Question. 5-02

Linear regression을 위한 Dataset이 다음과 같이 주어졌다.

$$\mathcal{D} = \left\{\left(x_2^{(1)}, x_1^{(1)}, y^{(1)}\right), \left(x_2^{(2)}, x_1^{(2)}, y^{(2)}\right), \cdots, \left(x_2^{(n)}, x_1^{(n)}, y^{(n)}\right)\right\}$$

이때, Dataset $y = ax_2 + bx_1 + c$ 에서부터 만들어졌다. 따라서 Linear regression을 위한 Prediction 모델은 $\hat{y} = \theta_2 x_2 + \theta_1 x_1 + \theta_0$, Loss는 Square Error를 사용할 수 있다.

θ_2, θ_1, θ_0를 Update하기 위해 n개의 data sample을 이용할 때,

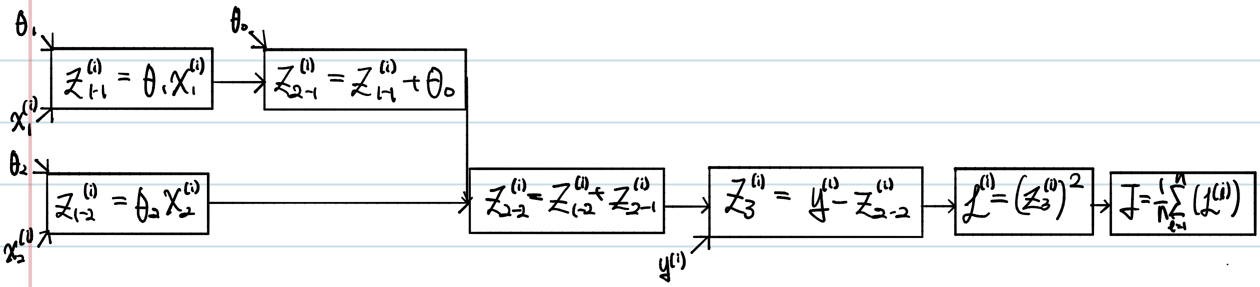
1번의 iteration동안 θ_2, θ_1, θ_0가 dataset을 잘 표현하는 θ_2, θ_1, θ_0로 Update되는 과정을 Vector notation을 이용하여 설명하시오.

단, forward/backward propagation을 설명하기 위해 각 연산은 basic building node를 이용하시오.

① model setting

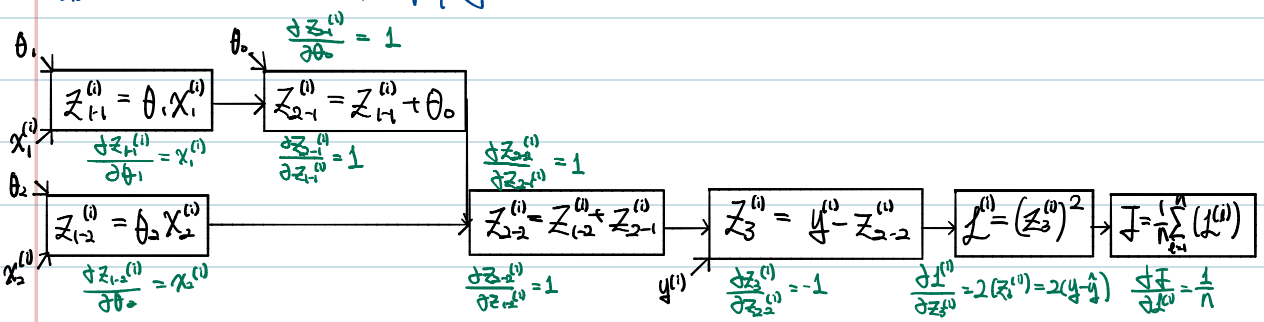
n개의 Dataset에 대해서 $\vec{\theta}$을 update하기 때문에 Cost에 대한 gradient descent method를 사용한다.

이에 i번째 input에 대해 basic building node를 이용하여 표현하면 다음과 같다.



주어진 building nodes에 대해 각각의 partial derivative를 구해보면 다음과 같다.

