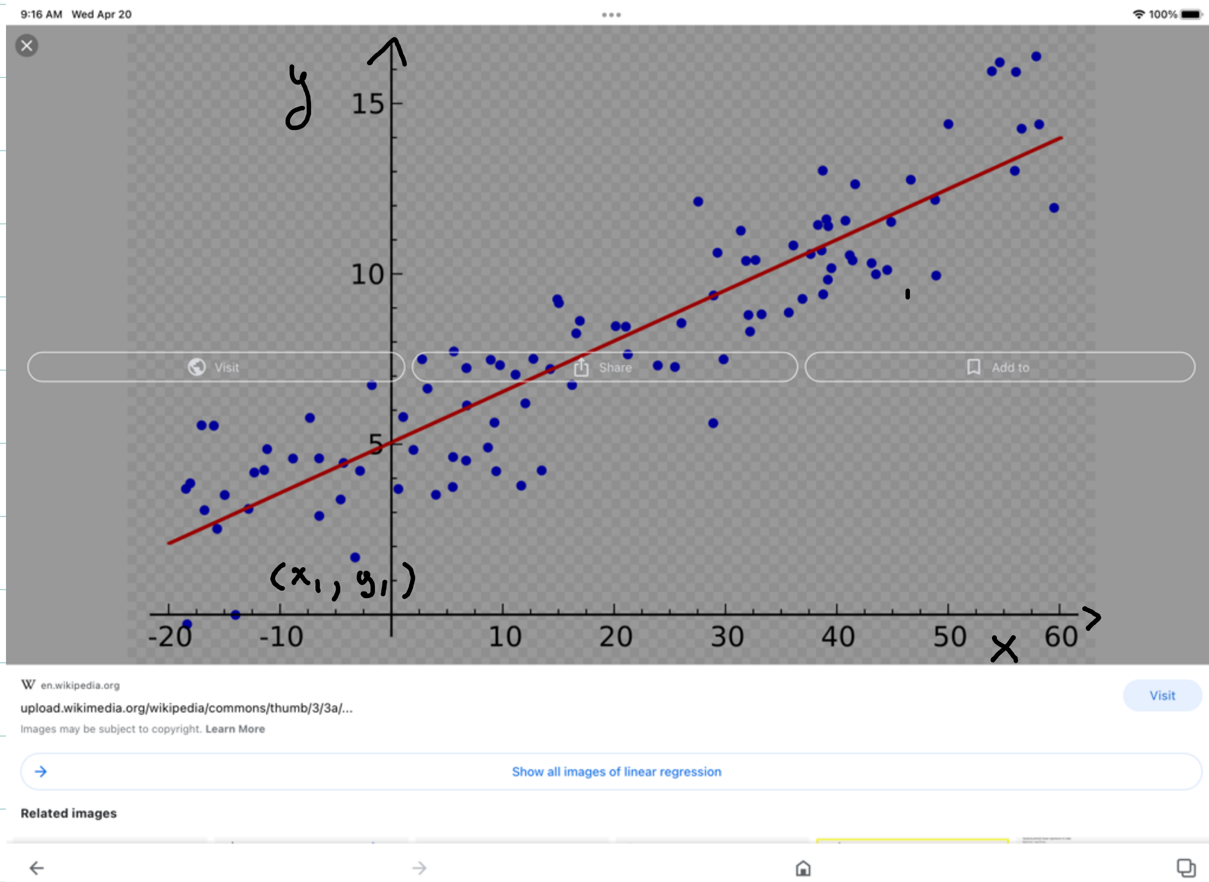
Lesson 12

Read chapter 4

Read chapter 5

Linear regression

Symmetrix matrices: spectral theorem



To find linear regression line: $c + dx = y$

$A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ $\quad b = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ $\quad A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b$ has no sol

x coordinates of data points

y coordinates of data point

Solve $A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b_V$ instead.

