SNE3002 Linear Algebra – 2025 Spring

Least Squares

April 9, 2025

INHA UNIVERSITY

Motivation

- Theory vs Practice gap

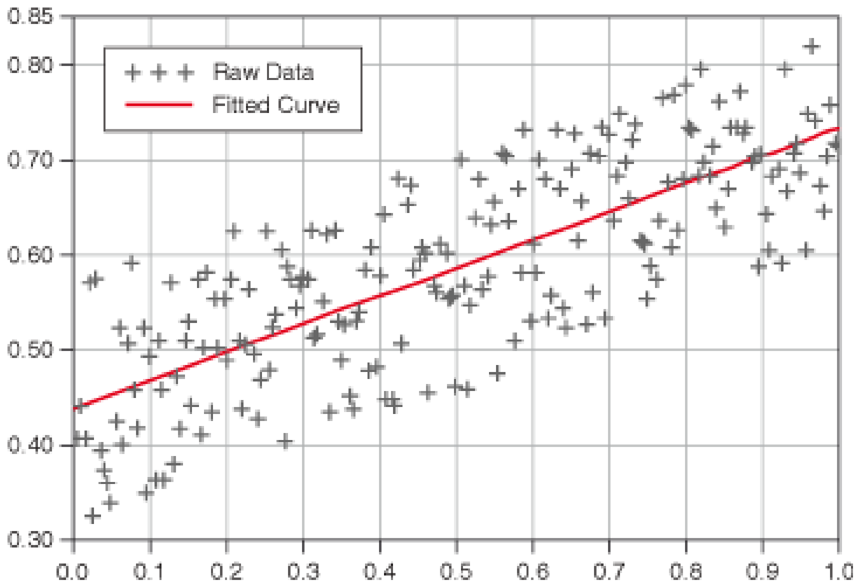
- In real-world, problems are often sloppy. We can solve clean and easy problem, but did you come to University to learn how to solve toy problems?

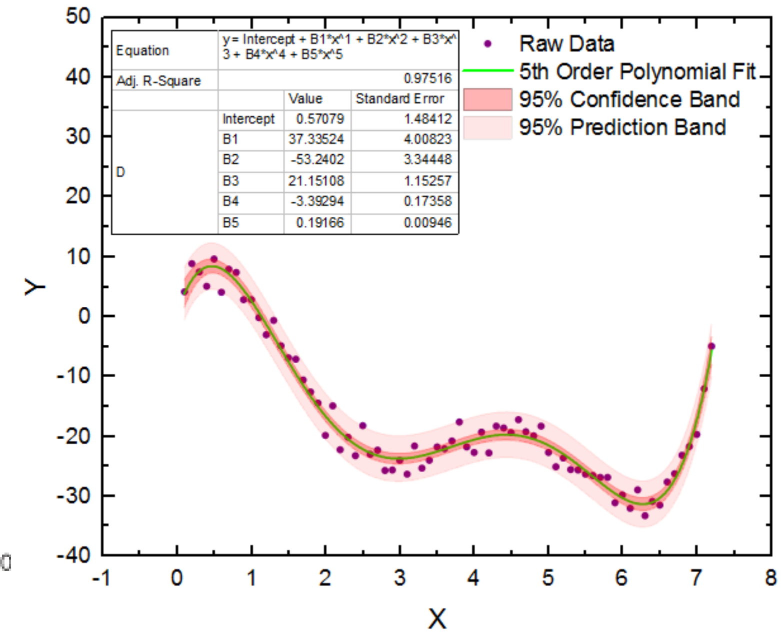
- We will learn how to solve $A\boldsymbol{x} = \boldsymbol{b}$ that do not have an exact solution
- Instead, we will be seeking an approximate answer that minimizes the error in the solution

$$e := A\boldsymbol{x} - \boldsymbol{b}$$

Real Problem

> Data fitting





Real Problem

- Line fitting $Ax = b \quad (y = ax + b)$

- Ex. $(1,7), (0,4), (-1,1)$ points are given
  - $y = 3x + 4$

- $(1,7), (0,4), (-1,1.00001)$points are given
  - No solution
  - Only approximate solution exists

Least Squares Approximate Problem

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + \cdots + a_{2n}x_n = b_2$$
$$\dots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

- $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$

- Then, $A\boldsymbol{x} - \boldsymbol{b} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n - b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n - b_m \end{bmatrix}$

- Therefore, minimizing $||A\boldsymbol{x} - \boldsymbol{b}||^2$ is the same as minimizing the sum of the square of the errors of all the equations.

