Question. 5-04

Question. 5-02로부터

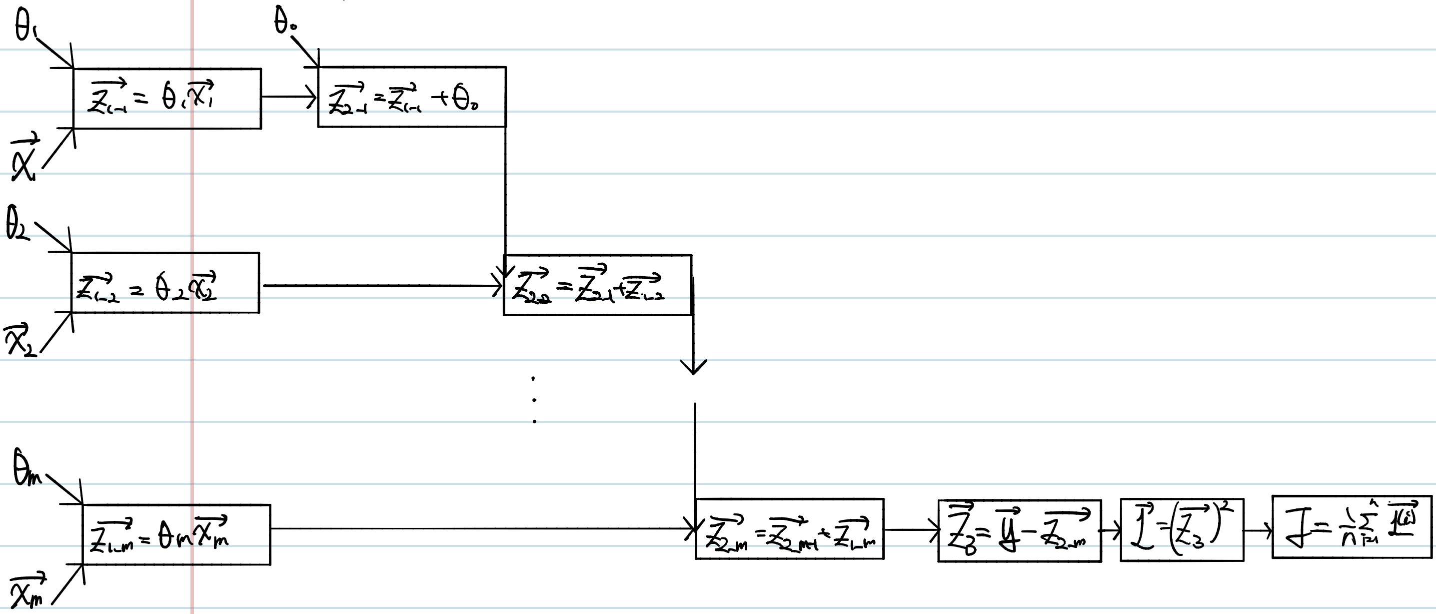
Dataset을 $y = ax_n + \cdots + bx_1 + c$로, Prediction 모델을 $\hat{y} = \theta_m x_m + \cdots + \theta_1 x_1 + \theta_0$로 확장하였다.

