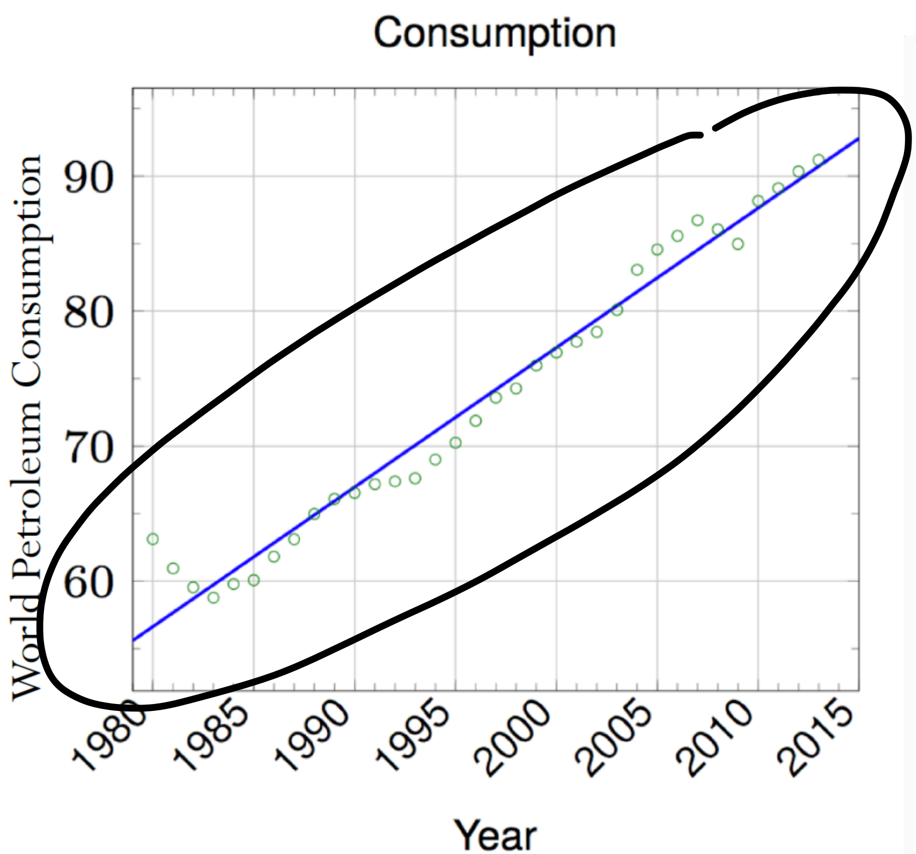
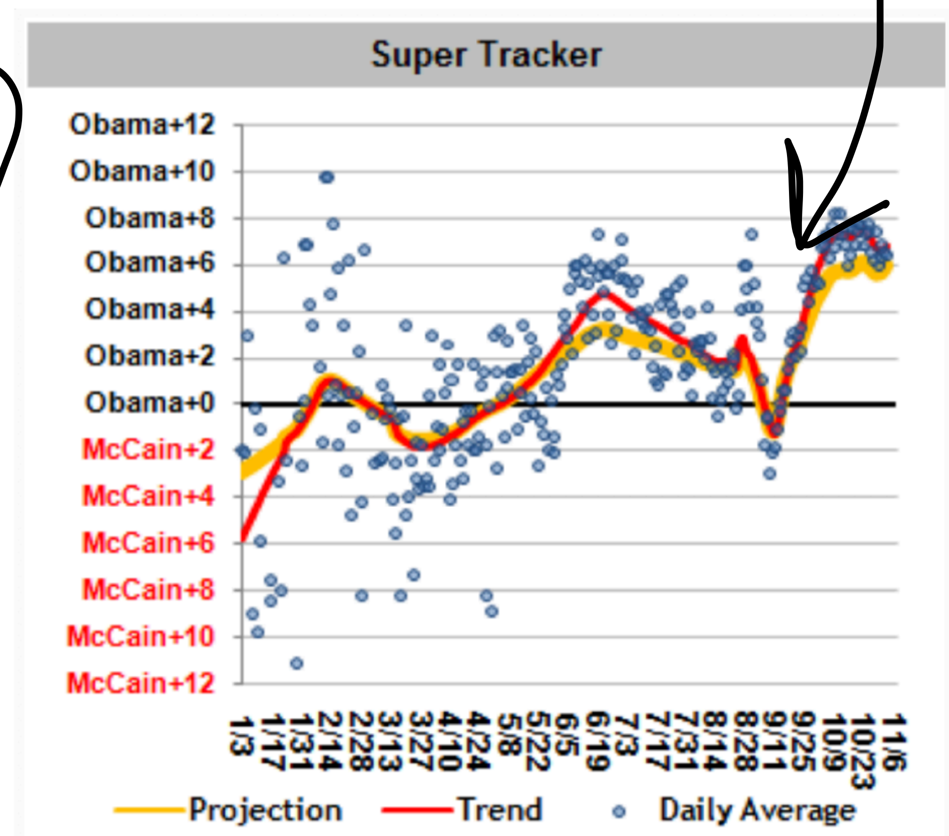
Least Squares and Data Fitting

Data fitting

How do we best fit a set of data points?





