Q1
(a)

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

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

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

$$= \begin{bmatrix} 3 & 10 \\ 10 & 46 \end{bmatrix}^{-1} \begin{bmatrix} 32 \\ 132 \end{bmatrix}$$

$$= \frac{1}{38} \begin{bmatrix} +46 & -10 \\ -10 & 3 \end{bmatrix} \begin{bmatrix} 32 \\ 132 \end{bmatrix}$$

$$= \frac{1}{38} \begin{bmatrix} 46 \times 32 - 1320 \\ -320 + 396 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$\theta_0 = 4, \quad \theta_1 = 2$

(b) $(\theta_0, \theta_1) = (0, 0)$

$E = (\theta_0 + \theta_1 - 6)^2 + (\theta_0 + 3\theta_1 - 10)^2 + [(\theta_0 + 6\theta_1) - 16]^2$

$\nabla E_0 = \frac{dE}{d\theta_0} = 2(\theta_0 + \theta_1 - 6) + 2(\theta_0 + 3\theta_1 - 16) + 2(\theta_0 + 6\theta_1 - 16)$

$\frac{dE}{d\theta_1} = 2(\theta_0 + \theta_1 - 6) + 6(\theta_0 + 3\theta_1 - 10) + 12(\theta_0 + 6\theta_1 - 16)$

$\frac{dE}{d\theta_2}$ at $(\theta_0, \theta_1) = (0, 0)$

$\frac{dE}{d\theta_0} = 2(-6) + 2(-10) + 2(-16) = -64$

$\frac{dE}{d\theta_1} = 2(-6) + 6(-10) + 12(-16) = -264$

$(\theta_0, \theta_1)_{new} = (\theta_0, \theta_1)_{old} - 0.1(-64, -264)$

$(\theta_0, \theta_1)_{new} = (6.4, 26.4)$

Iteration 2 :

$\frac{dE}{d\Theta_0} = 2[3\Theta_0 + 10\Theta_1 - 32] = 6\Theta_0 + 20\Theta_1 - 64$

$\frac{dE}{d\Theta_1} = 2[\Theta_0 + \Theta_1 - 6 + 3\Theta_0 + 9\Theta_1 - 30 + 6\Theta_0 + 36\Theta_1 - 96]$

$= 2[10\Theta_0 + 46\Theta_1 - 132]$

$= 20\Theta_0 + 92\Theta_1 - 264$

$\frac{dE}{d\Theta_0} = 6 \times 6.4 + 20 \times 26.4 - 64 = 502.4$

$\frac{dE}{d\Theta_1} = 20 \times 6.4 + 92 \times 26.4 - 264 = 2292.8$

$(\Theta_0, \Theta_1)_n = (6.4, 26.4) - 0.1(502.4, 2292.8)$

$(\Theta_0, \Theta_1) = (6.4, 26.4) - (50.24, 229.28)$

$= (-43.84, -202.88)$

Iteration 3 :

$\frac{dE}{d\Theta_0} = 6 \times -43.84 + 20(-202.88) - 64 = -4384.$

$\frac{dE}{d\Theta_1} = \cancel{20 \times -202.8 + 92}$

$20 \times (-43.84) + 92(-202.88) - 264 = -198$

$\frac{dE}{d\Theta_0}(\Theta_0, \Theta_1) = (-43.84, -202.88) - 0.1(-4384.64,$

$= (-43.84, -202.88) + (438.464, 198$

$= (394.624, 1777.696)$

Iteration 4 :

$\frac{dE}{d\theta_0} - 6 \times 394.624 + 1777.696 \times 20 - 64 = 37857.$

$\frac{dE}{d\theta_1} = 20 \times 394.624 + 92 \times 1777.696 - 264 = 171176.5$

$(\theta_0, \theta_1) = (394.624, 1777.696) - 0.1(37857.664, 171176.5$

$= (-3391.1424, -15339.9552)$

Iteration 5:

$\frac{dE}{d\theta_0} = 6 \times -3391.1424 + -15339.9552 \times 20 - 64$

$= -713800.192$

$\frac{dE}{d\theta_1} = 20 \times -3391.1424 + 92 \times -15339.9552 + -264$

$= 1e4(67822.848 + 1411275.8784 + 264)$

$= -1479422.7264$

$(\theta_0, \theta_1) = (-3391.1424, -15339.9552) + (71380.0192, 147942.27264)$

$= (67988.8768, 132,602.31744)$

$\boxed{\text{The } (\theta_0, \theta_1) \text{ does not converge after 5 iterations}}$

Page

$$\boxed{y = \theta_1 x + \theta_0}$$

Q1(c) $\theta_1 = \dfrac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}$

$$\mathrm{cov}(X,Y) = \frac{E[(X-\bar{X})(Y-\bar{Y})]}{\cancel{\mathrm{var}}\ \mathrm{var}(X)}$$

$$= \left[\left(1-\frac{10}{3}\right)\left(6-\frac{32}{3}\right) + \left(3-\frac{10}{3}\right)\left(10-\frac{32}{3}\right) + \left(6-\frac{10}{3}\right)\left(16-\frac{32}{3}\right)\right] / \mathrm{var}(X)$$

$$= \frac{\frac{98}{9} + \frac{2}{9} + \frac{128}{9}}{\left(\frac{10}{3}-1\right)^2 + \left(\frac{10}{3}-3\right)^2 + \left(\frac{10}{3}-6\right)^2}$$

$$= \frac{\left(\frac{228}{9}\right)}{\frac{49}{9} + \frac{1}{9} + \frac{64}{9}} = 2$$

$\cancel{\theta_2}$ $\theta_0 = E(Y) - \theta_1 E(X)$

$$= \frac{32}{3} - \frac{2 \times 10}{3}$$

$$= 4$$

~~Q2~~

~~Q2(a)~~

Q2(b) $\theta_{opt} = (X^T X)^{-1} X^T Y$

$$(X^T X)^{-1} = \left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{bmatrix} \right)^{-1}$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{pmatrix} = \begin{bmatrix} 30 & 60 \\ 60 & 120 \end{bmatrix}$$

Scanned by CamScanner

$\det(X^T X) = 0$

The matrix $X^T X$ is non-invertible.

The Scikit learn implements the pseudo inverse of $X^T X$ in order to obtain the solution.

classmate

Q3(b) No we cannot comment on the importance of different features, as an important feature may have ~~small~~ larger values & hence smaller coefficients, an a unimportant feature may have smaller values hence larger coefficients

Q3(c) Yes, now we can say that 'X6 longitude' is the most important feature

Q3(d) The distribution of the residuals is normal in nature.

Q3 (e) (i) The optimal feature size is 4

(ii) The greedy approach shows better performance than exhaustive search.