ML HW 1 313551170 王韋綎

Part 1

1. Closed-form LSE approach

N points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

The approximative formula:

$$f(x, w) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M$$

$$= \underbrace{\begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^M \\ \vdots & \vdots & & & \\ 1 & x_N^1 & x_N^2 & \cdots & x_N^M \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \end{bmatrix}}_{\vec{w}}$$

we need to find $\vec{w}$ to let $f(x)$ close to y.

Closed-form solution of LSE $= E(w) = \min \sum (f(x_i) - y_i)^2$

$= \min \| A\vec{w} - \vec{b} \|$, $\quad \vec{b} = (y_1, y_2, \ldots, y_N)^T$

$\| A\vec{w} - \vec{b} \|$

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

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

(since $\vec{b}^T A \vec{w}$ is scalar, $\vec{b}^T A \vec{w} = (\vec{b}^T A \vec{w})^T = \vec{w}^T A^T \vec{b}$)

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

To let $\|A\vec{w} - \vec{b}\|$ be minimal, we need to find $\vec{w}$ s.t.

differential of $\|A\vec{w} - \vec{b}\| = 0$

$$\frac{d\|A\vec{w} - \vec{b}\|}{d\vec{w}} = \frac{d\,\vec{w}^T A^T A \vec{w}}{d\vec{w}} - 2\frac{d\,\vec{b}^T A \vec{w}}{d\vec{w}}$$

since $\vec{b}^T A \vec{w}$ is a linear combination of the vector element of $\vec{w}$, $\frac{d\,\vec{b}^T A \vec{w}}{d\vec{w}} = (\vec{b}^T A)^T = A^T \vec{b}$

$$\frac{d\,\vec{w}^T A^T A \vec{w}}{d\vec{w}} = 2A^T A \vec{w}$$ (omit the process of proess)

$$2A^T A\vec{w} - 2A^T\vec{b} = 0$$

$$\Rightarrow A^T A \vec{w} = A^T \vec{b}$$

$$\Rightarrow \vec{w} = (A^T A)^{-1} A^T \vec{b}$$

2. Steepest descent method

Loss function with L1 norm:

$$E = \sum_{i=0}^{n} [y_i - (m_k x_i^k + \cdots + m_1 x_i + m_0)]^2 + \lambda \sum_{i=0}^{k} |m_i|$$

We need to find the derivative with respect to $m_0, m_1, \ldots, m_k$

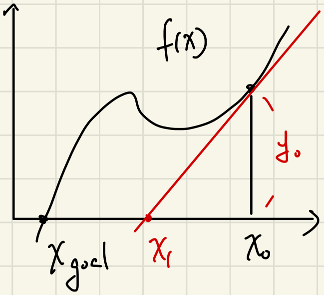
$$D_{m_k} = \sum_{i=0}^{n} 2[y_i - (m_k x_i^k + \cdots + m_1 x_i + m_0)](-x_i^k) + \lambda \frac{d|m_k|}{dm_k}$$

$$\vdots$$

$$D_{m_1} = \sum_{i=0}^{n} 2[y_i - (m_k x_i^k + \cdots + m_1 x_i + m_0)](-x_i) + \lambda \frac{d|m_1|}{dm_1}$$

$$D_{m_0} = \sum_{i=0}^{n} 2[y_i - (m_k x_i^k + \cdots + m_1 x_i + m_0)] + \lambda \frac{d|m_0|}{dm_0}$$

3. Newton's method



$$f'(x_0) = \frac{y_0}{x_0 - x_1} = f'(x_0) = \frac{f(x_0)}{x_0 - x_1}$$

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

Tyler Expansion:

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

$$\approx \sum \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n = g(x)$$

Newton's method:

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

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

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

$$= x_n - \underbrace{H^{-1}(f(x_n))}_{\text{Hessian matrix}} \underbrace{\nabla f(x_n)}_{\text{gradient}}$$

$$H(f(x)) = \begin{bmatrix} \frac{\partial f^2}{\partial x_1^2} & \frac{\partial f^2}{\partial x_1 x_2} & \cdots \\ \frac{\partial f^2}{\partial x_2 x_1} & \frac{\partial f^2}{\partial x_2^2} & \\ \vdots & & \ddots \end{bmatrix}$$

With LSE :

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

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

$$H(f(x)) = 2A^T A$$

$$\Rightarrow x_{n+1} = x_n - H^{-1}(f(x)) \cdot \nabla f(x)$$

$$= x_n - (2A^T A)^{-1} \cdot (2A^T A x_n - 2A^T b)$$