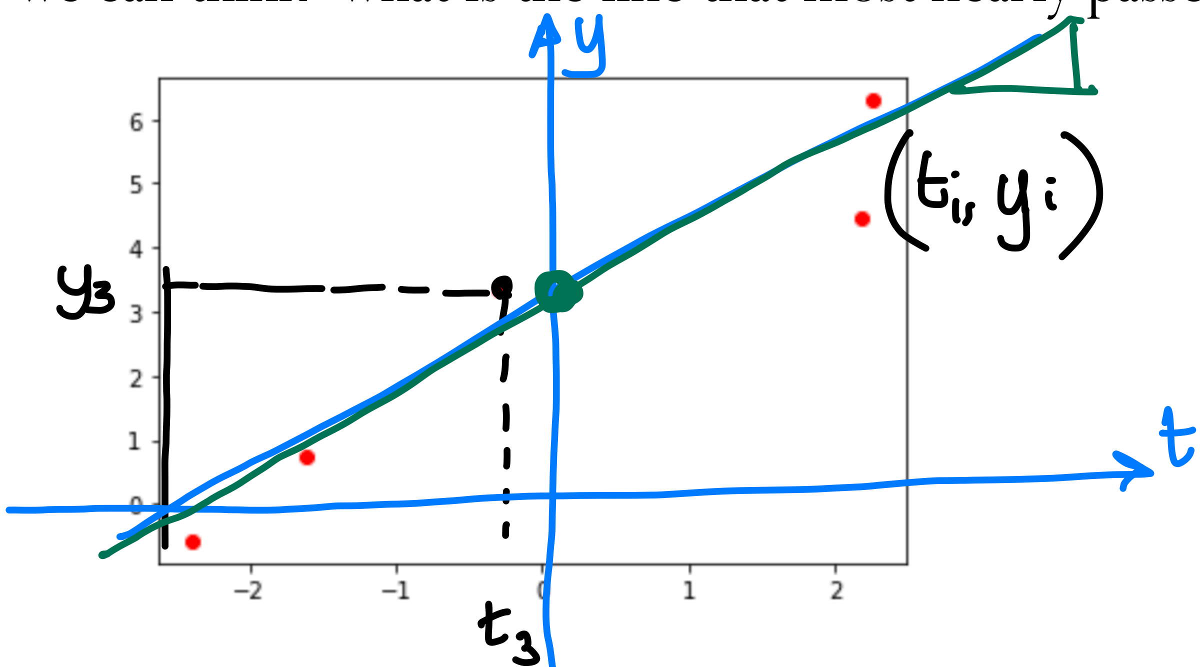
Linear Least Squares
1) Fitting with a line

Given m data points $\{\{t_1, y_1\}, \dots, \{t_m, y_m\}\}$, we want to find the function

$$y = x_o + x_1 t$$

that best fit the data (or better, we want to find the coefficients x_o, x_1).

Thinking geometrically, we can think "what is the line that most nearly passes through all the points?"



Given m data points $\{\{t_1, y_1\}, \ldots, \{t_m, y_m\}\}$, we want to find x_o and x_1 such that

(t_i, y_i)

$$\boxed{y_i = x_o + x_1\, t_i} \qquad \forall i \in [1, m]$$

$$\left.\begin{aligned} &y_1 = x_0 + x_1 t_1 \\ &y_2 = x_0 + x_1 t_2 \\ &y_3 = x_0 + x_1 t_3 \\ &\vdots \\ &y_m = x_0 + x_1 t_m \end{aligned}\right\} \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

$\underset{\sim}{b}$ $m \times 1$ $\qquad$ $\underset{\approx}{A}$ $m \times n$ $\qquad$ $\underset{\sim}{x}$ $n \times 1$

$\underset{\sim}{b} = \underset{\approx}{A}\,\underset{\sim}{x}$

overdetermined! $\boxed{m > n}$

$\underset{\sim}{b} = \underset{\approx}{A}\,\underset{\sim}{x}$

given $\qquad$ given $\qquad$ find

Given m data points $\{\{t_1, y_1\}, \dots, \{t_m, y_m\}\}$, we want to find x_o and x_1 such that

$$y_i = x_o + x_1\, t_i \qquad \forall i \in [1, m]$$

or in matrix form:

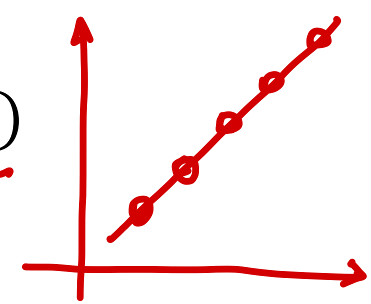
$$\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_o \\ x_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \qquad \boxed{\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}}$$

$\boldsymbol{m \times n} \quad \boldsymbol{n \times 1} \quad \boldsymbol{m \times 1}$

Note that this system of linear equations has more equations than unknowns – OVERDETERMINED SYSTEMS

