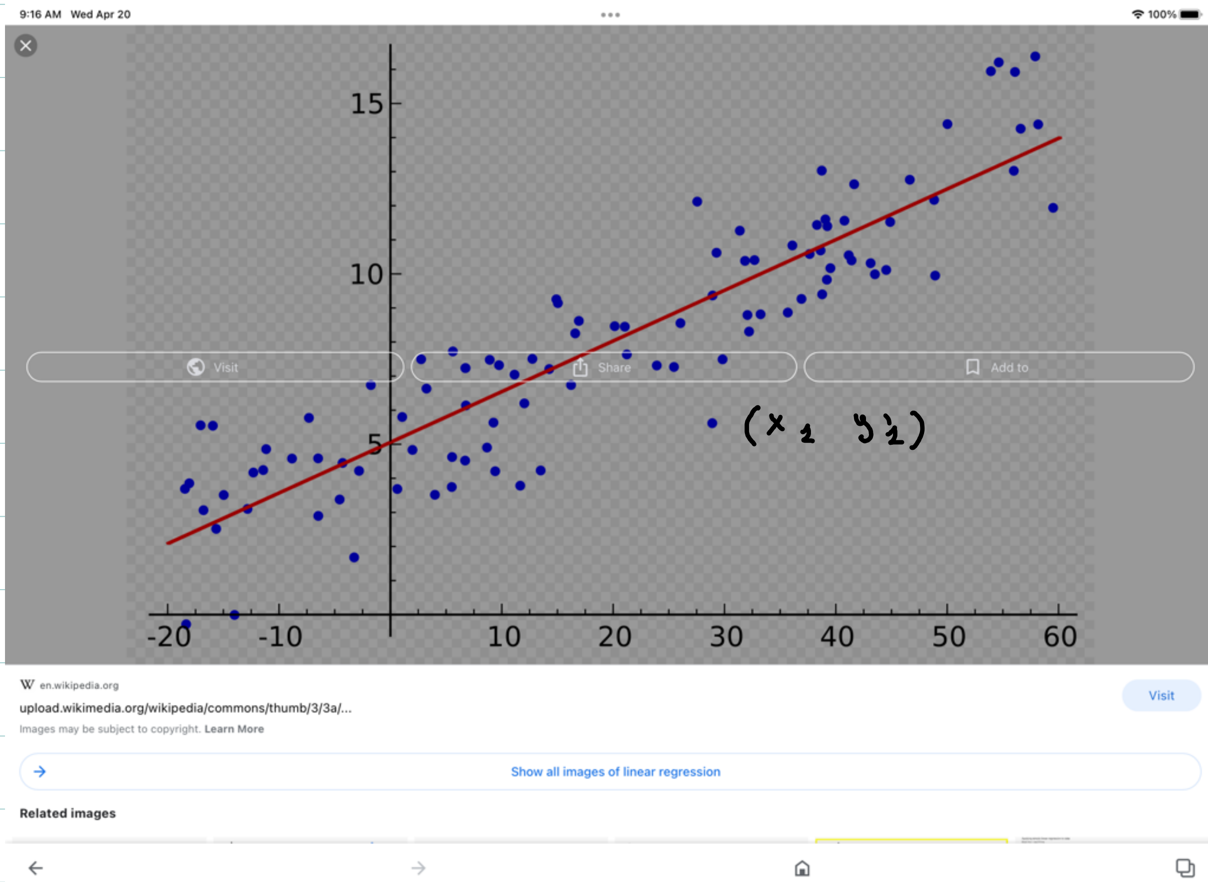
solution is $\hat{x}$

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

so $$\hat{x} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 8 \\ 10 \end{bmatrix} = \begin{bmatrix} 2.67 \\ 2.5 \end{bmatrix}$$

$$y = 2.67 + 2.5\,x$$


f
A
B
C
y = 2.67 + 2.5x



To find linear regression line

$A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ $\qquad b = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

x coordinates of data points

y coordinates of data point

Linear regression line is $y = c + dx$ with $\begin{bmatrix} c \\ d \end{bmatrix} = (A^T A)^{-1} A^T b$

It is line "close" to data points. What does it mean?

