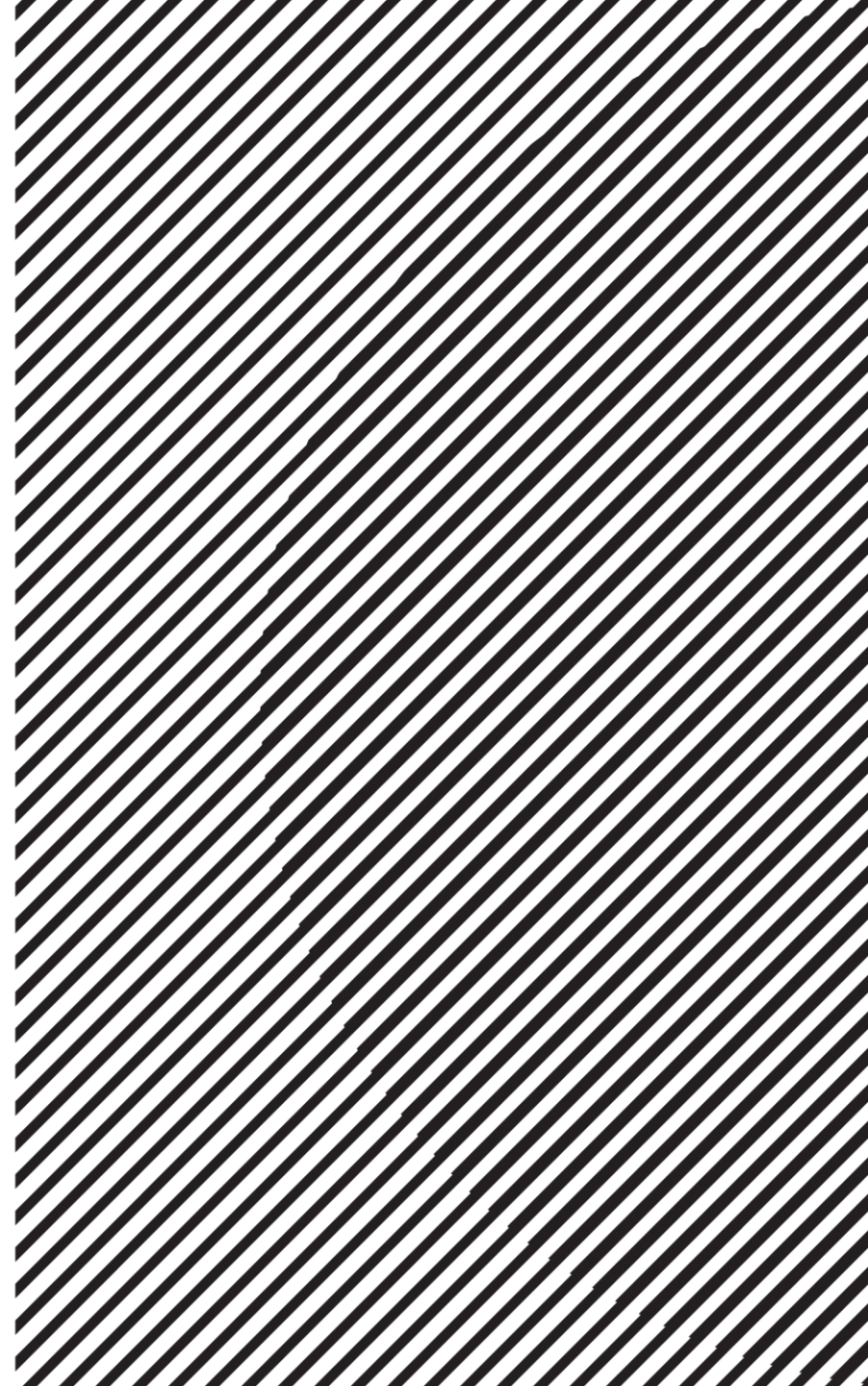


---

# Linear Algebra

주재걸  
고려대학교 컴퓨터학과



# Back to Over-Determined System

- Let's start with the original problem:

Person ID	Weight	Height	Is_smoking	Life-span
1	60kg	5.5ft	Yes (=1)	66
2	65kg	5.0ft	No (=0)	74
3	55kg	6.0ft	Yes (=1)	78

$$\begin{matrix} & \mathbf{A} & & \mathbf{x} & = & \mathbf{b} \\ \rightarrow & \begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \end{bmatrix} & & \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} & = & \begin{bmatrix} 66 \\ 74 \\ 78 \end{bmatrix} \end{matrix}$$

- Using the inverse matrix, the solution is  $\mathbf{x} = \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$

# Back to Over-Determined System

- Let's add one more example:

Person ID	Weight	Height	Is_smoking	Life-span
1	60kg	5.5ft	Yes (=1)	66
2	65kg	5.0ft	No (=0)	74
3	55kg	6.0ft	Yes (=1)	78
4	50kg	5.0ft	Yes (=1)	72

$$\begin{matrix} & \mathbf{A} & & \mathbf{x} & = & \mathbf{b} \\ \rightarrow & \begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 0 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} & = & \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix} \\ & & & & \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}
 \end{matrix}$$

- Now, let's use the previous solution  $\mathbf{x} = \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$

$$\begin{matrix} \mathbf{A} & \mathbf{x} & \neq \mathbf{b} & \text{Errors} \\ \begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix} & \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix} & = \begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix} & \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix} & \left| \begin{array}{c} 0 \\ 0 \\ 0 \\ -12 \end{array} \right.
 \end{matrix}$$

# Back to Over-Determined System

- How about using slightly different solution  $\mathbf{x} = \begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix}$ ?

$A$	$\mathbf{x}$	$\neq \mathbf{b}$	Errors $(\mathbf{b} - A\mathbf{x})$
$\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix}$	$\begin{bmatrix} 71.3 \\ 72.2 \\ 79.9 \\ 64.5 \end{bmatrix} \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}$	$\begin{bmatrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{bmatrix}$

# Which One is Better Solution?

$A$	$x$	$\neq$	$b$	Errors ( $b - Ax$ )
$\begin{bmatrix} 60 \\ 65 \\ 55 \\ 50 \end{bmatrix} \begin{matrix} 5.5 & 1 \\ 5.0 & 0 \\ 6.0 & 1 \\ 5.0 & 1 \end{matrix}$	$\begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix}$	$=$	$\begin{bmatrix} 71.3 \\ 72.2 \\ 79.9 \\ 64.5 \end{bmatrix} \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}$	$\begin{bmatrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{bmatrix}$

$\begin{bmatrix} 60 \\ 65 \\ 55 \\ 50 \end{bmatrix} \begin{matrix} 5.5 & 1 \\ 5.0 & 0 \\ 6.0 & 1 \\ 5.0 & 1 \end{matrix}$	$\begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$	$=$	$\begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix} \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ -12 \end{bmatrix}$
---	---	-----	--	--

# Least Squares: Best Approximation Criterion

- Let's use the squared sum of errors:

					Errors	Sum of squared errors
$A$	$\mathbf{x}$	$\neq \mathbf{b}$		$(\mathbf{b} - A\mathbf{x})$		
$\begin{bmatrix} 60 \\ 65 \\ 55 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 5.5 \\ 5.0 \\ 6.0 \\ 5.0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix}$	$= \begin{bmatrix} 71.3 \\ 69 \\ 79.9 \\ 64.5 \end{bmatrix} \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}$	$\begin{bmatrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{bmatrix}$	$\left( (-5.3)^2 + 1.8^2 + (-1.9)^2 + 7.5^2 \right)^{0.5}$ $= 9.55$ <div style="background-color: #d3d3d3; padding: 5px; display: inline-block;"><i>Better solution</i></div>

$\begin{bmatrix} 60 \\ 65 \\ 55 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 5.5 \\ 5.0 \\ 6.0 \\ 5.0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$	$= \begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix} \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ -12 \end{bmatrix}$	$(0^2 + 0^2 + 0^2 + (-12)^2)^{0.5}$ $= 12$
--	--	--	---	--	--	---



# Least Squares Problem

- Now, the sum of squared errors can be represented as  $\|\mathbf{b} - A\mathbf{x}\|$ .
- **Definition:** Given an overdetermined system  $A\mathbf{x} \simeq \mathbf{b}$  where  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^n$ , and  $m \gg n$ , a least squares solution  $\hat{\mathbf{x}}$  is defined as

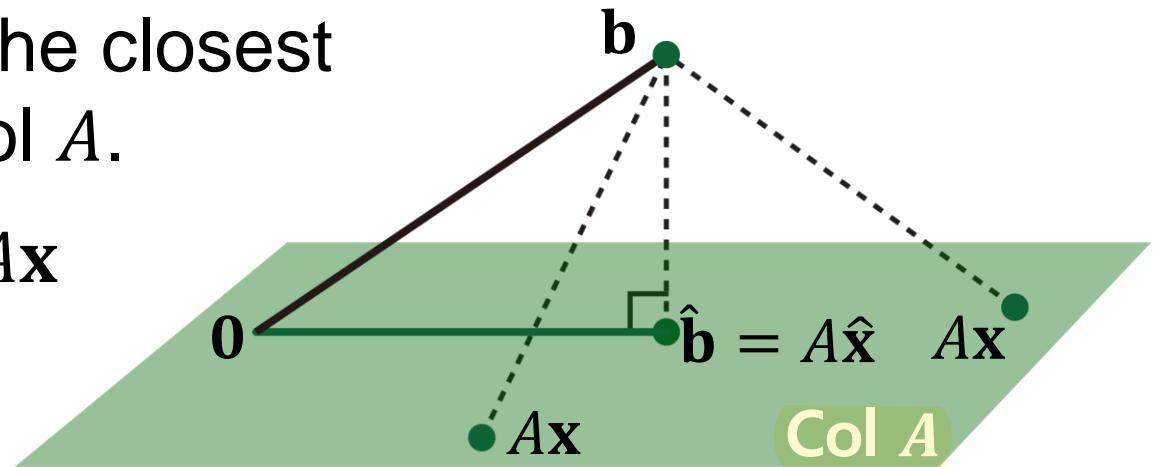
$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|$$

- The most important aspect of the least-squares problem is that no matter what  $\mathbf{x}$  we select, the vector  $A\mathbf{x}$  will necessarily be in the column space  $\text{Col } A$ .
- Thus, we seek for  $\mathbf{x}$  that makes  $A\mathbf{x}$  as the closest point in  $\text{Col } A$  to  $\mathbf{b}$ .

# Geometric Interpretation of Least Squares

- Consider  $\hat{\mathbf{x}}$  such that  $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$  is the closest point to  $\mathbf{b}$  among all points in  $\text{Col } A$ .
- That is,  $\mathbf{b}$  is closer to  $\hat{\mathbf{b}}$  than to  $A\mathbf{x}$  for any other  $\mathbf{x}$ .
- To satisfy this, the vector  $\mathbf{b} - A\hat{\mathbf{x}}$  should be orthogonal to  $\text{Col } A$ .
- This means  $\mathbf{b} - A\hat{\mathbf{x}}$  should be orthogonal to any vector in  $\text{Col } A$ :

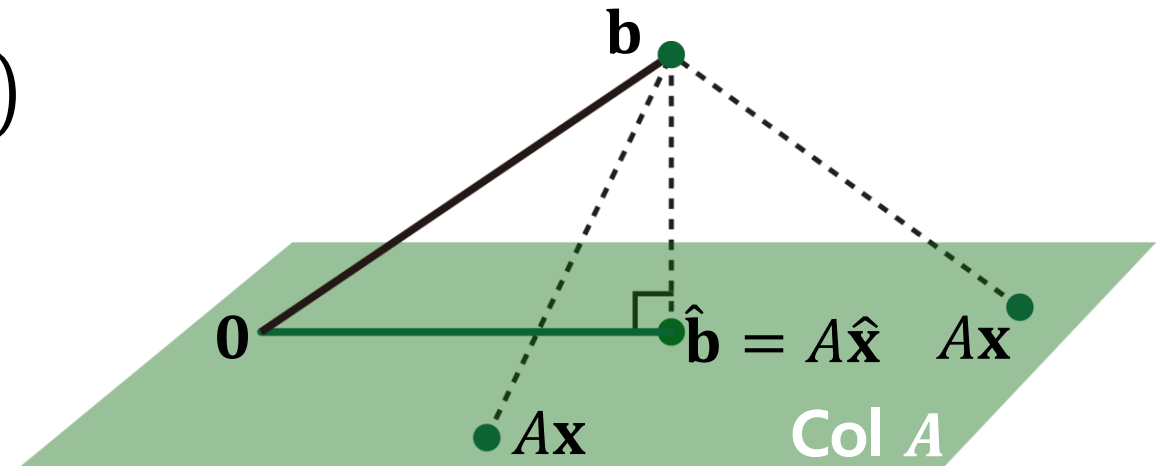
$$\mathbf{b} - A\hat{\mathbf{x}} \perp (x_1\mathbf{a}_1 + x_2\mathbf{a}_2 \cdots + x_p\mathbf{a}_n) \text{ for any vector } \mathbf{x}$$





# Geometric Interpretation of Least Squares

- $\mathbf{b} - A\hat{\mathbf{x}} \perp (x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 \cdots + x_p \mathbf{a}_n)$   
for any vector  $\mathbf{x}$



- Or equivalently,  
 $(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_1$   
 $(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_2$   
 $\vdots$   
 $(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_m$

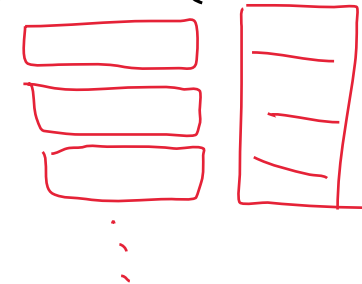
$$\mathbf{a}_1^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

$$\mathbf{a}_2^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

$$\mathbf{a}_m^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

$$A^T \mathbf{b} - A^T A \hat{\mathbf{x}} = \mathbf{0}$$

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





# Normal Equation

- Finally, given a least squares problem,  $A\mathbf{x} \simeq \mathbf{b}$ , we obtain

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

which is called a **normal equation**.

- This can be viewed as a new linear system,  $C\mathbf{x} = \mathbf{d}$ ,  
where a square matrix  $C = A^T A \in \mathbb{R}^{n \times n}$ , and  $\mathbf{d} = A^T \mathbf{b} \in \mathbb{R}^n$ .
- If  $C = A^T A$  is **invertible**, then the solution is computed as
$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

# Another Derivation of Normal Equation

$$\|x\|^2 = X^T X = \begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 9 + 4$$

$$\begin{bmatrix} 3 \\ 2 \end{bmatrix}^2 = 9 + 4 = 13$$

- $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\| = \arg \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|^2$

$$= \arg \min_{\mathbf{x}} (\mathbf{b} - A\mathbf{x})^T (\mathbf{b} - A\mathbf{x}) = \mathbf{b}^T \mathbf{b} - \mathbf{x}^T A^T \mathbf{b} - \mathbf{b}^T A \mathbf{x} + \mathbf{x}^T A^T A \mathbf{x}$$

- Computing derivatives w.r.t.  $\mathbf{x}$ , we obtain

$$-A^T \mathbf{b} - A^T \mathbf{b} + 2A^T A \mathbf{x} = \mathbf{0} \quad \Leftrightarrow \quad A^T A \mathbf{x} = A^T \mathbf{b}$$

- Thus, if  $C = A^T A$  is invertible, then the solution is computed as

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

$$\boxed{\phantom{f(x)}}$$

$$f(x) = A^T x = x A^T$$

$$\frac{df}{dx} = 1$$
$$= \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1 + 2x_2$$

$$f \cdot g = f'g + f \cdot g'$$

$$\boxed{\phantom{f(x)}}$$

# Life-Span Example

Person ID	Weight	Height	Is_smoking	Life-span
1	60kg	5.5ft	Yes (=1)	66
2	65kg	5.0ft	No (=0)	74
3	55kg	6.0ft	Yes (=1)	78
4	50kg	5.0ft	Yes (=1)	72

$$\begin{matrix} & A & \mathbf{x} \approx \mathbf{b} \\ \rightarrow & \begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}
 \end{matrix}$$

- The normal equation  $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$  is

$$\begin{bmatrix} 60 & 65 & 55 & 50 \\ 5.5 & 5.0 & 6.0 & 5.0 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 60 & 65 & 55 & 50 \\ 5.5 & 5.0 & 6.0 & 5.0 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}$$

$$\begin{bmatrix} 13350 & 1235 & 165 \\ 1235 & 116.25 & 16.5 \\ 165 & 16.5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 16600 \\ 1561 \\ 216 \end{bmatrix}$$

# What If $C = A^T A$ is NOT Invertible?

- Given  $A^T A \mathbf{x} = A^T \mathbf{b}$ , what if  $C = A^T A$  is NOT invertible?
- Remember that in this case, the system has either no solution or infinitely many solutions.
- However, the solution always exist for this “normal” equation, and thus infinitely many solutions exist.
- When  $C = A^T A$  is NOT invertible?  
If and only if the columns of  $A$  are linearly dependent. Why? feature
- However,  $C = A^T A$  is usually invertible. Why?

