1. Closed-form LSE approach.

$$\begin{cases} A: \text{design matrix} \\ x: \text{unknown parameters} \\ b: \text{response variable} \end{cases}$$

$$\|Ax-b\|^2 = f(x) = (Ax-b)^T(Ax-b)$$
$$= x^TA^TAx - 2x^TA^Tb + b^Tb$$

$$\nabla f(x) = 2A^TAx - 2A^Tb$$

$$2A^TAx - 2A^Tb = 0$$

$$x = (A^TA)^{-1}A^Tb$$

$$Hf(x) = 2A^TA$$

$$x_{n+1} = x_n \cdot H^{-1}f(x) \cdot \nabla f(x)$$
$$= 0 - (2A^TA)^{-1}(2A^TAx - 2A^Tb)$$
$$= -\frac{1}{2}(A^TA)^{-1} - 2A^Tb$$
$$= (A^TA)^{-1}A^Tb$$

2. Steepest descend

$$L(x) = \frac{1}{2}\|Ax - b\|^2 + \lambda \|x\|$$

$$\nabla L_{LSE}(x) = A^T(Ax - b)$$

$$\nabla L_{L1}(x) = \lambda \cdot \text{sign}(x) \quad , \text{ where } \text{sign}(x) = \begin{cases} 1, & x_i > 0 \\ -1, & x_i < 0 \\ 0, & x_i = 0 \end{cases}$$

$$\Rightarrow \nabla L(x) = A^T(Ax - b) + \lambda\, \text{sign}(x)$$

$$x_{n+1} = x_n - \alpha_n \nabla L(x)$$

$$= x_n - \alpha_n \left(A^T(Ax - b) + \lambda\, \text{sign}(x) \right) \quad , \text{ where } \alpha_n \text{ is learning rate}$$

3. Newton's method

$$令\ g(x) = f(x_0) + f'(x_0) \cdot \Delta x + \frac{1}{2!} f''(x_0) \cdot \Delta x^2$$

$$g'(x) = f'(x_0) + f''(x_0)\Delta x = 0$$

$$\Delta x = -\frac{f'(x_0)}{f''(x_0)}$$

$$x_{n+1} = x_n + \left(-\frac{f'(x_0)}{f''(x_0)} \right)$$

$$= x_n + \left(-H^{-1} f(x) \cdot \nabla f(x_n) \right)$$

$$= x_n - (2A^TA)^{-1} \cdot (2A^TAx_n - 2A^Tb)$$