Problem Set #1.

5. Locally weighted linear regression

(a) $J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)}\left(\theta^T x^{(i)} - y^{(i)}\right)^2 \qquad \underset{(n+1,1)}{x^{(i)} \in \mathbb{R}^{n+1}},\ y^{(i)} \in \mathbb{R},\ \underset{(n+1,1)}{\theta \in \mathbb{R}^{n+1}}$

i) Show that $J(\theta)$ can also be written as $J(\theta) = (X\theta - y)^T W (X\theta - y)$

$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - \dot{x})^2}{2}\right) \quad \dot{x} \in \mathbb{R}^{n+1}$$

$$w^{(i)} \in \mathbb{R}^1$$

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

$$= \frac{1}{2}\sum_{i=1}^{m} \underset{(1,1)}{w^{(i)}}\left(\underset{(1,n+1)}{\theta^T}\underset{(n+1,1)}{x^{(i)}} - y^{(i)}\right)\left(\underset{(1,n+1)}{\theta^T}\underset{(n+1,1)}{x^{(i)}} - y^{(i)}\right)$$

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

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

$(1,m)\ (m,1)$

$$X : \begin{bmatrix} - x^{(1)T} - \\ - x^{(2)T} - \\ \vdots \\ - x^{(m)T} - \end{bmatrix} \in \mathbb{R}^{m \times (n+1)} \qquad \theta : \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \\ \vdots \\ \theta^{(n+1)} \end{bmatrix} \in \mathbb{R}^{n+1} \qquad y : \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \in \mathbb{R}^m$$

$$\underset{(1,m)}{(X\theta - y)^T}\ \underset{(m,m)}{W}\ \underset{(m,1)}{(X\theta - y)}$$

$$W : \begin{bmatrix} \frac{1}{2}w^{(1)} & \cdots & & 0 \\ 0 & \frac{1}{2}w^{(2)} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \frac{1}{2}w^{(m)} \end{bmatrix} \in \mathbb{R}^{m \times m}$$

ii) $J(\theta) = \underbrace{(X\theta - y)^T}_{(1,m)} \underbrace{W}_{(m,m)} \underbrace{(X\theta - y)}_{(m,1)}$

$$\nabla_\theta J(\theta) = 0$$

$$\nabla_\theta J(\theta) = \nabla_\theta \{(X\theta - y)^T W (X\theta - y)\}$$

$$= \nabla_\theta \{((X\theta)^T - y^T) W (X\theta - y)\}$$

$$= \nabla_\theta \{(\theta^T X^T - y^T) W (X\theta - y)\}$$

$$= \nabla_\theta \{\theta^T X^T W - y^T W)(X\theta - y)\}$$

$$= \nabla_\theta \{\theta^T X^T W X\theta - \theta^T X^T W y - y^T W X\theta + y^T W y\}$$

$$= \{2X^T W X\theta - X^T W y - (y^T W X)^T\}$$

$$= \{2X^T W X\theta - X^T W y - X^T W^T y\}$$

$W^T = W$

(diagonal matrix)

$$= 2X^T W X\theta - 2X^T W y = 0$$

$$X^T W X\theta = X^T W y$$

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

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

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

$\begin{cases} \text{mean: } \theta^T x^{(i)} \\ \text{variance: } (\sigma^{(i)})^2 \end{cases}$ $\sigma^{(i)} \in \mathbb{R}$

$$\text{MLE}: \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} \exp\left(-\frac{(y^{(i)}-\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

$$\begin{aligned}
\log \mathcal{L} &= \log \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta) \\
&= \sum_{i=1}^{m} \log p(y^{(i)}|x^{(i)};\theta) \\
&= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} \exp\left(-\frac{(y^{(i)}-\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right) \\
&= \sum_{i=1}^{m} \left[\log \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} + \left(-\frac{(y^{(i)}-\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)\right] \\
&= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} + \sum_{i=1}^{m} \left(-\frac{(y^{(i)}-\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta} \log \mathcal{L} &= \frac{\partial}{\partial\theta} \sum_{i=1}^{m} \left(-\frac{(y^{(i)}-\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right) = 0 \\
&= \sum_{i=1}^{m} \left(-\frac{1}{2(\sigma^{(i)})^2} (\theta^T x^{(i)} - y)^2\right) \\
&= \frac{1}{2} \sum_{i=1}^{m} \left(-\frac{1}{(\sigma^{(i)})^2} (\theta^T x^{(i)} - y)^2\right)
\end{aligned}$$

locally weighted linear regression

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

$$w^{(i)} = -\frac{1}{(\sigma^{(i)})^2}$$ ✓