1a)

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

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

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

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

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

$$= 0$$

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

1 b)

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

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

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

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

$$= 0$$

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

3 a)

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

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

where,

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}_{n\times 1}, \quad X = \begin{bmatrix} 1 & x^{(1)(1)} & x^{(1)(2)} & \cdots & x^{(1)(d)} \\ 1 & x^{(2)(1)} & x^{(2)(2)} & \cdots & x^{(2)(d)} \\ \vdots & & \ddots & & \vdots \\ 1 & x^{(n)(1)} & x^{(n)(2)} & \cdots & x^{(n)(d)} \end{bmatrix}_{n\times(d+1)} \quad \theta = \begin{bmatrix} \theta^{(0)} \\ \theta^{(1)} \\ \theta^{(2)} \\ \vdots \\ \theta^{(d)} \end{bmatrix}_{(d+1)\times 1} \quad W = \begin{bmatrix} w^{(1)} & & 0 \\ & w^{(2)} & \\ & & \ddots & \\ 0 & & & w^{(n)} \end{bmatrix}_{n\times n}$$

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

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

(a)

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

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

(b)

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

$$= 2X^T W^T X$$

$$= 2X^T W X$$

(c)

$$\theta^{(t+1)} = \theta^{(t)} - 2\eta X^T W \left(X\theta^{(t)} - Y\right)$$

$$= \left(I_n - 2\eta X^T W X\right)\theta^{(t)} + 2\eta X^T W Y$$

(d)

$$\theta^{(t+1)} = \theta^{(t)} - \frac{1}{2}\left(X^T W X\right)^{-1} 2X^T W \left(X\theta^{(t)} - Y\right)$$

$$= \theta^{(t)} - \theta^{(t)} + \left(X^T W X\right)^{-1} X^T W Y$$

$$= \left(X^T W X\right)^{-1} X^T W Y$$

in

(4)

(1)

$$f(x) = \beta_0 + \beta_1 \sin(x) + \beta_2 \cos(x)$$

$$X = \begin{bmatrix} 1 & \sin(x_1) & \cos(x_1) \\ 1 & \sin(x_2) & \cos(x_2) \\ & \vdots & \\ 1 & \sin(x_n) & \cos(x_n) \end{bmatrix}$$

$\sin(x)$ and $\cos(x)$ are linearly independent. Therefore, we have design matrix with linearly independent columns. Hence, we can use least-squares method.

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\hat{y} = X\beta$$

$$\beta = \left(X^T X\right)^{-1} X^T Y$$

(2)

$$g(x) = \beta_0 + \sin(\beta_1 x) + \cos(\beta_2 x)$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad RSS = \sum_{i=1}^{n} \left(g(x^{(i)}) - y^{(i)} \right)^2$$

$$\beta^{(t+1)} = \beta^{(t)} - \eta \left(\frac{\partial (RSS)}{\partial \beta} \right)_{(t)}$$

learning rate

$$\frac{\partial (RSS)}{\partial \beta} = \begin{bmatrix} \frac{\partial (RSS)}{\partial \beta_0} \\ \frac{\partial (RSS)}{\partial \beta_1} \\ \frac{\partial (RSS)}{\partial \beta_2} \end{bmatrix}$$

$$\frac{\partial}{\partial \beta_j}(RSS) = \sum_{i=1}^{n} 2\left(g(x^{(i)}) - y^{(i)}\right) * \frac{\partial}{\partial \beta_j}\left(g(x^{(i)})\right)$$

$$\frac{\partial}{\partial \beta_0}\left(g(x^{(i)})\right) = 1$$

$$\frac{\partial}{\partial \beta_1}\left(g(x^{(i)})\right) = x^{(i)} \cos\left(\beta_1 x^{(i)}\right)$$

$$\frac{\partial}{\partial \beta_2}\left(g(x^{(i)})\right) = -x^{(i)} \sin\left(\beta_2 x^{(i)}\right)$$

Substitute the 3 terms in above equation to get derivatives with β_j $\forall\ j = 0, 1, 2$