$V = \text{col}(A) = \text{span}\left(\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right)$, $\quad b_V = \text{proj}_V(b)$

sol is $\begin{bmatrix} c \\ d \end{bmatrix} = \hat{x} = (A^T A)^{-1} A^T b$. $\quad y = c + dx$

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

$$\overbrace{\|\hat{A\hat{x}} - b\|}^{b_V} \quad V = \mathrm{Col}(A)$$

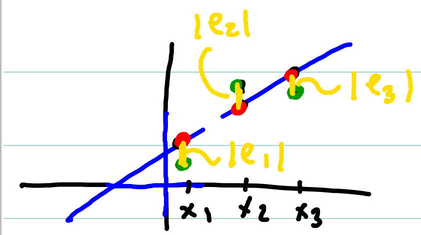
Recall $\|A\hat{x} - b\| \leq \|\overbrace{Ax}^{\text{vector in } V} - b\|$ for all x

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

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



line $y = cx + d$

Green: data points

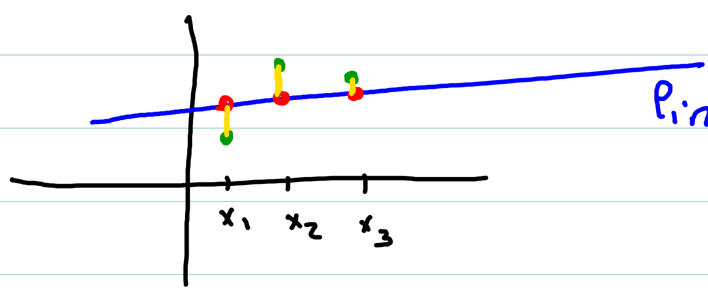
$(x_1\ y_1)$ $(x_2\ y_2)$ $(x_3\ y_3)$ ....

red points on regression line

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

same x coordinates

$$\|Ax - b\| = \left\| A \begin{bmatrix} g \\ f \end{bmatrix} - b \right\| = \left\| \begin{bmatrix} g + fx_1 - y_1 \\ g + fx_2 - y_2 \\ g + fx_n - y_n \end{bmatrix} \right\|$$

