1) $X = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}$ $\qquad Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$

$$Y = X\theta + \vec{\varepsilon}$$

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

$\cancel{X^T X = [1 \; 3 \; 6]} \begin{bmatrix} 1 \\ \; \end{bmatrix}$

Appending columns of 1 to X,

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

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

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

$$(X^T X)^{-1} = \frac{1}{(46 \times 3 - 100)} \begin{bmatrix} 46 & -10 \\ -10 & 3 \end{bmatrix} = \frac{1}{38} \begin{bmatrix} 46 & -10 \\ -10 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1.21 & -0.26 \\ -0.26 & 0.07 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix} = \begin{bmatrix} 32 \\ 132 \end{bmatrix}$$

$$\theta = (X^T X)^{-1} X^T Y = \begin{bmatrix} 1.211 & -0.263 \\ -0.263 & 0.079 \end{bmatrix} \begin{bmatrix} 32 \\ 132 \end{bmatrix} = \begin{bmatrix} 4.036 \\ 2.012 \end{bmatrix}$$

Gradient Descent

$$\text{Error}(\theta) = \cancel{(Y_i - X_i\theta)^2} \sum_{i=1}^{N} \left(y_i - (\theta_0 + \theta_1 x_i)\right)^2$$

$N = \#$ of Training examples

$$\text{Error}(\theta) = \left(6 - (\theta_0 + \theta_1)\right)^2 + \left(10 - (\theta_0 + 3\theta_1)\right)^2 + \left(16 - (\theta_0 + 6\theta_1)\right)^2$$

$$= 36 + (\theta_0^2 + \theta_1^2 + 2\theta_0\theta_1) - 12(\theta_0 + \theta_1)$$
$$+ 100 + (\theta_0^2 + 9\theta_1^2 + 6\theta_0\theta_1) - 20(\theta_0 + 3\theta_1)$$
$$+ 256 + (\theta_0^2 + 36\theta_1^2 + 12\theta_0\theta_1)$$
$$- 32(\theta_0 + 6\theta_1)$$

-72, -192, 264 ; $32 \times 6 = 192$

$$\text{Error} = 392 + 3\theta_0^2 + 46\theta_1^2 + 20\theta_0\theta_1 - 64\theta_0 - 264\theta_1$$

$$\theta_{0\,new} = \theta_{0\,old} - \alpha \frac{\partial \text{Error}}{\partial \theta_0} \qquad \theta_{1\,new} = \theta_{1\,old} - \alpha \frac{\partial \text{Error}}{\partial \theta_1}$$

$$\frac{\partial \text{Error}}{\partial \theta_0} = 6\theta_0 + 20\theta_1 - 64$$

$$\frac{\partial \text{Error}}{\partial \theta_1} = 92\theta_1 + 20\theta_0 - 264$$

Itⁿ 1:

θ_0 new $= 0 - 0.1(-64) = 6.4$

θ_1 new $= 0 - 0.1(-264) = 26.4$

Itⁿ 2:

θ_0 new $= 6.4 - 0.1 \times (6 \times 6.4 + 20 \times 26.4 - 64) = \cancel{50.2}$ -43.84

θ_1 new $= 26.4 - 0.1(92 \times 26.4 + 20 \times 6.4 - 264) = -177.28$

Itⁿ 3:

θ_0 new $= -43.84 - 0.1(6 \times (-43.84) + 20 \times (-177.28) - 64)$

$= 343.424$

$\theta_1 new = 1777.696$

It^n 4 $\theta_0 new = -3391.142$

$\theta_1 new = -15339.95$

It^n 5 $\theta_0 new = 29329.85344$

$\theta_1 new = 132596.31744$

By Covariance and Variance

$\bar{X} = \frac{10}{3}, \bar{Y} = \frac{32}{3}$

$\theta_1 = \frac{Cov(X,Y)}{Var(X)} =$

$Cov(X,Y) = \frac{1}{3} \times \left(1 - \frac{10}{3}\right)\left(3 - \frac{10}{3}\right)\left(6 - \frac{10}{3}\right)$

$= \frac{1}{3}\left(\frac{-7}{3}\right)\left(\frac{-1}{3}\right)\left(\frac{8}{3}\right) = \frac{56}{81}$

$Var(X) = E[X^2] - E[X]^2$

$E[X]^2 = \frac{100}{9}$

$E[X^2] = \frac{1^2+3^2+6^2}{3} = \frac{46}{3} = \cancel{\frac{100}{9}}$

$= \frac{138 - 100}{9} = \frac{38}{9} = \frac{38}{9} = 4.5$

$\cancel{\theta_1 = \frac{56 \times 9}{81 \times 38}}$

$Cov(X,Y) = \frac{1}{3} \times \left[\left(1 - \frac{10}{3}\right)\left(6 - \frac{32}{3}\right) + \left(3 - \frac{10}{3}\right)\left(10 - \frac{32}{3}\right) + \left(6 - \frac{10}{3}\right)\left(16 - \frac{32}{3}\right)\right]$

$= \frac{1}{3}\left[\frac{-7 \times (-14)}{9} + \frac{(-1)(-2)}{9} + \frac{(8)(16)}{9}\right]$

$= \frac{98 + 2 + 128}{27} = \frac{228}{27}$

$\theta_1 = \frac{\cancel{228}^{114} \times \cancel{9}}{3\cancel{27} \times \cancel{38}_{19}} = 2$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$= \frac{32}{3} - \frac{2 \times 10}{3} = \frac{12}{3} = 4.$$