Northeastern University, Department of Mathematics

MATH 5110: Applied Linear Algebra and Matrix Analysis. (Fall 2021)

• Instructor: **He Wang** Email: **he.wang@northeastern.edu**

§11 Least Squares and Data Fitting

Contents

Review:

1. Inner Product Space (vector space with an inner product)

2. Use orthogonal basis (to find orthogonal projection)

3. Find orthogonal basis (Gram-Schmidt process)

1. **Least Squares Problem**

Set up:

Question:

Answer:

Calculation:

Method (1)

Method (2)

2. **Approximate Solutions to Inconsistent Systems**

Set up:

Let A be an $n \times m$ matrix.

Let $\vec{b} \in \mathbb{R}^n$.

Suppose $A\vec{x} = \vec{b}$ has no solution.

Consider $\mathbb{R}^n$ with any inner product.

[Least-Squares Problem/Solution for $A\vec{x} = \vec{b}$]

Problem: Find the vector(s) $\vec{x}_* \in \mathbb{R}^m$ such that for all $x \in \mathbb{R}^m$,

$$||A\vec{x}_* - \vec{b}|| \leq ||A\vec{x} - \vec{b}||$$

Solutions:

Example 1. Find the least-squares solutions for $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix}$ and $\begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix}$

In particular, if we consider **dot product** on $\mathbb{R}^n$, we have the following formula.

Theorem 2. *(Normal Equation) The Least-Square solutions of $A\vec{x} = \vec{b}$ coincide with the solutions of of* ***normal equations***

$$(A^T A)\vec{x} = A^T\vec{b}.$$

More generally, we can also consider **weighted dot product** on $\mathbb{R}^n$,

$$\langle \vec{u}, \vec{v} \rangle_W := \vec{u}^T W \vec{v}$$

where W is a positive-definite symmetric matrix.

Example 3. Find the least-squares solutions for $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix}$

$$A^T A = \begin{bmatrix} -1 & 1 & -1 \\ 4 & 8 & 4 \end{bmatrix} \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 96 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} -1 & 1 & -1 \\ 4 & 8 & 4 \end{bmatrix} \begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} -18 \\ 24 \end{bmatrix}$$

Solve the normal equation $A^T A \vec{x} = A^T \vec{b}$

$$\left[\begin{array}{cc|c} 3 & 0 & -18 \\ 0 & 96 & 24 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 0 & -6 \\ 0 & 1 & 4 \end{array}\right]$$

So $\vec{x} = \begin{bmatrix} -6 \\ 4 \end{bmatrix}$ is the least-squares solution

(2) The image $\mathrm{im}(A)$ is a plane in $\mathbb{R}^3$ passing the origin. Find the distance from the vector $\vec{b}$ (or the point $(14, -4, 0)$) to the plane $\mathrm{im}(A)$.

The distance is given be the norm of $\vec{b}^{\perp} = \vec{b} - \mathrm{proj}_{\mathrm{im}(A)} \vec{b}$.

We know that $\mathrm{proj}_{\mathrm{im}(A)} \vec{b} = Ax_* = \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} -6 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \\ 7 \end{bmatrix}$.

So, $\vec{b}^{\perp} = \begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix} - \begin{bmatrix} 7 \\ -4 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix}$. So the distance is $||\vec{b}^{\perp}|| = 7\sqrt{2}$.

Example 4. Find the least-squares solutions for the system $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}$

Step 1. Construct the normal equation $A^TA\vec{x} = A^T\vec{b}$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ 4 \\ 6 \end{bmatrix}$$

Solve the normal equation

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 10 \\ 2 & 2 & 0 & 4 \\ 2 & 0 & 2 & 6 \end{array}\right] \rightarrow \cdots \longrightarrow \text{rref} = \left[\begin{array}{ccc|c} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$x_1 = 3 - x_3$

$x_2 = -1 + x_3$

x_3 free

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 - x_3 \\ -1 + x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

A technical property:

Proposition 5. *Let A be an $n \times m$ matrix.*

$$\ker(A) = \ker(A^T A)$$

Corollary 6. *If* $\operatorname{rank} A = m$*, the normal equation* $(A^T A)\vec{x} = A^T\vec{b}$ *has a unique solution:*

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}$$

QR factorization method Suppose A is $n \times m$ matrix with full column rank. Solve the least squares solution using QR factorization $A = QR$ where Q is an orthogonal matrix $n \times m$ and R is an $m \times m$ upper triangular matrix with rank m.