$\vec{\theta}$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

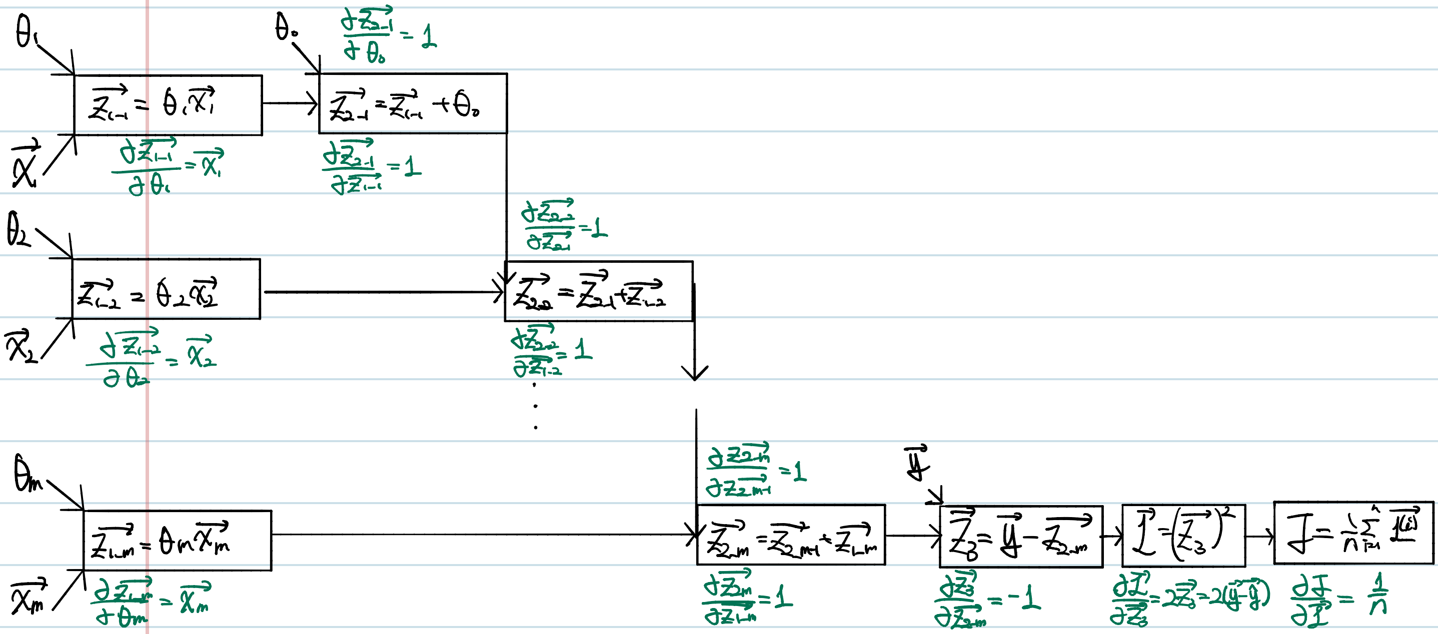
$\vec{\theta}$를 update하기 위해 n개의 Data Sample을 이용하기 때문에 Cost에 대한 Gradient Descent method를 사용한다. 이를 basic building node로 표현하면 다음과 같다.



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

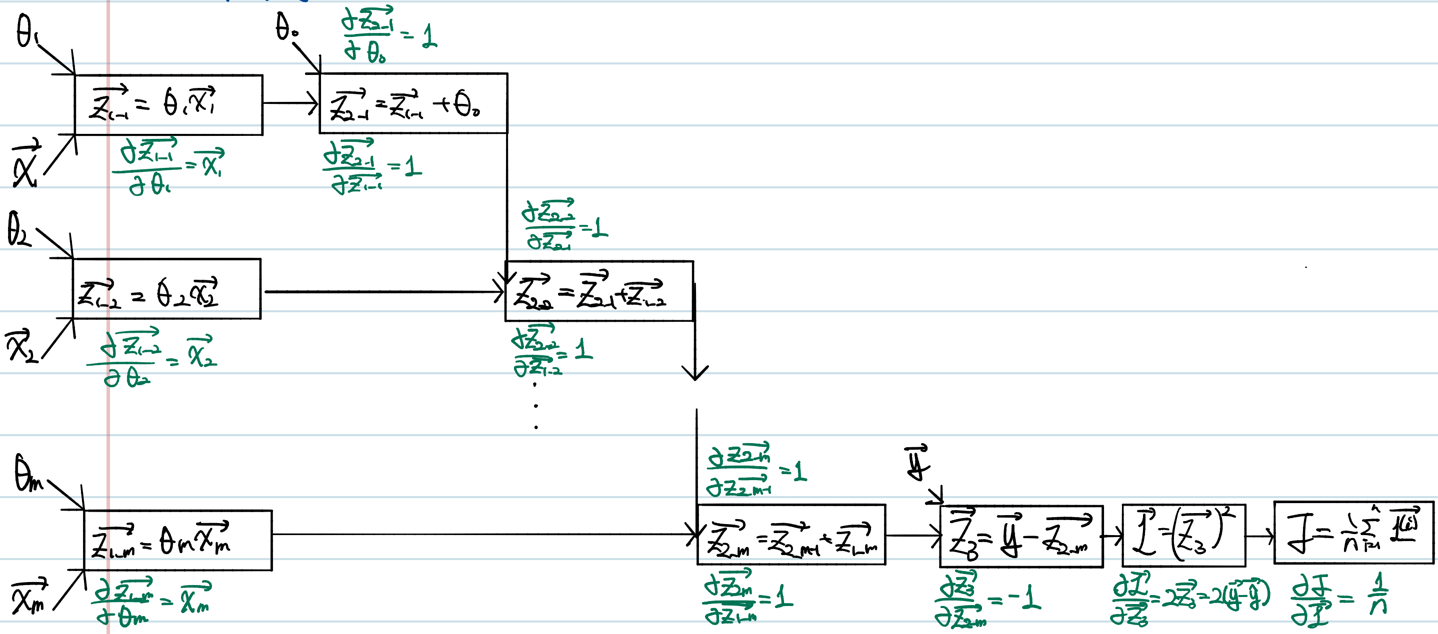
이때 $\vec{\theta}$의 update를 위한 partial derivative는 다음과 같다.

