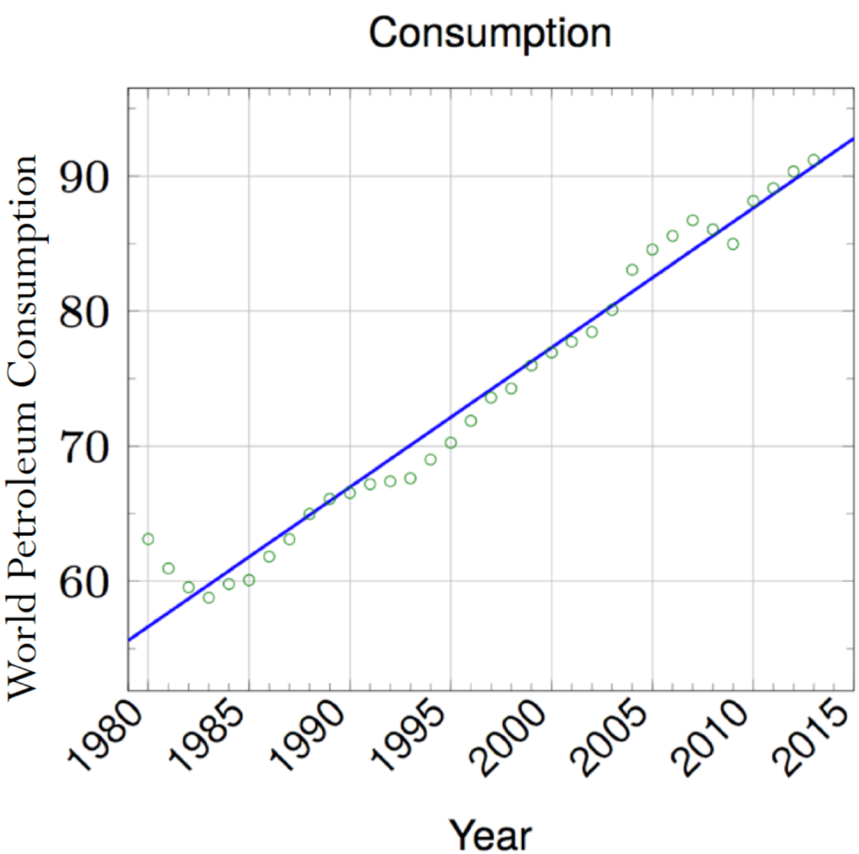
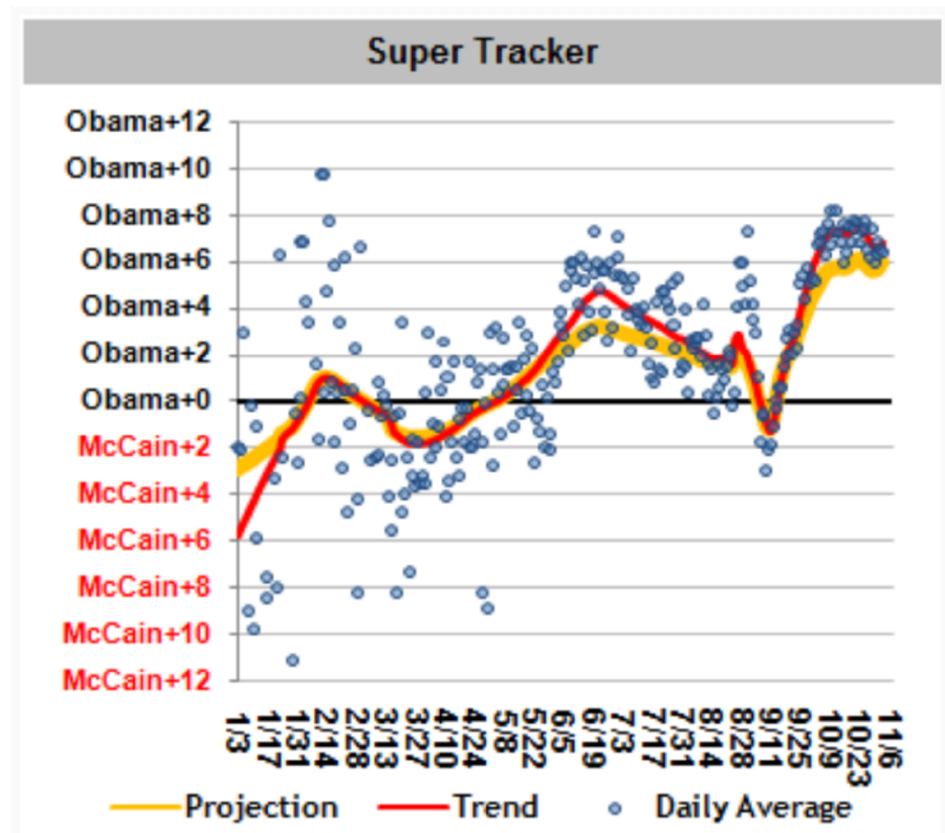
Least Squares and Data Fitting