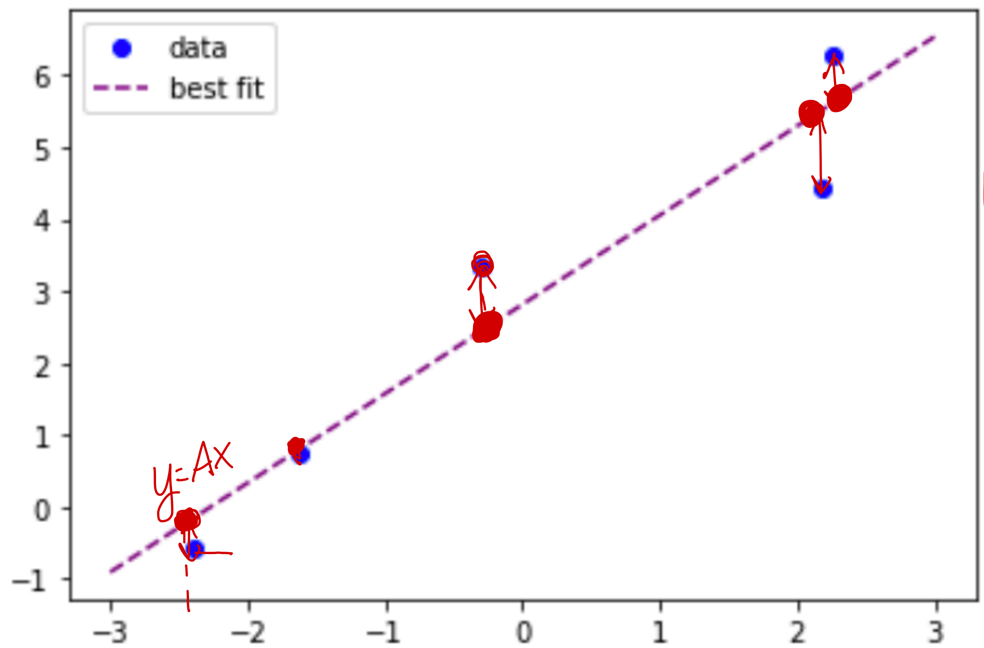
We want to find the appropriate linear combination of the columns of $\boldsymbol{A}$ that makes up the vector $\boldsymbol{b}$.

If a solution exists that satisfies $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ then $\boldsymbol{b} \in range(\boldsymbol{A})$



Linear Least Squares

- In most cases, $\boldsymbol{b} \notin range(\boldsymbol{A})$ and $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ **does not have an exact solution**!



we want to find $\underset{\sim}{x}$ s.t $\underset{\sim}{y} = \underset{=}{A}\underset{\sim}{x}$ better approximates $\underset{\sim}{b}$

- Therefore, an overdetermined system is better expressed as

$$\boxed{\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}} \qquad \boxed{m > n}$$

Linear Least Squares

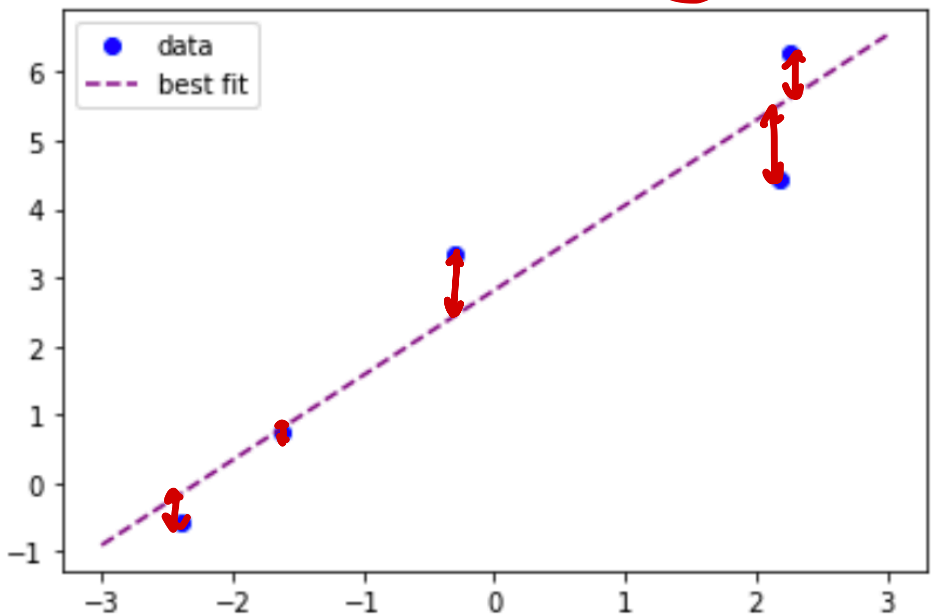
$\underset{\sim}{A}\,\underset{\sim}{x} \cong \underset{\sim}{b}$

$= \underset{\sim}{y} - \underset{\sim}{b}$

- **Least Squares**: find the solution $\boxed{\boldsymbol{x}}$ that minimizes the residual

min

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_5 \end{bmatrix} = \begin{bmatrix} b_1 - y_1 \\ b_2 - y_2 \\ \vdots \\ b_5 - y_5 \end{bmatrix}$$



$\underset{\sim}{r}$ is a vector

min $\|r\|$

- Let's define the function ϕ as the square of the 2-norm of the residual

$$\phi(\boldsymbol{x}) = \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2$$

Linear Least Squares

- **Least Squares**: find the solution $\boldsymbol{x}$ that minimizes the residual

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$$

- Let's define the function ϕ as the square of the 2-norm of the residual

$$\phi(\boldsymbol{x}) = \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2$$

- Then the least squares problem becomes

$$\min_{\boldsymbol{x}} \phi\,(\boldsymbol{x})$$

1d opt
1st order
2nd order

- Suppose $\phi: \mathcal{R}^m \rightarrow \mathcal{R}$ is a smooth function, then $\phi(\boldsymbol{x})$ reaches a (local) maximum or minimum at a point $\boldsymbol{x}^* \in \mathcal{R}^m$ only if

$$\boxed{\nabla\phi(\boldsymbol{x}^*) = 0}$$

How to find the minimizer?

