Linear Algebra

Least Squares

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu
Maryam Ramezani maryam.ramezani@sharif.edu

Table of contents

01

Introduction

02

Solution of Least squares problem

03

Least squares problem with constraints

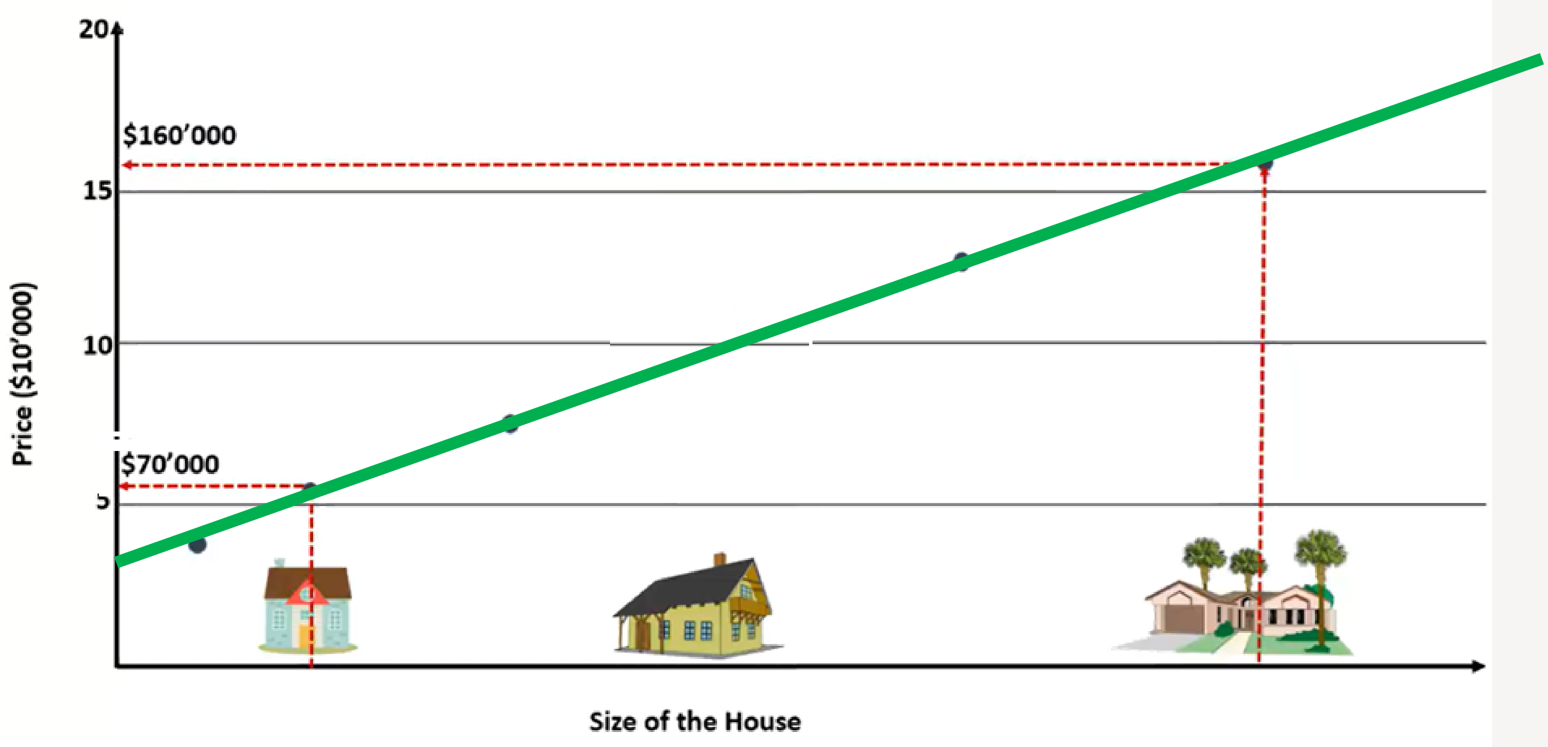
01

Introduction

Introduction

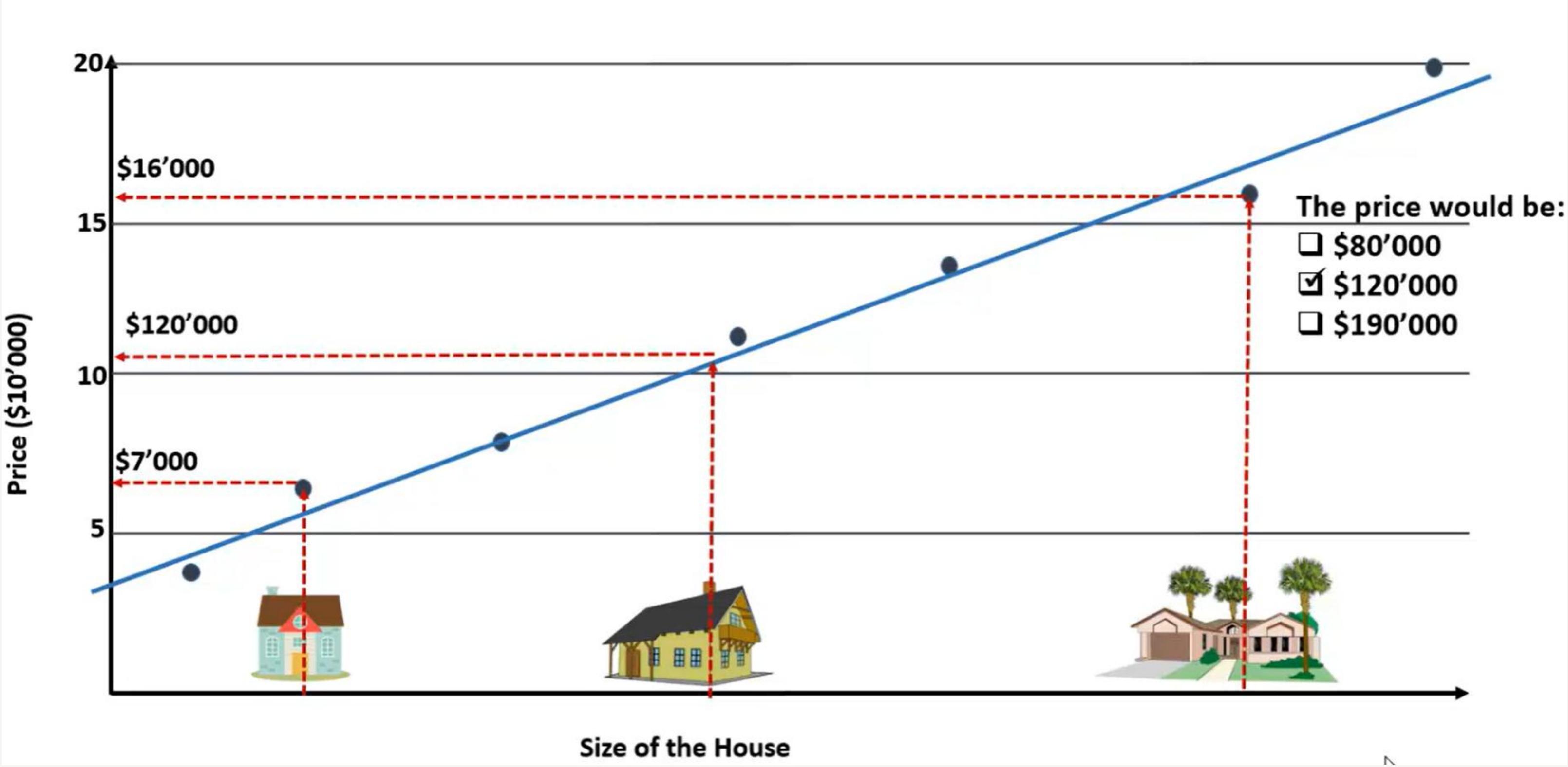
Linear Equation

$Ax = b$ has solution.



Least Squares Error Correction

$Ax = b$ has no solution.

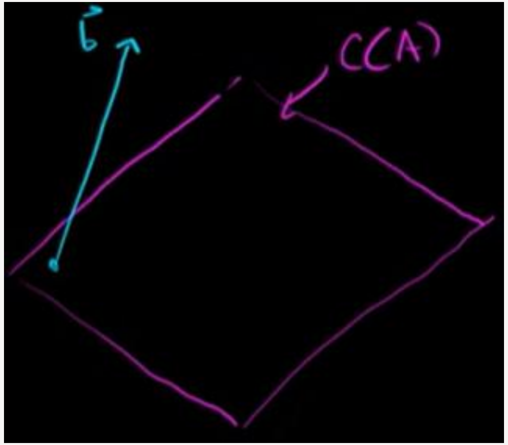


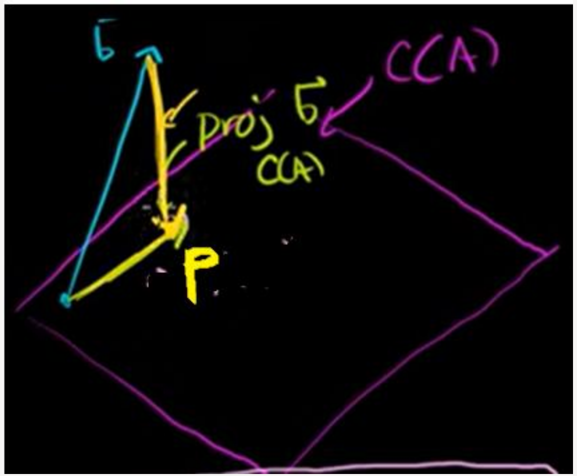
Introduction

- Fill this page with my notes on the board 😁
  - Least square in $\mathbb{R}^2$ and regression!!!
  - Error
  - Outlier

What is the problem?

- A is $m \times n$ matrix
- $Ax = b$ has no solution -> b is not in the $C(A)$ why?





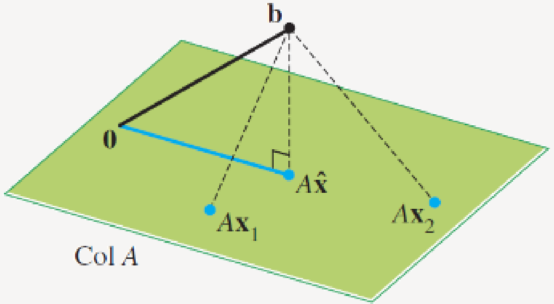
How to solve the problem?

- Bad News: $Ax = b$ has no solution
- Good News: $A\hat{x} = p$ has solution
  - Unique = Least Square
  - Many = SVD

Normal Equation (Method 1)

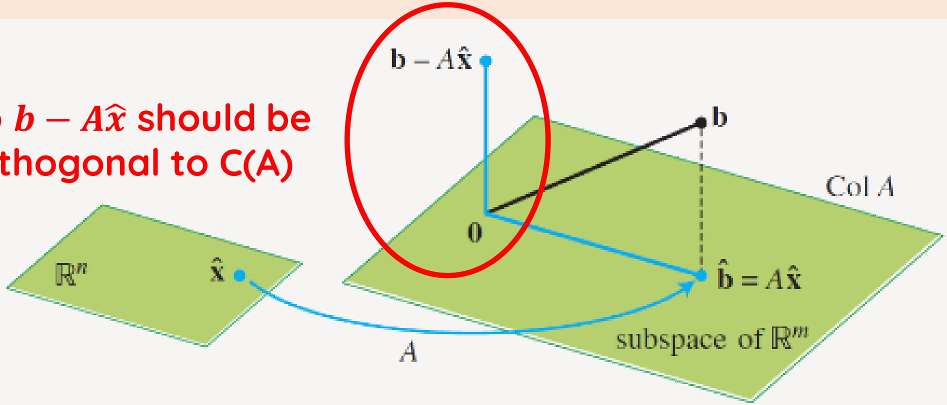
Note

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.



The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

So $b - A\hat{x}$ should be orthogonal to C(A)



The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Method 1

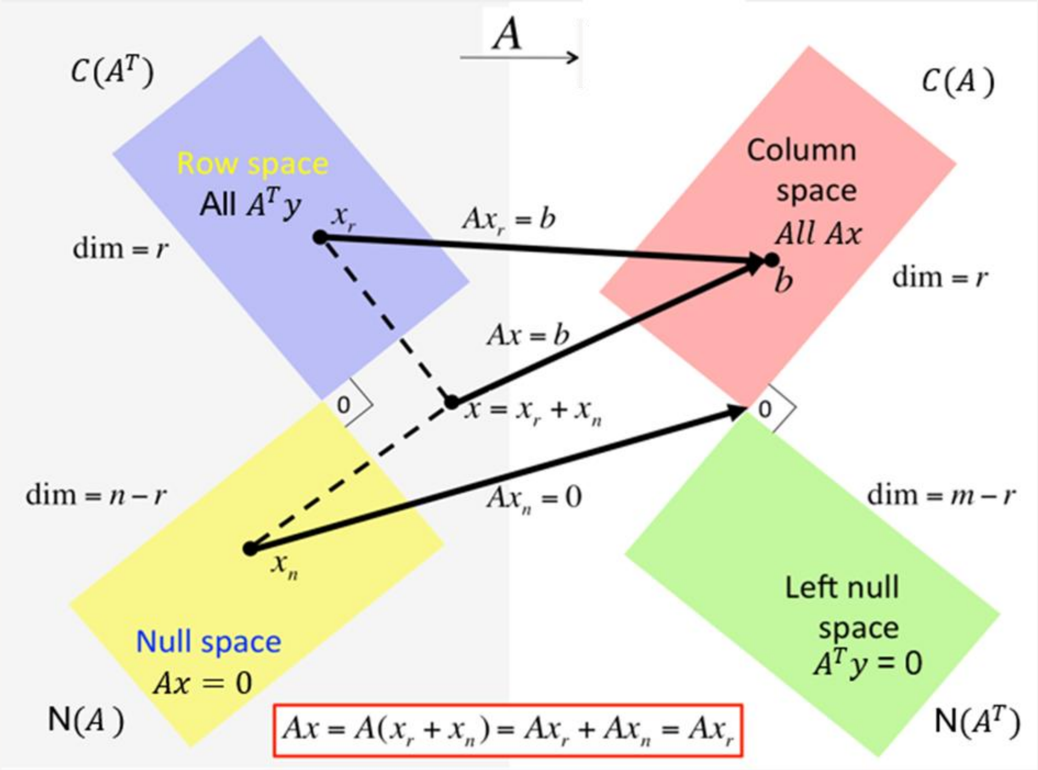
- $C(A^T) \perp N(A)$
- $C(A) \perp N(A^T)$

As we know from previous lectures, the subspace orthogonal to C(A) is $N(A^T)$. So, $(b - A\hat{x}) \in N(A^T)$.

Therefore, $A^T(b - A\hat{x}) = 0$

$$\Rightarrow A^T b = A^T A\, \hat{x}$$

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



Look another way!!

- When b is not in range(A)[C(A)], we can project b on C(A) and then find x where Ax=Pb
- $(A^T A)^{-1} A^T$ is the left inverse of A
- $A(A^T A)^{-1} A^T$ is the projection matrix on C(A)

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

What will happen when A is an invertible matrix?

A is square invertible matrix and solution is $(A)^{-1} b$

Least Squares Problem Unique Solution

Theorem

- A has linearly independent columns, then below vector is the unique solution of the least squares problem

$$\text{minimize } ||Ax - b||^2$$

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

$$= A^{\dagger} b$$

- Proof?

pseudo-inverse of a left-invertible matrix

Solving with Derivation (Method 2)

Example

- Normal equations of the least squares problem $A^T Ax = A^T b$
  - Coefficient matrix $A^T A$ is the ……
  - Equivalent to $\nabla f(x) = 0$ where $f(x) =$
  - All solutions of the least squares problem satisfy the normal equations

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

Let's write in vector and matrix form with derivation

Method 2

R1: $\frac{\partial(x^T B x)}{\partial x} = 2x^T B$ if B is symmetric

R2: $\frac{\partial(x^T a)}{\partial x} = a^T$

R3: $\frac{\partial(\mathrm{a}^T x)}{\partial x} = a^T$

$$\frac{\partial(Ax-b)^T(Ax-b)}{\partial x} = 0$$

$$\Rightarrow \frac{\partial(x^T A^T A x - x^T A^T b - b^T A x + b^T b)}{\partial x} = 0$$

$$\Rightarrow \underbrace{\frac{\partial(x^T A^T A x)}{\partial x}}_{R1} - \underbrace{\frac{\partial(x^T A^T b)}{\partial x}}_{R2} - \underbrace{\frac{\partial(b^T A x)}{\partial x}}_{R3} + 0 = 0$$

$A^T A$ is symmetric

$$\Rightarrow 2x^T A^T A - b^T A - b^T A = 0$$

$$\Rightarrow A^T A x = A^T b$$

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

Solving with QR Factorization (Method 3)

Example

- Rewrite least squares solution using QR factorization $A = QR$

- Complexity: $2mn^2$

Algorithm: Least squares via QR factorization

Input: A : $m \times n$ left-invertible
Input: b : $m \times 1$
output: x_{LS} : $n \times 1$
Find QR factorization $A = QR$
Compute $Q^T b$
Solve $Rx_{LS} = Q^T b$ using back substitution

- Identical to algorithm for solving $Ax = b$ for square invertible A, but when A is tall, gives least squares approximate solution

Another problem?

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

Theorem

❑ If A has linearly independent columns, then $A^T A$ is invertible.

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

$$= A^{\dagger} b$$

pseudo-inverse of a left-invertible matrix

Therefore, when $A^T A$ is invertible, $\hat{x}$ is the unique solution. This often happens when for D number of variables and N number of equations, we have $D \ll N$.

What will happen when $A^T A$ is not an invertible matrix? (when $\mathrm{N} < D$)

When $A^T A$ is not an invertible matrix?

$X^T X$ will not be invertible when $N < D$. To illustrate why we have infinite number of solutions, consider in a two-dimensional problem ($D = 2$) we have only one training sample $\boldsymbol{x}_1 = [1, -1], y_1 = 1$. We can see $\boldsymbol{w} = [a+1, a]$ for any $a \in \mathbb{R}$ will get 0 training error:

$$\boldsymbol{w}^T \boldsymbol{x}_1 = a + 1 - a = 1 = y_1.$$

This is true for any problem with $N < D$—in this case, you can always find a vector in the null space of X (a vector such that $X\boldsymbol{v} = 0$), and then for a solution $\boldsymbol{w}^*$, any vector with $\boldsymbol{w}^* + a\boldsymbol{v}$ with $a \in \mathbb{R}$ will get the same square error with $\boldsymbol{w}^*$. This case ($N < D$) is also called the **under-determined** problem, since you have too many degree of freedom in your problem and don't have enough constraints (data).

Good News!!! (When $A^T A$ is not an invertible matrix)

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

will have infinite number of solutions in this case

In fact, given any real $m \times n$-matrix A, there is always a unique x^+ of minimum norm that minimizes $\|Ax - b\|^2$, even when the columns of A are linearly dependent.

the following approach to find the **minimum-norm solution** $\boldsymbol{w}^+$: Let $\mathcal{W} = \operatorname{argmin}_{\boldsymbol{w}} \|X\boldsymbol{w} - \boldsymbol{y}\|^2$ denote the set of solutions, we aim to find the minimum norm solution that

$$\boldsymbol{w}^+ = \operatorname*{argmin}_{\boldsymbol{w} \in \mathcal{W}} \|\boldsymbol{w}\|_2. \tag{4}$$

SVD

SVD (Method 4)

- When $A^T A$ is not invertible, we can't apply the 3 mentioned methods!
- For $A_{N \times D}$ with $N < D$, by using SVD:

$$Ax = b \Rightarrow V\Sigma U^T x = b \Rightarrow x = U\Sigma^{+}\mathrm{V}^{\mathrm{T}}\mathrm{b}$$

where Σ^{+} is formed by taking the reciprocal of the non-zero singular values. But it is the minimum norm two solution.

How can we find all possible solutions?

If $\mathbf{x}_0 = \mathbf{A}^{+}\mathbf{b}$ is one solution (the minimum-norm one), then **all solutions** are of the form:

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{z} \quad \text{where } \mathbf{z} \in \text{null}(\mathbf{A})$$

You can extract the null space from the SVD:

- The last $n - r$ columns of $\mathbf{V}$ (where $r = \text{rank}(\mathbf{A})$) span the null space of $\mathbf{A}$
- So:

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=r+1}^{n} \alpha_i \mathbf{v}_i$$

where $\mathbf{v}_i$ are the right singular vectors in the null space and $\alpha_i \in \mathbb{R}$

Solving least squares problems

Example

a 3×2 matrix with "almost linearly dependent" columns

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 10^{-5} \\ 0 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 0 \\ 10^{-5} \\ 1 \end{bmatrix},$$

round intermediate results to 8 significant decimal digits

- ❑ Solve using learned methods
  - ❑ Which one is more stable? Why?

03

Least squares problem with constraints

Review: Linear-in-parameters model

Note

- we choose the model $\hat{f}(x)$ from a family models

$$\hat{f}(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \cdots + \theta_p f_p(x)$$

model parameters

scalar valued basis functions (chosen by us)

Solution of weighted least squares

Example

- weighted least squares is equivalent to a standard least squares problem

$$\text{minimize} \quad \left\| \begin{bmatrix} \sqrt{\lambda_1} A_1 \\ \sqrt{\lambda_2} A_2 \\ \vdots \\ \sqrt{\lambda_k} A_k \end{bmatrix} x - \begin{bmatrix} \sqrt{\lambda_1} b_1 \\ \sqrt{\lambda_2} b_2 \\ \vdots \\ \sqrt{\lambda_k} b_k \end{bmatrix} \right\|^2$$

- Solution is unique if the $\textit{stacked matrix}$ has linearly independent columns
- Each matrix A_i may have linearly dependent columns (or be a wide matrix)
- if the stacked matrix has linearly independent columns, the solution is

$$\hat{x} = \left(\lambda_1 A_1^T A_1 + \cdots + \lambda_k A_k^T A_k\right)^{-1} \left(\lambda_1 A_1^T b_1 + \cdots + \lambda_k A_k^T b_k\right)$$

Lagrange multiplier

Example

$$f(x) = \min(x_1 x_2)$$

$$g(x) = 1 - \mathrm{x}_1 - \mathrm{x}_2$$

$$g(x) = 0$$

$$L(x, \lambda) = f(x) + \lambda g(x)$$

$$\nabla_x f(x) = 0$$

Constrained Least Square

Example

- $\begin{cases} \min\limits_{x} \|Ax - b\|^2 & A: m \times n \\ s.t. \quad Cx = d & C: p \times n \end{cases}$

$$L(x, \lambda) = \|Ax - b\|^2 + \lambda^T (Cx - d)$$

$$\begin{cases} \nabla_x L = 2A^T Ax - 2A^T b + C^T \lambda = 0 \\ \nabla_\lambda L = Cx - d = 0 \end{cases} \rightarrow \begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix}$$