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
- <u>Least squares</u>
- Eigendecomposition
- Singular value decomposition

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

방정식의 수가 미지수의 수보다 더 많을 때

- 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

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

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

determine이 지나치다 → solution이 없음
제약 조건

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

span이 거의 불가

데이터↑

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
⋮	⋮	⋮	⋮	⋮

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

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

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

$m \gg n$: 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

해가 없으면 → 근사적으로라도 해를 구해보자.

- 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?

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 $\mathbf{u} \cdot \mathbf{v}$.

element 끼리 곱셈 $3+6+5=14$

- For $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = [14]$

내적

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

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) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ $\quad$ $u^T v$ 내적

  b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

  c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

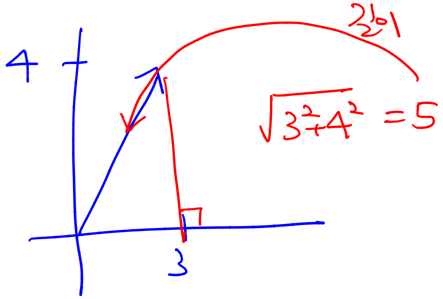
  d) $\mathbf{u} \cdot \mathbf{u} \geq \mathbf{0}$, and $\mathbf{u} \cdot \mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u} = \mathbf{0}$ $\quad$ $\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \underset{0}{x^2} + \underset{0}{y^2} = 0$

- 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 $\quad$ 나중 내적 $\quad$ 각각 내적 → 나중 선형결합

Vector Norm → 벡터의 길이



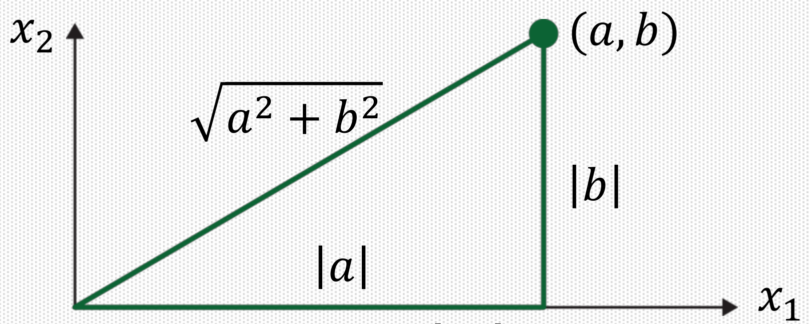
- For $\mathbf{v} \in \mathbb{R}^n$, with entries $v_1, \dots, v_n$, the square root of $\mathbf{v} \cdot \mathbf{v}$ is defined because $\mathbf{v} \cdot \mathbf{v}$ is nonnegative.

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

$$\|\mathbf{v}\| = \underbrace{\sqrt{\mathbf{v} \cdot \mathbf{v}}}_{\sqrt{\text{자기 자신 내적}}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \text{ and } \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

Geometric Meaning of Vector Norm

- 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 $||\mathrm{v}||$ as length

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

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

Unit Vector

$\frac{1}{5}\begin{bmatrix}3\\4\end{bmatrix}$

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

방향은 바꾸지 않고 길이만 1로 바꿈

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

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

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

벡터 간의 거리

$\begin{bmatrix}3\\4\end{bmatrix} \leftrightarrow \begin{bmatrix}1\\5\end{bmatrix}$ $\begin{bmatrix}2\\-1\end{bmatrix} \to \sqrt{2^2+(-1)^2}$