Data fitting

Demo "Linear Regression Examples #1"

How do we best fit a set of data points?





Linear Least Squares – Fitting with a line

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

$$y = \alpha + \beta\, t$$

that best fit the data (or better, we want to find the parameters α, β).

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

Linear Least Squares

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

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

- 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 $\boldsymbol{b} \in range(\boldsymbol{A})$

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

- Therefore, an overdetermined system is better expressed as

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

Linear Least Squares

- 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?

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

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

- Suppose $\phi: \mathcal{R}^m \to \mathcal{R}$ 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

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

How to find the minimizer?

- To minimize the 2-norm of the residual vector

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

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.

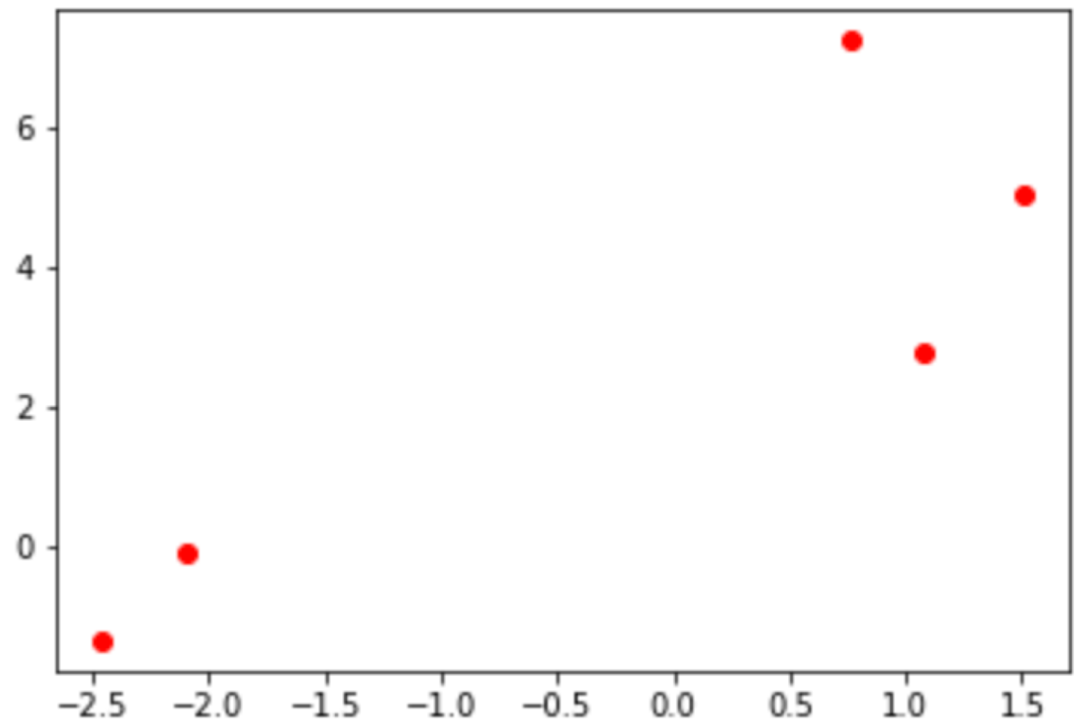
Example:

Demo: "Fit a line - Least Squares example"

$$\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \cong \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

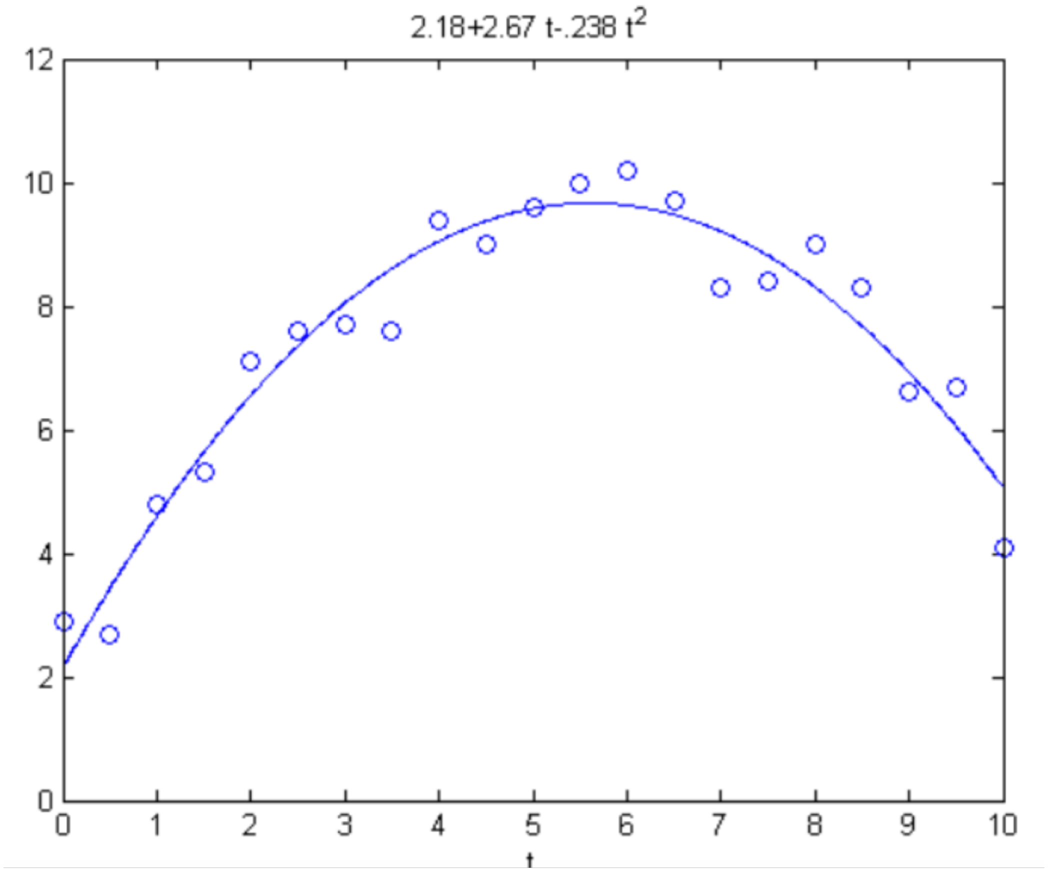
5×2 **2 ×1** **5 ×1**

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



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!**

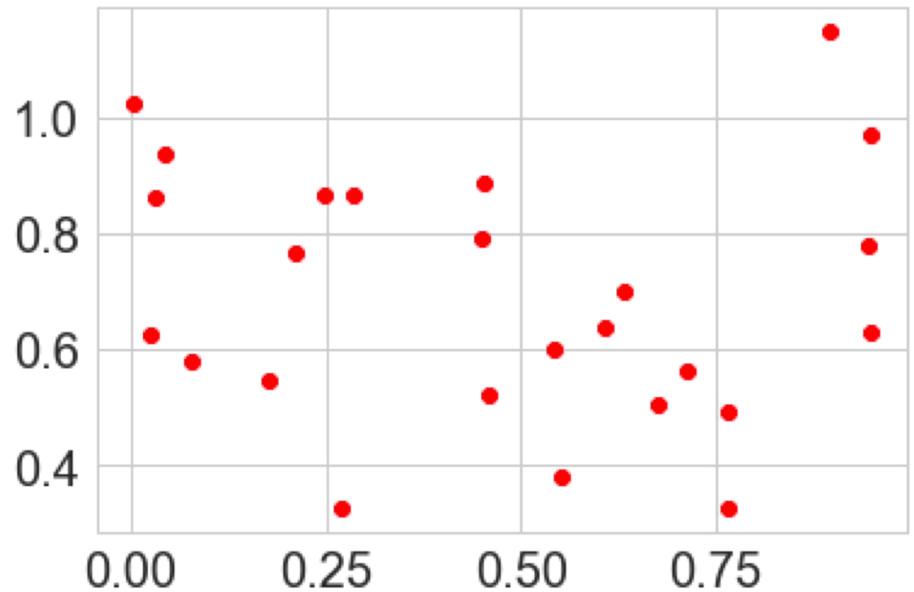
Which function is not suitable for linear least squares?

