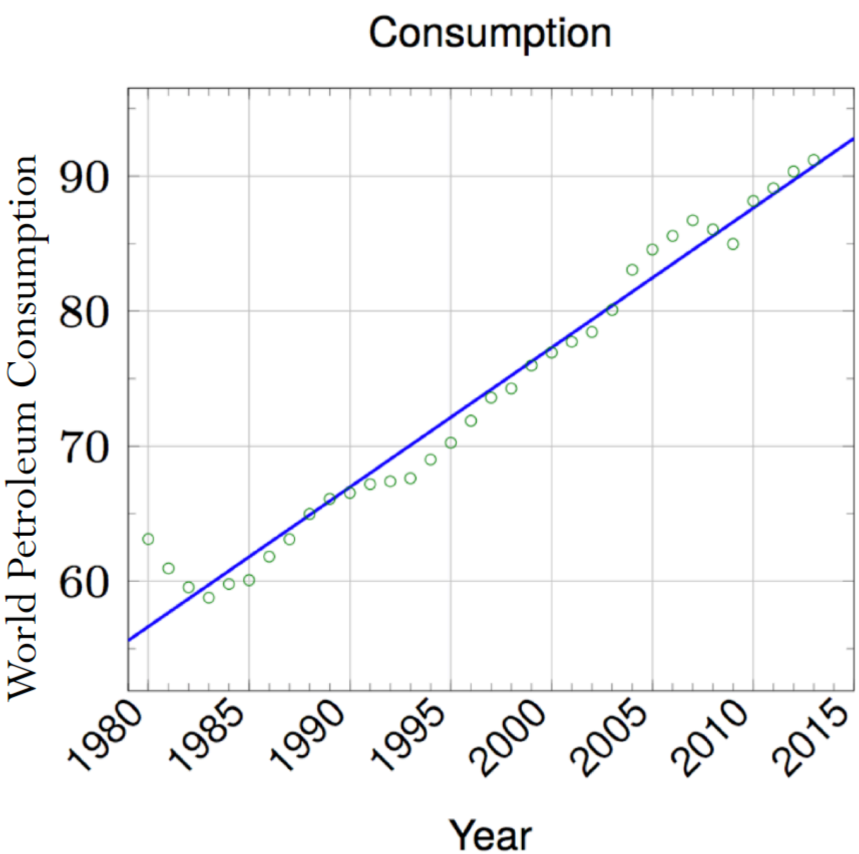
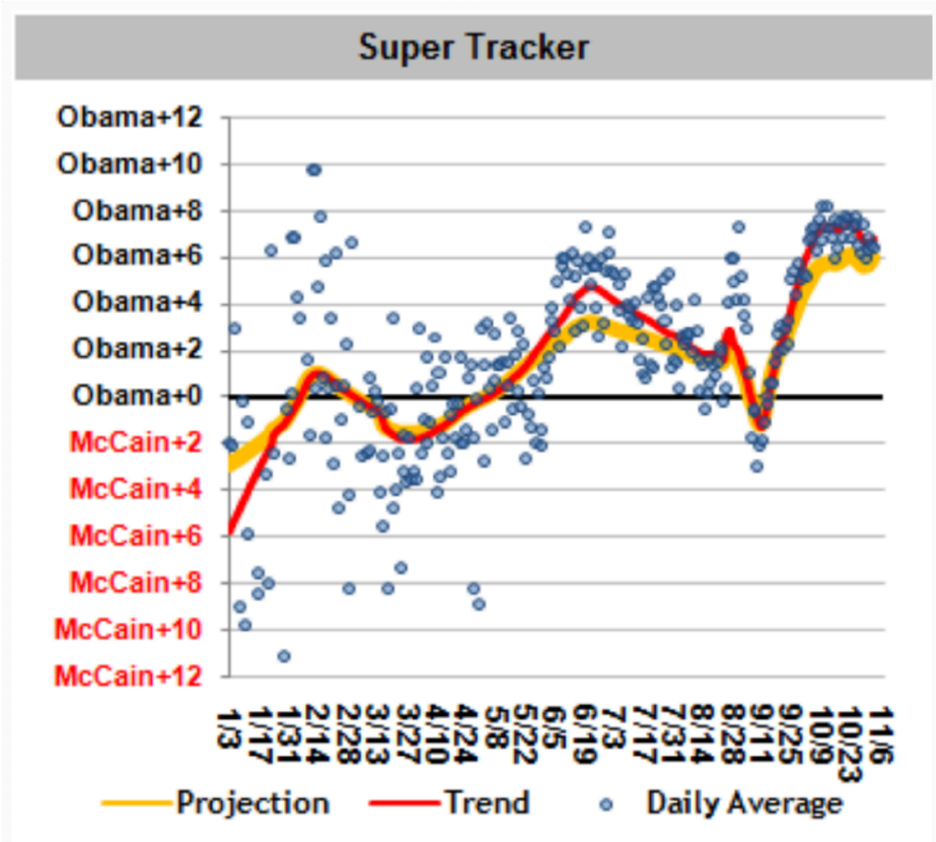
Least Squares and Data Fitting