I would like

$$A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$ that is $$\begin{aligned} c + d x_1 &= y_1 \\ c + d x_2 &= y_2 \\ &\vdots \\ c + d x_n &= y_n \end{aligned}$$

usually system has no solutions

solve $Ax = b_V$ (projection of b on $V = \mathrm{col}(A)$)

solution: $\hat{x} = \begin{bmatrix} c \\ d \end{bmatrix} = (A^T A)^{-1} A^T b$,

Linear regression line is $y = c + dx$

Recall $\|\overbrace{A\hat{x}}^{bv} - b\| \leq \|\overbrace{Ax}^{\text{vector in } V} - b\|$ for all x

Linear regression line $y = C + Dx$

$$A\hat{x} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} c + dx_1 \\ c + dx_2 \\ c + dx_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

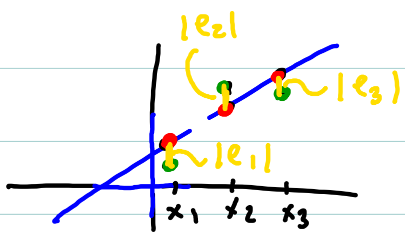
y coordinates of points on regression line

$$\|A\hat{x} - b\| = \left\| \begin{bmatrix} z_1 - y_1 \\ z_2 - y_2 \\ z_n - y_n \end{bmatrix} \right\| = \left\| \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \right\| = \sqrt{e_1^2 + e_2^2 + \cdots + e_n^2}$$

Least squares approx

$$\leq \qquad \leq$$

$$\|Ax - b\| = \left\| \begin{bmatrix} * - y_1 \\ * - y_2 \\ *_1 - y_n \end{bmatrix} \right\|$$



