Solution to Assignment 1

Shoeb Mohammed and Zhuo Chen

January 24, 2016

1 Locally weighted linear regression

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

1.1

Matrix X and vectors θ , y are

$$X=\begin{bmatrix} {x^{(i)}}^T \\ \vdots \\ {x^{(m)}}^T \end{bmatrix},\theta=\begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix},y=\begin{bmatrix} y^{(i)} \\ \vdots \\ y^{(m)} \end{bmatrix} \tag{2}$$

Let W be the $m\times m$ diagnoal matrix

$$W=\frac{1}{2}diag(w^{(1)},\cdots,w^{(m)}) \tag{3}$$

Using (2) and (3), equation (1) can be written as

$$J(\theta)=(X\theta-y)^T W(X\theta-y) \tag{4}$$

1.2

The normal equations for un-weighted linear regression are

$$X^T X\theta=X^T y \tag{5}$$

Equation 4 can be re-written as

$$\begin{aligned} J(\theta)&=(X\theta-y)^T\sqrt{W}\sqrt{W}(X\theta-y) \\ &=(\sqrt{W}X\theta-\sqrt{W}y)^T(\sqrt{W}X\theta-\sqrt{W}y) \\ &=(X'\theta-y')^T(X'\theta-y') \text{ where } X'=\sqrt{W}X \text{ and } y'=\sqrt{W}y \end{aligned} \tag{6}$$

Now, equation 6 is similar to unweighted $J(\theta)$. Thus, using equation 5, θ in closed form is

$$\theta=\left[(\sqrt{W}X)^T(\sqrt{W}X)\right]^{-1}\left(\sqrt{W}X\right)^T\sqrt{W}y \tag{7}$$

1.3

Locally weighted linear regression is a non-parametric model. To estimate y, given x

- first calculate the weights $w^{(i)}$. This gives the matrix W as defined in equation 3
- Start with random guess for θ
- In the i^{th} iteration of the algorithm, the θ is updated using the relation

$$\theta(i)\leftarrow\theta(i-1)-\alpha\left(\sqrt{W}X\right)^T\left(\sqrt{W}X\theta(i-1)-\sqrt{W}y\right) \tag{8}$$

- The above step is repeated until θ converges

2 Properties of the linear regression estimator

2.1

Vectors y,θ,ϵ and matrix X are related as

$$y=X\theta+\epsilon \tag{9}$$

ϵ is i.i.d $N(0,\sigma^2)$.

Thus, for the optimal value θ^* and fixed X

$$\begin{aligned} E[y]&=E[X\theta^*]+E[\epsilon] \\ &=E[X\theta^*]+0 \\ &=X\theta^* \end{aligned} \tag{10}$$

Given the normal equations

$$\begin{aligned} \theta&=(X^T X)^{-1}X^T y \\ \implies E[\theta]&=(X^T X)^{-1}X^T E[y] \\ &=(X^T X)^{-1}X^T X\theta^* \\ &=\theta^* \end{aligned} \tag{11}$$

2.2

$$
\begin{aligned}
Var(\theta) &= E\left[(\theta - E(\theta))(\theta - E(\theta))^T\right] \\
&= E\left[(\theta - \theta^*)(\theta - \theta^*)^T\right] \\
&= E[\theta\theta^T] - \theta^*\theta^{*T}
\end{aligned} \tag{12}
$$

Use the normal equations (linear regression estimator) to get

$$
\begin{aligned}
E[\theta\theta^T] &= E\left[(X^TX)^{-1}X^Tyy^T\left((X^TX)^{-1}\right)^T\right] \\
&= (X^TX)^{-1}X^TE[yy^T]\left((X^TX)^{-1}\right)^T
\end{aligned} \tag{13}
$$

Use equation 9 to get

$$
\begin{aligned}
E[yy^T] &= E[(X\theta + \epsilon)(X\theta + \epsilon) \\
&= E[X\theta^*\theta^{*T}X^T + X\theta^*\epsilon^T + \epsilon\theta^{*T}X^T + \epsilon\epsilon^T] \\
&= X\theta^*\theta^{*T}X^T + E[\epsilon\epsilon^T] \\
&= X\theta^*\theta^{*T}X^T + diag(\sigma^2) \\
&= X\theta^*\theta^{*T}X^T + \sigma^2 I
\end{aligned} \tag{14}
$$

Equations 13 and 14 imply

$$
\begin{aligned}
E[\theta\theta^T] &= (X^TX)^{-1}X^T\left[X\theta^*\theta^{*T}X^T + \sigma^2 I\right]X\left((X^TX)^{-1}\right)^T \\
&= \left[\theta^*\theta^{*T}X^T + (X^TX)^{-1}X^T\sigma^2 I\right]X\left((X^TX)^{-1}\right)^T \\
&= \theta^*\theta^{*T} + (X^TX)^{-1}X^T\sigma^2 IX\left((X^TX)^{-1}\right)^T \\
&= \theta^*\theta^{*T} + \sigma^2(X^TX)^{-1}X^TX\left((X^TX)^{-1}\right)^T \\
&= \theta^*\theta^{*T} + \sigma^2(X^TX)^{-1}
\end{aligned} \tag{15}
$$

Equations 12 and 15 imply

$$
Var(\theta) = \sigma^2(X^TX)^{-1} \tag{16}
$$