Data fitting

How do we best fit a set of data points?





Linear Least Squares

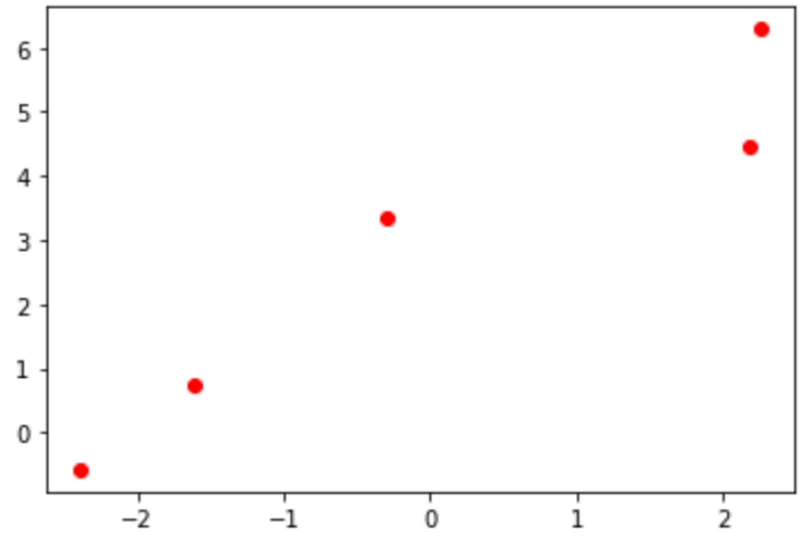
1) Fitting with a line

Given m data points $\{\{t_1, y_1\}, \ldots, \{t_m, y_m\}\}$, 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

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

Given m data points $\{\{t_1, y_1\}, \ldots, \{t_m, y_m\}\}$, 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:

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

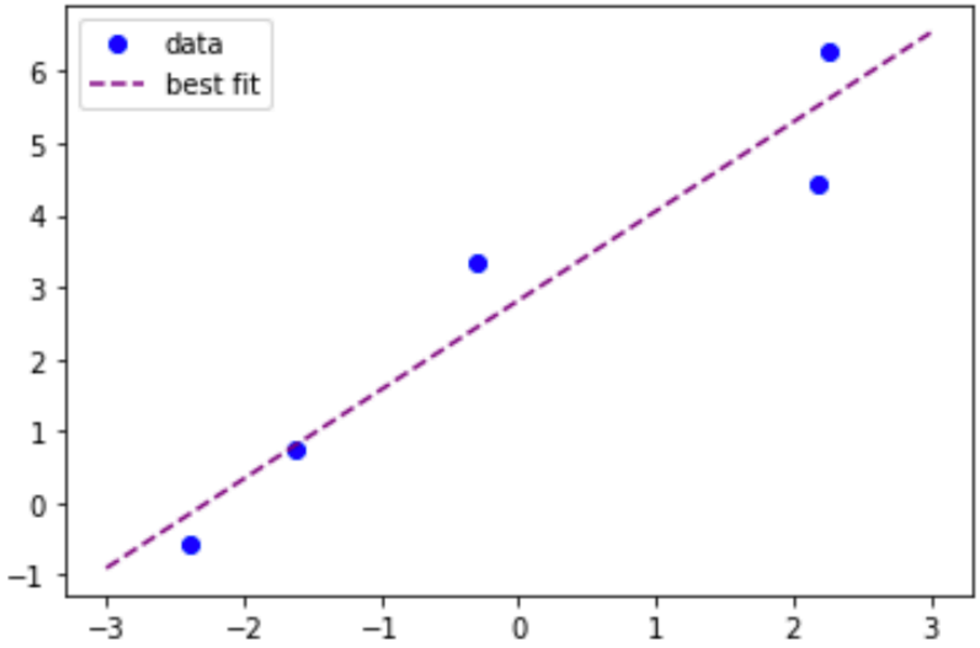
Note that this system of linear equations has more equations than unknowns – OVERDETERMINED SYSTEMS

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 $\boldsymbol{A\, x = b}$ 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**!



