

1a)

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

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

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

$$\frac{\partial}{\partial \theta} (RSS(\theta)) = -X^T Y - X^T Y + (X^T X + (X^T X)^T) \theta$$

$$= -2X^T Y + 2X^T X \theta$$

$$= 0$$

$$\theta_{\text{critical}} = (X^T X)^{-1} X^T Y$$

1b)

$$\text{Ridge}_\lambda(\theta) = \text{RSS}(\theta) + \lambda^2 \theta^T \theta$$

$$\frac{\partial}{\partial \theta} (\text{Ridge}_\lambda \theta) = \frac{\partial}{\partial \theta} (\text{RSS}(\theta)) + \lambda^2 \frac{\partial}{\partial \theta} (\theta^T \theta)$$

$$= -2X^T Y + 2X^T X \theta + 2\lambda^2 \theta$$

$$= -2X^T Y + 2(X^T X + \lambda^2 I) \theta$$

$$= 0$$

$$\Rightarrow \theta_{\text{critical}} = (X^T X + \lambda^2 I)^{-1} X^T Y$$

3a)

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

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

where,

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

 $n \times 1$

$$X = \begin{bmatrix} 1 & x^{(1)(1)} & x^{(1)(2)} & \dots & x^{(1)(d)} \\ 1 & x^{(2)(1)} & x^{(2)(2)} & \dots & x^{(2)(d)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x^{(n)(1)} & x^{(n)(2)} & \dots & x^{(n)(d)} \end{bmatrix}$$

 $n \times (d+1)$

$$\theta = \begin{bmatrix} \theta^{(0)} \\ \theta^{(1)} \\ \theta^{(2)} \\ \vdots \\ \theta^{(d)} \end{bmatrix}$$

 $(d+1) \times 1$

$$W = \begin{bmatrix} w^{(1)} & & & & \\ & w^{(2)} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & w^{(n)} \end{bmatrix}$$

 $n \times n$

$$J(\vec{\theta}, \vec{x}) = (Y^T - \theta^T x^T) (WY - WX\theta)$$

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

(a)

$$\begin{aligned} \frac{\partial}{\partial \theta} (J(\vec{\theta}, \vec{x})) &= 2 X^T W X \theta - 2 X^T W Y \\ &= 2 X^T W (X\theta - Y) \end{aligned}$$

(b)

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} (J(\vec{\theta}, \vec{x})) &= 2 (X^T W X)^T \\ &= 2 X^T W^T X \\ &= 2 X^T W X \end{aligned}$$

(c)

$$\begin{aligned}\theta^{(t+1)} &= \theta^{(t)} - 2\eta X^T W (X \theta^{(t)} - Y) \\ &= (I_n - 2\eta X^T W X) \theta^{(t)} + 2\eta X^T W Y\end{aligned}$$

(d)

$$\begin{aligned}\theta^{(t+1)} &= \theta^{(t)} - \frac{1}{2} (X^T W X)^{-1} 2 X^T W (X \theta^{(t)} - Y) \\ &= \theta^{(t)} - \theta^{(t)} + (X^T W X)^{-1} X^T W Y \\ &= (X^T W X)^{-1} X^T W Y\end{aligned}$$