MATH 7243 Homework 2

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

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & \cdots & x_{m,1} \\ \vdots & & \vdots \\ x_{1,n} & & x_{m,n} \end{bmatrix}_{m \times n}, \quad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}_{n \times 1}$$

$$X\theta = \begin{bmatrix} x_{1,1}\theta_1 + \cdots + x_{m,1}\theta_n & \cdots & x_{1,n}\theta_1 + \cdots + x_{m,n}\theta_n \end{bmatrix}$$

$$X\theta = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} x_{i,1}\theta_j \; \cdots \; \sum_{i=1}^{m} \sum_{j=1}^{n} x_{i,n}\theta_j \right]$$

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

$$\frac{d(RSS(\vec{\theta}))}{d\theta} = \frac{d}{d\theta} \sum_{i=1}^{n} \left(y^{(i)} - \sum_{j}^{m} x_{i,j}\theta_j \right)^2$$

$$= \sum_{i=1}^{n} 2 \left(y^{(i)} - \sum_{j=1}^{m} x_{i,j}\theta_j \right)(-x_{i,j})$$

$$= -(Y - X\theta)^T X = 0$$

$$= X^T (Y - X\theta) = 0$$

$$= X^T Y - X^T X\theta = 0$$

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

b) $$\vec{\theta}^{ridge} = \underset{\vec{\theta}}{\arg\min}\, J(\vec{\theta}^{ridge}) = \underset{\vec{\theta}}{\arg\min}\, RSS(\vec{\theta})$$

$$= (\vec{y} - X\vec{\theta})^T (\vec{y} - X\vec{\theta}) + \lambda^2 \vec{\theta}^T \vec{\theta}$$

$$\frac{d(\vec{\theta}^{ridge})}{d\theta} = -2(Y - X\theta)^T X + \lambda^2 \theta(2) = 0$$

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

$$\Rightarrow \theta^{ridge} = (X^T X + \lambda^2 I)^{-1} X^T Y = \boxed{(X^T X + \lambda^2 I)^{-1} X^T Y}$$

Q2)

$x^{(i)}$	1.2	3.2	5.1	3.5	2.6
$y^{(i)}$	7.8	1.2	6.4	2.6	8.1

$$X = \begin{bmatrix} 1 & 1.2 \\ 1 & 3.2 \\ 1 & 5.1 \\ 1 & 3.5 \\ 1 & 2.6 \end{bmatrix}, \quad X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.2 & 3.2 & 5.1 & 3.5 & 2.6 \end{bmatrix}$$

$$Y = \begin{bmatrix} 7.8 \\ 1.2 \\ 6.4 \\ 2.6 \\ 8.1 \end{bmatrix}$$

a) Using Critical point:

$$\theta = (X^T X)^{-1} X^T Y : \quad \overset{X^T}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.2 & 3.2 & 5.1 & 3.5 & 2.6 \end{bmatrix}} \overset{Y}{\begin{bmatrix} 7.8 \\ 1.2 \\ 6.4 \\ 2.6 \\ 8.1 \end{bmatrix}}$$

$$X^T Y = \begin{bmatrix} 26.1 \\ 76 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 5 & 15.6 \\ 15.6 & 56.7 \end{pmatrix}^{-1} = \begin{bmatrix} 1.412556 & -0.38864 \\ -0.38864 & 0.124564 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 1.412556 & -0.38864 \\ -0.38864 & 0.124564 \end{bmatrix} \begin{bmatrix} 26.1 \\ 76 \end{bmatrix} = \begin{bmatrix} 7.3310912 \\ -0.67663 \end{bmatrix}$$

$$Y = X\theta, \quad y = 7.3310912 - 0.67663x \quad — (1)$$

b) Using ridge:

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

$$\left(X^T X + \begin{pmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{pmatrix}\right)^{-1} \begin{pmatrix} 26.1 \\ 76 \end{pmatrix}$$