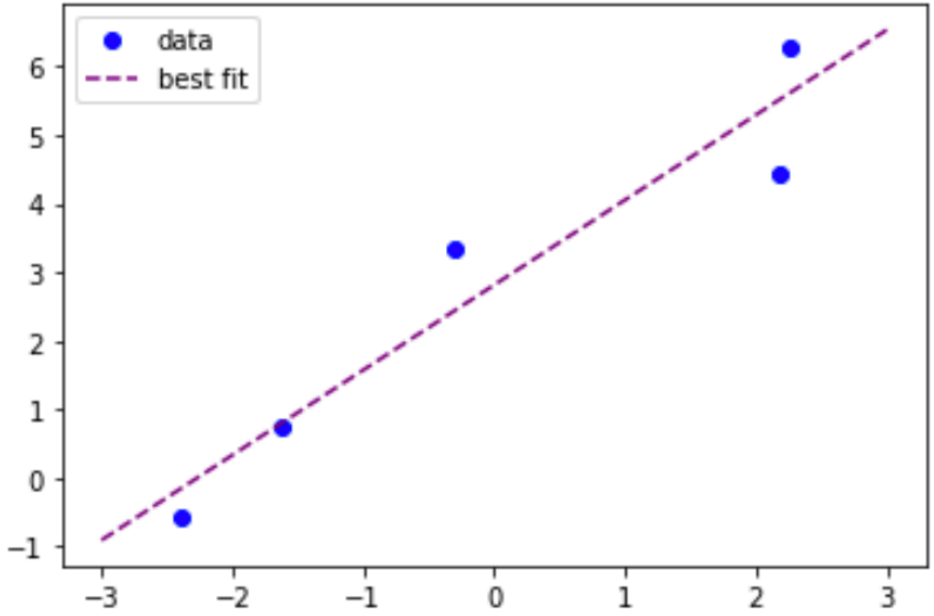
- Therefore, an overdetermined system is better expressed as

$$\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}$$

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

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

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 $\boldsymbol{\phi}$ 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

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

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

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

How to find the minimizer?

- To minimize the 2-norm of the residual vector

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

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?

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 $\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}$ ***always*** has solution.

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

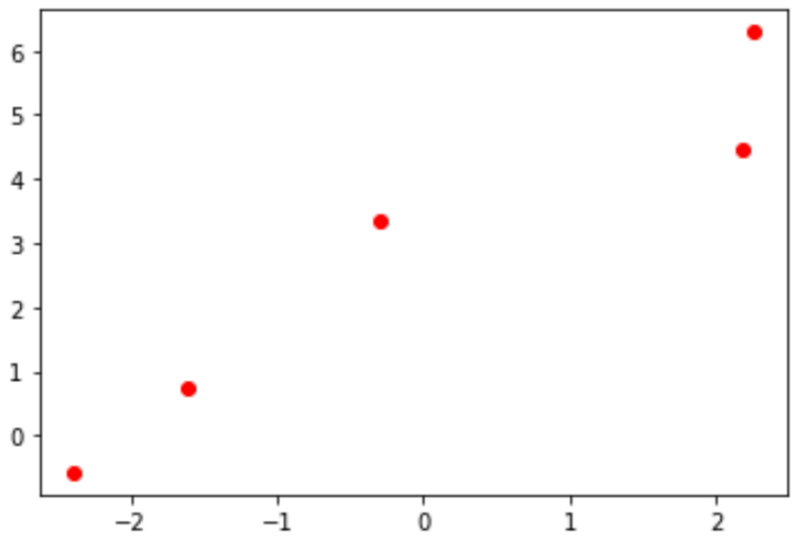
$$\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2$$