Green: data points

$(x_1\ y_1)\ (x_2\ y_2)\ (x_3\ y_3) \ldots$

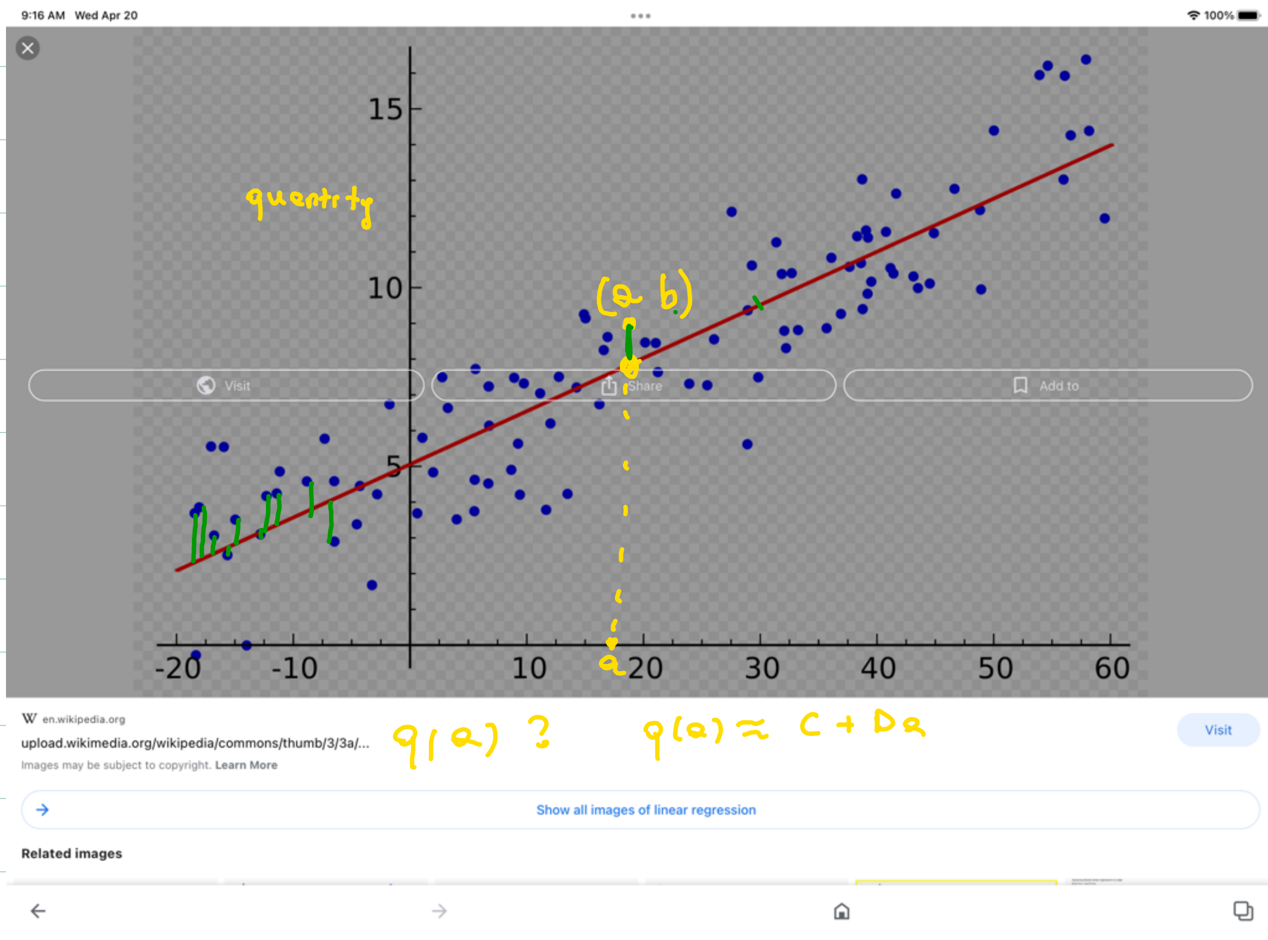
Red points on regression line

$(x_1\ z_1)\ (x_2\ z_2)\ (x_3\ z_3) \ldots$

same x coordinates

$$b = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

The line minimizes the sum of the squares of the (vertical) distances between original data points and points on line with same coordinates.



red line minimizes the sum of the squares of all vertical distances (in green).

I did not draw all vertical distances in the picture:

for any data point (x_1, y_1) there is a vertical distance $z_1 - y_1$ where (x_1, z_1) is the point on the regression line with x coordinate x_1.

It makes sense to use this line to estimate the results of an experiment with input data a. Input data a, result of the experiment b => data point (a, b)

without performing experiment estimate that $b \approx c + da$.

Distance between points $(a\ b)$ and $(a, c+da)$ should be small.

Generalizations:

from R^2 to R^{n+1}

points $(x_1^1 \cdots x_n^1 \;\; y^1)$

$(x_1^2 \cdots x_n^2 \;\; y^2)$

$(x_1^k \cdots x_n^k \;\; y^k)$

want linear function: $c + d_1 x_1 + \cdots d_n x_n = y$

$$c + d_1 x_1^l + d_2 x_2^l + \cdots + d_n x_n^l = y^l \quad \text{for } l = 1 \cdots k$$

$$\underbrace{\begin{pmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_n^1 \\ 1 & x_1^2 & x_2^2 & & x_n^2 \\ \vdots & & & & \vdots \\ 1 & x_1^k & x_2^k & & x_n^k \end{pmatrix}}_{A} \begin{bmatrix} c \\ d_1 \\ \vdots \\ d_n \end{bmatrix} = \underbrace{\begin{bmatrix} y^1 \\ \vdots \\ y^k \end{bmatrix}}_{b}$$

$$\hat{x} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{bmatrix} = A(A^T A)^{-1} A^T b$$

Is $A^T A$ always invertible?

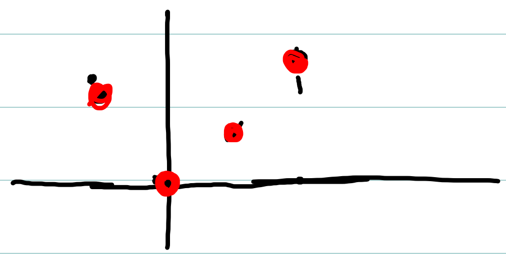
$w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$ is the best linear function for my data.

We know $\|A\hat{x} - b\| \leq \|Ax - b\|$ for all x in R^{n+1}. $w_0, w_1 \cdots w_n$ minimize

$$\sum_{l=1}^{k} (c + d_1 x_1^l + d_2 x_2^l + \cdots + d_n x_n - y^l)^2$$

Quadratic approximation

Fit a parabola through the points



(0,0) (1,1) (2,3) (-1, 2)

Equation of parabola $c+dx+fx^2=y$

want

$c = 0$

$c+d+f = 1$ $\qquad f=\frac{3}{2} \quad c=0 \quad d=-\frac{1}{2}$

$c+2d+4f=3$

$c+(-1)d+f=2$

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & -1 & 1 \end{pmatrix}}_{A}\begin{pmatrix} c \\ d \\ f \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 2 \end{pmatrix}$$

you can check A has rank 3

system has no solutions take

$$\begin{pmatrix} c \\ d \\ f \end{pmatrix} = \underset{3\times 3}{(A^TA)^{-1}}\ \underset{3\times 4}{A^T}\ \underset{4\times 1}{\begin{pmatrix} 0 \\ 1 \\ 3 \\ 2 \end{pmatrix}} = \begin{pmatrix} 0.3 \\ -0.6 \\ 1 \end{pmatrix}$$

$y = 0.3 - 0.6x + x^2$

best parabola through the given points. What does it mean?

$\left(\|A\hat{x} - b\| \le \|Ax - b\| \text{ for all } x \right)$

that

$$\sum_{i=1}^{4} \left(c + dx_i + fx_i^2 - y_i\right)^2 \quad \text{is minimized}$$