stationary

$\underset{\sim}{x}$

$\Uparrow$

- To minimize the 2-norm of the residual vector

1st order necessary cond

$$\min_{\boldsymbol{x}} \phi(\boldsymbol{x}) = \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2 \;=\; (\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x})^T(\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x})$$

$$\nabla \phi = -A^T(b - Ax) + (b - Ax)^T(-A)$$

$$= -A^Tb + A^TAx - A^Tb + A^TAx$$

$$\boxed{= 2(A^TAx - A^Tb)}$$

$$\nabla\phi = 0 \implies \boxed{\underbrace{A^T}_{n\times m}\underbrace{A}_{m\times n}\underbrace{x}_{n\times 1} = \underbrace{A^T}_{n\times m}\underbrace{b}_{m\times 1}}$$

n×1   n×1

Find

$$\underbrace{A^TA}_{\text{given}}\,x = \underbrace{A^Tb}_{\text{given}}$$

Normal Equations

$\Downarrow$

linsys

$$\nabla(\nabla\phi) = ?$$

$\underline{A}^T \underline{A} \underline{x} = \underline{A}^T \underline{b}$ → Normal equation

$\nabla \phi = 2(A^T A x - A^T b)$

$\boxed{H = 2A^T A}$

→ x is minimizer

$\underset{\sim}{x}$: solution ?

* if matrix A is full rank :

$A = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{m\times n}$ $m > n$ → $\text{rank}(A) = n$

$\text{rank}(A) = n \Rightarrow$ A has L.I. columns

$\Rightarrow$ has n singular values > 0

$A^T A$ has n eigenvalues > 0

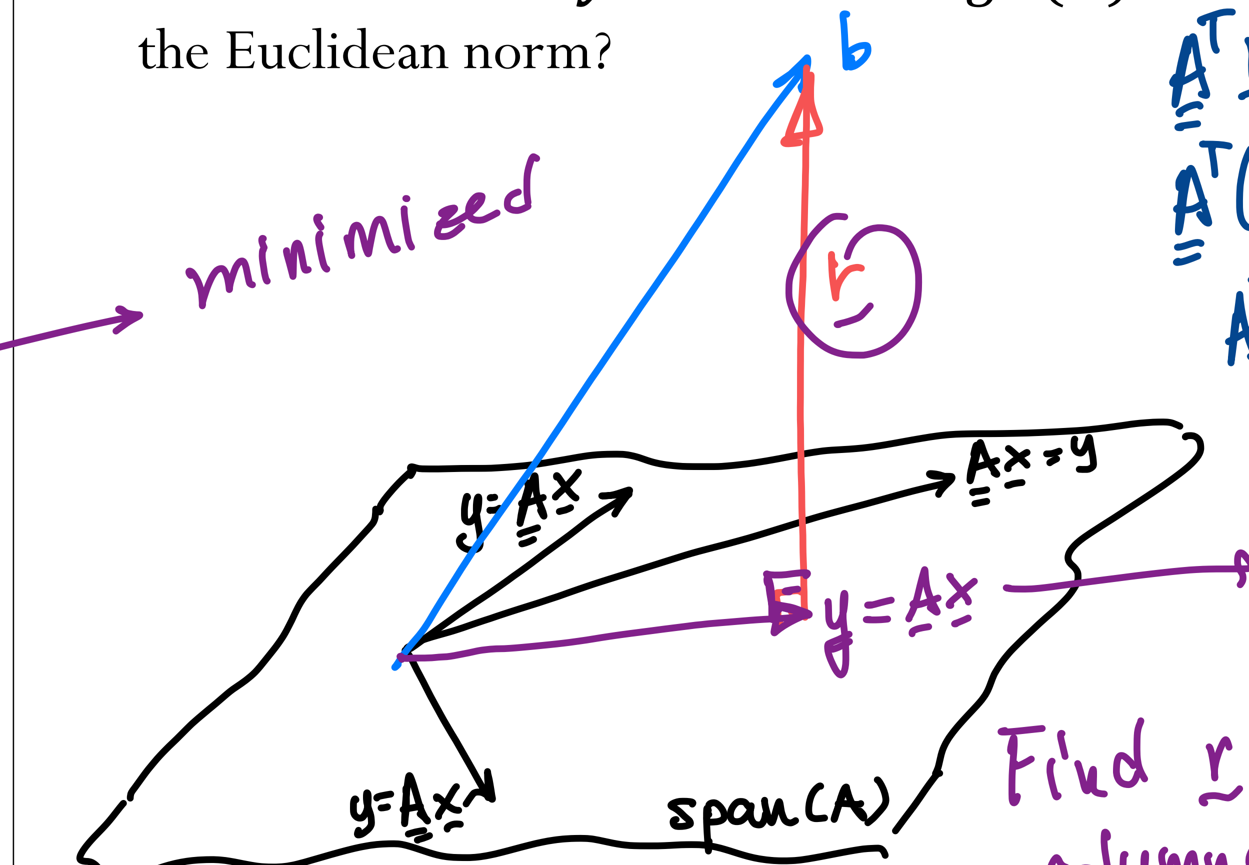
$y^T A^T A y > 0$ for any $y \neq 0$ $\Rightarrow$ $A^T A$ is positive def. & symmetric

$\hookrightarrow$ $A^T A$ is invertible

$\boxed{x = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \underline{b}}$ $\Rightarrow$ unique

Linear Least Squares (another approach)

- Find $\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}$ which is closest to the vector $\boldsymbol{b}$
- What is the vector $\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} \in range(\boldsymbol{A})$ that is closest to vector $\boldsymbol{y}$ in the Euclidean norm?