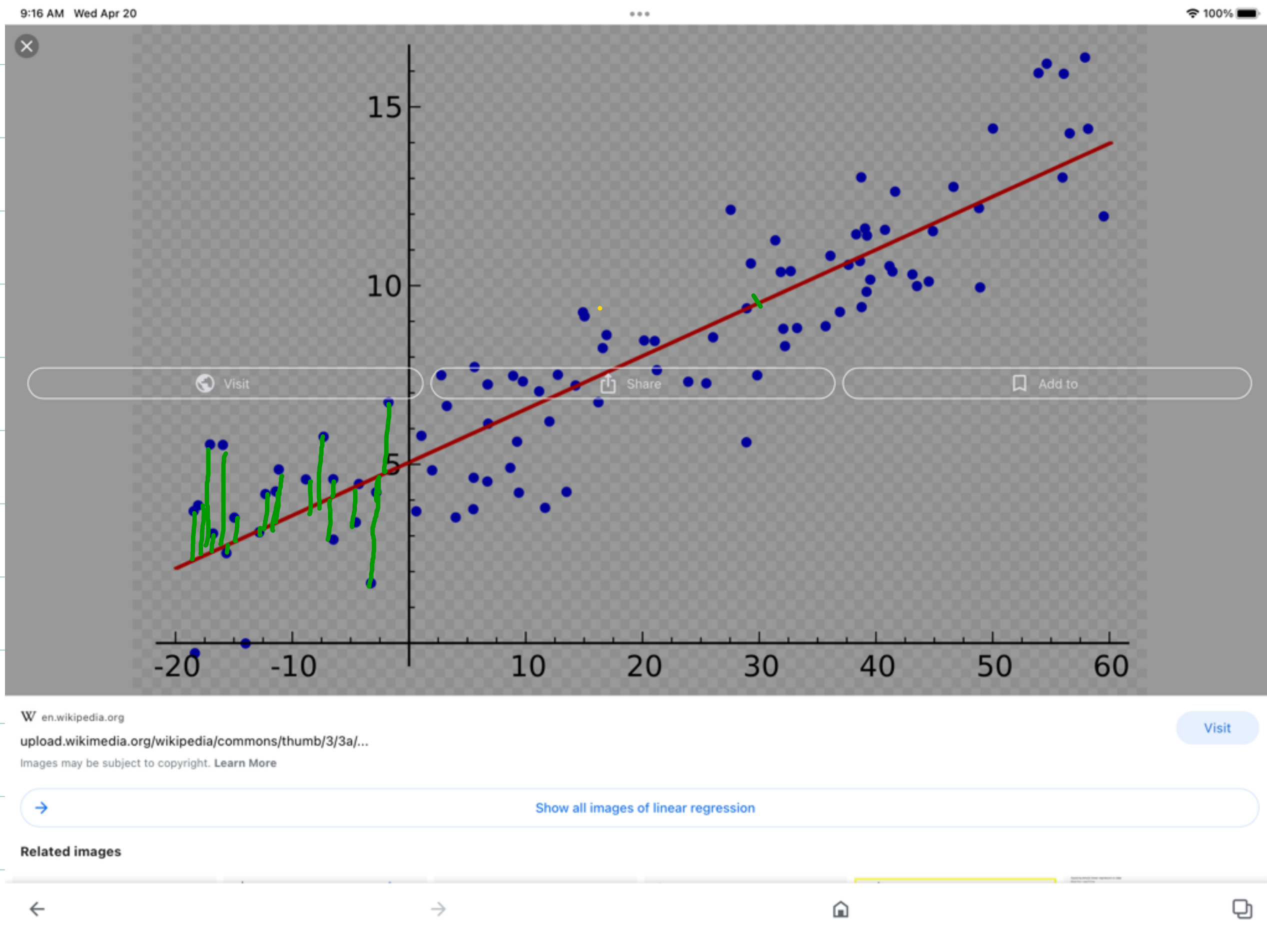


line $y = e + fx$

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

$$\sum_{i=1}^{n} (w_0 + w_1 x_i - y_i)^2$$

has minimum when $w_0 = c$
$w_1 = d$



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 $b \approx c + d a$.
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: $c + d_1 x_1 + \cdots + d_n x_n = y$

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

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

If $Ax = b$ has no ~~sol~~ solve $Ax = b_V$

b_V is the orthogonal projection

of b on $V = \mathrm{col}(A)$

$$\hat{x} = \begin{bmatrix} c \\ d_1 \\ \vdots \\ d_n \end{bmatrix} = A(A^T A)^{-1} A^T b$$

$c + d_1 x_1 + d_2 x_2 + \cdots + d_n x_n$ is the best linear approximation to my data.

What does it mean?

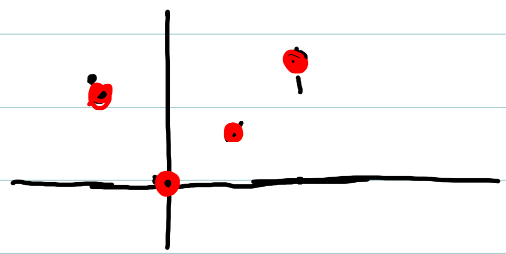
We know $\|A\hat{x} - b\| \leq \|Ax - b\|$ for all x in $\mathbb{R}^{n+1}$. $c, d_1, \ldots, d_n$ minimize

$$\sum_{l=1}^{K} (w_0 + w_1 x_1^l + w_2 x_2^l + \cdots + w_n x_n - y^l)^2$$

Is $A^T A$ always invertible?

