Linear Algebra

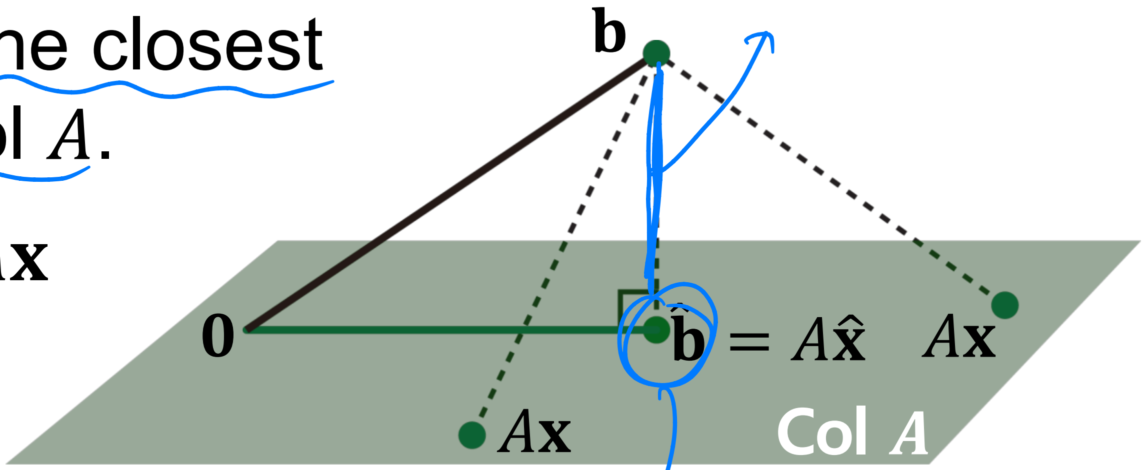
주재걸
고려대학교 컴퓨터학과

Geometric Interpretation of Least Squares

$\hat{x} = \underset{x}{\operatorname{argmin}} \|b-Ax\|$를 찾아야 함

최단 거리가 되려면 수직이어야 함

- Consider $\hat{\mathbf{x}}$ such that $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$ is the closest point to $\mathbf{b}$ among all points in 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 Col A.



Col A와 가장 가까워질 수 있는 Ax를 찾아야 함.

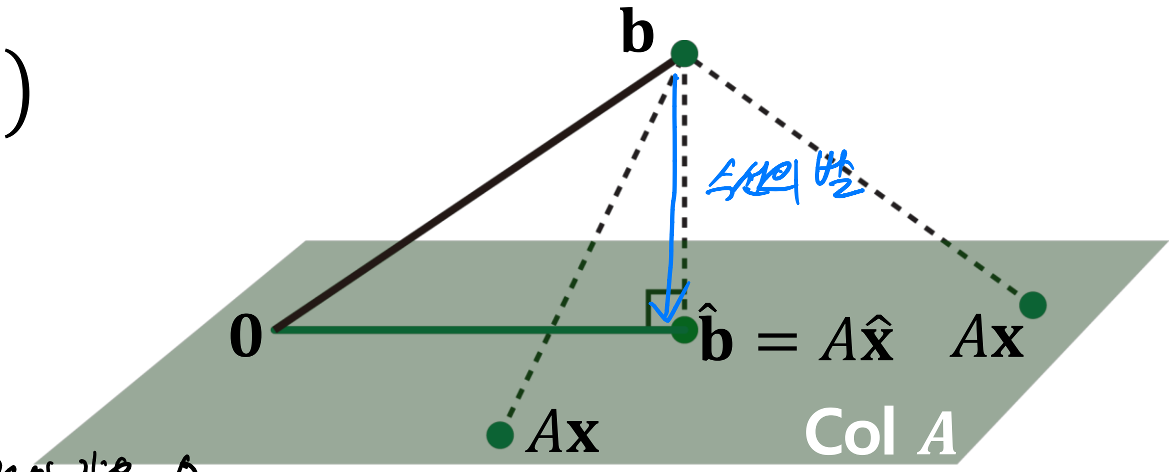
- This means $\mathbf{b} - A\hat{\mathbf{x}}$ should be orthogonal to any vector in Col A:

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

Geometric Interpretation of Least Squares

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

분배법칙



- Or equivalently,

수직이므로 각 내적의 값은 0

$$(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_1 \qquad \mathbf{a}_1^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$
$$(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_2 \quad \Rightarrow \quad \mathbf{a}_2^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \Rightarrow \quad A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$
$$\vdots \qquad\qquad \vdots$$
$$(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_m \qquad \mathbf{a}_m^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

$([\,][\,] \cdots [\,]) = A^T$

$\hat{x} = \arg\min_x (b - Ax|$

Normal Equation

(정규방정식)

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

$A^Tb - A^TA\hat{x} = 0$

$A^Tb = A^TA\hat{x}$

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

$$\underbrace{A^T A}_{C}\hat{\mathbf{x}} = \underbrace{A^T\mathbf{b}}_{d},$$

  which is called a normal equation.

- This can be viewed as a new linear system, $\boxed{C\mathbf{x} = \mathbf{d}}$, 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

역행렬이 있다면

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

$\underset{x}{\operatorname{argmin}} \| b - Ax \|$

Another Derivation of Normal Equation

제곱의
최솟값을 찾는 것과 같음

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

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

이리저리 계산 다 하면

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

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

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

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

결국 정규방정식은 같아

Life-Span Example

4,3 이므로 해 x

정규방정식으로 풀어야 함.

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

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

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

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

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

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

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

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

- Given $A^{\mathrm{T}}A\mathbf{x} = A^{\mathrm{T}}\mathbf{b}$, what if $C = \boxed{A^T A \text{ is NOT invertible?}}$

해가 무수히 많아짐

- Remember that in this case, the system has either no solution or infinitely many solutions.

Normal equation은 항상 해가 있다.
(1개 or 무수히 많다)

- However, the solution always exist for this "normal" equation, and thus infinitely many solutions exist.

증명은 생략,
그냥 받아들일 것

- When $C = A^T A$ is NOT invertible?
If and only if the columns of A are linearly dependent. Why?

- However, $C = A^T A$ is usually invertible. Why?