Sheet 6) Exercise 1)

Consider the following training data

$T = \{([-1,0]^T, 3), ([1,3]^T, -1), ([-2,1]^T, 0), ([0,4]^T, -2)\}$

Compute the predictor from linear regression by least squares. Use batch gradient descent to compute the coefficient vector β.

Choose the initial guess $\beta^0 = [0,0,1]^T$

Calculate the first two steps of batch gradient descent with learning rate $\eta = 0.25$

Solution :

$$X = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 3 \\ 1 & -2 & 1 \\ 1 & 0 & 4 \end{bmatrix} \Rightarrow X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -2 & 0 \\ 0 & 3 & 1 & 4 \end{bmatrix}, \quad y = \begin{bmatrix} 3 \\ -1 \\ 0 \\ -2 \end{bmatrix}$$

$$\beta^{(1)} = \beta^{(0)} + 2\eta X^T (y - X\beta^{(0)})$$

$$\beta^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + 2 \cdot (0.25) \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -2 & 0 \\ 0 & 3 & 1 & 4 \end{bmatrix} \cdot \left(\begin{bmatrix} 3 \\ -1 \\ 0 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 3 \\ 1 & -2 & 1 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} -4 \\ -2.5 \\ -17.5 \end{bmatrix}$$

$$\beta^{(2)} = \begin{bmatrix} -4 \\ -2.5 \\ -17.5 \end{bmatrix} + 2 \cdot (0.25) \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -2 & 0 \\ 0 & 3 & 1 & 4 \end{bmatrix} \cdot \left(\begin{bmatrix} 3 \\ -1 \\ 0 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 3 \\ 1 & -2 & 1 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} -4 \\ -2.5 \\ -17.5 \end{bmatrix} \right) = \begin{bmatrix} 71.5 \\ 7.75 \\ 221.75 \end{bmatrix}$$