- **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,

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

차이 벡터의 길이

- **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

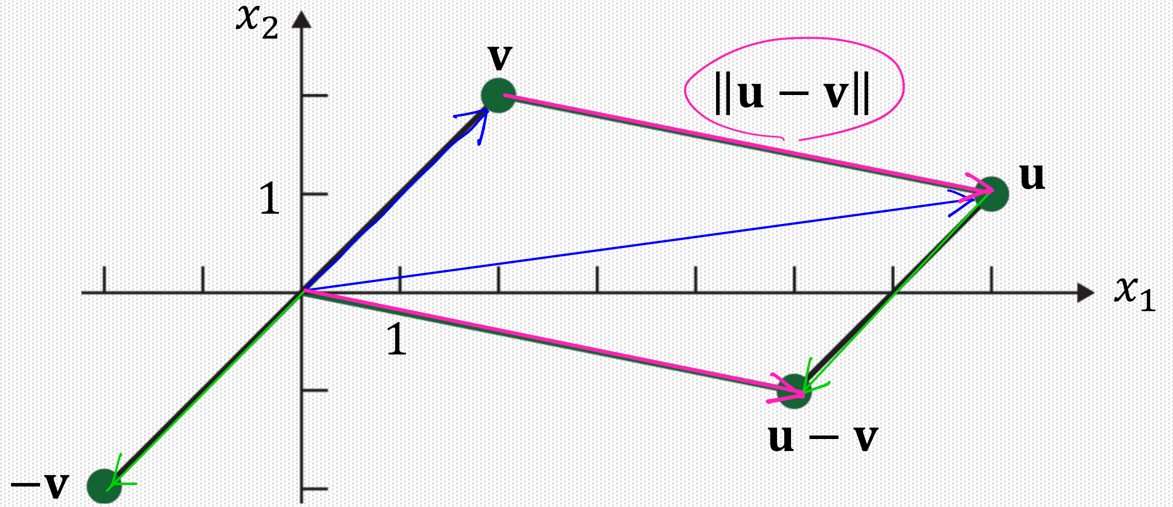
$$\mathbf{u} - \mathbf{v} = \begin{bmatrix}6\\1\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix} = \begin{bmatrix}3\\-1\end{bmatrix}$$

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

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.

$1 \cdot \vec{u} + (-1) \cdot \vec{v}$



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

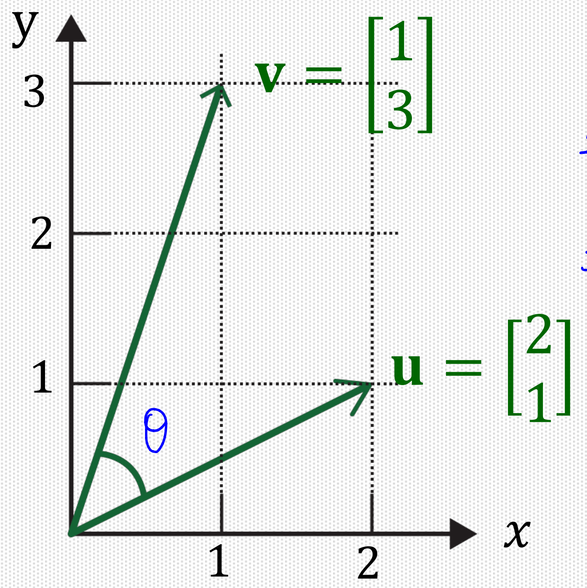
Inner Product and Angle Between Vectors

내적 & 두 벡터 사이의 각도

- 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**:



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

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

3) $\mathbf{u} \cdot \mathbf{v} = 5 = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta = \sqrt{5} \cdot \sqrt{10} \cos \theta$