Problem: Fitting a function of a certain type of data. We use the following three example to illustrate this application.

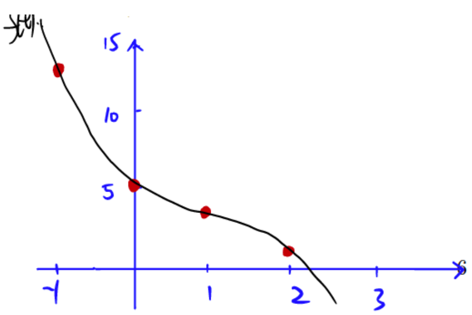
Example 7. Find a cubic polynomial $f(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$ whose graph passes through the points $(0, 5)$, $(1, 3)$, $(-1, 13)$, $(2, 1)$

Solution:

We need to solve the linear system $\begin{cases} c_0 & = 5 \\ c_0 + c_1 + c_2 + c_3 & = 3 \\ c_0 - c_1 + c_2 - c_3 & = 13 \\ c_0 + 2c_1 + 4c_2 + 8c_3 & = 1 \end{cases}$

$$[A|\vec{b}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 5 \\ 1 & 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & -1 & 13 \\ 1 & 2 & 4 & 8 & 1 \end{bmatrix} \to \cdots \to \textbf{rref}[A|\vec{b}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

So, the linear system has the unique solution $\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ 3 \\ -1 \end{bmatrix}$ So, the cubic polynomial is $f(t) = 5 - 4t + 3t^2 - t^3$.



perfect fit, but calculation is hard.

Example 8. Fit a quadratic function $g(t) = c_0 + c_1 t + c_2 t^2$ to the four data points $(0, 5)$, $(1, 3)$, $(-1, 13)$, $(2, 1)$

We need to solve the linear system

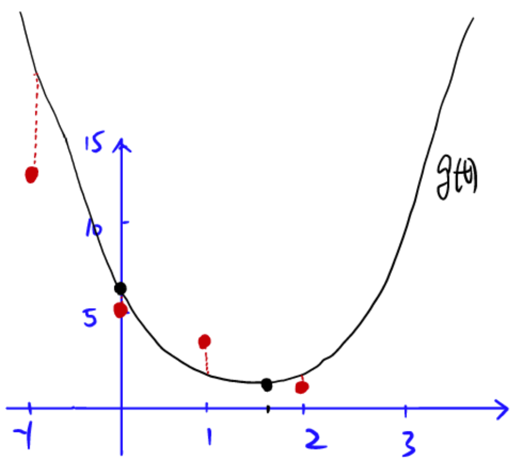
$$\begin{cases} c_0 & = 5 \\ c_0 + c_1 + c_2 & = 3 \\ c_0 - c_1 + c_2 & = 13 \\ c_0 + 2c_1 + 4c_2 & = 1 \end{cases}$$

As matrix equation $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 5 \\ 3 \\ 13 \\ 1 \end{bmatrix}$

$A^TA = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}$ $\quad A^T\vec{b} = \begin{bmatrix} 22 \\ -8 \\ 20 \end{bmatrix}$

Solve the normal equation $A^TA\vec{x} = A^T\vec{b}$ $\quad \vec{x} = \begin{bmatrix} 5.9 \\ -5.3 \\ 1.5 \end{bmatrix} = \vec{c}^*$

So, the quadratic function $g(t) = 5.9 - 5.3t + 1.5t^2$



$A\vec{c}^* = \begin{bmatrix} g(a_1) \\ g(a_2) \\ g(a_3) \\ g(a_4) \end{bmatrix}$ $\quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$

$$\|\vec{b} - A\vec{c}^*\|^2 = (b_1 - g(a_1))^2 + (b_2 - g(a_2))^2 + (b_3 - g(a_3))^2 + (b_4 - g(a_4))^2$$

The sum of the vertical distances between graph and data points is minimal.