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

More examples

Demo “Make some noise”

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



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

The data points were generated by adding random noise to the function

$$f(t) = 0.8 - t + t^2$$

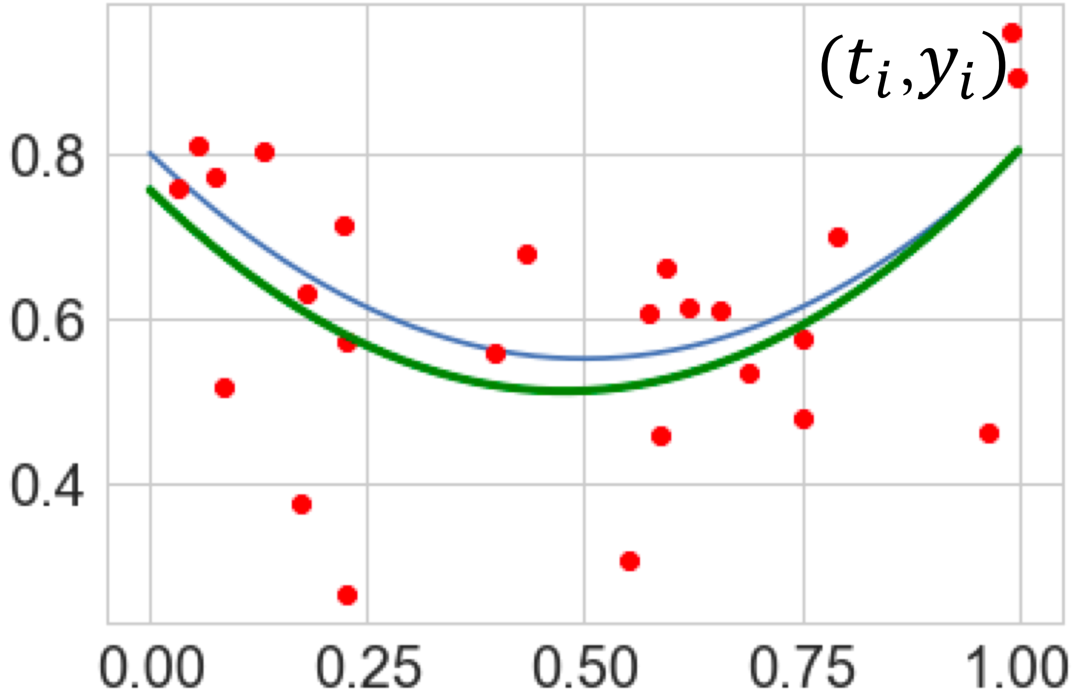
We would not want our “fit” curve to pass through the data points exactly as we are looking to model the general trend and not capture the noise.

