Mathematical Derivation

① closed-form LSE approach

$$f = \|A\vec{w}-\vec{b}\|^2 + \lambda\|w\|^2 = (A\vec{w}-b)^T(A\vec{w}-b) + \lambda w^T w$$

$$= \vec{w}^T A^T A\vec{w} - 2b^T A\vec{w} + b^T b + \lambda w^T w$$

$$\frac{\partial f}{\partial \vec{w}} = 2A^T A\vec{w} - 2A^T b + 2\lambda w$$

$$\frac{\partial f}{\partial \vec{w}} = 0 \qquad A^T b = (A^T A + \lambda I)\vec{w}$$

$$(A^T A + \lambda I)^{-1} A^T b = \vec{w}$$

example case: base=3

$$A = \begin{bmatrix} x_0^2 & x_0^1 & x_0^0 \\ x_1^2 & x_1^1 & x_1^0 \\ \vdots & \vdots & \vdots \\ x_{n-1}^2 & x_{n-1}^1 & x_{n-1}^0 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix}$$

$n \times 3$

$$A^T A = \underbrace{\begin{bmatrix} x_0^2 & x_1^2 & \cdots & x_{n-1}^2 \\ x_0^1 & x_1^1 & \cdots & x_{n-1}^1 \\ x_0^0 & x_1^0 & \cdots & x_{n-1}^0 \end{bmatrix}}_{\substack{A^T \\ 3\times n}} \underbrace{\begin{bmatrix} x_0^2 & x_0^1 & x_0^0 \\ x_1^2 & x_1^1 & x_1^0 \\ \vdots & \vdots & \vdots \\ x_{n-1}^2 & x_{n-1}^1 & x_{n-1}^0 \end{bmatrix}}_{\substack{A \\ n\times 3}} = \underbrace{\begin{bmatrix} a_{0,0} & a_{0,1} & a_{0,2} \\ a_{1,0} & a_{1,1} & a_{1,2} \\ a_{2,0} & a_{2,1} & a_{2,2} \end{bmatrix}}_{3\times 3}$$

$$A^T b = \underbrace{\begin{bmatrix} x_0^2 & x_1^2 & \cdots & x_{n-1}^2 \\ x_0^1 & x_1^1 & \cdots & x_{n-1}^1 \\ x_0^0 & x_1^0 & \cdots & x_{n-1}^0 \end{bmatrix}}_{3\times n} \underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} \sum_{i=0}^{n-1} x_i^2 y_i \\ \sum_{i=0}^{n-1} x_i^1 y_i \\ \sum_{i=0}^{n-1} x_i^0 y_i \end{bmatrix}}_{3\times 1}$$

Gaussian-Jorden elimination $(A^TA)^{-1}$

A square matrix A is invertible if and only if it is a non-singular i.e $\det(A) \neq 0$

$$\left[\begin{array}{ccc|ccc} a_{0,0} & a_{0,1} & a_{0,2} & 1 & 0 & 0 \\ a_{1,0} & a_{1,1} & a_{1,2} & 0 & 1 & 0 \\ a_{2,0} & a_{2,1} & a_{2,2} & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & a'_{0,0} & a'_{0,1} & a'_{0,2} \\ 0 & 1 & 0 & a'_{1,0} & a'_{1,1} & a'_{1,2} \\ 0 & 0 & 1 & a'_{2,0} & a'_{2,1} & a'_{2,2} \end{array}\right]$$

$$\vec{w} = \begin{bmatrix} a'_{0,0} & a'_{0,1} & a'_{0,2} \\ a'_{1,0} & a'_{1,1} & a'_{1,2} \\ a'_{2,0} & a'_{2,1} & a'_{2,2} \end{bmatrix} \begin{bmatrix} \sum_{i=0}^{n-1} x_i^2 y_i \\ \sum_{j=0}^{n-1} x_i^1 y_i \\ \sum_{j=0}^{n-1} x_i^0 y_i \end{bmatrix}$$

$$= \begin{bmatrix} a'_{0,0} \sum_{j=0}^{n-1} x_i^2 y_i + a'_{0,1} \sum_{j=0}^{n-1} x_i^1 y_i + a'_{0,2} \sum_{j=0}^{n-1} x_i^0 y_i \\ a'_{1,0} \sum_{j=0}^{n-1} x_i^2 y_i + a'_{1,1} \sum_{j=0}^{n-1} x_i^1 y_i + a'_{1,2} \sum_{j=0}^{n-1} x_i^0 y_i \\ a'_{2,0} \sum_{j=0}^{n-1} x_i^2 y_i + a'_{2,1} \sum_{j=0}^{n-1} x_i^1 y_i + a'_{2,2} \sum_{j=0}^{n-1} x_i^0 y_i \end{bmatrix} = (A^TA)^{-1}A^Tb = \vec{w}$$

② Steepest descent method with l_1 norm

minimize $\nabla g^T(v)\, d$，其中 $d = w - v$ $\|d_i\|_1 = 1$

$\langle d, \nabla g(v) \rangle$

最小化梯度與方向 d 的 inner product

找一個方向 d 使 $\nabla g^T(v)$ 在 d 上投影值最小

最佳解為梯度中最負的值

$\langle d, \nabla g(v) \rangle \quad w_1 d_1 + w_2 d_2 + \cdots + w_n d_n$，其中 $\nabla g(v) = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$

with $|d_1| + |d_2| + \cdots + |d_n| = 1$

⇒ 最佳解為某個最負的，使 $w_1 d_1 + \cdots + w_n d_n$ 最小

若 j 為 $\operatorname{argmax} \left| \frac{\partial}{\partial w_j} g(v) \right|$

③ newton's method

$$f(x) \approx f(x_0) + f'(x_0)(x-x_0) + \cdots \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

$\because$ when x close to x_0, $(x-x_0)^n$ is small enough

$$f(x) \approx f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2!} f''(x_0)(x-x_0)^2$$

f is loss function

then we need to find the smallest point which means $f'(x)=0$

$$0 = f'(x) = f'(x_0) + f''(x_0)(x-x_0)$$

$$f'(x_0) + f''(x_0)(x-x_0) = 0$$

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} \qquad x_{n+1} = x_n - H^{-1}(f(x_n)) \nabla f(x_n)$$

$$f(x) = \|Ax-b\|^2 = (Ax-b)^T(Ax-b)$$
$$= x^T A^T A x - 2b^T A x + b^T b$$

$$\nabla f(x) = 2A^T A x - 2A^T b$$

$$(H(f(x)))^{-1} = (2A^T A)^{-1}$$

$$x_{n+1} = x_n - (2A^T A)^{-1}\left[2A^T A x_n - 2A^T b\right]$$
$$= x_n - x_n + \frac{1}{2}(A^T A)^{-1} \cdot 2A^T b$$
$$= (A^T A)^{-1} A^T b$$