Algebraic Approach

- minimize $J = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (A\boldsymbol{x} - \boldsymbol{b})^\top (A\boldsymbol{x} - \boldsymbol{b})$

Algebraic Approach

Algebraic Approach

Algebraic Approach 2

- minimize $J = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (A\boldsymbol{x} - \boldsymbol{b})^\top (A\boldsymbol{x} - \boldsymbol{b})$
- $\frac{\mathrm{d}}{\mathrm{dx}}(A\boldsymbol{x} - \boldsymbol{b})^\top (A\boldsymbol{x} - \boldsymbol{b}) = 2A^\top (A\boldsymbol{x} - \boldsymbol{b}) = 0$
- $A^\top A\boldsymbol{x} = A^\top \boldsymbol{b}$
- We can solve it when $A^\top A$ is invertible.
  - $\boldsymbol{x} = (A^\top A)^{-1} A^\top \boldsymbol{b}$
  - $A^\top A$ is invertible = the columns of A are linearly independent
    = $rank(A)$ is n $(A\ is\ m \times n\ matrix)$
    = $A^\top A$ is positive definite

Recall Lec 5

Remark. Useful Theorems about Matrices

The following are equivalent for an $m \times n$ matrix A

1) $\operatorname{rank}(A) = n$
2) $row(A) = \mathbb{R}^n$, i.e., the rows of A span $\mathbb{R}^n$
3) The columns of A are linearly independent in $\mathbb{R}^m$
4) The $\mathrm{n} \times n$ matrix $A^\top A$ is invertible ($\det(A^\top A) \neq 0$)
5) The system $A\boldsymbol{x} = \mathbf{0}$ has only the trivial solution

And

1) $\operatorname{rank}(A) = m$
2) $\operatorname{col}(A) = \mathbb{R}^m$, i.e., the columns of A span $\mathbb{R}^m$
3) The rows of A are linearly independent in $\mathbb{R}^n$
4) The $m \times m$ matrix $AA^\top$ is invertible ($\det(AA^\top) \neq 0$)
5) The system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every $\boldsymbol{b} \in \mathbb{R}^n$

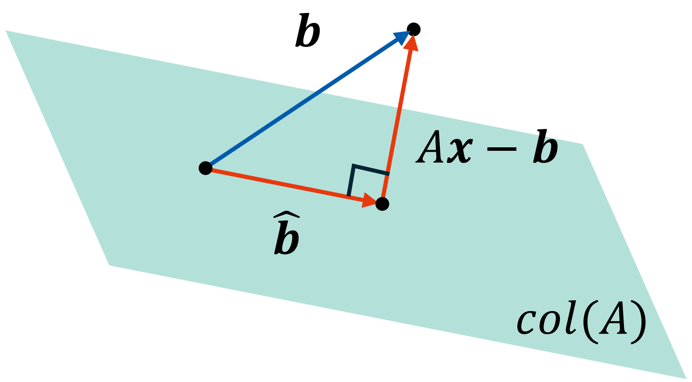
*consistent = there exists at least one solution

Recall

- Proof.

Geometric Approach

- $J = \|A\boldsymbol{x} - \boldsymbol{b}\|$, Let $\widehat{\boldsymbol{b}} = proj_{col(A)}\boldsymbol{b}$
  - We want to find the element of $col(A)$ that is closest to $\boldsymbol{b}$



$A\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{b}}$

  - This happens when $A\boldsymbol{x} - \boldsymbol{b}$ is orthogonal to $col(A)$

- $\boldsymbol{a}_i^\top(A\boldsymbol{x} - \boldsymbol{b}) = 0 \;\rightarrow A^\top(A\boldsymbol{x} - \boldsymbol{b}) = 0$
- $A^\top A\boldsymbol{x} = A^\top\boldsymbol{b} \rightarrow \quad x = (A^\top A)^{-1}A^\top\boldsymbol{b}$

