1 a)

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

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

For Critical point,

$$\frac{\partial}{\partial \theta}\left(RSS(\theta)\right) = 0$$

$$\Rightarrow -2X^T Y + 2X^T X\theta = 0$$

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

1 b)

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

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

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

For critical point,

$$\frac{\partial}{\partial \theta}\left(\text{Ridge}_\lambda(\theta)\right) = 0$$

$$\Rightarrow -2X^T Y + 2(X^T X + \lambda^2 I_n)\theta = 0$$

$$\theta = (X^T X + \lambda^2 I_n)^{-1} X^T Y$$

3a) Given,

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

$$= \underbrace{(Y - X\theta)^{T}}_{1\times n}\ \underbrace{W}_{n\times n}\ \underbrace{(Y - \overset{n\times(d+1)}{X}\underset{(d+1)\times 1}{\theta})}_{n\times 1}$$

Here,

$$W = \begin{bmatrix} w^{(1)} & & & 0 \\ & w^{(2)} & & \\ & & \ddots & \\ 0 & & & w^{(n)} \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta^{(0)} \\ \theta^{(1)} \\ \vdots \\ \theta^{(d)} \end{bmatrix}$$

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

$$J(\vec{\theta}; \vec{x}) = Y^T W Y - 2\theta^T X^T W Y + \theta^T X^T W X \theta$$

$\Rightarrow$

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

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

(b) $$\text{Hessian } (H) = \frac{\partial^2}{\partial \theta^2}\left(J(\vec{\theta}; \vec{x})\right)$$

$$= 2X^T W X$$

(c) Gradient descent

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

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

(d) Newton's method

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1} \times \left[\frac{\partial}{\partial\theta}\left(J(\vec{\theta}; \vec{x})\right)\right]^{(t)}$$

$$= \theta^{(t)} - \left(2X^TWX\right)^{-1}\left(2X^TW\left(X\theta^{(t)} - Y\right)\right)$$

$$= \theta^{(t)} - \theta^{(t)} + \left(X^TWX\right)^{-1} X^TWY$$

$$= \left(X^TWX\right)^{-1} X^TWY$$

closed form solution

4

1)

Given,

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

Input data matrix for "n" points,

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

These two columns are linearly independent as,

$$a \sin(x_i) + b\cos(x_i) = 0$$

$$\iff \sqrt{a^2+b^2}\left(\underbrace{\frac{a}{\sqrt{a^2+b^2}}}_{\cos(\alpha)}\sin(x_i) + \underbrace{\frac{b}{\sqrt{a^2+b^2}}}_{\sin(\alpha)}\cos(x_i)\right) = 0$$

$$\iff \sqrt{a^2+b^2}\,\sin(x_i+\alpha) = 0$$

$$\iff a = b = 0$$

$\therefore$ We can use least squares method

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

Using least squares,

we get,

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

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

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

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

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

$\longrightarrow$ (ii)

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

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

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

Substituting (iii), (iv) & (v) in (ii) and re-writing (i) gives,

$$\frac{\partial (RSS)}{\partial \beta} = \begin{bmatrix} \sum_{i=1}^{n} 2\left(g(x^{(i)}) - y^{(i)}\right) \times 1 \\ \\ \sum_{i=1}^{n} 2\left(g(x^{(i)}) - y^{(i)}\right) \times \left(x^{(i)} \cos\left(\beta_1 x^{(i)}\right)\right) \\ \\ \sum_{i=1}^{n} 2\left(g(x^{(i)}) - y^{(i)}\right) \times \left(-x^{(i)} \sin\left(\beta_2 x^{(i)}\right)\right) \end{bmatrix}$$

Look at Code for further analysis