② Partial derivatives



그리고 Backpropagation은 다음과 같이 이루어진다.

③ Backpropagation



θ_k 에 의한 Gradient $\frac{\partial J}{\partial \theta_k} = \frac{\partial J}{\partial L}\cdot\frac{\partial L}{\partial \vec{Z_3}}\cdot\frac{\partial \vec{Z_3}}{\partial \vec{Z_{2,m}}}\cdot\frac{\partial \vec{Z_{2,m}}}{\partial \vec{Z_{2,m-1}}}\cdots\frac{\partial \vec{Z_{2,k+1}}}{\partial \vec{Z_{2,k}}}\cdot\frac{\partial \vec{Z_{2,k}}}{\partial \vec{Z_{1,k}}}\cdot\frac{\partial \vec{Z_{1,k}}}{\partial \theta_k} = \frac{1}{n}\sum_{i=1}^{n}(\vec{y}-\hat{\vec{y}})\cdot(-1)\cdot 1\cdots 1\cdot\vec{x_k}$

$= -\frac{1}{n}\sum_{i=1}^{n} x_k^{(i)}(y^{(i)}-\hat{y}^{(i)})$

θ_0에 의한 Gradient $\frac{\partial J}{\partial \theta_0} = \frac{\partial J}{\partial L}\cdot\frac{\partial L}{\partial \vec{Z_3}}\cdot\frac{\partial \vec{Z_3}}{\partial \vec{Z_{2,m}}}\cdot\frac{\partial \vec{Z_{2,m}}}{\partial \vec{Z_{2,m-1}}}\cdots\frac{\partial \vec{Z_{2,2}}}{\partial \vec{Z_{2,1}}}\cdot\frac{\partial \vec{Z_{2,1}}}{\partial \theta_0} = \frac{1}{n}\sum_{i=1}^{n}(\vec{y}-\hat{\vec{y}})\cdot(-1)\cdot 1\cdots 1$

$= -\frac{1}{n}\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})$ 이다.

따라서 위와같이 θ_k는 $\vec{x_k}$의 평균과 Loss에 비례하여 update되고
θ_0는 Loss에만 비례하여 update된다.

④ gradient descent method

따라서 weight와 bias의 update equation은 다음과 같다.

$$\theta_k := \theta_k + \frac{\alpha}{n} \sum_{i=1}^{n} x_k^{(i)} (y^{(i)} - \hat{y}^{(i)}) \quad \text{Where } k \geq 1$$

$$\theta_0 := \theta_0 + \frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})$$