Least Squares

> Q. Find a least-squares solution of the inconsistent system $A\boldsymbol{x} = \boldsymbol{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Least Squares

- Alternative calculation using orthogonality
- Q. Find a least-squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

EXAMPLE 4 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

SOLUTION Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \tag{5}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

Now that $\hat{\mathbf{b}}$ is known, we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. But this is trivial, since we already know what weights to place on the columns of A to produce $\hat{\mathbf{b}}$. It is clear from (5) that

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$ ■

Recall Modified Gram-Schmidt

- Find an orthogonal basis of a subspace
- Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ be a basis for some subspace W

$$\boldsymbol{u}_1 = \boldsymbol{v}_1$$
$$\boldsymbol{u}_1 \leftarrow \boldsymbol{u}_1 / \|\boldsymbol{u}_1\|$$

$$\boldsymbol{u}_2 = \boldsymbol{v}_2 - (\boldsymbol{u}_1 \cdot \boldsymbol{v}_2)\boldsymbol{u}_1$$
$$\boldsymbol{u}_2 \leftarrow \boldsymbol{u}_2 / \|\boldsymbol{u}_2\|$$

$$\boldsymbol{u}_3 = \boldsymbol{v}_3 - (\boldsymbol{u}_1 \cdot \boldsymbol{v}_3)\boldsymbol{u}_1 - (\boldsymbol{u}_2 \cdot \boldsymbol{v}_3)\boldsymbol{u}_2$$
$$\boldsymbol{u}_3 \leftarrow \boldsymbol{u}_3 / \|\boldsymbol{u}_3\|$$
$$\dots$$

$$\boldsymbol{u}_k = \boldsymbol{v}_k - (\boldsymbol{u}_1 \cdot \boldsymbol{v}_k)\boldsymbol{u}_1 - (\boldsymbol{u}_2 \cdot \boldsymbol{v}_k)\boldsymbol{u}_2 - \cdots - (\boldsymbol{u}_{k-1} \cdot \boldsymbol{v}_k)\boldsymbol{u}_{k-1}$$
$$\boldsymbol{u}_k \leftarrow \boldsymbol{u}_k / \|\boldsymbol{u}_k\|$$

QR Factorization

$$\boldsymbol{u}_1 = \boldsymbol{v}_1$$
$$\boldsymbol{u}_1 \leftarrow \boldsymbol{u}_1/\|\boldsymbol{u}_1\|$$

$$\boldsymbol{u}_2 = \boldsymbol{v}_2 - (\boldsymbol{u}_1 \cdot \boldsymbol{v}_2)\boldsymbol{u}_1$$
$$\boldsymbol{u}_2 \leftarrow \boldsymbol{u}_2/\|\boldsymbol{u}_2\|$$

$$\boldsymbol{u}_3 = \boldsymbol{v}_3 - (\boldsymbol{u}_1 \cdot \boldsymbol{v}_3)\boldsymbol{u}_1 - (\boldsymbol{u}_2 \cdot \boldsymbol{v}_3)\boldsymbol{u}_2$$
$$\boldsymbol{u}_3 \leftarrow \boldsymbol{u}_3/\|\boldsymbol{u}_3\|$$

$$\Rightarrow$$

$$\widehat{\boldsymbol{q}}_1 = \boldsymbol{v}_1$$
$$\boldsymbol{q}_1 = \frac{\widehat{\boldsymbol{q}}_1}{\|\widehat{\boldsymbol{q}}_1\|} = \frac{1}{r_{1,1}}\widehat{\boldsymbol{q}}_1$$

$$\begin{aligned}\widehat{\boldsymbol{q}_2} &= \boldsymbol{v}_2 - (\boldsymbol{q}_1^\top \cdot \boldsymbol{v}_2)\boldsymbol{q}_1 \\ &= \boldsymbol{v}_2 - r_{1,2}\boldsymbol{q}_1\end{aligned}$$

$$\boldsymbol{q}_2 = \frac{\widehat{\boldsymbol{q}_2}}{\|\widehat{\boldsymbol{q}_2}\|} = \frac{1}{r_{2,2}}\widehat{\boldsymbol{q}_2}$$

$$\begin{aligned}\widehat{\boldsymbol{q}_3} &= \boldsymbol{v}_3 - (\boldsymbol{q}_1^\top \boldsymbol{v}_3)\boldsymbol{q}_1 - (\boldsymbol{q}_2^\top v_3)\boldsymbol{q}_2 \\ &= \boldsymbol{v}_3 - r_{1,3}\boldsymbol{q}_1 - r_{2,3}\boldsymbol{q}_2\end{aligned}$$
$$\boldsymbol{q}_3 = \frac{\widehat{\boldsymbol{q}_3}}{\|\widehat{\boldsymbol{q}_3}\|} = \frac{1}{r_{3,3}}\widehat{\boldsymbol{q}}_3$$