② Partial derivatives & Backpropagation



위에서 구한 Partial derivatives를 이용하여 $\frac{\partial J}{\partial \theta_2}$, $\frac{\partial J}{\partial \theta_1}$, $\frac{\partial J}{\partial \theta_0}$를 구한다.

$$\frac{\partial J}{\partial \theta_2} = \frac{\partial J}{\partial \mathcal{L}^{(i)}} \cdot \frac{\partial \mathcal{L}^{(i)}}{\partial z_3^{(i)}} \cdot \frac{\partial z_3^{(i)}}{\partial z_{2\text{-}2}^{(i)}} \cdot \frac{\partial z_{2\text{-}2}^{(i)}}{\partial z_{1\text{-}2}^{(i)}} \cdot \frac{\partial z_{1\text{-}2}^{(i)}}{\partial \theta_2} = \frac{1}{n} \cdot 2 \cdot (y - \hat{y}) \cdot (-1) \cdot 1 \cdot x_2 = -\frac{1}{n} \cdot (2x_2^{(i)}(y^{(i)} - \hat{y}^{(i)})) \rightarrow \frac{\partial J}{\partial \theta_2} = -\frac{1}{n}\sum_{i=1}^{n}(2x_2^{(i)}(y^{(i)} - \hat{y}^{(i)}))$$

$$\frac{\partial J}{\partial \theta_1} = \frac{\partial J}{\partial \mathcal{L}^{(i)}} \cdot \frac{\partial \mathcal{L}^{(i)}}{\partial z_3^{(i)}} \cdot \frac{\partial z_3^{(i)}}{\partial z_{2\text{-}2}^{(i)}} \cdot \frac{\partial z_{2\text{-}2}^{(i)}}{\partial z_{2\text{-}1}^{(i)}} \cdot \frac{\partial z_{2\text{-}1}^{(i)}}{\partial z_{1\text{-}1}^{(i)}} \cdot \frac{\partial z_{1\text{-}1}^{(i)}}{\partial \theta_1} = \frac{1}{n} \cdot 2 \cdot (y - \hat{y}) \cdot (-1) \cdot 1 \cdot 1 \cdot x_1 = -\frac{1}{n} \cdot (2x_1^{(i)}(y^{(i)} - \hat{y}^{(i)})) \rightarrow \frac{\partial J}{\partial \theta_1} = -\frac{1}{n}\sum_{i=1}^{n}(2x_1^{(i)}(y^{(i)} - \hat{y}^{(i)}))$$

$$\frac{\partial J}{\partial \theta_0} = \frac{\partial J}{\partial \mathcal{L}^{(i)}} \cdot \frac{\partial \mathcal{L}^{(i)}}{\partial z_3^{(i)}} \cdot \frac{\partial z_3^{(i)}}{\partial z_{2\text{-}2}^{(i)}} \cdot \frac{\partial z_{2\text{-}2}^{(i)}}{\partial z_{2\text{-}1}^{(i)}} \cdot \frac{\partial z_{2\text{-}1}^{(i)}}{\partial \theta_0} = \frac{1}{n} \cdot 2 \cdot (y - \hat{y}) \cdot (-1) \cdot 1 \cdot 1 = -\frac{1}{n} \cdot (2(y^{(i)} - \hat{y}^{(i)})) \rightarrow \frac{\partial J}{\partial \theta_0} = -\frac{1}{n}\sum_{i=1}^{n}(2(y^{(i)} - \hat{y}^{(i)}))$$

이렇게 구한 gradient method를 이용하여 $\theta_2, \theta_1, \theta_0$의 update equation을 구한다.

3 gradient descent method

$$\theta_2 := \theta_2 - \alpha \cdot \frac{\partial J}{\partial \theta_2} = \theta_2 + \frac{\alpha}{n}\sum_{i=1}^{n}\left(2x_2^{(i)}\left(y^{(i)} - \hat{y}^{(i)}\right)\right)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial J}{\partial \theta_1} = \theta_1 + \frac{\alpha}{n}\sum_{i=1}^{n}\left(2x_1^{(i)}\left(y^{(i)} - \hat{y}^{(i)}\right)\right)$$

$$\theta_0 := \theta_0 - \alpha \frac{\partial J}{\partial \theta_0} = \theta_0 + \frac{\alpha}{n}\sum_{i=1}^{n}\left(2\left(y^{(i)} - \hat{y}^{(i)}\right)\right)$$