Part 1.

1. Closed form LSE method

goal: $\arg\min \left(\sum_i (y_i - f(x_i))^2 \right) = \arg\min \|A\vec{x} - \vec{b}\|^2$ where $A \in \mathbb{R}^{m\times n}$, $\vec{b} \in \mathbb{R}^m$, $\vec{x} \in \mathbb{R}^n$

$$\|A\vec{x}-\vec{b}\|^2 = (A\vec{x}-\vec{b})^T(A\vec{x}-\vec{b})$$

$$= (\vec{x}^T A^T - \vec{b}^T)(A\vec{x}-\vec{b})$$

$$= \vec{x}^T A^T A \vec{x} - \vec{x}^T A^T \vec{b} - \vec{b}^T A \vec{x} + \vec{b}^T \vec{b}$$

(since $\underset{1\times n}{\vec{x}^T} \underset{n\times m}{A^T} \underset{m\times 1}{\vec{b}}$ and $\underset{1\times m}{\vec{b}^T} \underset{m\times n}{A} \underset{n\times 1}{\vec{x}}$ is 1×1 (scalar)

so $(A\vec{x})^T b = (b^T(A\vec{x}))^T = b^T A\vec{x}$)

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

Let $g = \vec{x}^T A^T A \vec{x} - 2\vec{b}^T A \vec{x} + \vec{b}^T \vec{b}$

$$\frac{\partial g}{\partial x} = 2A^T A \vec{x} - 2A^T \vec{b} \triangleq 0$$

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}$$ #

加入 L2 norm

$$g := \|A\vec{x} - b\|^2 + \lambda \|\vec{x}\|^2 = \vec{x}^T A^T A \vec{x} - 2\vec{b} A \vec{x} + \vec{b}^T b + \|\vec{x}\|^2$$

$$\frac{\partial g}{\partial x} = 2A^T A \vec{x} - 2A^T \vec{b} + 2\lambda \vec{x} \triangleq 0$$

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

2. Steepest descent method

Let initial point x_0

learning rate α

$$x_1 = x_0 - \alpha \cdot \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$$

$\vdots$ continue iterating until $\nabla f(x_0) \approx 0$

$$x_n = x_{n-1} - \alpha \cdot \frac{\nabla f(x_{n-1})}{\|\nabla f(x_{n-1})\|}$$

" steepest "

$f: \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}$ and $dir(\nabla f(x_0)) = v$

$$\left.\frac{\partial f(x)}{\partial v}\right|_{x=x_0} := (\nabla f(x_0))^T \frac{v}{\|v\|}$$

By Cauchy - Schwarz inequality $(|x^T y| \leq \|x\| \|y\|)$

$$\left| (\nabla f(x_0))^T \frac{v}{\|v\|} \right| \leq \|\nabla f(x_0)\| \frac{\|v\|}{\|v\|}$$

if $v = \nabla f(x_0)$

$$\left| (\nabla f(x_0))^T \cdot \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|} \right| \leq \|\nabla f(x_0)\|$$

$$\Rightarrow \frac{\|\nabla f(x_0)\|^2}{\|\nabla f(x_0)\|} \leq \|\nabla f(x_0)\|$$

when the equation holds, $v = \nabla f(x_0)$

So, $\nabla f(x_0)$ is the steepest.

In this homework,

f is LSE + L1 norm

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

$$= (Ax-b)^T (Ax-b) + \lambda \|x\|$$

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

$$\nabla f = 2A^T A \vec{x} - 2A^T \vec{b} + \lambda \cdot sgn(x)$$

3. Newton's Method

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = x_0 - Hf(x_0)^{-1} \nabla f(x_0)$$

where $Hf(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_0^2} & \frac{\partial^2 f}{\partial x_0 \partial x_1} & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_0} & \ddots & \\ \vdots & & \end{bmatrix}$

$$x_2 = x_1 - \frac{f'(x_1)}{f''(x_1)}$$

$\vdots$ continue iterating until the difference $\Delta x = x_n - x_{n-1}$ is small enough

$$x_n = x_{n-1} - \frac{f'(x_{n-1})}{f''(x_{n-1})}$$

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

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

$$Hf(\vec{x}) = 2A^T A$$

$$\vec{x}_1 = \vec{x}_0 - (2A^T A)^{-1}(2A^T A \vec{x}_0 - 2A^T \vec{b})$$

$$\vec{x}_2 = \vec{x}_1 - (2A^T A)^{-1}(2A^T A \vec{x}_1 - 2A^T \vec{b})$$

$\vdots$

$$\vec{x}_n = \vec{x}_{n-1} - (2A^T A)^{-1}(2A^T A \vec{x}_{n-1} - 2A^T \vec{b})$$