QR Factorization

$$\hat{\boldsymbol{q}}_1 = \boldsymbol{v}_1$$

$$\boldsymbol{q}_1 = \frac{\hat{\boldsymbol{q}}_1}{\|\hat{\boldsymbol{q}}_1\|} = \frac{1}{r_{1,1}}\hat{\boldsymbol{q}}_1$$

$$\widehat{\boldsymbol{q}_2} = \boldsymbol{v}_2 - (\boldsymbol{q}_1^\top \cdot \boldsymbol{v}_2)\boldsymbol{q}_1$$
$$= \boldsymbol{v}_2 - r_{1,2}\boldsymbol{q}_1$$

$$\boldsymbol{q}_2 = \frac{\widehat{\boldsymbol{q}_2}}{\|\widehat{\boldsymbol{q}_2}\|} = \frac{1}{r_{2,2}}\widehat{\boldsymbol{q}_2}$$

$$\widehat{\boldsymbol{q}_3} = \boldsymbol{v}_3 - (\boldsymbol{q}_1^\top \boldsymbol{v}_3)\boldsymbol{q}_1 - (\boldsymbol{q}_2^\top v_3)\boldsymbol{q}_2$$
$$= \boldsymbol{v}_3 - r_{1,3}\boldsymbol{q}_1 - r_{2,3}\boldsymbol{q}_2$$

$$\boldsymbol{q}_3 = \frac{\widehat{\boldsymbol{q}_3}}{\|\widehat{\boldsymbol{q}_3}\|} = \frac{1}{r_{3,3}}\hat{\boldsymbol{q}}_3$$

$$r_{1,1}\boldsymbol{q}_1 = \boldsymbol{v}_1$$

$$r_{2,2}\boldsymbol{q}_2 = \boldsymbol{v}_2 - r_{1,2}\boldsymbol{q}_1$$

$$r_{3,3}\boldsymbol{q}_3 = \boldsymbol{v}_3 - r_{1,3}\boldsymbol{q}_1 - r_{2,3}\boldsymbol{q}_2$$

$$\boldsymbol{v}_1 = r_{1,1}\boldsymbol{q}_1 + 0\boldsymbol{q}_2 + 0\boldsymbol{q}_3$$

$$\boldsymbol{v}_2 = r_{1,2}\boldsymbol{q}_1 + r_{2,2}\boldsymbol{q}_2 + 0\boldsymbol{q}_3$$

$$\boldsymbol{v}_3 = r_{1,3}\boldsymbol{q}_1 + r_{2,3}\boldsymbol{q}_2 + r_{3,3}\boldsymbol{q}_3$$

QR Factorization

> $$V = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \boldsymbol{v}_3] = \underbrace{[\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \boldsymbol{q}_3]}_{Q} \underbrace{\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ 0 & r_{2,2} & r_{2,3} \\ 0 & 0 & r_{3,3} \end{bmatrix}}_{R}$$

Remark. QR Factorization

Suppose $V \in \mathbb{R}^{m \times n}$ is any matrix with linearly independent columns. Then there exists an orthogonal basis $Q \in \mathbb{R}^{m \times n}$ and an invertible upper triangular matrix $\mathrm{R} \in \mathbb{R}^{n \times n}$ such that

$$V = QR$$

QR Factorization

> Since $Q^\top Q = I$, $Q^\top = Q^{-1}$

> If $A = QR$, a least square solution is

$$
\begin{aligned}
x &= (A^\top A)^{-1} A^\top \boldsymbol{b} \\
&= ((QR)^\top QR)^{-1} (QR)^\top \boldsymbol{b} \\
&= (R^\top Q^\top QR)^{-1} R^\top Q^\top \boldsymbol{b} \\
&= (R^\top R)^{-1} R^\top Q^\top \boldsymbol{b} \\
&= R^{-1} (R^\top)^{-1}\, R^\top Q^\top b \\
&= R^{-1} Q^\top \boldsymbol{b}
\end{aligned}
$$

QR Factorization

- Note. Cheloesky Decomposition

- $A^\top A = \tilde{A} = R^\top R$
- In general, decomposing into a lower triangular matrix $A = LL^\top$ or an upper triangular matrix $A = U^\top U$.