Example 9. Fit a linear function $h(t) = c_0 + c_1 t$ to the four data points $(0, 5)$, $(1, 3)$, $(-1, 13)$, $(2, 1)$

We need to solve the linear system

$$\begin{cases} c_0 & = 5 \\ c_0 + c_1 & = 3 \\ c_0 - c_1 & = 13 \\ c_0 + 2c_1 & = 1 \end{cases}$$

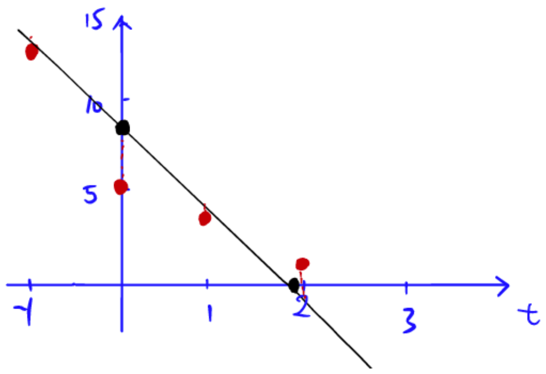
As matrix equation $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 5 \\ 3 \\ 13 \\ 1 \end{bmatrix}$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 13 \\ 1 \end{bmatrix} = \begin{bmatrix} 22 \\ -8 \end{bmatrix}$$

Solve the normal equation $A^TA\vec{x} = A^T\vec{b}$ $\quad \vec{x} = \begin{bmatrix} 7.4 \\ -3.8 \end{bmatrix}$

So the linear function is $h(t) = 7.4 - 3.8t$

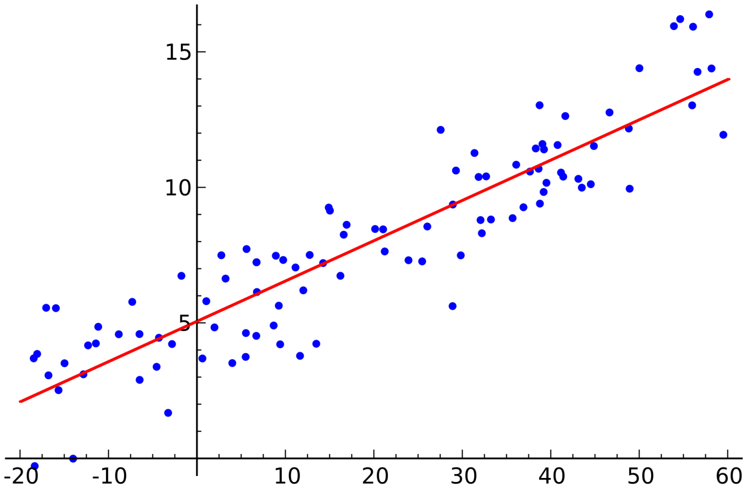


Remark: More generally, we can consider n-points $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)$.

- Find a linear function $h(t) = c_0 + c_1 t$ fits the data by the least squares.

More generally, the following question is very standard in statistics.

