Problem 1

We assume that the feed-forward neural network has L hidden layers, where each layer applies a linear activation function.

Assume the input is x, then we have:

$$h^{(1)} = W_1 x + b_1$$

$$h^{(2)} = W_2 h^{(1)} + b_2 = W_2 W_1 x + W_2 b_1 + b_2$$

$$h^{(3)} = W_3 h^{(2)} + b_3 = W_3 (W_2 W_1 x + W_2 b_1 + b_2) + b_3$$

$$= W_3 W_2 W_1 x + \underbrace{W_3 W_2 b_1 + W_3 b_2 + b_3}_{\text{bias terms}}$$

$$\vdots$$

Then, $h^{(L)} = W_L W_{L-1} \cdots W_1 x + \text{bias terms}$

$$= W_{\text{combined}} \cdot x + b_{\text{combined}}$$

where $W_{\text{combined}} = W_L W_{L-1} \cdots W_1$ and b_{combined} is a combination of weighted bias.

Therefore, even though we have multiple layers, since the activation function is linear, the network behaves exactly like a single linear transformation with no hidden layer.