Linear Algebra

주재걸
고려대학교 컴퓨터학과

Lecture Overview

- Elements in linear algebra
- Linear system
- Linear combination, vector equation,
  Four views of matrix multiplication
- Linear independence, span, and subspace
- Linear transformation
- Least squares
- Eigendecomposition
- Singular value decomposition

Over-determined Linear Systems (#equations ≫ #variables)

방정식 수 > 미지수 수

- Recall a linear system:

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

$$
\begin{aligned}
60x_1+5.5x_2+1\cdot x_3 &= 66\\
65x_1+5.0x_2+0\cdot x_3 &= 74\\
55x_1+6.0x_2+1\cdot x_3 &= 78
\end{aligned}
$$

↓

방정식의 수 : 3개
미지수의 수 : 3개) ⊜

Over-determined Linear Systems (#equations ≫ #variables)

제약조건이 지나치게 多

- Recall a linear system:
- What if we have much more data examples?

미지수의 수 : 3개
방정식의 수 : ∞

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
⋮	⋮	⋮	⋮	⋮

⇨

$$\begin{array}{l} 60x_1+5.5x_2+1\cdot x_3 = 66 \\ 65x_1+5.0x_2+0\cdot x_3 = 74 \\ 55x_1+6.0x_2+1\cdot x_3 = 78 \\ \quad\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \end{array}$$

- Matrix equation: $\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 66 \\ 74 \\ 78 \\ \vdots \end{bmatrix}$

$$A \qquad \mathbf{x} = \mathbf{b}$$

$m \gg n$: more equations than variables

⇨ Usually no solution exists

Vector Equation Perspective

- Vector equation form:

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

$$\mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \mathbf{a}_3 x_3 = \mathbf{b}$$

- Compared to the original space $\mathbb{R}^n$, where $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{b} \in \mathbb{R}^n$, Span $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ will be a thin hyperplane, so it is likely that $\mathbf{b} \notin$ Span $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$

➡ No solution exists.

Motivation for Least Squares

- Even if no solution exists, we want to approximately obtain the solution for an over-determined system.

- Then, how can we define the best approximate solution for our purpose?

근사적으로 best 답을 찾아보자

Inner Product

내적

- Given $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we can consider $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices.
- The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a scalar without brackets.
- The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** or **dot product** of $\mathbf{u}$ and $\mathbf{v}$, and it is written as $\mathbf{u} \cdot \mathbf{v}$.
- For $\mathbf{u} = \begin{bmatrix}3\\2\\1\end{bmatrix}, \mathbf{v} = \begin{bmatrix}1\\3\\5\end{bmatrix}, \mathbf{u} \underset{\downarrow \text{내적의 의미}}{\boxed{\cdot}} \mathbf{v} = \boxed{\mathbf{u}^T}\mathbf{v} = \underline{\begin{bmatrix}3 & 2 & 1\end{bmatrix}}\begin{bmatrix}1\\3\\5\end{bmatrix} = [14]$

$(1 \times 3)(3 \times 1) = 1 \times 1$

Properties of Inner Product

- **Theorem:** Let $\mathbf{u}, \mathbf{v},$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
  - a) $\mathbf{u} \cdot \overset{= u^T v}{\mathbf{v}} = \mathbf{v} \cdot \mathbf{u}$ → 교환법칙
  - b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$ → 분배법칙
  - c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$
  - d) $\mathbf{u} \cdot \mathbf{u} \geq \mathbf{0}$, and $\mathbf{u} \cdot \mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u} = \mathbf{0}$

- Properties (b) and (c) can be combined to produce the following useful rule:

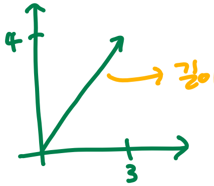
$$(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{w} = c_1(\mathbf{u}_1 \cdot \mathbf{w}) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{w})$$

선형결합 후 w 내적 $\textcircled{=}$ w를 먼저 내적 후 선형결합

Vector Norm

→ 벡터의 길이

ex)



→ 길이 : $\sqrt{4^2+3^2}$
$= 5$

- For $\mathbf{v} \in \mathbb{R}^n$, with entries $v_1, \ldots, v_n$, the square root of $\mathbf{v} \cdot \mathbf{v}$ is defined because $\mathbf{v} \cdot \mathbf{v}$ is nonnegative.

