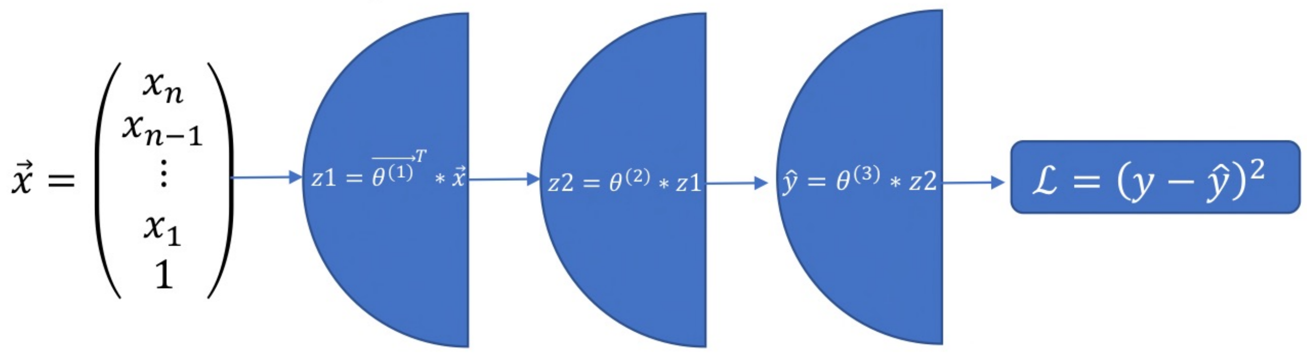
Question. 7-01

다음과 같이 Linear Regression에서 사용했던 Neuron으로 Multi-Layer를 쌓았다고 가정하자.
이때, Activation Function이 없는 Multi-Layer가 Single-Layer와 같음을 보여라. (단, 모든 $\theta \in \mathbb{R}$)
이로 인한 Linear Regression의 단점은 무엇인가.



$$\overrightarrow{\theta^{(1)}} = \begin{pmatrix} \theta_n^{(1)} \\ \theta_{n-1}^{(1)} \\ \vdots \\ \theta_1^{(1)} \\ \theta_0^{(1)} \end{pmatrix}$$

1) Linear Regression에서 Multi-Layer와 Single-Layer의 연산과정 중 $\vec{x}$, y가 동일하므로 $\hat{y}$의 비교를 통해 같음을 보인다.

Single-Layer의 경우

$$\hat{y} = \overrightarrow{\theta^{(1)}}^T * \vec{x} = \begin{pmatrix} \theta_n^{(1)} \\ \theta_{n-1}^{(1)} \\ \vdots \\ \theta_1^{(1)} \\ \theta_0^{(1)} \end{pmatrix}^T * \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \\ 1 \end{pmatrix} = \theta_n^{(1)} x_n + \theta_{n-1}^{(1)} x_{n-1} + \cdots + \theta_1^{(1)} x_1 + \theta_0^{(1)}$$

Multi-layer의 경우

$$z_1 = \overrightarrow{\theta^{(1)}}^T * \vec{x} = \begin{pmatrix} \theta_n^{(1)} \\ \theta_{n-1}^{(1)} \\ \vdots \\ \theta_1^{(1)} \\ \theta_0^{(1)} \end{pmatrix}^T * \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \\ 1 \end{pmatrix} = \theta_n^{(1)} x_n + \theta_{n-1}^{(1)} x_{n-1} + \cdots + \theta_1^{(1)} x_1 + \theta_0^{(1)}$$

$$z_2 = \theta^{(2)} * z_1 = \theta^{(2)}(\theta_n^{(1)} x_n + \theta_{n-1}^{(1)} x_{n-1} + \cdots + \theta_1^{(1)} x_1 + \theta_0^{(1)})$$

$$\hat{y} = \theta^{(3)} * z_2 = \theta^{(3)}\theta^{(2)}(\theta_n^{(1)} x_n + \theta_{n-1}^{(1)} x_{n-1} + \cdots + \theta_1^{(1)} x_1 + \theta_0^{(1)})$$

$$= \theta^{(3)}\theta^{(2)}\theta_n^{(1)} x_n + \theta^{(3)}\theta^{(2)}\theta_{n-1}^{(1)} x_{n-1} + \cdots + \theta^{(3)}\theta^{(2)}\theta_1^{(1)} x_1 + \theta^{(3)}\theta^{(2)}\theta_0^{(1)}$$

이때, 모든 $\theta \in \mathbb{R}$ 이므로 $0 \le k \le n$ 인 임의의 k에 대해서 $\theta^{(3)}\theta^{(2)}\theta_k^{(1)} \in \mathbb{R}$ 이다.

Single-Layer에서 x_k의 계수 $\theta_k^{(1)}$와 Multi-Layer에서 x_k의 계수 $\theta^{(3)}\theta^{(2)}\theta_k^{(1)}$ 모두 arbitrary constant 이므로 두 경우의 Layer 모두 $\hat{y}$이 같다.

2) Linear Regression에서 Single-Layer와 Multi-Layer 모두 Layer의 개수와 상관없이 학습 대상인 $\hat{y}$이 같아지므로 Layer를 많이 쌓아도 Single-Layer같이 단순한 모델만 Regression 가능하고 복잡한 형태의 모델은 구현할 수 없다.