QR Factorization

- Note. QR Iteration Algorithm

- $A = QR$

  $A_k \leftarrow RQ$

  $A_{k+1} = QR$

  $A_{k+2} \leftarrow RQ$

- If A has only real eigenvalues, this procedure usually converges to an upper triangular matrix, whose eigenvalues are displayed along its main diagonal.

  Proof.

Least Squares using SVD

> $A = U\Sigma V^\top$, U and V are orthogonal matrices.

> $x = (A^\top A)^{-1} A^\top \boldsymbol{b}$
$= ((U\Sigma V^\top)^\top (U\Sigma V^\top))^{-1} (U\Sigma V^\top)^\top \boldsymbol{b}$
$= (V\Sigma U U^\top \Sigma V^\top)^{-1} (U\Sigma V^\top)^\top \boldsymbol{b}$
$= (V\Sigma\Sigma V^\top)^{-1} (U\Sigma V^\top)^\top \boldsymbol{b}$
$= V\Sigma^{-1}\Sigma^{-1} V^\top V \Sigma U^\top \boldsymbol{b}$ (?)
$= V\Sigma^{-1} U^\top \boldsymbol{b}$ (?)

> $Ax = b, x = A^{-1}\boldsymbol{b} = V\Sigma^{-1}U^\top \boldsymbol{b}$

> $x = V\Sigma^{+} U^\top \boldsymbol{b}$

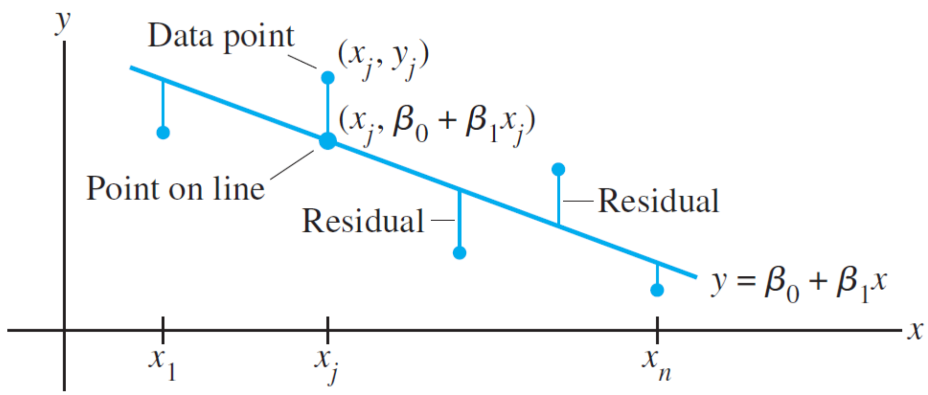
Least Squares Computation Summary

> Methods for solving LS problems

Method	Computational Cost	Numerical Stability	Notes
Normal Eqn. $\boldsymbol{x} = (A^\top A)^{-1} A^\top \boldsymbol{b}$	$O(mn^2 + n^3)$	Poor	Fast but can be unstable; Not recommended if A is ill-conditioned or rank-deficient
QR Decomposition $\boldsymbol{x} = R^{-1} Q^\top \boldsymbol{b}$	$O(mn^2)$	Good	Fast, numerically stable, commonly used in practice
SVD $\boldsymbol{x} = V \Sigma^{+} U^\top \boldsymbol{b}$	$O(mn^2 + n^3)$	Excellent	Most stable; works even when A is rank-deficient Gives minimum-norm solution