$A^T r = 0$

$A^T(b - Ax) = 0$

$A^T b - A^T A x = 0$

$\boxed{A^T A x = A^T b}$ ✓

projection of b in the column space of A

Find r orthogonal to ALL columns of A.

$\Leftrightarrow A^T r = 0$

Summary:

- $\boldsymbol{A}$ is a $m \times n$ matrix, where $m > n$.
- m is the number of data pair points. n is the number of parameters of the "best fit" function.

- Linear Least Squares problem $\underline{\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}}$ ***always*** has solution.

- The Linear Least Squares solution $\boldsymbol{x}$ minimizes the square of the 2-norm of the residual:

$$\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2$$

- One method to solve the minimization problem is to solve the system of **Normal Equations**

$$\boxed{\boldsymbol{A}^T \boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{A}^T\,\boldsymbol{b}} \rightarrow$$

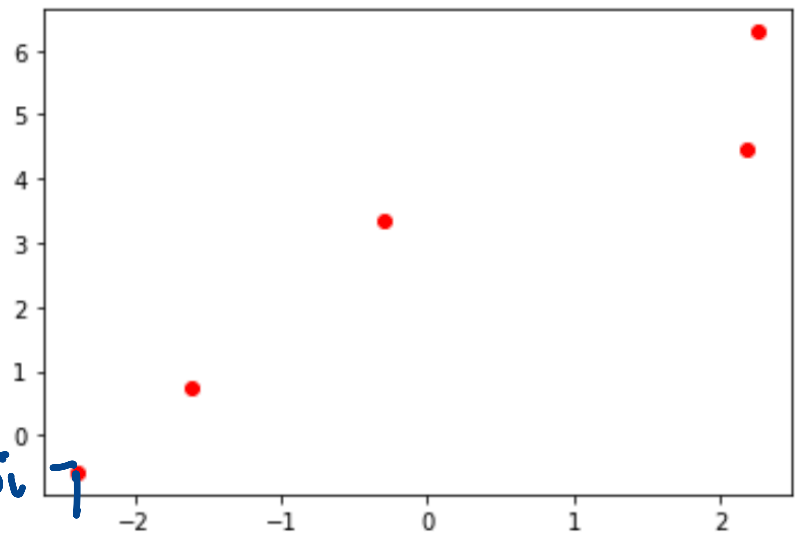
x is unique
matrix is full
rank

- Let's see some examples and discuss the limitations of this method.

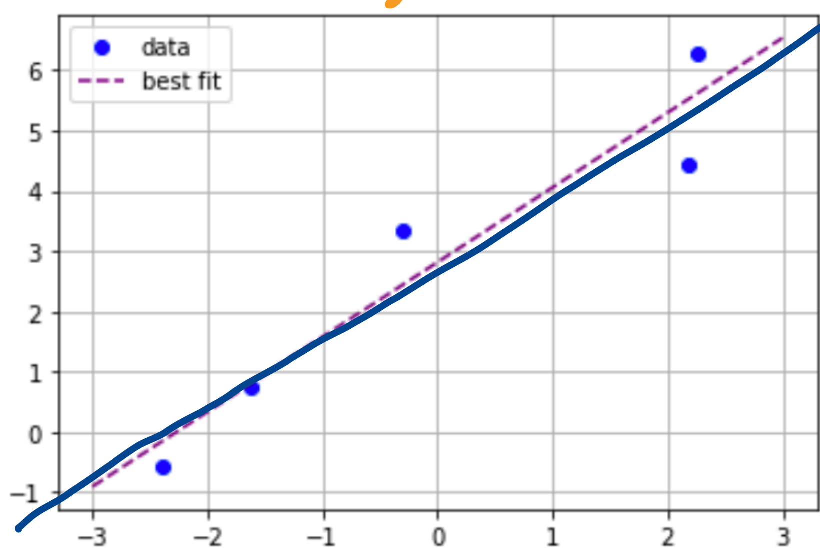
Example:

$y = x_0 + x_1 t$

b_i



t_i




```
t
```

```
array([-1.61477467, -2.3970584 , -0.30372944,  2.26304537,  2.188127  ])
```

```
b
```

```
array([ 0.74112251, -0.57768693,  3.33523097,  6.29377547,  4.44786481])
```

Solve: $\boldsymbol{A}^T \boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{A}^T\, \boldsymbol{b}$