When $\lambda = 1$:

$$\left(\begin{pmatrix} 5 & 15.6 \\ 15.6 & 56.7 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 26.1 \\ 76 \end{pmatrix} = \begin{pmatrix} 3.115228 \\ 0.474912 \end{pmatrix}$$

$$y = 3.115228 + 0.474912x \quad - ②$$

When $\lambda = 10$

$$\left(\begin{pmatrix} 5 & 15.6 \\ 15.6 & 56.7 \end{pmatrix} + \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix} \right)^{-1} \begin{pmatrix} 26.1 \\ 76 \end{pmatrix} = \begin{pmatrix} 0.179164 \\ 0.467167 \end{pmatrix}$$

$$y = 0.179164 + 0.467167x \quad - ③$$

After adding the penalty term and reducing the variance, the ② & ③ equations show an upward trend with positive θ_1 unlike ①. As λ increases value of y decreases since the parameters are shrinking!
It is hard to tell with such less data which one of the three equations is the best fit. but ② is probably the most reasonable.

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

$$w^{(i)} = \exp\left(-\frac{\|\vec{x}^{(i)} - \vec{x}\|^2}{2\tau^2}\right)$$

a) $$\frac{d}{d\theta_j} J(\vec{\theta}; \vec{x}) = 2\left[\sum_{i=1}^{n} w^{(i)} \left(\vec{\theta}^T \vec{x}^{(i)} - y^{(i)}\right) \cdot x_j^{(i)}\right] = \nabla J(\vec{\theta})$$

b) $$\text{Hessian} = \frac{d^2}{d\theta_j^2} J(\vec{\theta}; \vec{x}) = 2\left[\sum_{i=1}^{n} w^{(i)} \left(x_j^{(i)}\right)\left(x_k^{(i)}\right)\right] = \nabla^2 J(\vec{\theta})$$

c) $\vec{\theta}^{next} = \vec{\theta} - \alpha \nabla J(\vec{\theta})$; $\alpha = \eta$ = step size

$$\vec{\theta}_{K+1} = \vec{\theta}_K - \eta_K \nabla J(\vec{\theta})$$

d) $$\vec{\theta}_{K+1} = \theta_K - \frac{\nabla J(\vec{\theta})}{\nabla^2 J(\vec{\theta})} \Rightarrow \vec{\theta}_{K+1} = \theta_K - H^{-1} \cdot \nabla J(\vec{\theta})$$

$\vec{\theta}_{K+1} = \vec{\theta}_K - \eta_K H^{-1} \cdot \nabla J(\vec{\theta})$; H^{-1} = Inverse Hessian,
$H = \nabla^2 J(\vec{\theta})$

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

$$Y = X\beta = \begin{bmatrix} 1 & \sin(x_1) & \cos(x_1) \\ \vdots & \vdots & \vdots \\ 1 & \sin(x_n) & \cos(x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

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

we can solve using least squares method;

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

b) $RSS = \sum_{i=1}^{n} \left(y^{(i)} - g(x^{(i)}) \right)^2 = J(\vec{\beta})$;

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

$$\frac{d}{d\beta_0} J(\vec{\beta}) = 2 \left[\sum_{i=1}^{n} \left(y^{(i)} - g(x^{(i)}) \right) \right] (-1)$$

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

$$\frac{d}{d\beta_1} J(\vec{\beta}) = 2 \sum_{i=1}^{n} \left(y^{(i)} - g(x^{(i)}) \right) \left(-\cos(\beta_1 x^{(i)}) \cdot x^{(i)} \right)$$

$$\frac{d}{d\beta_2} J(\vec{\beta}) = 2 \sum_{i=1}^{n} \left(y^{(i)} - g(x^{(i)}) \right) \left(\sin(\beta_2 x^{(i)}) \cdot x^{(i)} \right)$$

$$\vec{\beta}^{next} = \beta - \eta_k \nabla J \Rightarrow \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$