LS Examples

> Least-squares line



Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n

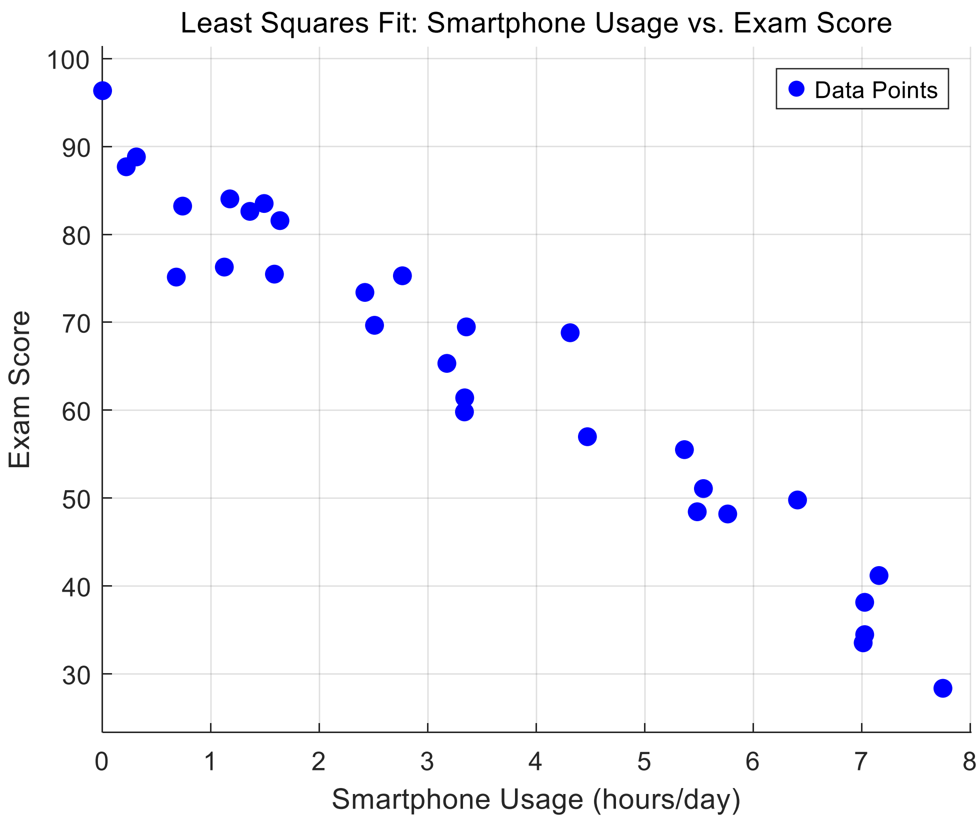
$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

LS Examples

- Least-squares line
- Q. Find the equation $y = \beta_1 x + \beta_0$ of the least-squares line that best fits the data points (2,1), (5,2), (7,3), and (8,3)

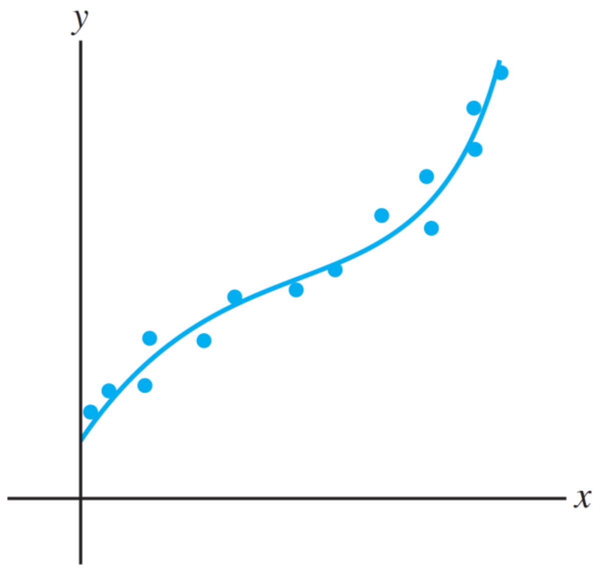
LS Examples

- Least-squares line (MATLAB)



