Q1 # $X = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}$ # $Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$

$X^* = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$

$Y = \begin{bmatrix} 6 \\ 10 \\ 6 \end{bmatrix}$

~~Cost~~
Cost $= (\theta_0 + \theta_1 - 6)^2 + (\theta_0 + 3\theta_1 - 10)^2 + (\theta_0 + 6\theta_1 - 16)^2$

$= \theta_0^2 + \theta_1^2 + 36 - 12\theta_1 - 12\theta_0 + \theta_0^2 + 9\theta_1^2 + 100 - 60\theta_1 - 20\theta_0 + 6\theta_0\theta_1 + \theta_0^2 + 36\theta_1^2 + 256 + 12\theta_0\theta_1 - 192\theta_1 - 32\theta_0$

Cost $= 3\theta_0^2 + 46\theta_1^2 + 392 - 264\theta_1 - 64\theta_0 + 18\theta_0\theta_1$

(1) Now fix $\theta_0 = 0$

cost $= 46\theta_1^2 + 392 - 264\theta_1$

$\frac{dC}{d\theta_1} = 92\theta_1 - 264 = 0 \Rightarrow \theta_1 = 2.87$

(2) Now fix $\theta_1 = 2.87$

Cost $= 3\theta_0^2 + 46(2.87)^2 + 392 - 264(2.87) - 64\theta_0 + 18\theta_0 \times 2.87$

~~$d\theta$~~
$\frac{dC}{d\theta} = 6\theta_0 - 64 + 18 \times 2.87 = 0$

$= \theta_0 = ?$

① Now fix $\theta_0 = 2$

$$\frac{dc}{d\theta_1} = 92\theta_1 - 264 + 18\theta_0 = 0$$

$= 92\theta_1 - 264 + 36 = 0$

$= 92\theta_1 = 264 - 36$

\# $\theta_1 = 2.47$.

Q1(b) # Stochastic gradient descent with initial value of θ_0 & θ_1 & a learning rate of 0.01

① # First fix $\theta_0 = 0$ & $\theta_1 = 0$

~~Now the~~ we take first training sample.

\# $C = \cancel{\theta_0 + 6\theta} \ (\theta_0 + \theta_1 - 6)^2$

→ # $\frac{dc}{d\theta_0} = 2(\theta_0 + \theta_1 - 6) = -12$

→ # $\frac{dc}{d\theta_1} = 2(\theta_0 + \theta_1 - 6) = -12$

$$(\theta_0, \theta_1)_{new} = (\theta_0, \theta_1)_{old} - \alpha \left(\frac{dc}{d\theta_0}, \frac{dc}{d\theta_1} \right)$$

$(\theta_0, \theta_1) \Rightarrow (0.12, 0.12)$

② $\theta_0 = 0.12$ & $\theta_1 = 0.12$

\# $C = (\theta_0 + 3\theta_1 - 10)^2$

\# $\frac{dc}{d\theta_0} = 2(\theta_0 + 3\theta_1 - 10) = 2(0.48 - 10) = -19.04$

\# $\frac{dc}{d\theta_1} = 6(\theta_0 + 3\theta_1 - 10) = 6(0.48 - 10) = -57.12$

Scanned by CamScanner

$$(\theta_0, \theta_1) = (0.12,\ 0.12) - 0.01(-19.04,\ -57.12)$$
$$= (0.12,\ 0.12) + (0.1904,\ 0.5712)$$
$$= (0.3104,\ 0.6912)$$

Iteration 3

$\theta_0 = 0.3104$, $\theta_1 = 0.6912$

$(\cancel{\theta_0, \theta_1})$

$$C = (\cancel{\theta_0 + 3\theta_1}(\theta_0 + 6\theta_1 - 16)^2$$

$\frac{\partial C}{\partial \theta_0} = 2(\theta_0 + 6\theta_1 - 16)$ $\quad \frac{\partial C}{\partial \theta_1} = 12(\theta_0 + 6\theta_1 - 16)$

$$-23.08$$

$\frac{\partial C}{\partial \theta_0} = \cancel{8.9152}$ $\quad \frac{\partial C}{\partial \theta_1} = 1\cancel{8}(-138.5088$

$$(\theta_0, \theta_1) = (0.3104,\ 0.6912) - 0.01(-23.08,\ -138.5088)$$
$$= (0.5412,\ 2.07)$$

The final value of θ_0 & θ_1 is 0.5412 & 2.07

θ_0

Q (c)

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

$$= [\ \ ] \quad \# \ X = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \quad \# \ Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 10 \\ 10 & 46 \end{bmatrix}$$

$X^T X + I = \begin{bmatrix} 4 & 10 \\ 10 & 47 \end{bmatrix}$

Scanned by CamScanner

$$\theta_{opt} = \begin{bmatrix} 4 & 10 \\ 10 & 47 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix} \right)$$

$$= \frac{1}{88} \begin{bmatrix} 47 & -10 \\ -10 & 4 \end{bmatrix} \begin{bmatrix} 32 \\ 132 \end{bmatrix}$$

$$= \frac{1}{88} \begin{bmatrix} 184 \\ 208 \end{bmatrix} = \begin{bmatrix} 2.09 \\ 2.36 \end{bmatrix}$$

The optimal value obtained by Ridge regression is $\theta_0 = 2.09$ & $\theta_1 = 2.36$

Q5(c)

The solution obtained by lasso regression is sparse compared to the one obtained by Ridge regression.

Q4 (b)

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

$$= \left[\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{bmatrix} + \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right]^{-1} X^T Y$$

$$\Rightarrow \left[\begin{bmatrix} 30 & 60 \\ 60 & 120 \end{bmatrix} + \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right]^{-1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

$$\theta_{opt} = \begin{bmatrix} 31 & 60 \\ 60 & 121 \end{bmatrix}^{-1} \begin{bmatrix} 40 \\ 80 \end{bmatrix}$$

$$\theta_{opt} = \begin{bmatrix} 121 & -60 \\ -60 & 31 \end{bmatrix}\begin{bmatrix} 40 \\ 80 \end{bmatrix} \times \frac{1}{150}$$

$$\Rightarrow \begin{bmatrix} 4840 - 4800 \\ -2400 + 2460 \end{bmatrix} \times \frac{1}{150} = \cancel{\begin{bmatrix} 40 \\ -80 \end{bmatrix} \frac{1}{150}}$$

Scanned by CamScanner

$\theta_{opt} = \begin{bmatrix} 0.26 \\ +0.53 \end{bmatrix}$

Yes we can obtain a solution here because of the regularization imposed on variable θ