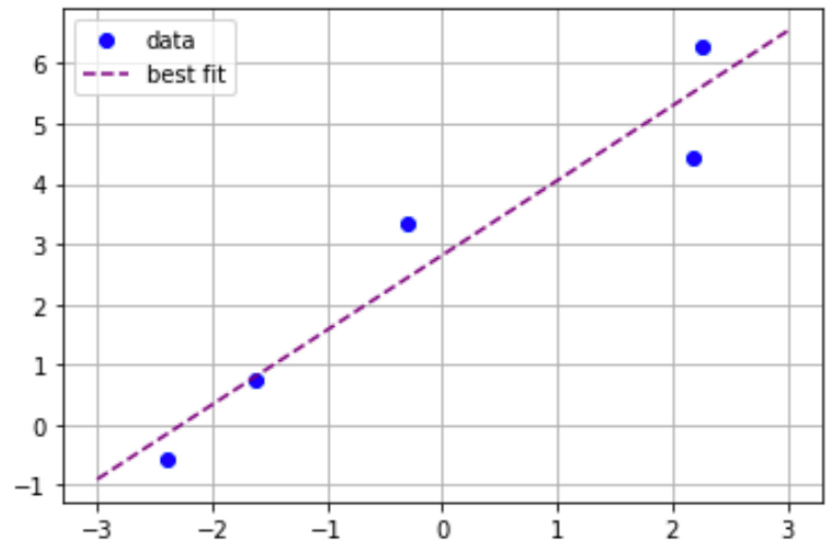
- One method to solve the minimization problem is to solve the system of **Normal Equations**

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

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

Example:






```
t
```

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

```
b
```

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

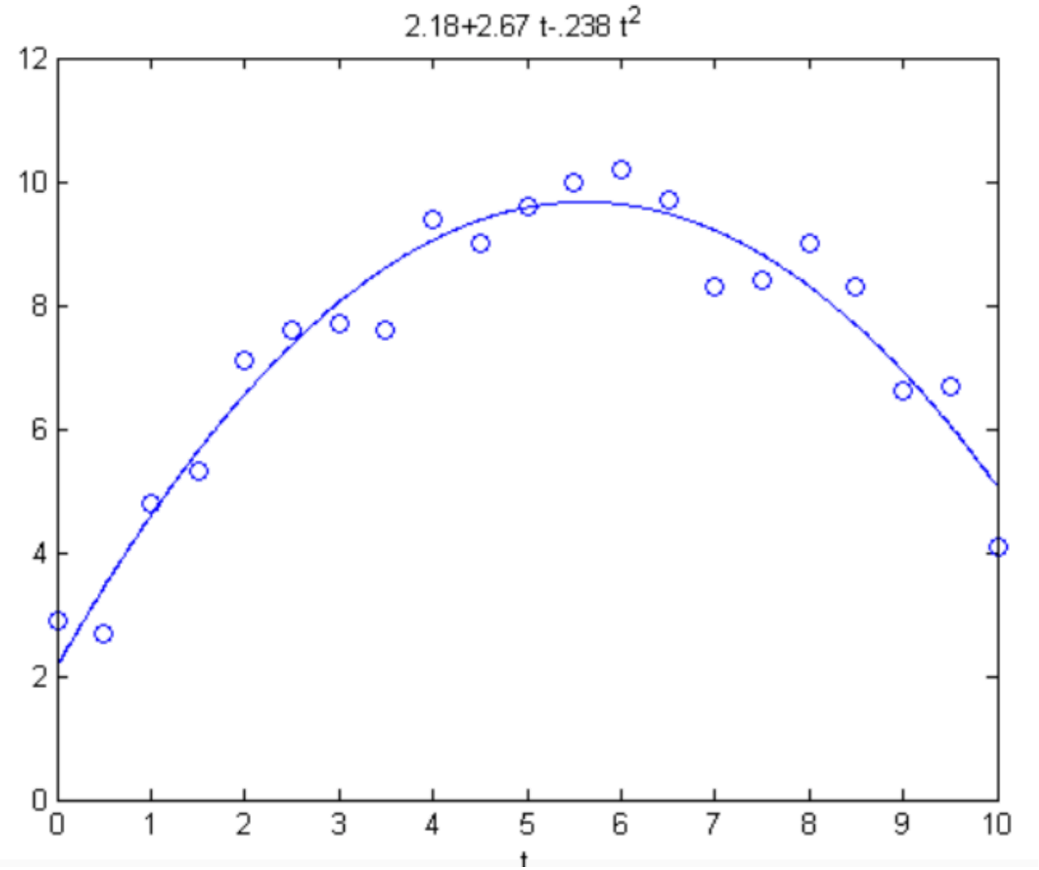
Solve: $\boldsymbol{A^T A\, x = A^T\, b}$

```
x
```

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

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.

