

example) If

$$\begin{pmatrix} x_1 + x_2 = 2 \\ x_2 + x_3 = 2 \\ x_1 + x_2 + x_3 = 3 \\ x_1 + x_2 + 2x_3 = 4 \approx 4.1 \end{pmatrix}$$

- If we consider "predicted values" and "target value", we haven't predicted 100%. However, we could consider it as "mostly predicted".

- Since there are many problems that we can't solve due to  $N \neq M$ . However, we could try to solve the problems by minimizing differences between "predicted values" and "targets".

↳ The difference between "predicted values" and "target values" is called "residual".

$$e = Ax - b$$

↳ Since the residual ( $e$ ) is a vector, we could solve problem by minimizing the vector's norm.

↳ Previously, minimizing "norm" is equal to minimizing "square of norm"

$$e^T e = \|e\|^2 = (Ax - b)^T (Ax - b)$$

"Minimizing"

$$x = \arg \min_x e^T e = \arg \min_x (Ax - b)^T (Ax - b)$$

↳ We call it as "Least Square Problem".