- **Definition**: The length (or norm) of $\mathbf{v}$ is the non-negative scalar $\|\mathbf{v}\|$ defined as the square root of $\mathbf{v} \cdot \mathbf{v}$:

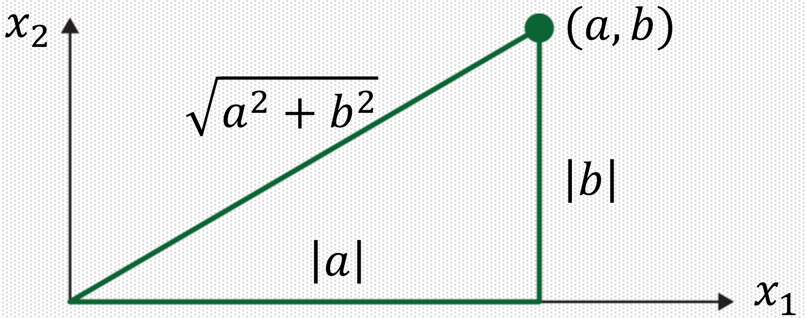
$$\boxed{\|\mathbf{v}\|} = \boxed{\sqrt{\mathbf{v} \cdot \mathbf{v}}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \text{ and } \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

→ 벡터의 길이는 자기 자신을 내적하고 루트 씌운 값

Geometric Meaning of Vector Norm

if) $\left\|\begin{bmatrix}a\\b\end{bmatrix}\right\| = c$, $\left\|\begin{bmatrix}2a\\2b\end{bmatrix}\right\| = 2c$

- Suppose $\mathbf{v} \in \mathbb{R}^2$, say, $\mathbf{v} = \begin{bmatrix}a\\b\end{bmatrix}$.
- $\|\mathbf{v}\|$ is the length of the line segment from the origin to $\mathbf{v}$.
- This follows from Pythagorean Theorem applied to a triangle such as the one shown in the following figure:



Interpretation of ||v|| as length

- For any scalar c, the length $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$ That is,

$$\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$

↳ 음수여도 적용될 수 있도록

Unit Vector

→벡터의 방향은 바꾸지 않고 길이만 1인 벡터

- A vector whose length is 1 is called a unit vector.

(단위 벡터)

- Normalizing a vector: Given a nonzero vector $\mathbf{v}$, if we divide it by its length, we obtain a unit vector $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$.

→ 벡터를 길이로 나누어주면
단위 벡터가 됨

- $\mathbf{u}$ is in the same direction as $\mathbf{v}$, but its length is 1.

Distance between Vectors in $\mathbb{R}^n$

- **Definition:** For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance between $\mathbf{u}$ and $\mathbf{v}$**, written as dist $(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$ That is,

$$\text{dist}\ (\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

벡터 u, 벡터 v의 거리

- **Example**: Compute the distance between the vector

$$\mathbf{u} = \begin{bmatrix} 6 \\ 1 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