LS Examples

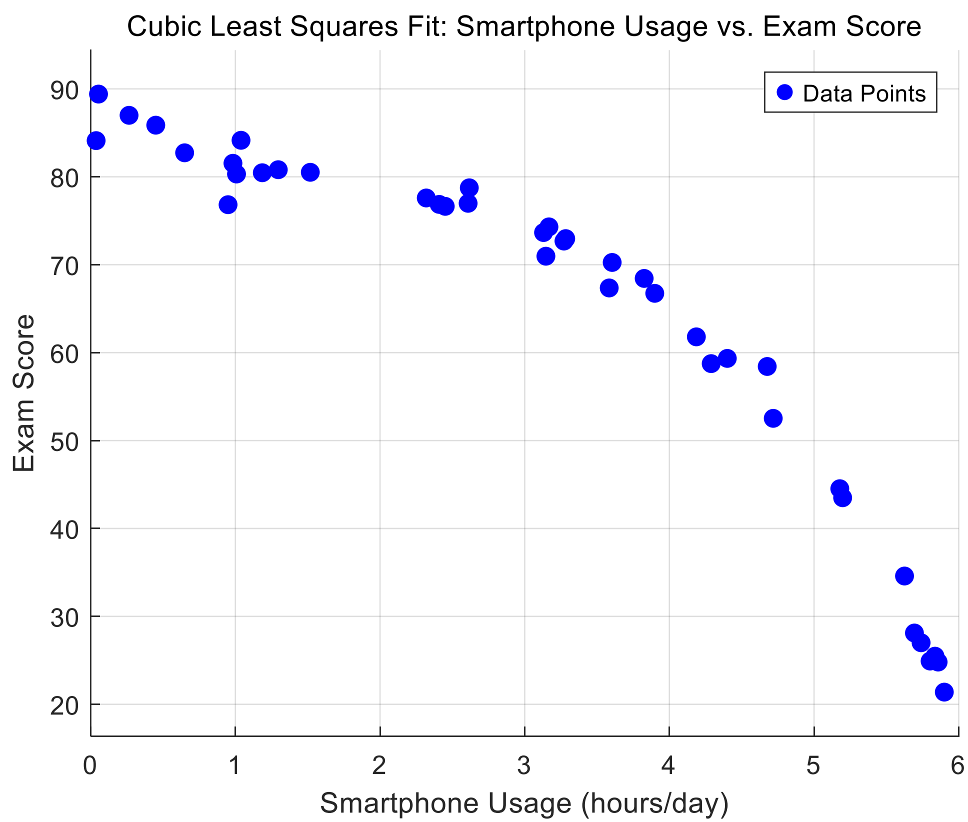
- LS fitting of other curves
- $y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x)$
  - Ex. $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$



$$\begin{array}{ccc} \text{Observation vector} & \text{Design matrix} & \text{Parameter vector} \\ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, & X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}, & \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \end{array}$$

