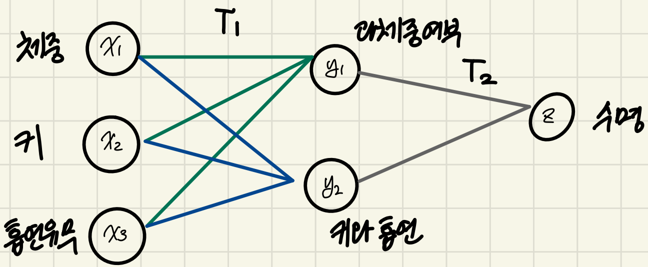
* fully connected layers.



$$y_1 = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3$$ fit

if $y_1 = 1 \cdot x_1 - 1 \cdot y_2$,

키 180, 몸무게 80 $\quad y_1 = 100$

키 170 몸무게 70 $\quad y_1 =$ 〃

⋮

not one-to-one.

* Least square problem

: no solution of equation.

⇒ approximately obtain solution.

• dot-product

- $\underline{u} \cdot \underline{v} = \underline{u}^T \underline{v} = \underline{v} \cdot \underline{u}$

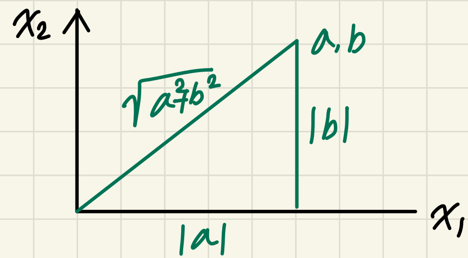
- $(\underline{u} + \underline{v}) \cdot \underline{w} = \underline{u} \cdot \underline{w} + \underline{v} \cdot \underline{w}$

- $(c_1 \underline{u_1} + c_2 \underline{u_2} + \cdots c_p \underline{u_p}) \cdot \underline{w}$

$= c_1 (\underline{u_1} \cdot \underline{w}) + c_2 (\underline{u_2} \cdot \underline{w}) + \cdots c_p (\underline{u_p} \cdot \underline{w})$.

* Vector norm. = vec의 길이.

• $\|V\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \cdots v_n^2}$



* Unit vector

: length 1 ⟶ normalize

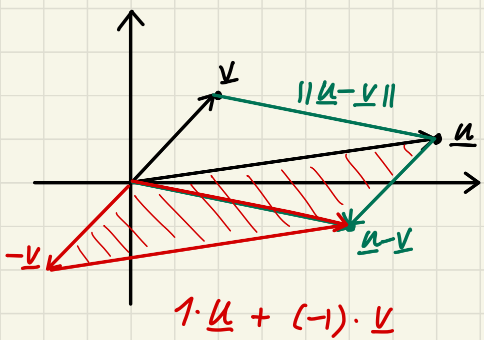
: 방향을 유지하며 길이를 1로.

Vector's norm → $\frac{\text{vector}}{\text{norm}}$. $\frac{1}{\|V\|} \underline{V}$

* distance betw vectors.

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

$\begin{bmatrix} 3 \\ 4 \end{bmatrix} \longleftrightarrow \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. $\sqrt{2^2 + (-1)^2}$



* angle between vectors.

- $\underline{u} \cdot \underline{v} = \|u\| \|v\| \cos\theta$

* orthogonal vectors.
(수직).

orthogonal $\underline{u}$, $\underline{v}$, $\underline{u} \cdot \underline{v} = 0$.

* Over determined system.

: which one is a better solution?