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

$c = 0$ using $(0,0)$

$c + d + f = 1$ using $(1,1)$

$c + 2d + 4f = 3$ using $(2,3)$

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

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

you can check A has rank 3

system has no solutions take

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

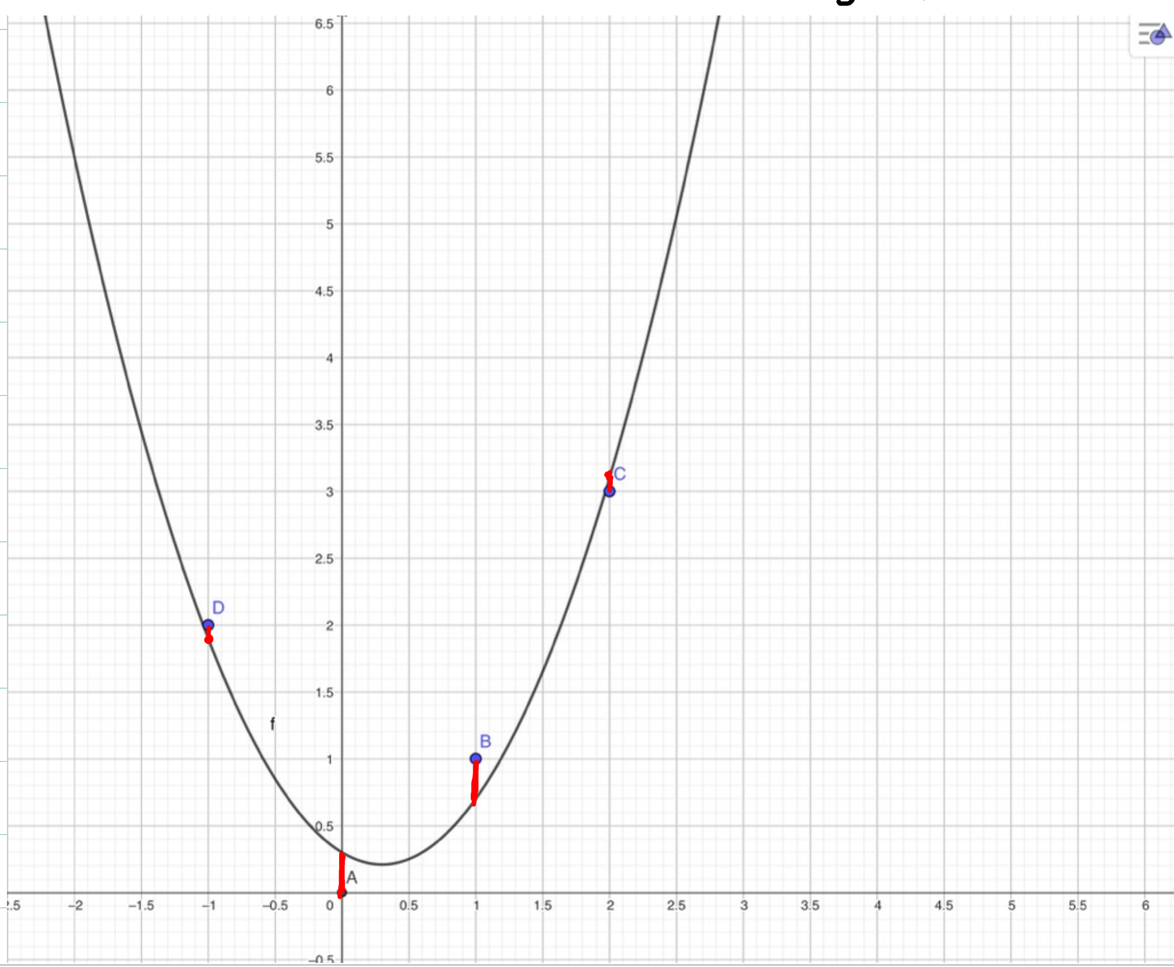
$$y = 0.3 - 0.6x + x^2$$

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

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

that

$$\sum_{i=1}^{4} \left(c + d x_i + f x_i^2 - y_i \right)^2$$ is minimized when $c = 0.3$, $d = -0.6$, $f = 1$



5

Ch 5 Symmetric metrices

Def : $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$

$$\text{Ex} \quad \begin{bmatrix} 1 & 5 & 7 \\ 5 & 2 & 8 \\ 7 & 8 & 3 \end{bmatrix} \quad \text{is symmetric}$$

Def $A \in \mathbb{C}^{n \times n}$ is hermitian (self-adjoint) if $A = \bar{A}^T$ (A^*)

where $\overline{a+ib} = a - ib$ and $\bar{A} = \overline{(a_{ij})} = (\bar{a}_{ij})$

3 Special properties of a Symmetric matrix A

1) All eigenvalues of A are real i.e solutions to $p(x) = 0$ where $p(x)$ is the characteristic polynomial of A are real.