LS Examples

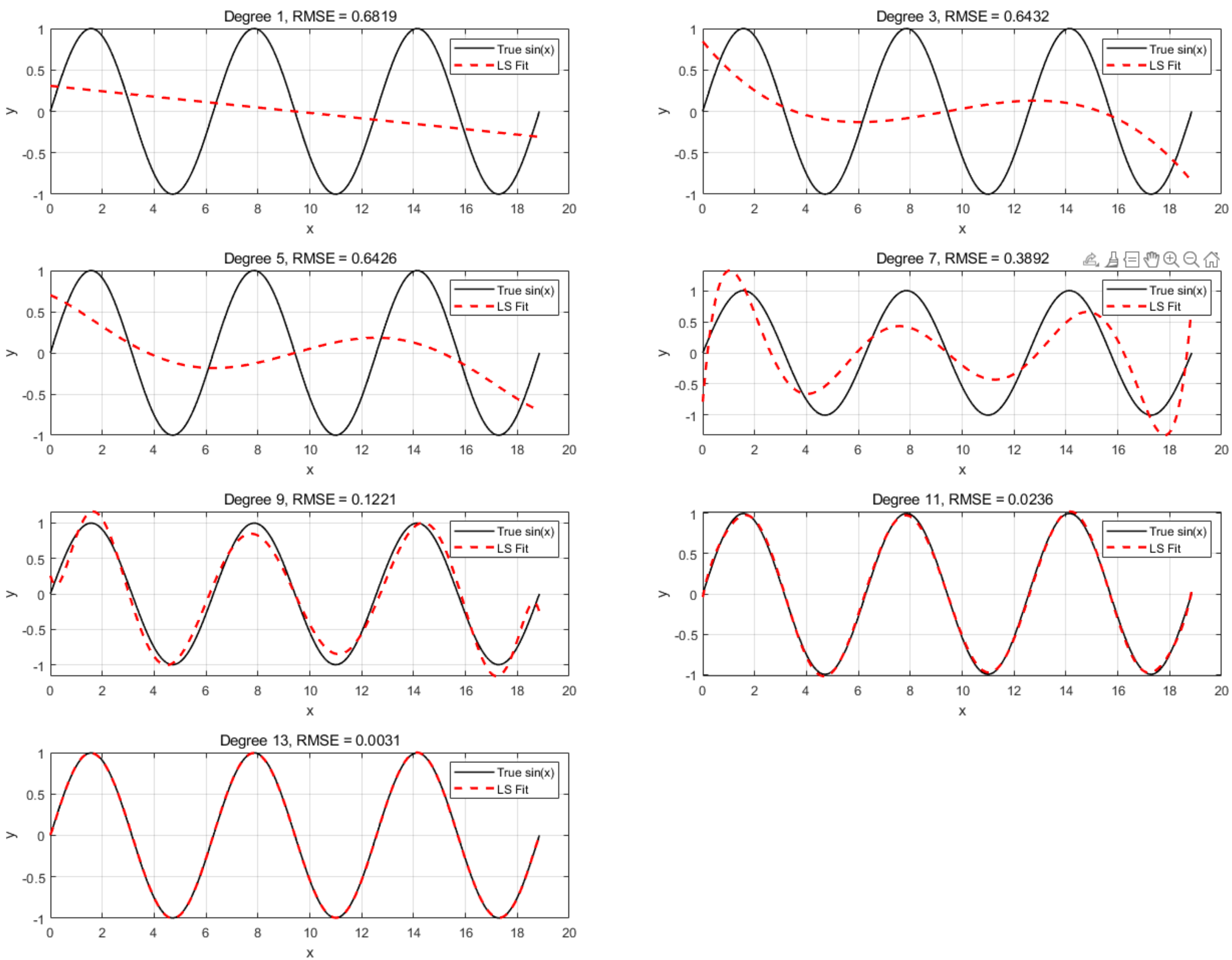
> Least-squares (MATLAB)



LS Examples

- Taylor series:
  - We know that $\sin(\mathrm{x}) = \mathrm{x} - \frac{\mathrm{x}^3}{3!} + \frac{\mathrm{x}^5}{5!} - \cdots$
  - We can approximate this as $\sin(x) = \Sigma_{i=1}^{\infty} \beta_i x^{2i-1}$
  - Can we do that with LS?

LS Examples



Degree 1, RMSE = 0.6819
Degree 3, RMSE = 0.6432
Degree 5, RMSE = 0.6426
Degree 7, RMSE = 0.3892
Degree 9, RMSE = 0.1221
Degree 11, RMSE = 0.0236
Degree 13, RMSE = 0.0031
True sin(x)
LS Fit
x
y

Moore-Penrose Inverse

> Moore-penrose Inverse (Pseudo Inverse)

> $x = A^{-1}\boldsymbol{b}$ or $x = (A^{\top}A)^{-1}A^{\top}\boldsymbol{b}$

> $A^{+} = (A^{\top}A)^{-1}A^{\top}\boldsymbol{b}$, when $A^{\top}A$ is invertible

Def. Moore-Penrose Inverse

For $A \in \mathbb{R}^{m\times n}$, a Moore-Penrose Inverse of A satisfies all of the following four criteria,

1. $AA^{+}A = A$
2. $A^{+}AA^{+} = A^{+}$
3. $(AA^{+})^{\top} = AA^{+}$
4. $(A^{+}A)^{\top} = A^{+}A$

Moore-Penrose Inverse

Remark. Moore-Penrose Inverse

Moore-Penrose Inverse always exists and is unique.

Expanding LS Problems

> Tikhonov Regularization
  - minimize $J = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \lambda^2\|\boldsymbol{x}\|^2$

> Constrained Least Squares problem (CLS)
  - minimize $J = \|A\boldsymbol{x} - \boldsymbol{b}\|^{\mathbf{2}}$
    subject to $Cx = d$