$x = \text{la.solve}(A^T A, A^T b)$

```
x
```

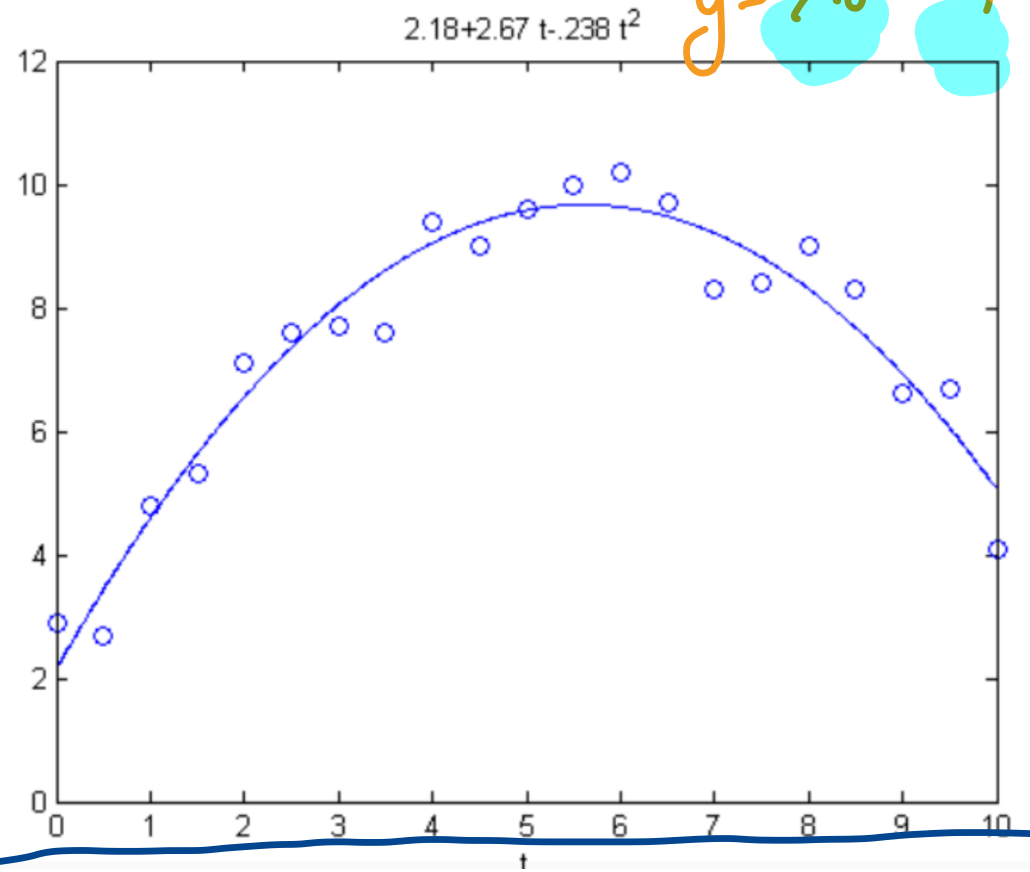
```
array([2.81441707, 1.24048133])
```

$\underset{\sim}{A} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_5 \end{bmatrix}_{5\times 2} \qquad \underset{\sim}{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_5 \end{bmatrix}$

Data fitting - not always a line fit!

- Does not need to be a line! For example, here we are fitting the data using a quadratic curve.

$y = x_0 + x_1 t + x_2 t^2$



$$\left[\quad \right]$$

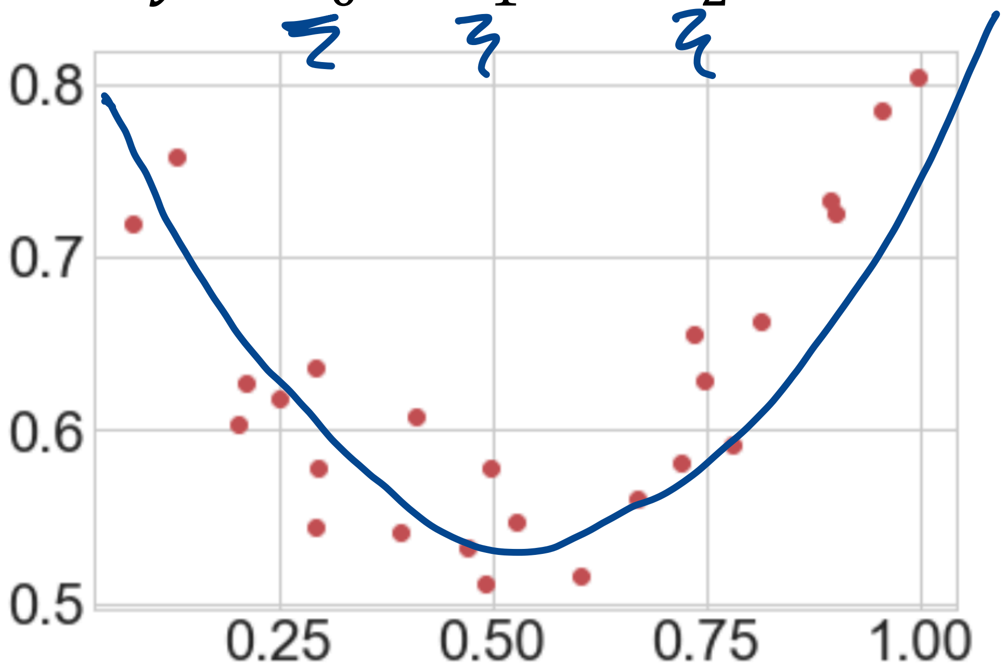
$m \times 3$

Linear Least Squares: The problem is **linear in its coefficients!**

Another example

We want to find the coefficients of the quadratic function that best fits the data points:

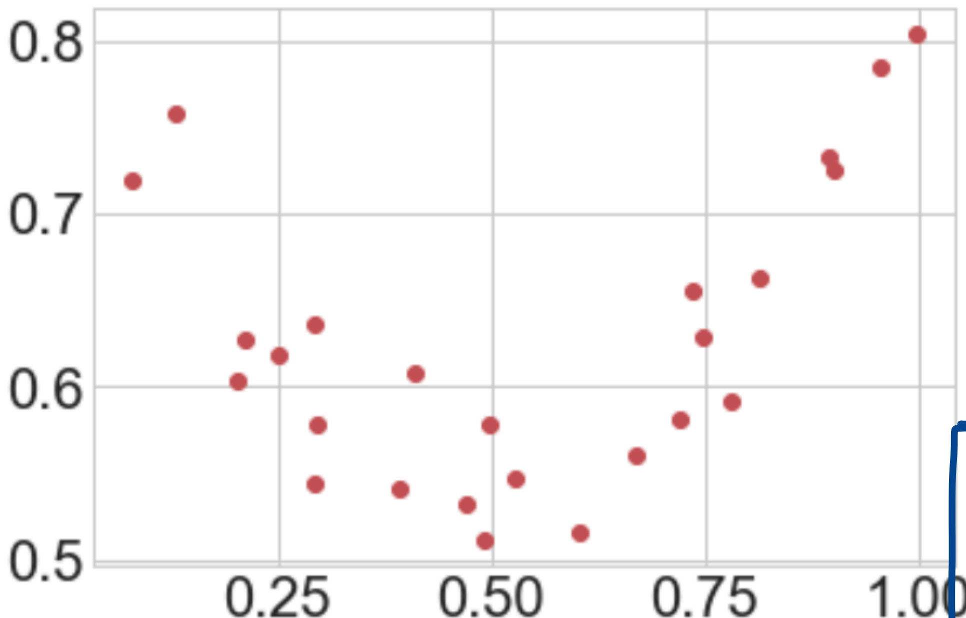
$$y = x_0 + x_1\, t + x_2\, t^2$$



We would not want our "fit" curve to pass through the data points exactly as we are looking to model the general trend and not capture the noise.

Data fitting

$$y = x_0 + x_1\, t + x_2\, t^2$$



$$
\begin{aligned}
y_1 &= x_0 + x_1 t_1 + x_2 t_1^2\\
y_2 &= x_0 + x_1 t_2 + x_2 t_2^2\\
y_3 &= x_0 + x_1 t_3 + x_2 t_3^2\\
&\vdots\\
y_m &= x_0 + x_1 t_m + x_2 t_m^2
\end{aligned}
$$

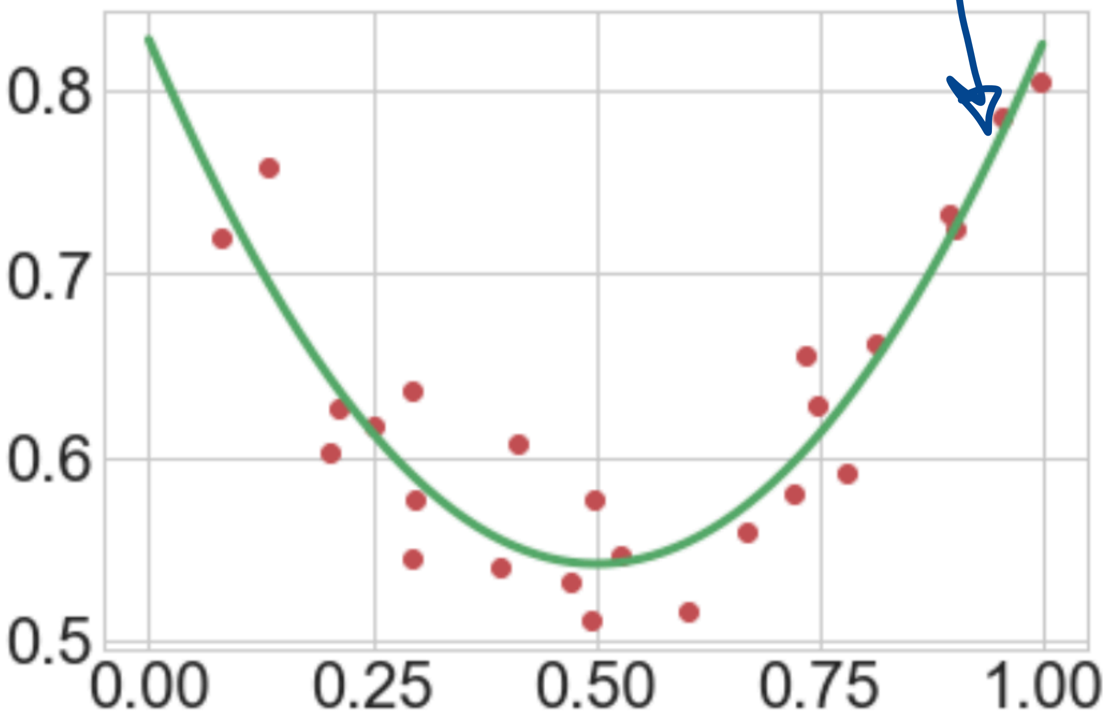
$$
\underset{m\times 1}{\underset{\sim}{b}} = \begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_m \end{bmatrix} = \overset{\underset{=}{A}}{\begin{bmatrix} 1 & t_1 & t_1^2\\ 1 & t_2 & t_2^2\\ \vdots & t_3 & t_3^2\\ \vdots & \vdots & \vdots\\ 1 & t_m & t_m^2 \end{bmatrix}} \overset{\underset{\sim}{x}}{\underset{3\times 1\ (n\times 1)}{\begin{bmatrix} x_0\\ x_1\\ x_2 \end{bmatrix}}}
$$

Data fitting

$\underline{\underline{A}}$ $\underset{\sim}{b}$ $\rightarrow$ $\underline{\underline{A}}\,\underline{x} \cong b$

$$\begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix} \begin{bmatrix} x_o \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Solve: $\boxed{\boldsymbol{A}^T \boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{A}^T\, \boldsymbol{b}}$



Which function is not suitable for linear least squares?