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

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

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,

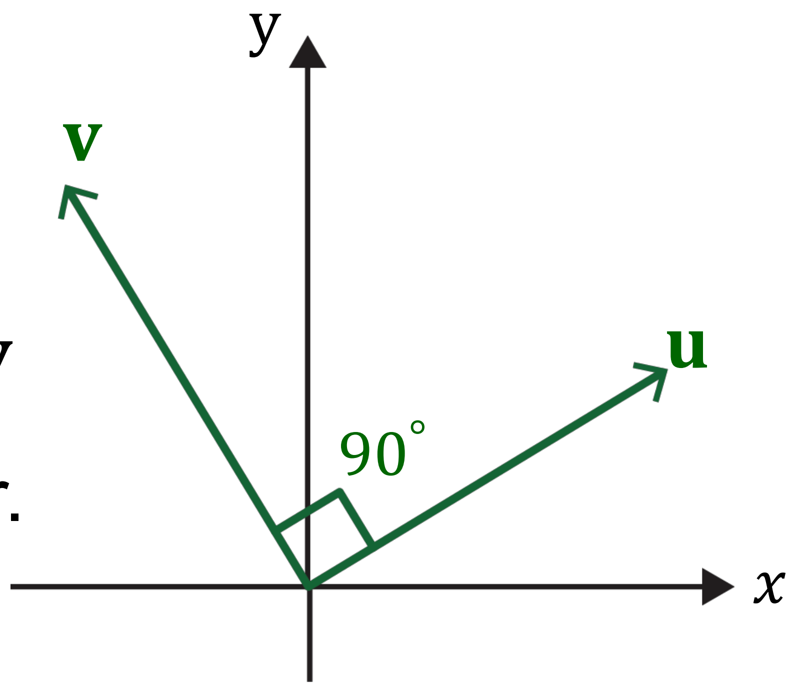
  $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos \theta = 0.$ 길이와 상관없이

  90°

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

  ⇒ $\theta = 90°$ $(\mathbf{u} \perp \mathbf{v})$.

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



$\begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -\frac{8}{3} \end{bmatrix} = 0$ 이면 $\theta = 90°$

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

- 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

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

- 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

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

- Now, let's use the previous solution $\mathbf{x} = \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$ → 3개만 주어졌을 때의 solution.

$$\underbrace{\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix} \neq \underbrace{\begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}}_{\mathbf{b}}$$

예측 ≠ 실제

Errors

$$(\mathbf{b} - A\mathbf{x}) = \begin{matrix} 0 \\ 0 \\ 0 \\ +12 \end{matrix}$$

perfect

+12 → 얼마만큼 틀렸는지

Back to Over-Determined System

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

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

Which One is Better Solution?

$$
\begin{array}{cccccc}
A & \mathbf{x} & & \neq \mathbf{b} & & \text{Errors } (\mathbf{b} - A\mathbf{x})
\end{array}
$$

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

둘 중 어느게 더 나을까?

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

Least Squares: Best Approximation Criterion

- Let's use the squared sum of errors:

SSE

	Errors	Sum of squared errors
$$\begin{array}{c}A\\ \begin{bmatrix}60 & 5.5 & 1\\ 65 & 5.0 & 0\\ 55 & 6.0 & 1\\ 50 & 5.0 & 1\end{bmatrix}\end{array}\begin{array}{c}\mathbf{x}\\ \begin{bmatrix}-0.12\\ 16\\ -9.5\end{bmatrix}\end{array} = \begin{bmatrix}71.3\\ 69\\ 79.9\\ 64.5\end{bmatrix} \neq \begin{array}{c}\mathbf{b}\\ \begin{bmatrix}66\\ 74\\ 78\\ 72\end{bmatrix}\end{array}$$	$(\mathbf{b}-A\mathbf{x})$ -5.3 1.8 -1.9 7.5	$\left((-5.3)^2 + 1.8^2 + (-1.9)^2 + 7.5^2\right)^{0.5}$ $= 9.55$ 더 적은 error *Better solution*
$$\begin{bmatrix}60 & 5.5 & 1\\ 65 & 5.0 & 0\\ 55 & 6.0 & 1\\ 50 & 5.0 & 1\end{bmatrix}\begin{bmatrix}-0.4\\ 20\\ -20\end{bmatrix} = \begin{bmatrix}66\\ 74\\ 78\\ 60\end{bmatrix} \neq \begin{bmatrix}66\\ 74\\ 78\\ 72\end{bmatrix}$$	0 0 0 -12	$(0^2 + 0^2 + 0^2 + (-12)^2)^{0.5}$ $= 12$

error를 최소화하는 최적의 solution?

Least Squares Problem

i) 목적 함수 정의

- 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

실제 값 − 예측값 = error

error를 최소화하게 만드는 x 값

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

norm

이 error를 최소화하는 가중치를 찾자☆

x를 조절하면서 알려진 feature 값 가중치

- 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 $\mathbf{b}$.