Closed-form LSE approach:

$$f: \|Aw-b\|^2 + \lambda \|w\|^2$$

$$\frac{df}{dw} = 0 \Rightarrow w = (A^TA + \lambda I)^{-1} A^T b$$

Steepest descent method:

$$\|Aw-b\|^2 = f(w) \Rightarrow \nabla f(w) = 2A^TAw - 2A^Tb$$

$$w_{k+1} = w_k - L \cdot \nabla f(w_k)$$

$L$: learning rate

Newton's method:

$$w_{k+1} = w_k - Hf(w_k)^{-1} \cdot \nabla f(w_k)$$

$$Hf(w) = 2A^TA$$
$$\nabla f(w) = 2A^TAw - 2A^Tb$$