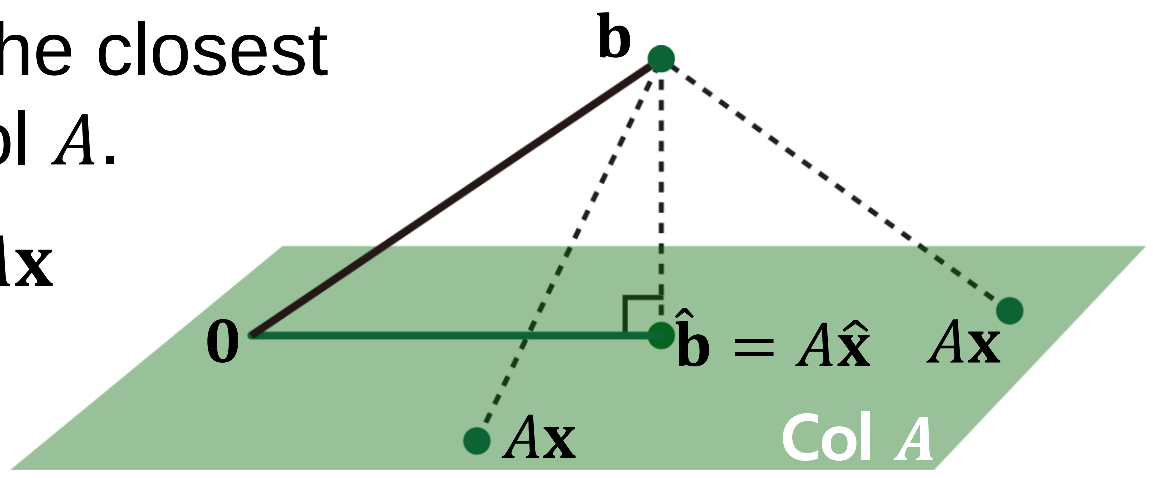
Geometric Interpretation of Least Squares

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



- That is, $\mathbf{b}$ is closer to $\hat{\mathbf{b}}$ than to $A\mathbf{x}$ for any other $\mathbf{x}$.

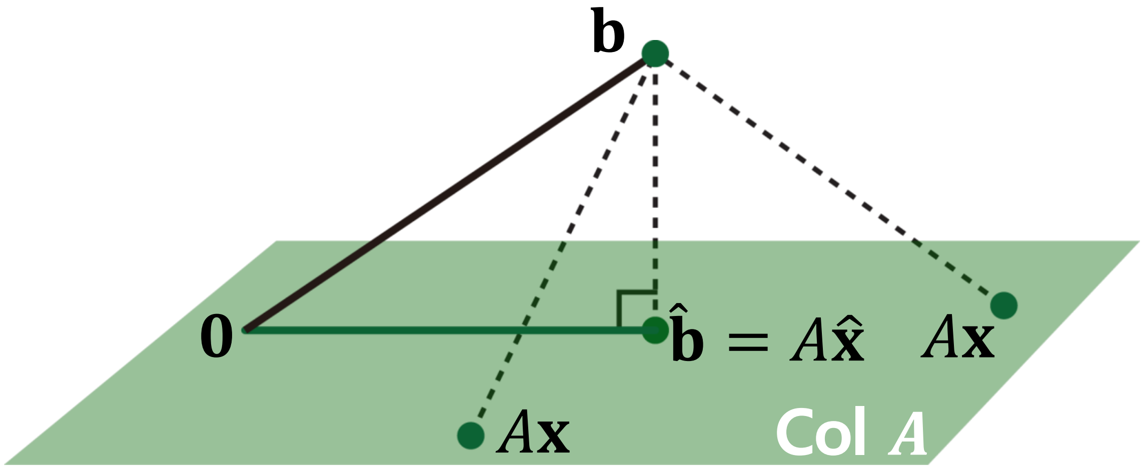
- To satisfy this, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ should be orthogonal to Col A.

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

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

Geometric Interpretation of Least Squares

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



- Or equivalently,

$$
\begin{array}{l}
(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_1 \\
(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_2 \\
\vdots \\
(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathbf{a}_m
\end{array}
\Rightarrow
\begin{array}{l}
\mathbf{a}_1^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \\
\mathbf{a}_2^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \\
\vdots \\
\mathbf{a}_m^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0
\end{array}
\Rightarrow
A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0
$$