

Suppose that we are finding $\hat{\beta}$ in

$$y = X\beta + \text{noise}$$

$y \in \mathbb{R}^n$
not in $C(X)$

$X\hat{\beta}$ is
closest to y

$$\min_{\beta} \|y - X\beta\|^2$$

$$R = y - X\hat{\beta}$$

↑ residuals

$R \perp$ columns of X

