1(a) Coordinate Descent

$\begin{bmatrix}1\\3\\6\end{bmatrix}$ $\begin{bmatrix}6\\10\\16\end{bmatrix}$ $4+2\times1$, $4+2\times3$, $4+2\times6$

$$\text{Error} = \sum_i (y_i - (\theta_0 + \theta_1 x_i))^2$$

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

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

$$= 392 + 3\theta_0^2 + 46\theta_1^2 + 20\theta_0\theta_1 - 64\theta_0 - 264\theta_1$$

Coordinate descent.

I1 $\text{Error}(0,\theta_0) = 392 + 3\theta_0^2 - 64\theta_0$

$\frac{\partial \varepsilon}{\partial \theta_0} = 0 \Rightarrow 6\theta_0 = 64, \quad \theta_0 = \frac{32}{3}$.

$\varepsilon(\theta_1, 0) = 392 + 46\theta_1^2 - 264\theta_1 + 20 \times \frac{32}{3} \times \theta_1$

$\frac{\partial \varepsilon}{\partial \theta_1} = 92\theta_1 - 264 \quad \theta_1 = \frac{66}{23}$

$\frac{640}{3} = 213.3$

$\frac{\partial \varepsilon}{\partial \theta_1} = 0 \Rightarrow 92\theta_1 = 264 - \frac{320 \times 2}{3}$

$\theta_1 = \frac{264 - 213.3}{92} = \frac{51}{92} \simeq 0.55$

$(\theta_0, \theta_1) = (10.67, 0.55)$

I2 $\varepsilon = 392 + 3\theta_0^2 + 46\theta_1^2 + 20\theta_0\theta_1 - 64\theta_0 - 264\theta_1$

$\varepsilon = 392 + 3\theta_0^2 + 46 \times 0.55^2 + 20 \times \theta_0 \times 0.55 - 64\theta_0 - 264 \times 0.55$

$\frac{\partial \varepsilon}{\partial \theta_0} = 0 \Rightarrow 6\theta_0 + 20 \times 0.55 \times \theta_0 = 64\theta_0$

$\theta_0 = \frac{64 - \frac{51}{92} \times 20}{6} = \frac{64 - 11.5}{6}$

$= \frac{52.5}{6} \simeq 8.83$

$\frac{51}{23}$

23 | 57 | 2.21

46

50

46

4

$\frac{\partial \varepsilon}{\partial \theta_1}, \theta_0 = 8.83 \Rightarrow 392 + 3(8.83)^2 + 46\theta_1^2 + 20 \times 8.83 \times \theta_1 - 64 \times 8.83 - 264\theta_1$

$\Rightarrow 92\theta_1 + 20 \times 8.83 = 264 \Rightarrow \theta_1 = \frac{87.4}{92} = 0.95$

classmate
Date
Page

3) $\frac{\partial E}{\partial \theta_0} \Rightarrow 392 + 3\theta_0^2 + 46 \times 0.95^2 + 20 \times \theta_0 \times 0.95 - 64 \times \theta_0 - 264 \times 0.95$

$(\theta_1 = 0.95) \Rightarrow \frac{\partial E}{\partial \theta_0} = 6\theta_0 + 20 \times 0.95 = 64$

$6\theta_0 = 45 \Rightarrow \theta_0 = \frac{45}{6} = 7.5$

$\frac{\partial E}{\partial \theta_1}, \theta_0 = 7.5 \Rightarrow 392 + 3(7.5)^2 + 46 \times \theta_1^2 + 20 \times 7.5 \times \theta_1 - 64 \times 7.5 - 264\theta_1 = 0$

$\Rightarrow 30 \times \frac{\partial E}{\partial \theta_1} \Rightarrow 92\theta_1 + 150 - 264 = 0$

$\Rightarrow 92\theta_1 = 114$

$\theta_1 = \frac{57}{46} \approx 1.24.$

$\therefore \quad \theta_0 = 7.5, \quad \theta_1 = 1.24.$

$\theta_i = \theta_i - \alpha \frac{\partial}{\partial \theta_i} \varepsilon^2$

$= -\alpha \frac{\partial}{\partial \theta_i} 2\varepsilon$

1 6
3 10
6 16

SGD

It 1

$\theta_0 = \theta_0 - 0.01(6-(0))$

$\theta_0 = 0 + 0.01(2)(6-(0.+0.))(1)$

$\theta_0 = 0 + 0.02 \times (6) = 0.12$

~~$\theta_1 = 0 - 0.01 \times (10 - (\theta_0 + \theta_1 3))$~~

$\theta_1 = \theta_1(0) - 0.01(6-0) \times 2(-1)$

$\Rightarrow \theta_1 = 0.12$

3 1
9.52
6
57 1 2

It 2

$\theta_0 = 0.12 - 2 \times (10 - (0.12 + 0.12 \times 3) \times (-1)$

$= 0.12 + 2 \times (10 - (0.48)) \times 0.01$

$= 0.12 + 2 \times 9.52 \times 0.01$

$= 0.12 + 0.1904 = 0.3104$

$\theta_1 = 0.12 - 2 \times (10 - (0.12 \times 0.12 \times 3) \times (-1)$

$= 0.12 + 2 \times 9.52 \times 0.01 \times 3$

$= 0.12 + 57.12 \times 0.01$

$= 0.69$

It 3

$\theta_0 = 0.31 - 2 \times (16 - (0.31 * 0.69 \times 6))(-1)$ x0.01

$= 0.31 - 0.02 \times (16 - (5.34 + 0.31))$

$= 0.31 - 0.02 \times (10.34)$

$= 0.540992$

$\theta_1 = 0.69 - 2 \times 6 \times (-1) \times 0.01$

$\times (0.31 + 0.69 \times 6)$

$= 0.69 + 12 \times 0.01 \times (5.6)$

$= 0.69 + 1.386$

$= 2.076$

(c) Normal Equation

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

$$= \begin{bmatrix} 3 & 10 \\ 10 & 46 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\theta = \left(X^T X + \delta^2 I\right)^{-1} X^T y \qquad = \begin{bmatrix} 4 & 10 \\ 10 & 47 \end{bmatrix}$$

$$M = X^T X + \delta^2 I = \begin{bmatrix} 4 & 10 \\ 10 & 47 \end{bmatrix}$$

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

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

$$\Rightarrow \frac{1}{86} \begin{bmatrix} 47 & -10 \\ -10 & 4 \end{bmatrix} \begin{bmatrix} 32 \\ 132 \end{bmatrix}$$

$$= \frac{1}{88} \begin{bmatrix} 47 \times 32 - 132 \times 10 \\ -320 + 132 \times 4 \end{bmatrix} = \begin{bmatrix} 184/88 \\ 208/88 \end{bmatrix}$$

$$= \begin{bmatrix} 2.09 \\ 2.36 \end{bmatrix}$$