Exercise 2

$T = \{((2,-4)^T, 1), ((1,1)^T, 2), ((6,2)^T, 10), ((3,-1)^T, -2)\}$

using a polynomial model

$f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2$

a) $\nabla_\beta L_2(y_i, f(\mathbf{x}_i)) = ?$

The coefficient vector $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$

Let $\mathbf{z}_i = \begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix} = \begin{bmatrix} 1 \\ x_{i_1} \\ x_{i_2} \\ x_{i_2}^2 \end{bmatrix}$

$$L_2(y_i, f(\mathbf{x}_i)) = (y_i - f(\mathbf{x}_i))^2$$
$$= (y_i - \mathbf{z}_i^T \boldsymbol{\beta})^2$$

$$\nabla_\beta L_2(y_i, f(\mathbf{x}_i)) = \nabla_\beta ((y_i - \mathbf{z}_i^T \boldsymbol{\beta})^2)$$

Chain rule

$$= \underbrace{2(y_i - \mathbf{z}_i^T \boldsymbol{\beta})}_{\text{Outer}} \underbrace{(-\mathbf{z}_i)}_{\text{Inner}} = -2(y_i - \mathbf{z}_i^T \boldsymbol{\beta})\mathbf{z}_i$$

Inner derivative (w.r.t vector $\boldsymbol{\beta}$)

$$\frac{\partial (y_i - \mathbf{z}_i^T \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{\partial (\mathbf{z}_i^T \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{\partial (\boldsymbol{\beta}^T \mathbf{z}_i)}{\partial \boldsymbol{\beta}} = -\mathbf{z}_i$$

Gradient descent:

$$\beta^{(t+1)} = \beta^{(t)} - \eta \nabla J(\beta^{(t)})$$

where η is the learning step

and the functional J

$J(\beta^{(t)}) = \sum_{i=1}^{N} L_2(y_i, f(\mathbf{x}_i))$ for the full batch gradient descent

$$\nabla J(\beta^{(t)}) = \nabla\left(\sum_{i=1}^{N} L_2(y_i, f(\mathbf{x}_i))\right)$$

$$= \sum_{i=1}^{N} \nabla L_2(y_i, f(\mathbf{x}_i))$$

$$= \sum_{i=1}^{N} (-2(y_i - \mathbf{z}_i^T \beta^{(t)})\mathbf{z}_i)$$

$$= -2 \sum_{i=1}^{N} (y_i - \mathbf{z}_i^T \beta^{(t)})\mathbf{z}_i$$

So: $\beta^{(t+1)} = \beta^{(t)} + 2\eta \sum_{i=1}^{N} (y_i - \mathbf{z}_i^T \beta^{(t)})\mathbf{z}_i$

In stochastic gradient descent only one point is used per iteration.

In mini-batch gradient descent just perform the summation for the points that are being used in that iteration.

b) $\beta^{(0)} = \vec{0}$ $\quad \eta = 0.2$ $\quad$ Stochastic

① $\beta^{(1)} = \beta^{(0)} + 2\eta(y_1 - \mathbf{z}_1^T \beta^{(0)})\mathbf{z}_1$

$$\beta^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 2 \cdot 0.2 \left(1 - \begin{bmatrix} 1 & 2 & -4 & 16 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) \begin{bmatrix} 1 \\ 2 \\ -4 \\ 16 \end{bmatrix} =$$

$$= 0.4 \begin{bmatrix} 1 \\ 2 \\ -4 \\ 16 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.8 \\ -1.6 \\ 6.4 \end{bmatrix}$$

② $\beta^{(2)} = \beta^{(1)} + 2\eta\,(y_2 - \mathbf{z}_2^T \beta^{(1)})\,\mathbf{z}_2$

$$\beta^{(2)} = \begin{bmatrix} 0.4 \\ 0.8 \\ -1.6 \\ 6.4 \end{bmatrix} + 0.4\left(2 - [1\ 1\ 1\ 1]\begin{bmatrix} 0.4 \\ 0.8 \\ -1.6 \\ 6.4 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4 \\ 0.8 \\ -1.6 \\ 6.4 \end{bmatrix} + 0.4\begin{bmatrix} -4 \\ -4 \\ -4 \\ -4 \end{bmatrix} = \boxed{\begin{bmatrix} -1.2 \\ -0.8 \\ -3.2 \\ 4.8 \end{bmatrix}}$$

c) $\beta^{(0)} = \vec{0}$ $\quad \eta = 0.2$ $\quad$ <u>Mini-batch $N_b = 2$</u>

① $\beta^{(1)} = \beta^{(0)} + 2\eta \sum_{i=1}^{2} (y_i - \mathbf{z}_i^T \beta^{(0)})\,\mathbf{z}_i$

$$(y_1 - \mathbf{z}_1^T \beta^{(0)})\,\mathbf{z}_1 = \left(1 - [1\ 2\ -4\ 16]\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 2 \\ -4 \\ 16 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -4 \\ 16 \end{bmatrix}$$

$$(y_2 - \mathbf{z}_2^T \beta^{(0)})\,\mathbf{z}_2 = \left(2 - [1\ 1\ 1\ 1]\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

$$\beta^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 0.4\left(\begin{bmatrix} 1 \\ 2 \\ -4 \\ 16 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}\right) = \boxed{\begin{bmatrix} 1.2 \\ 1.6 \\ -0.8 \\ 7.2 \end{bmatrix}}$$

② $\beta^{(2)} = \beta^{(1)} + 2\eta \sum_{i=3}^{4} (y_i - \mathbf{z}_i^T \beta^{(1)})\,\mathbf{z}_i$

$$(y_3 - \mathbf{z}_3^T \beta^{(1)})\,\mathbf{z}_3 = \left(10 - [1\ 6\ 2\ 4]\begin{bmatrix} 1.2 \\ 1.6 \\ -0.8 \\ 7.2 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 6 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -28 \\ -168 \\ -56 \\ -112 \end{bmatrix}$$

$$(y_4 - \mathbf{z}_4^T \beta^{(1)})\,\mathbf{z}_4 = \left(-2 - [1\ 3\ -1\ 1]\begin{bmatrix} 1.2 \\ 1.6 \\ -0.8 \\ 7.2 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -16 \\ -48 \\ 16 \\ -16 \end{bmatrix}$$

$$\beta^{(2)} = \begin{bmatrix} 1.2 \\ 1.6 \\ -0.8 \\ 7.2 \end{bmatrix} + 0.4\left(\begin{bmatrix} -28 \\ -168 \\ -56 \\ -112 \end{bmatrix} + \begin{bmatrix} -16 \\ -48 \\ 16 \\ -16 \end{bmatrix}\right) = \boxed{\begin{bmatrix} -16.4 \\ -84.8 \\ -16.8 \\ -44 \end{bmatrix}}$$