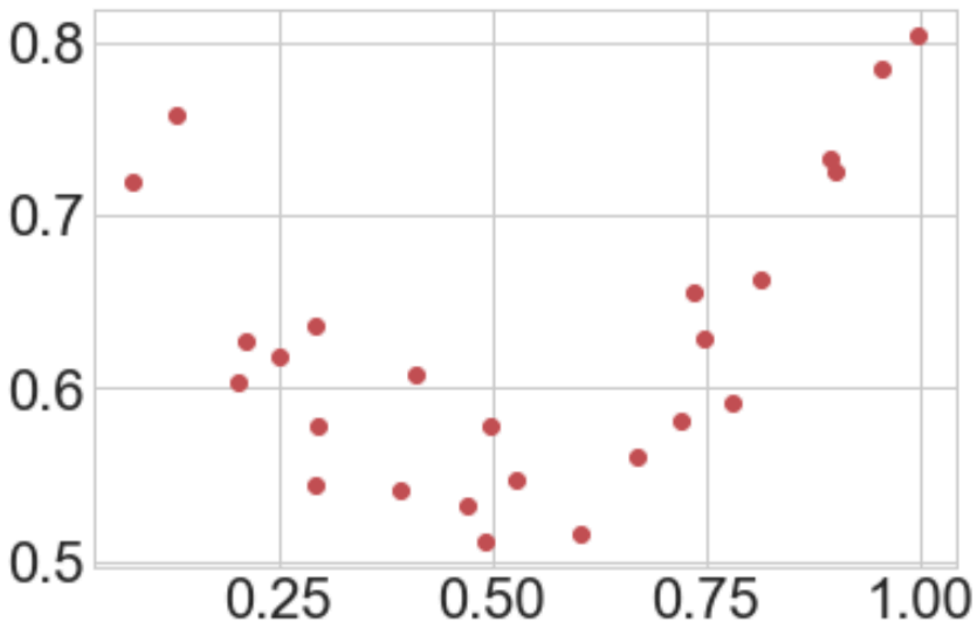


<u>Linear</u> Least Squares: The problem is **linear in its coefficients!**

Another example

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

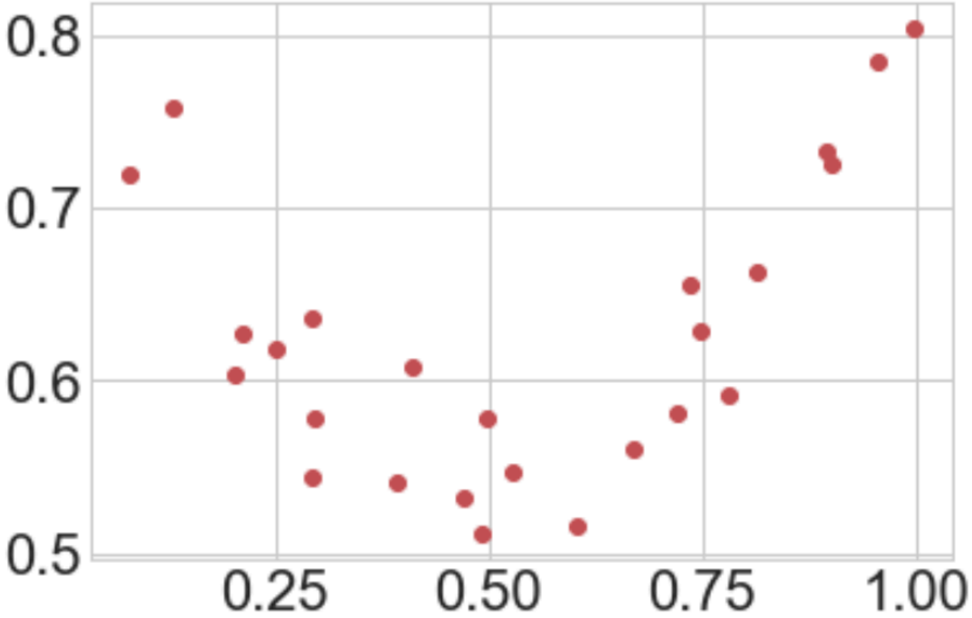
$$y = x_0 + x_1\, t + x_2\, t^2$$



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

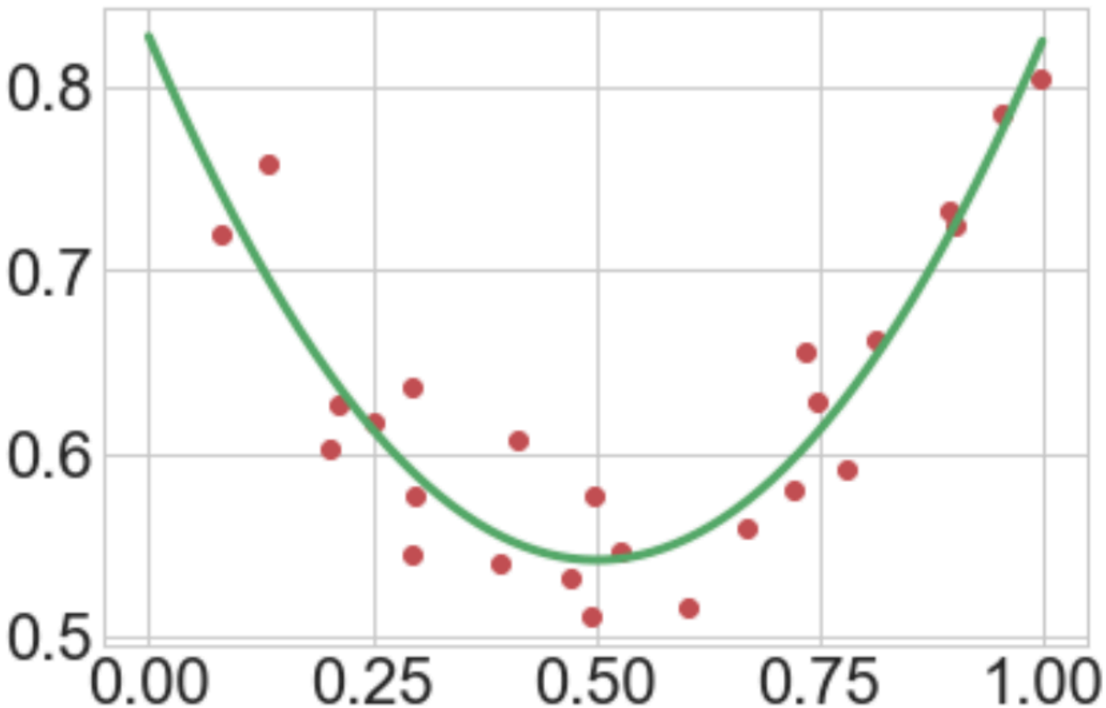
$$y = x_0 + x_1\, t + x_2\, t^2$$



Data fitting

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

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



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

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

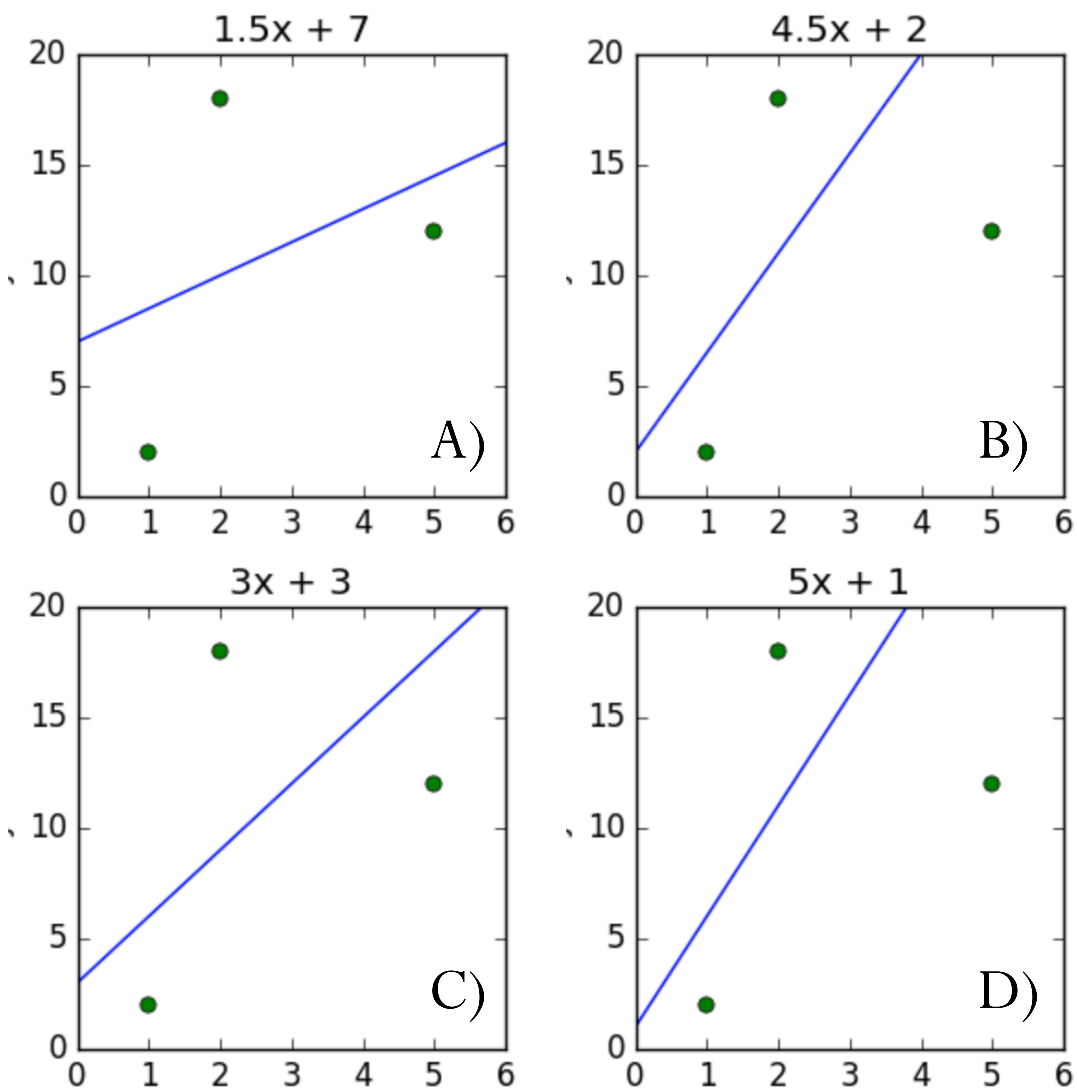
Computational Cost

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

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

Condition number for Normal Equations

