GRADIENT DESCENT EXERCISE

a)

Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^6$ be the function that transforms the vector $x^t = (x_1 \;\; x_2)$ into $\Phi(x)^t = (1 \;\; x_1 \;\; x_2 \;\; x_1 x_2 \;\; x_1^2 \;\; x_2^2)$.

Now, let $\mathbf{X}$ be the matrix with the transformed samples in its rows.

$$X = \begin{bmatrix} \Phi(x_1) \\ \vdots \\ \Phi(x_n) \end{bmatrix}$$

The matrix is now set up for regression.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{n} \left(\Phi(x_i)\beta - y_i \right)^2}_{J(\beta)}$$

Applying Gradient Descent yields the following:

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

Now Computing the gradient w.r.t. β of $L_2(f_\beta(x_i), y_i)$.

$$\nabla_\beta L_2(f_\beta(x_i), y_i) = 2\left(\Phi(x_i)\beta - y_i\right) \begin{bmatrix} 1 \\ x_i^{(1)} \\ x_i^{(2)} \\ x_i^{(1)} x_i^{(2)} \\ \left(x_i^{(1)}\right)^2 \\ \left(x_i^{(2)}\right)^2 \end{bmatrix}$$

where $x_i^{(j)}$ indicates feature j of sample i from the data-set.

b)

$$\beta^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \eta = 0.1$$

$$\beta^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} - 0.1 \cdot 2(0-2) \begin{bmatrix} 1 \\ 4 \\ 1 \\ 4 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0.4 \\ 1.6 \\ 0.4 \\ 1.6 \\ \vdots \end{bmatrix}$$

$$\rightarrow \quad \beta \;\; \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 \cdot 2(\ldots) \begin{bmatrix} 4 \\ 16 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 1.6 \\ 6.4 \\ 0.4 \end{bmatrix}$$

$$\rightarrow \beta^{(2)} = \begin{bmatrix} 0.4 \\ 1.6 \\ 0.4 \\ 1.6 \\ 6.4 \\ 0.4 \end{bmatrix} - 0.1 \cdot 2 \Big(\underbrace{0.4 + 1.6 \cdot 2 + 0.4 \cdot 8 + 1.6 \cdot 16 + 6.4 \cdot 4 + 0.4 \cdot 64 + 14}_{97.6}\Big) \begin{bmatrix} 1 \\ 2 \\ 8 \\ 16 \\ 4 \\ 64 \end{bmatrix}$$

$$\beta^{(2)} = \begin{bmatrix} 0.4 \\ 1.6 \\ 0.4 \\ 1.6 \\ 6.4 \\ 0.4 \end{bmatrix} - 19.52 \begin{bmatrix} 1 \\ 2 \\ 8 \\ 16 \\ 4 \\ 64 \end{bmatrix} = \begin{bmatrix} -19.12 \\ -37.44 \\ -155.76 \\ -310.72 \\ -71.68 \\ -1248.88 \end{bmatrix}$$

c)

$$\beta^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} - 0.2 \sum_{i=1}^{2} \left(\Phi(x_i) \beta^{(0)} - y_i \right) \vec{\Phi}(x_i)$$

$$\beta^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} - 0.2 \left(\begin{bmatrix} -2 \\ -8 \\ -2 \\ -8 \\ -32 \\ -2 \end{bmatrix} + \begin{bmatrix} 14 \\ 28 \\ 112 \\ 224 \\ 56 \\ 896 \end{bmatrix} \right)$$

$$= \begin{bmatrix} -2.4 \\ -4 \\ -22 \\ -43.2 \\ -4.8 \\ -178.8 \end{bmatrix}$$

$$\begin{bmatrix} -2.4 \\ \vdots \end{bmatrix} \quad \beta \quad \begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix}$$

$$\beta^{(2)} = \begin{bmatrix} -4 \\ -22 \\ -43.2 \\ -4.8 \\ -178.8 \end{bmatrix} - 0.2 \Bigg\{ \Big(\underbrace{-2.4 - 4 - 4.8 - 1}_{-12.2} \Big) \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \Big(-2.4 - 12 - 44 -$$

$$\underbrace{259.2 - 43.2 - 715.2 + 1}_{1075} \Big) \begin{bmatrix} 1 \\ 3 \\ 2 \\ 6 \\ 9 \\ 4 \end{bmatrix} \Bigg\}$$

$$\beta^{(2)} = \begin{bmatrix} -2.4 \\ -4 \\ -22 \\ -43.2 \\ -4.8 \\ -178.8 \end{bmatrix} + \begin{bmatrix} 2.44 \\ 2.44 \\ 0 \\ 0 \\ 2.44 \\ 0 \end{bmatrix} + \begin{bmatrix} 215 \\ 645 \\ 430 \\ 1290 \\ 1935 \\ 860 \end{bmatrix}$$

$$\beta^{(2)} = \begin{bmatrix} 215.04 \\ 643.44 \\ 408 \\ 1246.8 \\ 1932.64 \\ 681.2 \end{bmatrix}$$