5. Locally weighted linear regression

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)}\left[\theta^T x^{(i)} - y^{(i)}\right]^2$$

ⓐ (i) $\cancel{\text{Let } z^{(i)} = \theta^T x^{(i)} - y^{(i)} = \begin{bmatrix} z_0^{(i)} \\ z_1^{(i)} \\ \vdots \\ z_n^{(i)} \end{bmatrix}}$

$$X = \begin{bmatrix} x_0^{(1)} & x_1^{(1)} & \cdots & x_n^{(1)} \\ x_0^{(2)} & x_1^{(2)} & \cdots & x_n^{(2)} \\ \vdots & & & \\ x_0^{(m)} & x_1^{(m)} & & x_n^{(m)} \end{bmatrix} \in R^{m \times n}$$

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

Let $Z = (X\theta - Y) = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$

$$J(\theta) = \frac{1}{2}\left(w^{(1)} z_1^2 + w^{(2)} z_2^2 + \cdots + w^{(m)} z_m^2\right)$$

$$J(\theta) = \frac{1}{2}\begin{bmatrix} z_1 & z_2 & \cdots & z_m \end{bmatrix}\begin{bmatrix} w^{(1)} & & 0 \\ & w^{(2)} & \\ 0 & & \ddots \; w^{(m)} \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

$$= Z^T \begin{bmatrix} \frac{w^{(1)}}{2} & & \\ & \frac{w^{(2)}}{2} & \\ & & \ddots \; \frac{w^{(m)}}{2} \end{bmatrix} Z$$

$$= (X\theta - Y)^T W (X\theta - Y)$$

Where, $W = \frac{1}{2}\begin{bmatrix} w^{(1)} & & \\ & w^{(2)} & \\ & & \ddots \; w^{(m)} \end{bmatrix}$

(ii) $J(\theta) = (X\theta - Y)^T W (X\theta - Y)$

$$= (Y^T - \theta^T X^T) W (X\theta - Y)$$

$$= Y^T W X\theta - \theta^T X^T W X\theta - Y^T W Y + \theta^T X^T W Y$$

$\{as\ W^T = W\}$ $Y^T W X\theta \leftarrow Y^T W^T X\theta$

$$J(\theta) = -\theta^T (X^T W X)\theta + 2Y^T W X\theta - Y^T W Y$$

$$\nabla_\theta J(\theta) = -2(X^T W X)\theta + 2(Y^T W X)^T = 0$$

$$(X^T W X)\theta = (Y^T W X)^T$$

$$\boxed{\theta = (X^T W X)^{-1}(X^T W Y)}$$

iii) $\{(x^{(i)}, y^{(i)});\ i = 1 \dots m\}$

$$P(y^{(i)} \mid x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

$$l(\theta) = \log\left(\prod_{i=1}^{m} P(y^{(i)} \mid x^{(i)}; \theta)\right)$$

$$= \sum_{i=1}^{m} \log\left(\frac{1}{\sqrt{2\pi}\,\sigma^{(i)}}\right) - \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}$$

$\Rightarrow$ Maximizing $l(\theta)$ is equivalent to minimizing $J(\theta)$.

$$\boxed{J(\theta) = \sum_{i=1}^{m} \frac{1}{2(\sigma^{(i)})^2} \left(y^{(i)} - \theta^T x^{(i)}\right)^2}$$

So $\boxed{\omega^{(i)} = \frac{1}{2(\sigma^{(i)})^2}}$