Exercise 2: From Logistic Regression to Neural Networks

4) For such a network with the identity as the activation function the network output is:

$$z^{(L)} = W^{(L)}\left(W^{(L-1)}\left(\dots\left(W^{(2)}\left(W^{(1)}x + b^{(1)}\right) + b^{(2)}\right) + \dots + b^{(L-1)}\right) + b^{(L)}$$

This can be rewritten using:

$$\tilde{b} = b^{(L)} + W^{(L)} b^{(L-1)} + W^{(L)} W^{(L-1)} b^{(L-2)} + \dots + W^{(L)} \dots W^{(2)} b^{(1)} = \sum_{l=1}^{L} \left(\prod_{k=1}^{L-l} W^{(L-k)}\right) b^{(l)}$$

$$\tilde{W} = W^{(L)} W^{(L-1)} \dots W^{(1)} x = \prod_{l=1}^{L} W^{(L-l)} x$$

Yielding

$$z = \tilde{W} x + \tilde{b}$$

a linear 1-layer network.

In order to extend the expressiveness of the network beyond linear transformation, non-linearities such as sigmoid activation are required, since however intricate the layer structure might be, if only linear operations are performed within it, it can always be reduced to a simple 1-layer net.