1. LSE

goal : $\arg\min \sum_i (y_i - f(x_i))^2$

$\Rightarrow \min \|Ax - b\|^2$

$\Rightarrow (Ax-b)^T(Ax-b)$

scalar

$\Rightarrow x^TA^TAx - \underbrace{x^TA^Tb} - \underbrace{b^TAx} + b^Tb$

$\Rightarrow x^TA^TAx - 2x^TA^Tb + b^Tb$

$\therefore \frac{df}{dx} = 2A^TAx - 2A^Tb \overset{\text{let}}{=} 0$

$A^TAx = A^Tb$

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

2. Steepest Descent

use LSE $\Rightarrow E(w) = \frac{1}{2}(y-Aw)^T(y-Aw)$

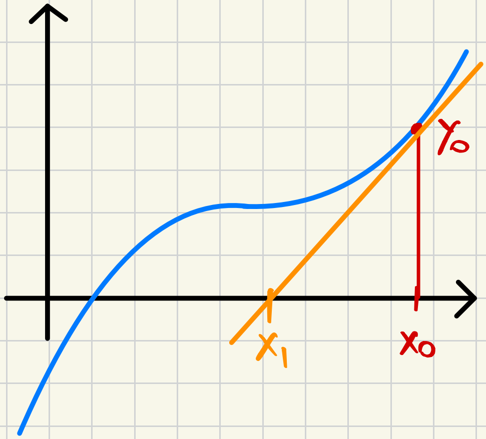
$$\Rightarrow \nabla E(w) = \frac{\partial E(w)}{\partial w} = A^TAw - A^Ty$$

$$\Rightarrow w^{(t+1)} = w^{(t)} - \eta \nabla E(w^{(t)})$$

$$= w^{(t)} - \eta(A^TAw^{(t)} - A^Ty)$$

$$= w^{(t)} - \eta A^T(Aw^{(t)} - y)$$

3. Newton's method



$$f'(x_0) = \frac{f(x_0)}{x_0 - x_1} \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

In matrix form

$$w^{(t+1)} = w^{(t)} - \frac{\nabla E(w^{(t)})}{\frac{\partial E(w^{(t)})}{\partial w}} = w^{(t)} - H^{-1} \nabla E(w^{(t)})$$

LSE :

Let $w^{(0)} = 0$

$H = A^T A$

$$\Rightarrow w^{(1)} = 0 - H^{-1} \nabla E(w^{(0)}) \rightarrow \nabla E(w) = (A^T A w - A^T y)$$

$$= 0 - (A^T A)^{-1} (0 - A^T y)$$

$$= (A^T A)^{-1} A^T y$$