Question. 4−11

Linear regression을 위한 Dataset이 다음과 같이 주어졌다.

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(n)}, y^{(n)})\}$$

이때, Dataset은 $y = ax + b$에서부터 만들어졌다. 따라서 Linear regression을 위한

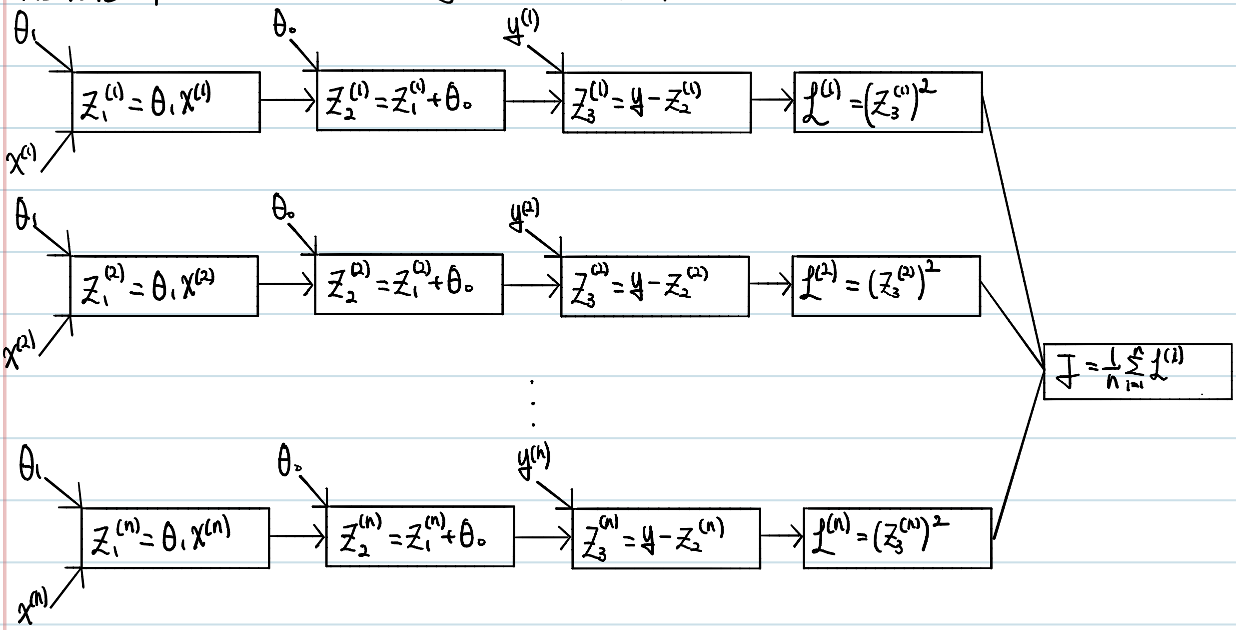
Prediction 모델은 $\hat{y} = \theta_1 x + \theta_0$, Loss는 Square Error, Cost는 MSE를 사용할 수 있다.

$\vec{\theta}$를 Update하기 위해 n개의 data sample을 이용할 때, 1번의 iteration동안 $\vec{\theta}$가 dataset을 잘 표현하는 $\vec{\theta}$로 Update되는 과정을 설명하시오.

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

① model Setting

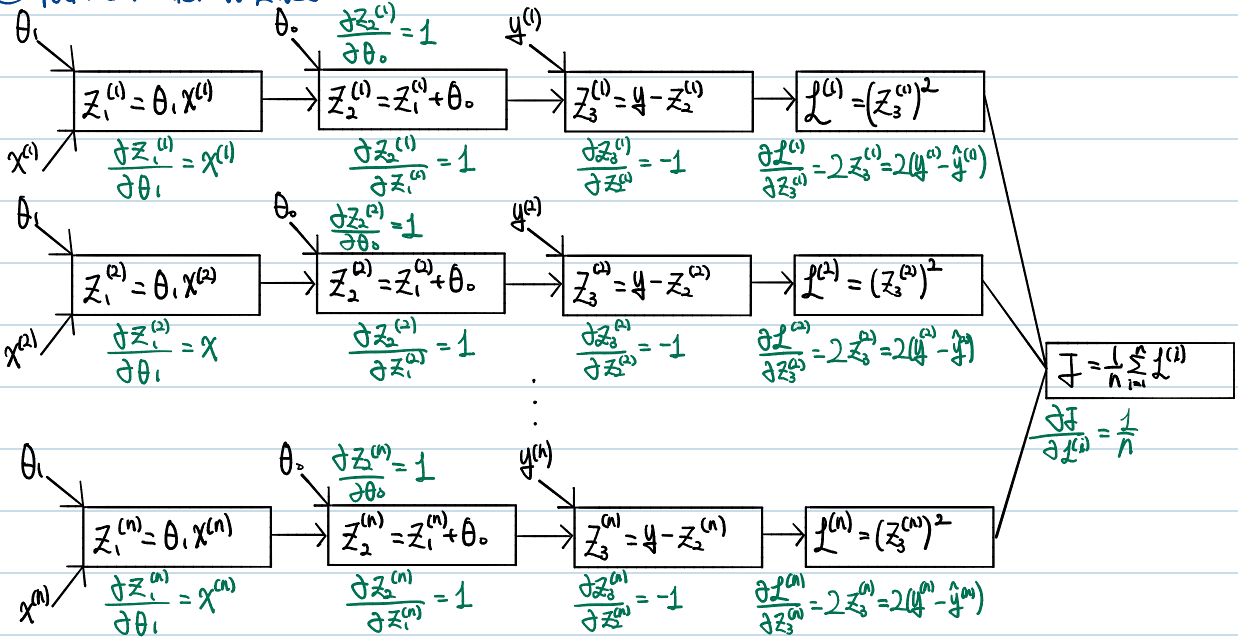
θ_1과 θ_0를 update하기 위해 n개의 Data Sample을 이용하기 때문에 Cost에 대한 Gradient Descent method를 사용한다. 이를 basic building node로 표현하면 다음과 같다.