A) $y = a + b\,x + c\,x^2 + d\,x^3$

B) $y = x(a + b\,x + c\,x^2 + d\,x^3)$

C) $y = a\sin(x) + b/\cos(x)$

D) $\boxed{y = a\sin(x) + x/\cos(bx)}$

E) $y = a\,e^{-2x} + b\,e^{2x}$

$y = x_0 + x_1 t$

$y = x_0 + x_1 t + x_2 t^2$

$Ax = b$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} e^{-2x_1} & e^{2x_1} \\ e^{-2x_2} & e^{2x_2} \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

Computational Cost

$A_{m\times n}$

inner product $O(m)$

$$\boxed{A^T A x = A^T b}$$

① Construct A^TA → $[\, n\times m \,]\,[\, m\times n \,] = [\, n\times n \,]$

m

m

$O(m)$

n^2 times

$\boxed{O(mn^2)}$

② Factorize $\underbrace{A^TA}_{n\times n}$ → $\boxed{O(n^3)}$

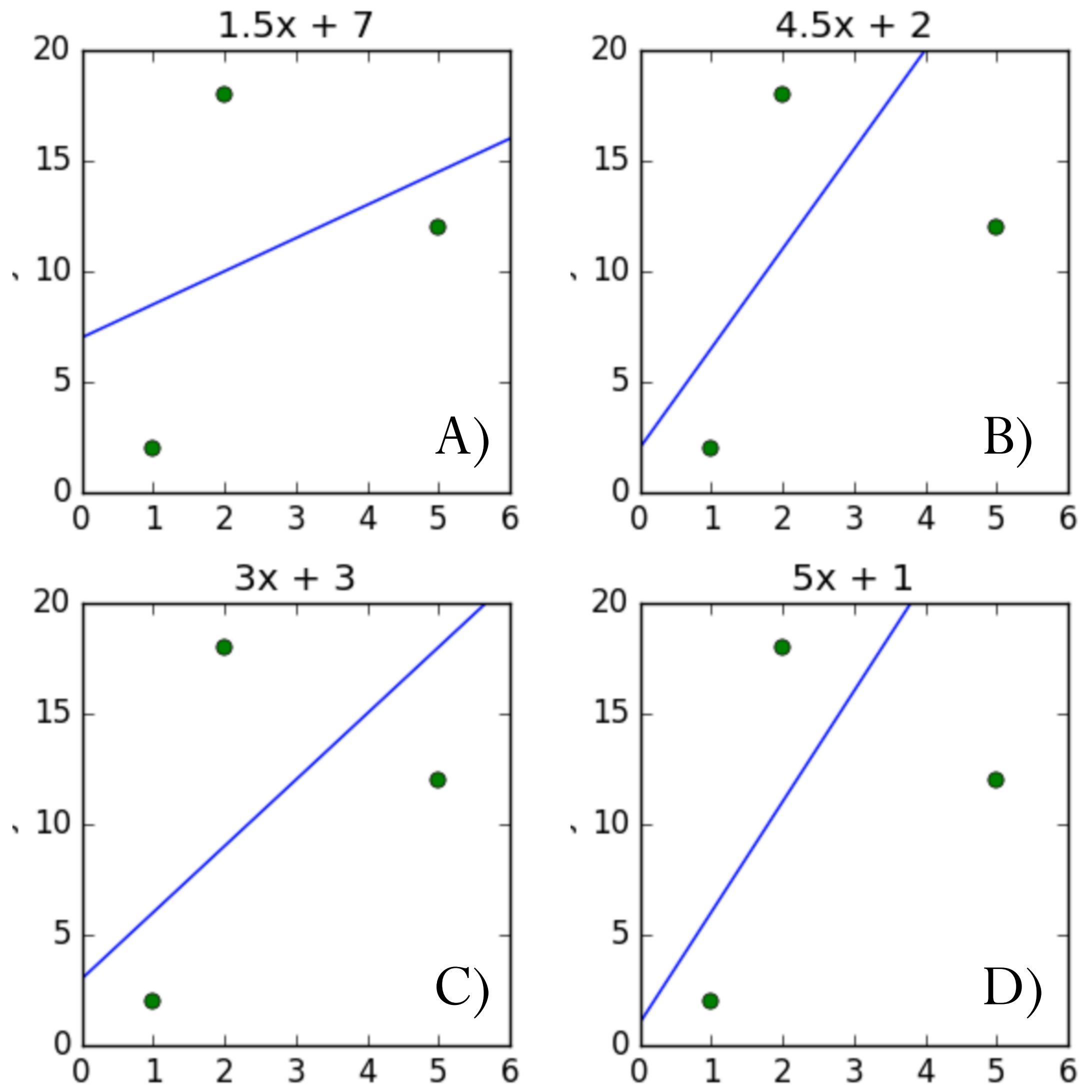
③ Solve → $\boxed{O(n^2)}$

overall cost: $O(mn^2)$ $m > n$

Short questions

Given the data in the table below, which of the plots shows the line of best fit in terms of least squares?

x	1	2	5
y	2	18	12



Short questions

Given the data in the table below, and the least squares model

$$y = c_1 + c_2 \sin(t\pi) + c_3 \sin(t\pi/2) + c_4 \sin(t\pi/4)$$

written in matrix form as

$$A \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \cong \mathbf{y}$$

determine the entry A_{23} of the matrix $\boldsymbol{A}$.
Note that indices start with 1.

A) -1.0
B) 1.0
C) -0.7
D) 0.7
E) 0.0

t_i	y_i
0.5	0.72
1.0	0.79
1.5	0.72
2.0	0.97
2.5	1.03
3.0	0.96
3.5	1.00