2. output = $(0.001, 0.9, 0.001, 0.098)^T$ 이고 label = $(0,0,0,1)^T$ 일때 MSE, CE, LL 구하라

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - t_i)^2 \qquad \frac{(0.001-0)^2+(0.9-0)^2+(0.001-0)^2+(0.098-1)^2}{4} = 0.4059$$

$$CEE = -\sum_{i=1}^{n} \ln(y_i) \times t_i \qquad -(\log(0.001)\times 0 + \log(0.9)\times 0 + \log(0.001)\times 0 + \log(0.098)\times 1) = 2.3227$$

$$LL = \Sigma[t_i \ln y_i + (1-t_i)\ln(y_i)]$$

$$= (0\cdot\ln 0.001 + 1\times\ln 0.001) + (0\times\ln 0.9 + 1\times\ln 0.9) + (0\times\ln 0.001 + 1\times\ln 0.001) + (1\times\ln 0.098 + 0\times\ln 0.098)$$

$$= -16.2436$$

※ y_i = predict t_i = target

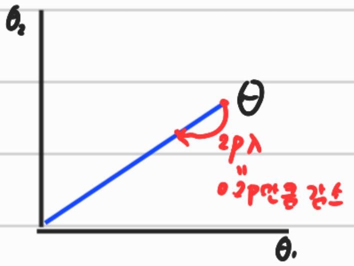
3. Ridge Regression 에서 $\lambda=0.1$, $\lambda=0.5$ 일때 계산하고 λ에 따른 효과 설명

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 3 \end{pmatrix}, \quad y = \begin{pmatrix} 3.0 \\ 7.0 \\ 8.8 \end{pmatrix} \qquad \hat{\omega} = (X^TX + \lambda I)^{-1} X^T Y$$

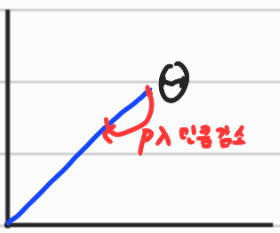
$\lambda=0.1$ 일때
$$\hat{\omega} = \left(\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 3 \end{pmatrix} + \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}\right)^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}\begin{pmatrix} 3 \\ 7 \\ 8.8 \end{pmatrix}$$
$$= \left(\begin{pmatrix} 14 & 16 \\ 16 & 19 \end{pmatrix} + \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}\right)^{-1}\begin{pmatrix} 43.4 \\ 50.4 \end{pmatrix} = \begin{pmatrix} 1.6154 \\ 1.2788 \end{pmatrix} \Rightarrow y = 1.6154x_1 + 1.2788x_2$$

$\lambda=0.5$ 일때
$$\hat{\omega} = \left(\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}\begin{pmatrix} 3 \\ 7 \\ 8.8 \end{pmatrix}$$
$$= \begin{pmatrix} 14 & 16 \\ 16 & 19 \end{pmatrix}^{-1}\begin{pmatrix} 43.4 \\ 50.4 \end{pmatrix} = \begin{pmatrix} 1.4 \\ 1.4 \end{pmatrix} \qquad y = 1.4x_1 + 1.4x_2$$

$\lambda=0.1$일때



$\lambda=0.5$



λ가 클수록 최종해를 원점가까이 당겨서
가중치를 더 작게 유지하게 규제를 가한다.

5. 혈압, 키, 몸무게 특징벡터 이룸

$$\begin{pmatrix} 121 \\ 1.72 \\ 69.0 \end{pmatrix}, \begin{pmatrix} 140 \\ 1.62 \\ 63.2 \end{pmatrix}, \begin{pmatrix} 120 \\ 1.70 \\ 59.0 \end{pmatrix}, \begin{pmatrix} 131 \\ 1.8 \\ 82 \end{pmatrix}, \begin{pmatrix} 101 \\ 1.78 \\ 73.5 \end{pmatrix}$$

5-1 weight vector = $(-0.01, 0.5, -0.23)$, bias = 0

① gradient $g_1 = -(-0.01)\times 121 - (-0.01)\times 1.72 - (-0.01)\times 69 \doteq 1.917$

$g_2 = -0.5\times 121 - 0.5\times 1.72 - 0.5\times 69 = -95.86$

$\frac{\partial J}{\partial u_{ij}} = -\delta_i^{t} z_j^{t-1}$ $\quad g_3 = -(-0.23)\times 121 - (-0.23)\times 1.72 - (-0.23)\times 69 = 44.09$

⇒ 가중치가 음수인 것은 모두 양수로 가중치가 양수인 것은 음수로 update, update의 크기 차이가 심함

② $g_1 = 0.01 \times 140 + 0.01 \times 1.62 + 0.01 \times 63.2 = 2.048$

$g_2 = -0.5 \times 140 - 0.5 \times 1.62 - 0.5 \times 63.2 = -102.41$

$g_3 = 0.23 \times 140 + 0.23 \times 1.62 + 0.23 \times 63.2 = 47.1086$

=> ①번과 동일하게 가중치음수는 모두양수로 가중치 양수는 모두 음수로 update

∴ 가중치가 뭉치로 update되어 수렴 속도의 저하

\+ g_1의 update와 g_2의 update 크기차이가 크게 난다. 따라서 느리게 학습됨! 규모의문제 발생

5.2 전처리 적용한 후의 훈련 집합은?

혈압의 평균: $\frac{121+140+120+131+101}{5} = 122.6$

$$\frac{x_i - \mu_i}{\sigma_i}$$

키의평균(m): $\frac{1.72+1.62+1.7+1.8+1.78}{5} = 1.724$

몸무게의평균: $\frac{69+63.2+59+82+73.5}{5} = 69.38$

혈압의 표준편차 : 13.03 키의표준편차 : 0.06 몸무게의표준편차 : 8.02

$$\therefore \begin{pmatrix} -0.123 \\ -0.06 \\ -0.04 \end{pmatrix} \begin{pmatrix} 1.34 \\ -1.63 \\ -0.76 \end{pmatrix} \begin{pmatrix} -0.2 \\ -0.38 \\ -1.29 \end{pmatrix} \begin{pmatrix} -1.66 \\ 0.88 \\ 0.52 \end{pmatrix}$$

5.3 규모문제가 완화되는가?

가중치 : $(-0.01\ 0.5\ -0.23)^T$, bias = 0 이라고 동일하게 가정

① $g_1 = -(-0.01) \times (-0.123) - (-0.01) \times (-0.06) - (-0.01) \times (-0.04) = -0.0023$

$g_2 = -0.5 \times (-0.123 - 0.06 - 0.04) = 0.1115$

$g_3 = -(-0.23) \times (-0.123 - 0.06 - 0.04) = -0.05129$

② $g_1 = -(-0.01)(1.34 - 1.63 - 0.76) = -0.01$

$g_2 = -0.5(1.34 - 1.63 - 0.76) = 0.524$

$g_3 = -(-0.23)(1.34 - 1.63 - 0.76) = -0.24$

g_1, g_2, g_3를 비교하였을때 gradient의 크기차이가 크지 않고 비슷한 크기로 update 된다. 따라서 전처리로 인해 규모의 문제는 완화되었다.