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

<case1>

$$\underline{x} = \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix} \rightarrow \underline{b} = \begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix}$$

Calculate errors

$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 12 \end{bmatrix}}_{\underline{b}-A\underline{x}} \Rightarrow \sqrt{(0^2+0^2+0^2+(12)^2)} = 12 = \|\underline{b}-A\underline{x}\|.$$

<case2>

$$\underline{x} = \begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix} \rightarrow \underline{b} = \begin{bmatrix} 71.3 \\ 69 \\ 79.9 \\ 64.5 \end{bmatrix}$$

errors

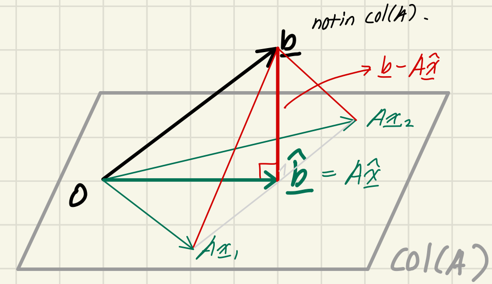
$$\underbrace{\begin{bmatrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{bmatrix}}_{b-A\underline{x}} \Rightarrow \sqrt{(-5.3)^2+(1.8)^2+\cdots} = 9.55.$$

Better solution.

* Least square problem.

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

→ 최소의 error를 만들어낸 $\underline{x}$ 가 누구냐.



$$A\underline{x} = (x_1\ \underline{a_1}) + (x_2\ \underline{a_2}) + (x_3\ \underline{a_3})$$

$$\perp \hat{b} \qquad \perp \hat{b} \qquad \perp \hat{b}$$

$$(b-A\hat{x}) \perp (x_1\underline{a_1} + x_2\underline{a_2} + x_3\underline{a_3})$$

$\Rightarrow$ dot-p $= 0$.

$b-A\hat{x} \cdot a_1 = 0 \longrightarrow \underline{a_1}^T(b-A\hat{x}) = 0. = A^T(b-A\hat{x})$

$b-A\hat{x} \cdot a_2 = 0 \qquad a_2^T \qquad = \underline{0}.$

$\vdots \qquad \vdots$

$\cdot\ a_m = 0.$

→ * normal equation

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

$$\Rightarrow \quad A^TA\underline{\hat{x}} = A^T\underline{b}$$

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

- $f(\underline{x}) = \underline{a}^T\underline{x}$

$$= [3 \quad 2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1 + 2x_2 .$$

$$\frac{df}{d\underline{x}} \Rightarrow 3x_1 + 2x_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\therefore \frac{df}{d\underline{x}}\underline{a}^T\underline{x} = \underline{a} = \frac{df}{d\underline{x}}\underline{x}^T\underline{a}$$

* Another Derivation of normal equation.

$$\underset{\underline{x}}{\operatorname{argmin}} \|\underline{b} - A\underline{x}\| = \underset{\underline{x}}{\operatorname{argmin}} \|\underline{b} - A\underline{x}\|^2 ,$$

$$\|\underline{x}\|^2 = \underline{x}^T\underline{x} \quad \therefore \underset{\underline{x}}{\operatorname{argmin}} (\underline{b} - A\underline{x})^T(\underline{b} - A\underline{x})$$

$$= \underline{b}^T\underline{b} - \underline{x}^T\underline{A^T\underline{b}} - \underbrace{\underline{b}^TA\underline{x}}_{\underline{x}^TA^T\underline{b}} + \underline{x}^T\underline{A^TA\underline{x}}$$

$$\underline{x}^T(A^TA\underline{x})$$

$$\frac{d}{d}(\underline{x}^T(A^TA\underline{x})) = A^TA\underline{x} + A^TA\underline{x}$$

↓ derivatives

$$\frac{\partial \underset{\underline{x}}{\operatorname{argmin}} \|b - A\underline{x}\|}{\partial \underline{x}} = 0 - A^T\underline{b} - A^T\underline{b} + A^TA\underline{x} + A^TA\underline{x}$$

$$\Rightarrow \quad 2A^T\underline{b} = 2A^TA\underline{x}$$

→ normal equation.

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

* Orthogonal projection.

normal equation.

$$\hat{b} = f(b) = A\hat{x} = A(A^TA)^{-1}A^Tb$$

> Orthogonal sets & orthonormal sets.
unit v.

(행렬에 주어진 열vec 모두 서로 수직.