위와 같이 Forward Propagation을 통해 n개의 Data Sample Loss의 평균을 구해 Cost를 연산한다.

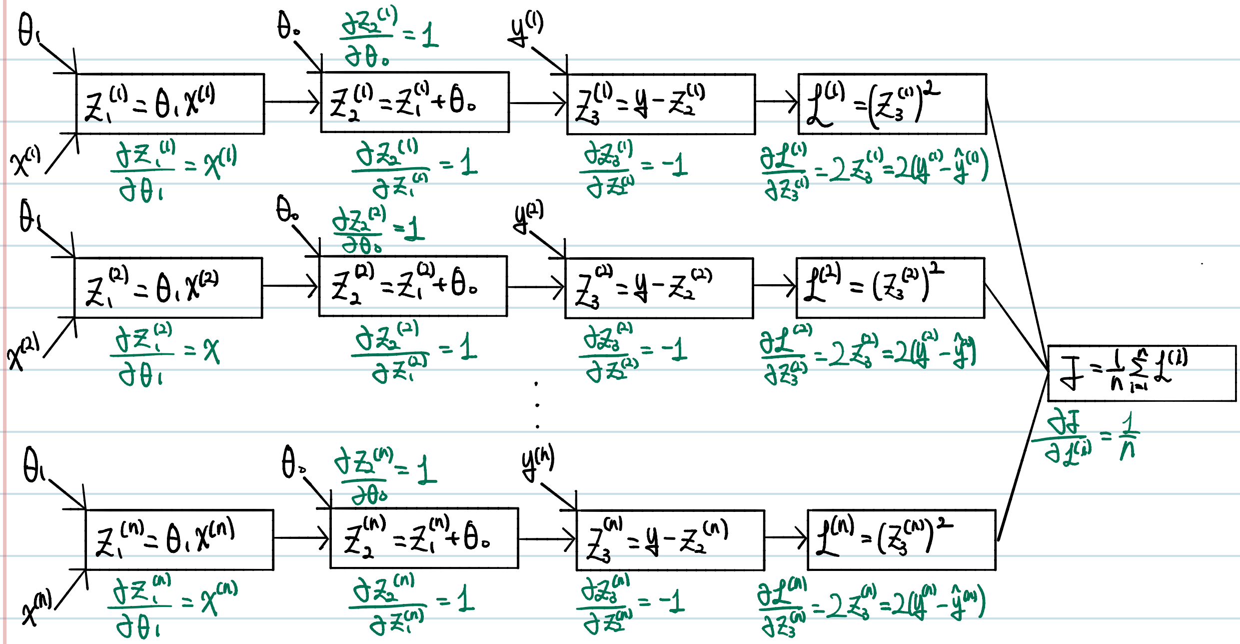
이때 θ_1, θ_0의 update를 위한 partial derivative는 다음과 같다.

② Partial derivatives



그러므로 Back propagation은 다음과 같이 이루어진다.

③ Backpropagation



$$\frac{\partial J}{\partial \theta_1} = \sum_{i=1}^{n} \frac{\partial J}{\partial \mathcal{L}^{(i)}} \cdot \frac{\partial \mathcal{L}^{(i)}}{\partial z_3^{(i)}} \cdot \frac{\partial z_3^{(i)}}{\partial z_2^{(i)}} \cdot \frac{\partial z_2^{(i)}}{\partial z_1^{(i)}} \cdot \frac{\partial z_1^{(i)}}{\partial \theta_1}$$

$$= \sum_{i=1}^{n} \frac{1}{n} \cdot 2(y^{(i)} - \hat{y}^{(i)}) \cdot (-1) \cdot (1) \cdot x^{(i)} = -\frac{2}{n} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})\, x^{(i)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathcal{L}^{(i)}}{\partial \theta_1}$$

$$\frac{\partial J}{\partial \theta_0} = \sum_{i=1}^{n} \frac{\partial J}{\partial \mathcal{L}^{(i)}} \cdot \frac{\partial \mathcal{L}^{(i)}}{\partial z_3^{(i)}} \cdot \frac{\partial z_3^{(i)}}{\partial z_2^{(i)}} \cdot \frac{\partial z_2^{(i)}}{\partial \theta_0}$$

$$= \sum_{i=1}^{n} \frac{2}{n} \cdot 2(y^{(i)} - \hat{y}^{(i)})\,(-1) \cdot (1) \cdot (1) = -\frac{2}{n} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathcal{L}^{(i)}}{\partial \theta_0}$$

따라서 위와같이 θ_1은 $\vec{x}$의 평균과 loss에 비례하여 update되고
θ_0는 $\vec{x}$의 loss에만 비례하여 update된다.

(4) gradient descent method

위에서 구한 $\frac{\partial J}{\partial \theta_1}$, $\frac{\partial J}{\partial \theta_0}$와 gradient descent method를 이용하여 θ_1, θ_0의 update equation를 표현하면

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

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

가 되어 θ_1과 θ_0이 학습되게된다.