Finding the least square solution of $\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}$ (where $\boldsymbol{A}$ is full rank matrix) using the Normal Equations

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

has some advantages, since we are solving a square system of linear equations with a symmetric matrix (and hence it is possible to use decompositions such as Cholesky Factorization)

However, the normal equations tend to worsen the conditioning of the matrix.

$$cond(\boldsymbol{A}^T \boldsymbol{A}) = (cond(\boldsymbol{A}))^2$$

How can we solve the least square problem without squaring the condition of the matrix?

Solving Linear Least Squares with SVD

What we have learned so far…

$\boldsymbol{A}$ is a $m \times n$ matrix where $m > n$

(more points to fit than coefficient to be determined)

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

- The solution $\boldsymbol{A}\, \boldsymbol{x} \cong \boldsymbol{b}$ is unique if and only if $rank(\mathbf{A}) = n$

($\boldsymbol{A}$ is full column rank)

- $rank(\mathbf{A}) = n \;\rightarrow$ columns of $\boldsymbol{A}$ are ***linearly independent*** $\rightarrow n$ non-zero singular values $\rightarrow \boldsymbol{A}^T\, \boldsymbol{A}$ has only positive eigenvalues $\rightarrow\; \boldsymbol{A}^T \boldsymbol{A}$ is a symmetric and positive definite matrix $\rightarrow\; \boldsymbol{A}^T \boldsymbol{A}$ is invertible

$$\boldsymbol{x} = (\boldsymbol{A}^T \boldsymbol{A})^{-\mathbf{1}} \boldsymbol{A}^T\, \boldsymbol{b}$$

- If $rank(\mathbf{A}) < n$, then $\boldsymbol{A}$ is rank-deficient, and solution of linear least squares problem is ***not unique***.