Data fitting

Demo "Make some noise"

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

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



Computational Cost

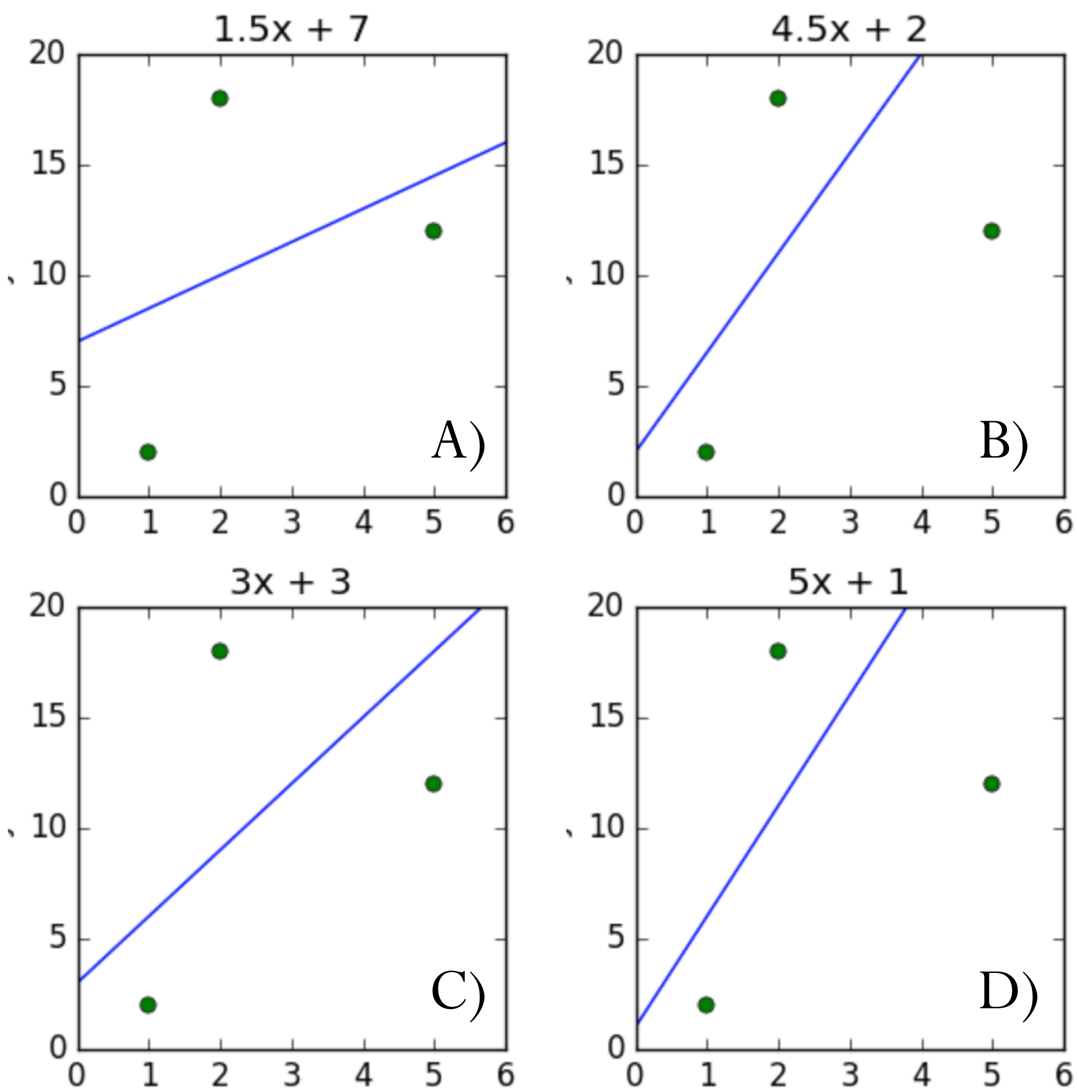
$$\boldsymbol{A}^T\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{A}^T\,\boldsymbol{b}$$

- Compute $\boldsymbol{A}^T\boldsymbol{A}$: $O(mn^2)$
- Factorize $\boldsymbol{A}^T\boldsymbol{A}$: LU $\rightarrow O\left(\frac{2}{3}n^3\right)$, Cholesky $\rightarrow O\left(\frac{1}{3}n^3\right)$
- Solve $O(n^2)$
- Since $m > n$ the overall cost is $O(mn^2)$

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

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

Write the matrix A

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

Demo: “Ice example”

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?