Example 10. Consider the data with n points (a_1, b_1), (a_2, b_2), $\ldots$, (a_n, b_n). Find a linear function $h(t) = c_0 + c_1 t$ fits the data by the least squares. (Suppose $a_1 \neq a_2$)



We need to solve the least-squares problem for $A\vec{x} = \vec{b}$, for $A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_n \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

$$A^T A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_n \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n a_i \\ \sum_{i=1}^n a_i & \sum_{i=1}^n a_i^2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n b_i \\ \sum_{i=1}^n a_i a_i \end{bmatrix}$$

Since $a_1 \neq a_2$, we know that rank $A = 2$.
The normal equation $A^T A\vec{x} = A^T\vec{b}$ has a unique solution

$$\vec{x}_* = (A^T A)^{-1} A^T \vec{b} = \frac{1}{n\sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2} \begin{bmatrix} \sum_{i=1}^n a_i^2 & -\sum_{i=1}^n a_i \\ -\sum_{i=1}^n a_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n b_i \\ \sum_{i=1}^n a_i a_i \end{bmatrix}$$

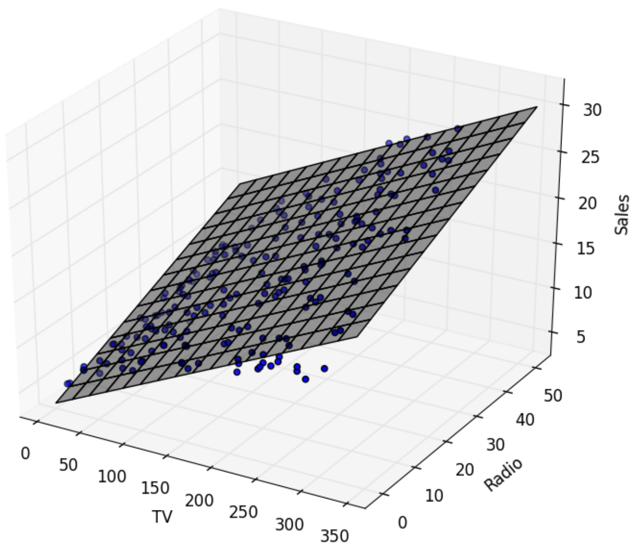
$$= \frac{1}{n\sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2} \begin{bmatrix} (\sum_{i=1}^n a_i^2)(\sum_{i=1}^n b_i) - (\sum_{i=1}^n a_i)(\sum_{i=1}^n a_i a_i) \\ -(\sum_{i=1}^n a_i)(\sum_{i=1}^n b_i) + n\sum_{i=1}^n a_i a_i \end{bmatrix}$$

Example 11. Consider the data with m inputs and 1 output:

$(a_{11}, a_{12}, ..., a_{1m}, b_1),\ (a_{21}, a_{22}, ..., a_{2m}, b_2),\ \ldots,\ (a_{n1}, a_{n2}, ..., a_{nm}, b_n)$.

Find a linear function $h(t_1, t_2, ..., t_n) = c_0 + c_1 t_1 + c_2 t_2 + \cdots + c_n t_n$ fits the data by the least squares.

For example, when $m = 2$,



We need to solve the least-squares problem for $A\vec{x} = \vec{b}$, for $A = \begin{bmatrix} 1 & a_{11} & a_{12} & ... & a_{1m} \\ 1 & a_{21} & a_{22} & ... & a_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & a_{n1} & a_{n2} & ... & a_{nm} \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

Example 12. Consider the data with m inputs and s outputs:

$(a_{11}, a_{12}, ..., a_{1m}, b_{11}, ..., b_{1s}),\ (a_{21}, a_{22}, ..., a_{2m}, b_{21}, ..., b_{2s}),\ \ldots,\ (a_{n1}, a_{n2}, ..., a_{nm}, b_{n1}, ..., b_{ns})$.

Find a linear function $H(\vec{t}) = \vec{c}_0 + C\vec{t}$ fits the data by the least squares.

4. **Best Approximation for Functions**

Set up:

Let V be the vector space of continuous functions.

Consider the inner product on V:

Let W be the subspace of polynomials of degree $\leq n$.

Consider a function $f(x) \in V$. (e.g., $f(x) = e^x$)

Question:

Find the best degree n polynomial approximation of $f(x)$.

Example: