Lecture 16

Least Squares Problem I

Instructor: Ruoyu Sun

香港中文大學(深圳) The Chinese University of Hong Kong, Shenzhen | 数据科学学院 School of Data Science

Today's Lecture: Outline

Today: Least Squares Problem

1. Motivation: Linear Regression

2. Least Squares Problem

3. Solving Least Squares Problem

Strang's book: Sec

Today's Lecture: Learning Goals

After today's lecture, you shall be able to:

1. Model certain real-world problems as linear regression problems
2. Tell the relation/difference of least squares and linear equations
3. Solve a least squares problem

Roadmap of Linear Systems

Theme: Solving Linear Systems	Math Tools	What is Learned? (On linear system)	
Part 1 Preparation (Lec 3,4,5)	Matrix mulplication Row operations	Linear system: $Ax = b$	
Part 2 Solving Square system (Lec 6,7,8,9,10)	LU Decomposition Inverse	"Good" square system: Solution $A^{-1}b$	Method & Analytical Expressions
Part 3 Solving rectangular system (Lec 11,12, 13)	Linear space Span	Rectangular system: Solution $x_p + N(A)$	Method & Analytical Expressions
Part 4 Solution set structure (Lec 13,14,15)	Linear independence Basis Dimension Rank	Solution set dim (n-r) Column space dim r	Deeper understanding
Part 5 Approximate solutions (Lec 16,17)	Least square	Find an approximate solution	

Today's topic

Roadmap of MATH2041

Segment 1 Solving linear systems (Lec 3-15)

Segment 2 **[Lec 16-20]:** Three relatively independent parts:

—Solving least squares problem (Lect 16,17)

—Determinant. [Important tool!] (Lec 18)

—Linear transformation. [Advanced math perspective of matrix] (Lec 19,20)

Segment 3 **Lec 21-27: Eigenvalues and related.**

—Eigenvalues. Lec 21-24

—**Singular values**. Lec 25-26

—Quadratic forms. Lec 27.

Part I Motivation: Linear Regression

Motivation: Salary Prediction

Example (Salary prediction)

	Score	Assist	Salary (million dollars)
Paul	16	10	35
Harden	25	8	45
Capela	18	3	15
Tom	12	14	???

Suppose I’m the Boss of Rocket.
Now I want to hire a new guy Tom.

How much shall I pay Tom?

Motivation: Salary Prediction

	Score	Assist	Salary (million
Paul	16	10	35
Harden	25	8	45
Capela	18	3	15
Tom	12	14	???

Linear model:

Assume Salary = $a \cdot$ Score + $b \cdot$ Assist.

Equations:

Is there a solution?

How to Proceed?

Is there a solution?

Business is not math class.
You cannot say “no solution”.

You have to find a solution.
So… what to do?

Find an approximate solution!

Find w s.t. ____________________

But….

What Approximate Solution?

Find an approximate solution!

Find w s.t. __$\|Ax - b\|$ is small______

Still not clear enough: how small is small?

What Approximate Solution?

Find an approximate solution!

Find w s.t. __$\|Ax - b\|$ is small______

Still not clear enough: how small is small?

One trial: Set the goal as find x s.t. $\|Ax - b\|^2 \leq 10^{-8}$

Issue: Anyways, Boss needs a solution.

Method: Make error as small as possible.

Minimization Problem

Method: Make error as small as possible.

Find x s.t. $\|Ax - b\|$ is the smallest among all.

More precisely, find x^* such that

$$\|Ax^* - b\| \le \|Ax - b\|, \quad \forall x\,.$$

Equivalently, solve the following problem:

$$\min_{w} \|Aw - b\|^2$$

Part II Least Squares: Definition

Definition
Row Interpretation: Linear Regression
Column Interpretation: Residual

Least Squares

What if a linear system $A\mathbf{x} = \mathbf{b}$ has no solution?

One method: Find the best "approximation"!

Definition (Least Squares Problem)

Given a matrix $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^{m\times 1}$, the **least squares problem** is

$$\min_{\mathbf{x}\in\mathbb{R}^{n\times 1}} \|A\mathbf{x} - \mathbf{b}\| \quad (*)$$

Remark: $\|\cdot\|$ is the ℓ_2 **(Euclidean) norm** between vectors (lecture 2)

$\|\mathbf{y}\|$ is a **solution** of the problem (*) (or "the least square problem") if

$$\|A\mathbf{y} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\| \text{ for any } \mathbf{x}$$

Do not call this $\mathbf{y}$ a "solution of the linear system $A\mathbf{x} = \mathbf{b}$"!

Least Squares

Question

Are the following problems equivalent?

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\| \qquad \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2$$

Rewrite what they mean:

Find their relation:

Least Squares: Row Interpretation

Suppose $\tilde{\mathbf{a}}_1^\top, \ldots \tilde{\mathbf{a}}_m^\top$ are rows of A

$$\|A\mathbf{x} - \mathbf{b}\|^2 = (\tilde{\mathbf{a}}_1^\top \mathbf{x} - b_1)^2 + \cdots + (\tilde{\mathbf{a}}_n^\top \mathbf{x} - b_n)^2$$

Remark: The least square objective is a sum of squares of residual components

Remark: so least squares minimizes sum of squares of residuals

- solving $A\mathbf{x} = \mathbf{b}$ is making all residuals zero
- least squares attempts to make them all small

Application of Least Squares: Linear Regression

Examples (Salary prediction)

	Score	Assist	Salary (million dollars)
Paul	16	10	35
Harden	25	8	43
Capela	18	3	16
Tom	12	14	???

How much shall I pay Tom?

Step 1: (estimate the value of "score" and "assist")

Compute $\mathbf{x}$ s.t. the $\sum_{i=1}^{m} (b_i - \mathbf{x}^\top \mathbf{a}_i)^2 = \|A\mathbf{x} - \mathbf{b}\|^2$ is the smallest

(among all possible choices of $\mathbf{x}$), where $A = \begin{bmatrix} 16 & 10 \\ 25 & 8 \\ 18 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 35 \\ 43 \\ 16 \end{bmatrix}$

Denote the solution as $\mathbf{y}$.

Step 2: (estimate salary)

Compute $\mathbf{y}^\top \mathbf{a}_{\text{Tom}}$.

Linear Regression

Consider $\langle \mathbf{x}, \mathbf{a}_i \rangle + r_i = b_i, \quad i = 1,2,\ldots,m\,.$

Task (informal):
Compute x s.t. the approximation error $r_i \triangleq b_i - \mathbf{x}^\top \mathbf{a}_i, \forall i$ are "small".

Task (informal):
Compute $\mathbf{x}$ s.t. the $\sum_{i=1}^{m} (b_i - \mathbf{x}^\top \mathbf{a}_i)^2 = \|A\mathbf{x} - \mathbf{b}\|^2$ is the smallest
(among all possible choices of $\mathbf{x}$)

This problem is called linear regression.
The first model in machine learning.

Linear Regression is Fundamental

Machine learning:

Linear regression is the first model.

Deep learning (深度学习):

Linear regression is the foundation.

Statistics:

Will spend many lectures on it.

Ecometrics (计量经济学):

Will spend many lectures on it.

…

Don’t be too surprised (and bored) if you see it again next year.

Over-Determined Systems (Tall Matrix)

Proposition 13.1 (Column space and Solvability)

$A\mathbf{x} = \mathbf{b}$ has a solution iff $\mathbf{b} \in C(A)$

No solution happens quite often for over-determined linear system

In real-world applications, for “tall" system $A\mathbf{x} = \mathbf{b}$, it is likely that $\mathbf{b} \notin C(A)$

Definition (Residual)

A residual of a linear system $A\mathbf{x} = \mathbf{b}$ is $r(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$

LS problem: find an $\mathbf{x}$ that makes the residual as small as possible, if not 0

Least Squares: Column Interpretation

$$\|A\mathbf{x} - \mathbf{b}\| = \|x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n - \mathbf{b}\|$$

Remark: The least squares problem is to find a linear combination of columns of A that is closest to $\mathbf{b}$

Remark: If $\mathbf{y}$ is a solution of least squares problem, then:

the **vector $A\mathbf{y}$ is closest to $\mathbf{b}$** among all linear combinations of columns of A.

Part III Solving Least Squares

Toy Problem

Find $x \in \mathbb{R}$ s.t. $(x-1)^2$ is minimized.

Namely, find $y \in \mathbb{R}$ s.t.

$(y-1)^2 \le (x-1)^2$ for any x.

Answer: $y = 1$.

Verify:

Example of e_1, e_2

$$\min \|A\mathbf{x} - \mathbf{b}\| = \|x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n - \mathbf{b}\|$$

Let's start from a simple example.

$\mathbf{a}_1 = \mathbf{e}_1, \mathbf{a}_2 = \mathbf{e}_2.$

Exercise: How to solve it? Try both geometry and algebra.

Find x_1, x_2, s.t. $\left\| x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\|^2 = f(x)$ is smallest.

Find y_1, y_2, s.t $f(x_1, x_2) \geq f(y_1, y_2)$ $\forall x_1, x_2 \in \mathbb{R}$.

Hint: Algebra or Geometry.

Solution: Algebra Method

__________ (easy/hard) to extend

Solution: Geometry Method

Find $x \in \mathbb{R}^2$ s.t. $\|x_1 e_1 + x_2 e_2 - b\|$ is the smallest, i.e.

Find $y \in \mathbb{R}^2$ s.t. $\|y_1 e_1 + y_2 e_2 - b\| \leq \|x_1 e_1 + x_2 e_2 - b\|, \forall x . \quad (*)$

Geometry: Draw ___, ____, _____; Translate (*) to geometry.

Claim: Suppose p satisfies _________________________, then

$$\|b - p\| \leq \|b - u\|, \quad \forall u \in \text{span}(e_1, e_2) .$$

Proof (by geometry):

Example of Two Points

$$\min \|A\mathbf{x} - \mathbf{b}\| = \|x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n - \mathbf{b}\|$$

Let's make it slightly more complicated.

What if $n = 2$ and $\mathbf{a}_1, \mathbf{a}_2$ are two arbitrary vectors?

Exercise: How to solve it? Try geometry.

Claim: Suppose p satisfies ________________________, then

$$\|b - p\| \leq \|b - u\|, \quad \forall u \in \text{span}(a_1, a_2)\,.$$

General Case: Geometry

$$\min \|A\mathbf{x} - \mathbf{b}\| = \|x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n - \mathbf{b}\|$$

Let's consider the general case.

Exercise: How to solve it? Try geometry.

Geometry problem:

Find a point in ________, such that __________ to $\mathbf{b}$ is the smallest.

Claim: Suppose p satisfies ______________________, then

$$\|b - p\| \leq \|b - u\|, \quad \forall u \in \text{span}(a_1, a_2).$$

Projection: Lemma and Definition

Lemma 16.1 (Orthogonal Projection)

Suppose S is a subspace of $\mathbb{R}^m$. Suppose $\mathbf{p} \in S$, then

(1) $\|\mathbf{b}-\mathbf{p}\| \geq \|\mathbf{b}-\mathbf{z}\|, \forall \mathbf{z} \in S$

$\Longleftrightarrow$

(2) $\mathbf{b}-\mathbf{p} \perp S$.

Definition 16.1

Suppose S is a subspace of $\mathbb{R}^m$.

Suppose $\mathbf{p} \in S$ and $\mathbf{b}-\mathbf{p} \perp S$, then we say $\mathbf{p}$ is the projection of $\mathbf{b}$ onto S.

Proof (Skipped)

We skip the proof of Lemma 16.1.

We reserve a blank page if you want space to prove it.

Next: Make It Solvable

Lemma 16. says, just need to find $p \in S$ such that

$$\mathbf{b} - \mathbf{p} \perp S.$$

In our problem, $S =$ ____________________

Recall: find $\mathbf{y}$ s.t. $\|A\mathbf{y} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$ for any $\mathbf{x}$

$\Leftrightarrow$ Find $\mathbf{p} \in \mathrm{C}(A)$ s.t. $\|\mathbf{p} - \mathbf{b}\| \leq \|\mathbf{u} - \mathbf{b}\|$ for any $\mathbf{u} \in \mathrm{C}(A)$

Just need to find $\mathbf{y} \in \mathbb{R}^{n\times 1}$ such that

$$\mathbf{b} - \mathbf{y} \perp \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^{n\times 1}\}$$

$\Leftrightarrow$

$\Leftrightarrow$

Small Result Used in Last Page

Math problem. What $z \in \mathbb{R}^n$ satisfies $z \perp x, \quad \forall x \in \mathbb{R}^n$?

Lemma If $z \in \mathbb{R}^n$ satisfies $z \perp x, \quad \forall x \in \mathbb{R}^n$, then ________.

Proof:

Another View (Orthogonal Complement):

Lemma 2

Lemma 16.2

Suppose $S = C(A)$ is the column space of matrix A. Then

(1) $\mathbf{b} - A\mathbf{y} \perp S$

$\Longleftrightarrow$

(2) $A^\top A\mathbf{y} = A^\top \mathbf{b}$

Recall: $C(A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^{n\times 1}\}$

Solving Least Squares: Characterization

The least squares solution can be found by solving a linear system!

Theorem 16.1 (LS solution and normal equation)

Consider a least squares problem. The following statements are equivalent:

1. $\mathbf{y}$ minimizes $||A\mathbf{x} - \mathbf{b}||$

2. $A^\top A\mathbf{y} = A^\top \mathbf{b}$

Proof: Directly combine two lemmas.

Solving Least Squares: Characterization

The least squares solution can be found by solving a linear system!

Theorem 16.1 (LS solution and normal equation)

The following statements are equivalent:

1. $\mathbf{y}$ minimizes $||A\mathbf{x} - \mathbf{b}||$
2. $A^\top A\mathbf{y} = A^\top \mathbf{b}$

This is an $n \times n$ linear system!

Remark: The linear system $A^\top A\mathbf{y} = A^\top \mathbf{b}$ is called the **normal equation**

Solving Least Squares: Solution

Proposition (Existence)

The linear system $A^\top A\mathbf{y} = A^\top\mathbf{b}$ has at least one solution.

Hint: Use the following fact:
"Bx=z" is solvable iff $z \in \mathrm{C}(B)$.

Proof:

Solving Least Squares: Expression for Full-Rank A

Corollary (Characterization)

Suppose A has **linearly independent columns** (i.e., has full column rank).

Then solution of the least square problem $\min_{\mathbf{x}\in\mathbb{R}^{n\times 1}} \|A\mathbf{x}-\mathbf{b}\|$ is

$$\mathbf{y} = (A^\top A)^{-1}A^\top \mathbf{b}$$

Remark: A has linearly independent columns $\Longleftrightarrow \mathrm{rank}(A) = n$

Proof: Step 1: show the following lemma.

Lemma: If A has full column rank, then $A^\top A$ is invertible.

Step 2:

Reading: An Alternative Proof via Calculus

(1) ==> (2)

Derivation via calculus

- define

$$f(x) = \|Ax - b\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)^2$$

- solution $\hat{x}$ satisfies

$$\frac{\partial f}{\partial x_k}(\hat{x}) = \nabla f(\hat{x})_k = 0, \quad k = 1, \ldots, n$$

- taking partial derivatives we get $\nabla f(x)_k = \left(2A^T(Ax - b)\right)_k$
- in matrix-vector notation: $\nabla f(\hat{x}) = 2A^T(A\hat{x} - b) = 0$
- so $\hat{x}$ satisfies *normal equations* $(A^TA)\hat{x} = A^Tb$
- and therefore $\hat{x} = (A^TA)^{-1}A^Tb$

(Source: Stephen Boyd's book)

Reading: An Alternative Proof via Calculus

(2) ==> (1)

Direct verification

- let $\hat{x} = (A^T A)^{-1} A^T b$, so $A^T(A\hat{x} - b) = 0$
- for any n-vector x we have

$$\begin{aligned} \|Ax - b\|^2 &= \|(Ax - A\hat{x}) + (A\hat{x} - b)\|^2 \\ &= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 + 2(A(x - \hat{x}))^T(A\hat{x} - b) \\ &= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 + 2(x - \hat{x})^T A^T(A\hat{x} - b) \\ &= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \end{aligned}$$

- so for any x, $\|Ax - b\|^2 \geq \|A\hat{x} - b\|^2$
- if equality holds, $A(x - \hat{x}) = 0$, which implies $x = \hat{x}$ since columns of A are linearly independent

(Source: Stephen Boyd's book)

Summary Today (write Your Own)

One sentence summary:

Detailed summary:

Summary Today (of Instructor)

One sentence summary:

We have studied least squares and analyzed it solution

Detailed summary:

1. What can we do when a linear system has no solutions?

   Solve least squares!

2. Applications: linear regression

   e.g. predict salary

3. Characterization of its solution and uniqueness condition

   Solution satisfies orthogonality condition.

   Solution satisfies the Normal equation.

   Full-column-rank case: Unique expression $\mathbf{y} = (A^\top A)^{-1} A^\top \mathbf{b}$