Rank of a matrix

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

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

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \dots & \vdots \\ \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \\ \vdots & \dots & \vdots \end{pmatrix} \begin{pmatrix} \dots & \sigma_1\, \mathbf{v}_1^T & \dots \\ \vdots & \vdots & \vdots \\ \dots & \sigma_n\, \mathbf{v}_n^T & \dots \end{pmatrix} = \sigma_1 \boldsymbol{u}_1 \mathbf{v}_1^T + \sigma_2 \boldsymbol{u}_2 \mathbf{v}_2^T + \cdots + \sigma_n \boldsymbol{u}_n \mathbf{v}_n^T$$

Rank of a matrix

For general rectangular matrix $\boldsymbol{A}$ with dimensions $m \times n$, the reduced SVD is:

$$\underbrace{\boldsymbol{A}}_{m\times n} = \underbrace{\boldsymbol{U_R}}_{m\times k}\ \underbrace{\boldsymbol{\Sigma_R}}_{k\times k}\underbrace{\boldsymbol{V^T}}_{k\times n}$$

$$k = \min(m,n)$$

$$\boldsymbol{A} = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \mathbf{v}_i^T$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \\ 0 & & 0 \\ & \ddots & \vdots \\ & & 0 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_k & 0 & \dots & 0 \end{pmatrix}$$

If $\sigma_i \neq 0\ \forall i$, then $\text{rank}(\boldsymbol{A}) = k$ (Full rank matrix)

In general, $\text{rank}(\boldsymbol{A}) =$ number of non-zero singular values σ_i (Rank deficient)

Rank of a matrix

- The rank of $\mathbf{A}$ equals the number of non-zero singular values which is the same as the number of non-zero diagonal elements in $\boldsymbol{\Sigma}$.
- Rounding errors may lead to small but non-zero singular values in a rank deficient matrix, hence the rank of a matrix determined by the number of non-zero singular values is sometimes called "effective rank".
- The right-singular vectors (columns of $\boldsymbol{V}$) corresponding to vanishing singular values span the null space of $\mathbf{A}$.
- The left-singular vectors (columns of $\boldsymbol{U}$) corresponding to the non-zero singular values of $\mathbf{A}$ span the range of $\mathbf{A}$.

Back to least squares…

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

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:

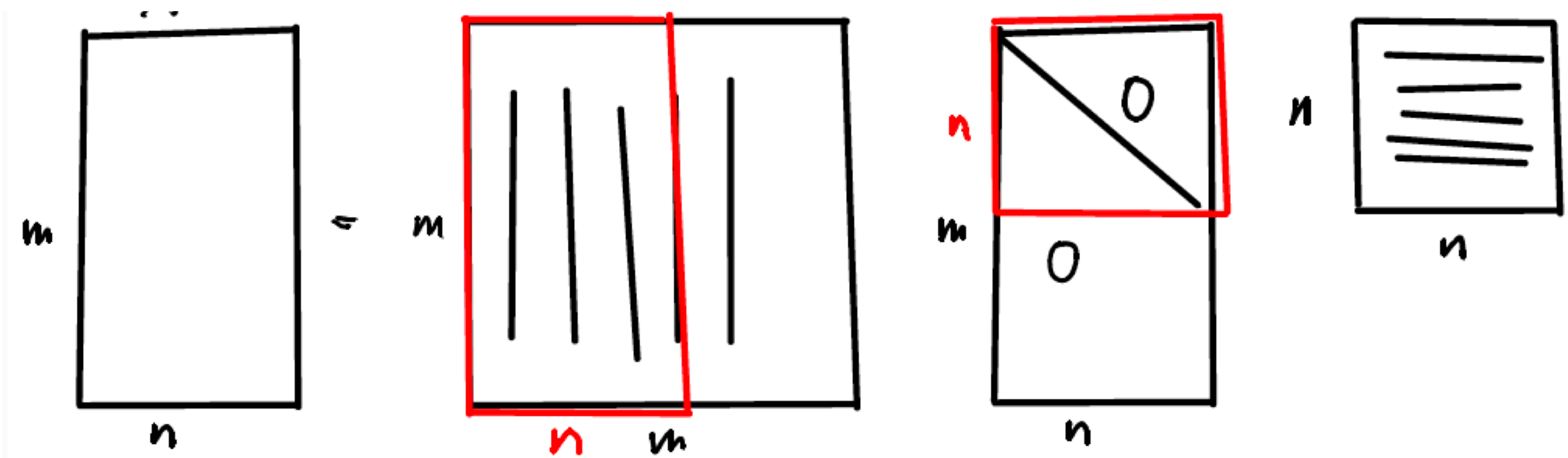
$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots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