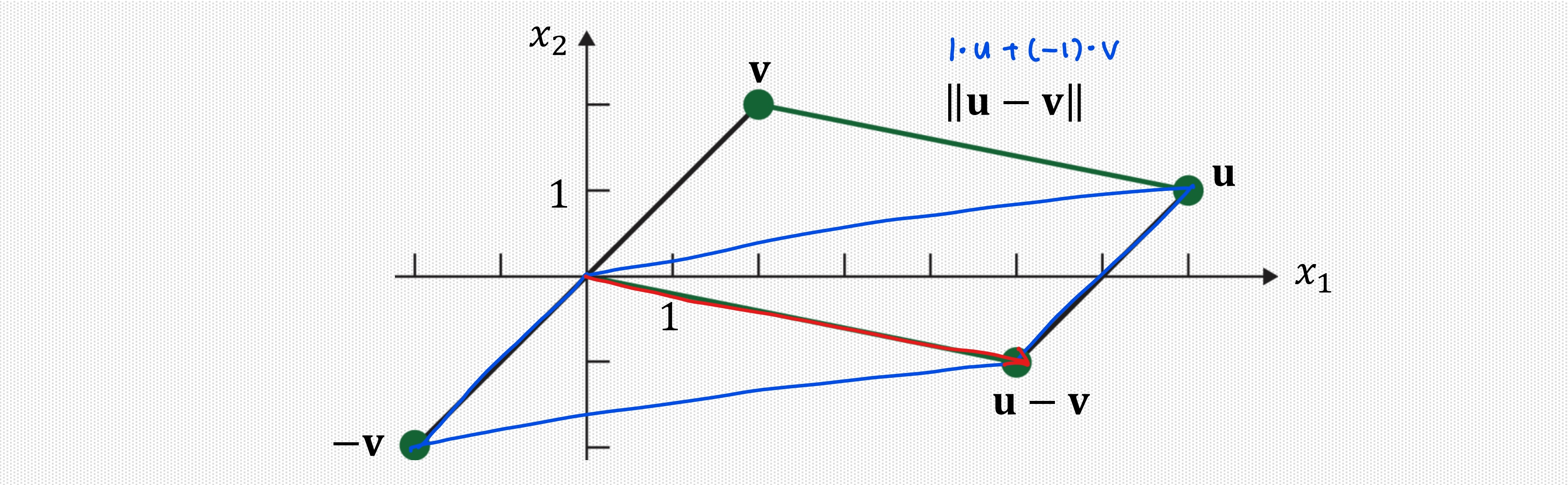
- **Solution**: Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 6 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{3^2 + (-1)^2} = \sqrt{10}$$

Distance between Vectors in $\mathbb{R}^n$

- The distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u} - \mathbf{v}$ to 0.



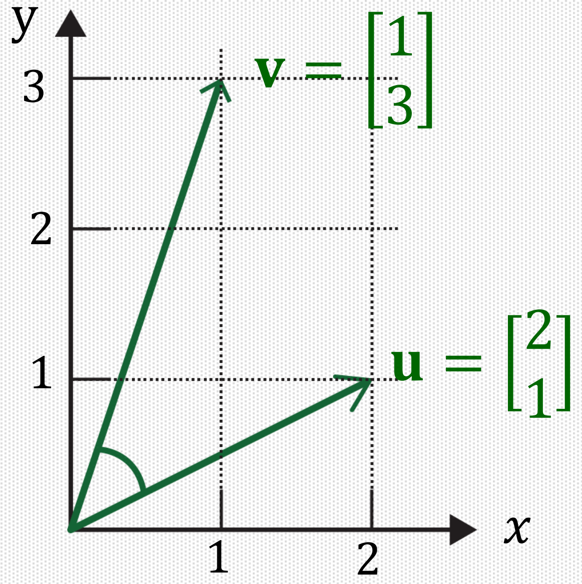
The distance between $\mathbf{u}$ and $\mathbf{v}$ is the length of $\mathbf{u} - \mathbf{v}$

Inner Product and Angle Between Vectors

- Inner product between $\mathbf{u}$ and $\mathbf{v}$ can be rewritten using their norms and angle:

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

- **Example**:



$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix}2\\1\end{bmatrix} \cdot \begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}2 & 1\end{bmatrix} \begin{bmatrix}1\\3\end{bmatrix} = 5$$

$$\boxed{\|\mathbf{u}\|} = \sqrt{2^2 + 1^2} = \underline{\sqrt{5}} \quad \boxed{\|\mathbf{v}\|} = \sqrt{1^2 + 3^2} = \underline{\sqrt{10}}$$

$$\boxed{\mathbf{u} \cdot \mathbf{v}} = \underline{5} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta = \underline{\sqrt{5} \cdot \sqrt{10} \cos\theta}$$