SVD to solve linear least squares problems

$\boldsymbol{A}$ is a $m \times n$ rectangular matrix where $m > n$, and hence the SVD decomposition is given by:

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}$$

We want to find the least square solution of $\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}$, where $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$

or better expressed in reduced form: $\boldsymbol{A} = \boldsymbol{U}_R\ \boldsymbol{\Sigma}_{\boldsymbol{R}}\ \boldsymbol{V}^{\boldsymbol{T}}$

Recall Reduced SVD $m > n$

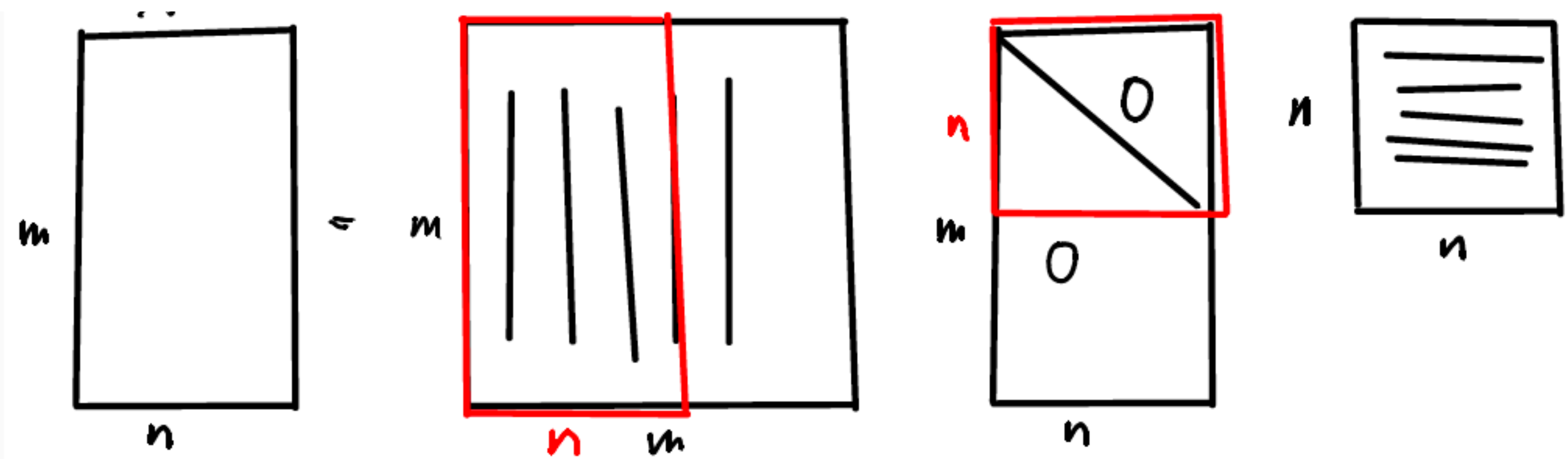
$$\boldsymbol{A} = \boldsymbol{U}_R\ \boldsymbol{\Sigma}_R\ \boldsymbol{V}^T$$

$\boldsymbol{A}$: $m \times n$

$\boldsymbol{U}_R$: $m \times n$

$\boldsymbol{\Sigma}_R$: $n \times n$

$\boldsymbol{V}^T$: $n \times n$



Shapes of the Reduced SVD

Suppose you compute a reduced SVD $A = U\Sigma V^T$ of a 10×14 matrix A. What will the shapes of U, Σ, and V be?
Hint: Remember the transpose on V!

The shape of U will be ____ $\times$ ____ .

The shape of Σ will be ____ $\times$ ____ .

The shape of V will be ____ $\times$ ____ .

SVD to solve linear least squares problems

$$\boldsymbol{A} = \boldsymbol{U}_R \ \boldsymbol{\Sigma}_R \ \boldsymbol{V}^T$$

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}$$

Example:

Consider solving the least squares problem $\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}$, where the singular value decomposition of the matrix $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{x}$ is:

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} \cong \begin{bmatrix} 12 \\ 9 \\ 9 \\ 10 \end{bmatrix}$$

Determine $\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2$

Example

Suppose you have $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{x}$ calculated. What is the cost of solving

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

A) $O(n)$
B) $O(n^2)$
C) $O(mn)$
D) $O(m)$
E) $O(m^2)$