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 $m > n$

$$\underset{m\times n}{\boldsymbol{A}} = \underset{m\times n}{\boldsymbol{U}_R}\ \underset{n\times n}{\boldsymbol{\Sigma}_R}\ \underset{n\times n}{\boldsymbol{V}^T}$$



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

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

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

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

Solving Least Squares Problem with SVD (summary)

Cost of SVD:

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

- Find $\boldsymbol{y}$ that satisfies $\min_{\boldsymbol{y}} \| \boldsymbol{\Sigma_R}\, \boldsymbol{y} - \boldsymbol{U}_R{}^T \boldsymbol{b} \|_2^2$

- Propose $\boldsymbol{y}$ that is solution of $\boldsymbol{\Sigma_R}\, \boldsymbol{y} = \boldsymbol{U}_R{}^T \boldsymbol{b}$

Cost:

- Evaluate: $\boldsymbol{z} = \boldsymbol{U}_R{}^T \boldsymbol{b}$ $\longrightarrow$

- Set: $y_i = \begin{cases} \frac{z_i}{\sigma_i}, & \text{if } \sigma_i \neq 0 \\ 0, \text{otherwise} \end{cases} \quad i = 1, \dots, n$ $\longrightarrow$

- Then compute $\boldsymbol{x} = \boldsymbol{V}\,\boldsymbol{y}$ $\longrightarrow$

Solving Least Squares Problem with SVD (summary)

- If $\sigma_i \neq 0$ for $\forall i = 1, \ldots, n$, then the solution $\boldsymbol{y} = \boldsymbol{V}\ (\boldsymbol{\Sigma_R}\ )^{-1}\boldsymbol{U}_R{}^T\ \boldsymbol{b}$ is unique (and not a "choice").

- If at least one of the singular values is zero, then the proposed solution $\boldsymbol{y}$ is the one with the smallest 2-norm ($\|\boldsymbol{y}\|_2$ is minimal) that minimizes the 2-norm of the residual $\left\| \boldsymbol{\Sigma_R}\ \boldsymbol{y} - \boldsymbol{U}_R{}^T\boldsymbol{b} \right\|_2$

- Since $\|\boldsymbol{x}\|_2 = \|\boldsymbol{V}\ \boldsymbol{y}\|_2 = \|\boldsymbol{y}\|_2$, then the solution $\boldsymbol{x}$ is also the one with the smallest 2-norm ($\|\boldsymbol{x}\|_2$ is minimal) for all possible $\boldsymbol{x}$ for which $\|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ is minimal.

Pseudo-Inverse

- **Problem:** $\boldsymbol{\Sigma}$ may not be invertible
- **How to fix it:** Define the Pseudo Inverse
- **Pseudo-Inverse of a diagonal matrix**:

$$(\boldsymbol{\Sigma}^{+})_i = \begin{cases} \frac{1}{\sigma_i}, & \text{if } \sigma_i \neq 0 \\ 0, & \text{if } \sigma_i = 0 \end{cases}$$

- **Pseudo-Inverse of a matrix $\boldsymbol{A}$**:

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

Solving Least Squares Problem with SVD (summary)

Solve $\boldsymbol{A}\,\boldsymbol{x} \cong \boldsymbol{b}$ or $\boldsymbol{U}_R\,\boldsymbol{\Sigma}_{\boldsymbol{R}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{x} \cong \boldsymbol{b}$

$$\boldsymbol{x} \cong \boldsymbol{V}\,(\boldsymbol{\Sigma}_{\boldsymbol{R}})^{+}\,{\boldsymbol{U}_R}^{T}\,\boldsymbol{b}$$

Demo: Least Squares – all together

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

Iclicker question

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) $O(m^2)$