2) If v and w are eigenvectors of A corresponding to different eigenvalues λ_1 and λ_2, then $v \perp w$ (and v, w are real)

3) For any eigenvalue λ, $AM(\lambda) = GM(\lambda)$

orthonormal

Def $P \in \mathbb{R}^{n\times n}$ is orthogonal if $P = [c_1 \cdots c_n]$

$\|c_i\| = 1$ for $i = 1 \cdots n$ $\quad c_i^T c_j = 0$ for $1 \le i \ne j \le n$

ie if the set $c_1 \cdots c_n$ is orthonormal.

Def $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable if $A = PDP^{-1}$ for some orthogonal matrix P and diagonal matrix D.

In This case $A = PDP^T$ (see next th)

orthonormal

Th: P is orthogonal $\Leftrightarrow$ $P^T \cdot P = I$

Proof: Let $P = [c_1 \cdots c_n]$

$$P^T P = \begin{bmatrix} c_1^T \\ c_2^T \\ c_n^T \end{bmatrix} \begin{bmatrix} c_1 \cdots c_n \\ \end{bmatrix} = \begin{bmatrix} c_1^T c_1 & c_1^T c_2 & \cdots & c_1^T c_n \\ c_2^T c_1 & c_2^T c_2 & \cdots & c_2^T c_n \\ \cdots & \cdots & \cdots & \cdots \\ c_n^T c_1 & c_n^T c_2 & & c_n^T c_n \end{bmatrix}$$

$$= (c_i^T c_j)$$

$$P^T P = I \Leftrightarrow c_i^T c_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{o/w} \end{cases}$$

$$P^T P = I \Leftrightarrow \boxed{P^{-1} = P^T} \Leftrightarrow P \cdot P^T = I$$

Note 1: if A, B are square matrices s.t. $BA = I$ then it must be true that $AB = I$

Note 2: An orthonormal set of vectors $c_1 \ldots c_m$ is always independent.

Note 3: an orthonormal set $c_1 \ldots c_n$ consisting of n vectors in $\mathbb{R}^n$ is an <u>orthonormal basis</u> for $\mathbb{R}^n$: $P = [c_1 \ldots c_n]$ is orthogonal so it is invertible so its columns are linearly independent and span $\mathbb{R}^n$

Th. $A \in \mathbb{R}^{n \times n}$ is orthogonally diagonalizable $\iff$ A is symmetric.

(spectral theorem)

Proof: $\Leftarrow$ If A is symmetric with k distinct eigenvalues $\lambda_1 \dots \lambda_k$ with $AM(\lambda_i) = h_i$ then $h_1 + h_2 + \cdots + h_k = n$ (True for any matrix since $p(x) = (-1)^n (\lambda-\lambda_1)^{h_1} \cdots (\lambda-\lambda_k)^{h_k}$) and $GM(\lambda_i) = AM(\lambda_i) = h_i$ (true for symmetric matrices). Let B_i be a basis for E_{λ_i}. The set $B_1 \cup B_2 \cup \cdots \cup B_k = B$ is linearly independent and has $h_1 + h_2 + \cdots + h_k = n$ vectors so it is a basis for $\mathbb{R}^n$.

A is diagonalizable i.e $A = PDP^{-1}$

P: has for columns vectors in B

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Why is P orthogonal? We will finish proof next time.

$\Rightarrow$: assume $A = PDP^{-1}$ with P orthogonal then $A^T = (PDP^{-1})^T = (PDP^T)^T = PD^TP^T = PDP^T = PDP^{-1} = A$