= linearly independent.

ex) $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $L = \text{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \right\}$ $\left(\sqrt{1^2+(-1)^2+2^2}\right)$

$$= \text{span}\left\{ \begin{array}{c} u \\ \begin{bmatrix} 1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} \end{array} \right\}$$

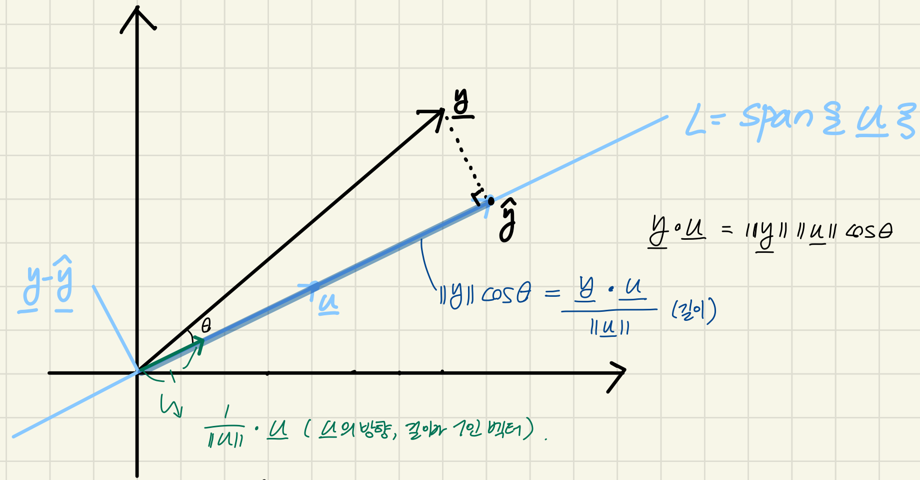
↓

$$y \circ u = \frac{1-2+6}{\sqrt{6}} = \frac{5}{\sqrt{6}}$$

$\hat{y}$의 길이.

$$\frac{5}{\sqrt{6}} \times u = \frac{5}{\sqrt{6}} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix} = \hat{y}$$

$\hat{y}$ 방향.



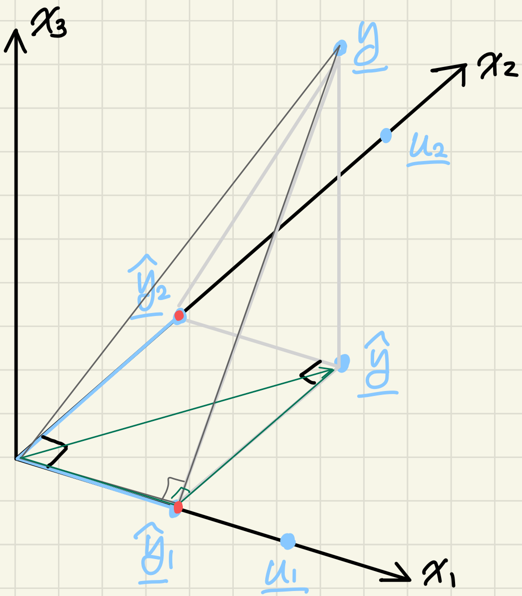
$y \cdot u = \|y\|\,\|u\| \cos\theta$

$$\therefore\ \hat{y} = \frac{y \cdot u}{\|u\|} \times \frac{u}{\|u\|} = \frac{y \cdot u}{u \cdot u} u$$

if u is a unit vector, $(y \cdot u)u$

$\hat{y}$의 길이 ($\because \|u\| = 1$)

$W = \text{Span}\{\underline{u}_1, \underline{u}_2\}$



$$\underline{\hat{y}} = \frac{\underline{y} \cdot \underline{u}_1}{\underline{u}_1 \cdot \underline{u}_1} \underline{u}_1 + \frac{\underline{y} \cdot \underline{u}_2}{\underline{u}_2 \cdot \underline{u}_2} \underline{u}_2$$

$$= \underline{\hat{y}}_1 + \underline{\hat{y}}_2$$

$\downarrow$

$\therefore$ projection is done independently on each orthogonal basis vector.