Machine Learning exercise 2

Task 2

(1) $Z_0 = X$

(2) $\tilde{Z}_l = Z_{l-1} \cdot B_l + b_l$

(3) $Z_l = \phi_l(\tilde{Z}_l)$

if the activation function ϕ_l is the identity function, any neural network with depth $L > 1$ is equivalent to a 1-layer neural network

$\Rightarrow$ output of this network has to be represented as a single affine transformation of the inputs

ϕ_l is identity function $\Rightarrow$ $\phi_l(z) = z$

For $L > 1$:

Layer 1 $\quad \tilde{Z}_1 = Z_0 \cdot B_1 + b_1 = X \cdot B_1 + b_1$

$Z_1 = \phi_1(\tilde{Z}_1) = \tilde{Z}_1 = X \cdot B_1 + b_1$

Layer 2: $\quad \tilde{Z}_2 = Z_1 \cdot B_2 + b_2 = (X \cdot B_1 + b_1) \cdot B_2 + b_2$

$Z_2 = \phi_2(\tilde{Z}_2) = \tilde{Z}_2$

$= (X \cdot B_1 + b_1) \cdot B_2 + b_2$

$= X \cdot (B_1 \cdot B_2) + b_1 \cdot B_2 + b_2$

Generally with (2) and (3), lets assume for layer $l-1$ we have:

$$Z_{l-1} = X \cdot \underbrace{W_{l-1}}_{\substack{\text{product of} \\ \text{weight matrices} \\ \text{up to } l-1}} + \underbrace{b'_{l-1}}_{\text{cumulative bias}}$$

Now prove for l:

$\tilde{Z}_l = Z_{l-1} \cdot B_l + b_l = (X \cdot W_{l-1} + b'_{l-1}) \cdot B_l + b_l$

$Z_l = \phi(\tilde{Z}_l) = \tilde{Z}_l$

$= (X \cdot W_{l-1} + b'_{l-1}) \cdot B_l + b_l$

$= X \cdot (W_{l-1} \cdot B_l) + b'_{l-1} \cdot B_l + b_l$

Define: $\quad W_l = W_{l-1} \cdot B_l$

$b'_l = b'_{l-1} \cdot B_l + b_l$

So: $\quad Z_l = X \cdot W_l + b'_l \quad (4)$

$\square$

From (1), (2), (3) and (4), it is proven that: if the activation function ϕ_l is the identity function, any neural network with depth $L > 1$ is equivalent to a 1-layer neural network