$$\Rightarrow \cos\theta = \frac{5}{\sqrt{50}} = \frac{1}{\sqrt{2}}$$

$$\Rightarrow \theta = 45^\circ$$

Orthogonal Vectors

수직

- **Definition:** $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v}=0$

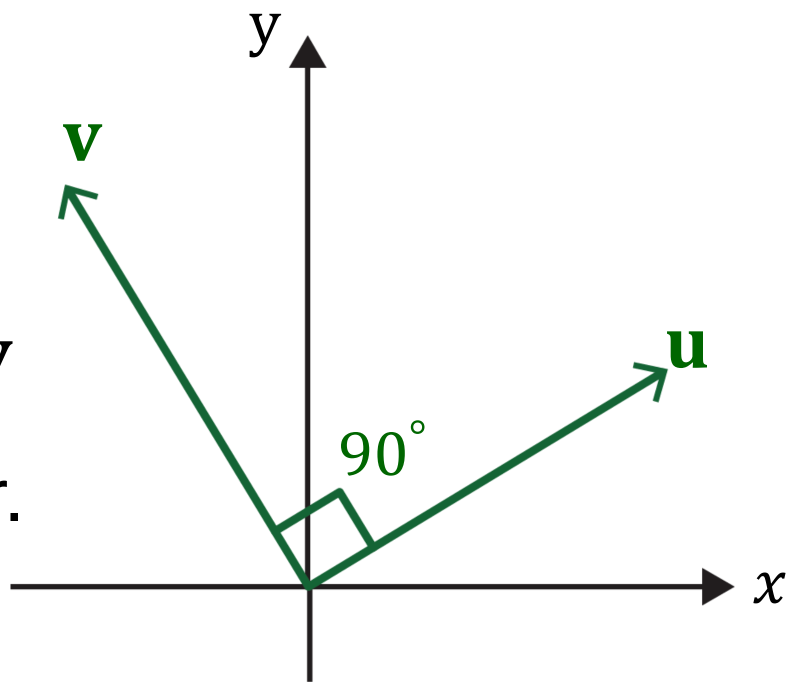
  That is,

  $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos \overset{\to 90^\circ}{\theta} = 0.$

  ➡ $\cos\theta = 0$ for nonzero vectors $\mathbf{u}$ and $\mathbf{v}$

  ➡ $\theta = 90^\circ$ ($\mathbf{u} \perp \mathbf{v}$).

  ➡ $\mathbf{u}$ and $\mathbf{v}$ are perpendicular each other.



ex) $\begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -\frac{8}{3} \end{bmatrix} = 0$ 이므로 $\theta = 90^\circ$

Summary So Far

- Linear transformation
  - Properties of linear transformation
  - Standard matrix
  - One-to-one
  - Onto

- Vector norm, distance, and inner product

- Intro to least squares

Back to Over-Determined System

아직 over-determined ⓧ

↑

- 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

➡

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

- 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

→

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

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

Errors

$$\overset{A}{\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix}} \overset{\mathbf{x}}{\begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}} = \underset{\text{틀림}}{\overset{\text{정답}}{\begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix}}} \neq \overset{\neq\ \mathbf{b}}{\begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}} \left| \overset{(\mathbf{b} - A\mathbf{x})}{\begin{matrix} 0 \\ 0 \\ 0 \\ +12 \end{matrix}} \right.$$

→ 수치화 (정량화)

Back to Over-Determined System

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

Errors

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

Which One is Better Solution?

Errors

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

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

Least Squares: Best Approximation Criterion

→ 작을수록 better solution

Error의 정량화

- Let’s use the squared sum of errors:

Errors **Sum of squared errors**

$$\begin{array}{ccccc|c|c} A & \mathbf{x} & & \neq & \mathbf{b} & (\mathbf{b} - A\mathbf{x}) & \| \mathbf{b} - A\mathbf{x} \| \\ \begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix} & \begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix} & = \begin{bmatrix} 71.3 \\ 69 \\ 79.9 \\ 64.5 \end{bmatrix} & \neq & \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix} & \begin{matrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{matrix} & \begin{matrix} \left((-5.3)^2 + 1.8^2 + (-1.9)^2 + 7.5^2\right)^{0.5} \\ = 9.55 \end{matrix} \end{array}$$

Better solution

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

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

$$\boxed{\hat{\mathbf{x}} = \underline{\arg}\,\overset{\text{최솟값}}{\underset{\mathbf{x}\text{를 뽑고 싶을 때}}{\min}} \|\mathbf{b} - A\mathbf{x}\|}$$

- 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 Col A.
- Thus, we seek for $\mathbf{x}$ that makes $A\mathbf{x}$ as the closest point in Col A to $\mathbf{b}$.