5.

a.

5 a) $g(v_1 + V_2 z_1 + V_2 z_2) = g(g(x \cdot w_1) \cdot w_2)$

$V_1 + V_2 z_1 + V_2 z_2 = g(x \cdot w_1) \cdot w_2$

$$X \cdot w_1 = \begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \\ w_5 & w_6 \end{bmatrix}$$

$$= \begin{bmatrix} w_1 + w_3 x_1 + w_5 x_2 & w_2 + w_4 x_1 + w_6 x_2 \end{bmatrix}$$

add intercept for $g(x \cdot w_1)$

$$g(x \cdot w_1) = \begin{bmatrix} 1 & w_1 + w_3 x_1 + w_5 x_2 & w_2 + w_4 x_1 + w_6 x_2 \end{bmatrix}$$

$$g(x \cdot w_1) \cdot w_2 = \begin{bmatrix} 1 & w_1 + w_3 x_1 + w_5 x_2 & w_2 + w_4 x_1 + w_6 x_2 \end{bmatrix} \cdot \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$$

$$g(x \cdot w_1) \cdot w_2 = V_1 + \underbrace{(w_1 + w_3 x_1 + w_5 x_2)}_{z_1} V_2 + \underbrace{(w_2 + w_4 x_1 + w_6 x_2)}_{z_2} V_3$$

$$g(V_1 + (w_1 + w_3 x_1 + w_5 x_2) V_2 + (w_2 + w_4 x_1 + w_6 x_2) V_3)$$

b.

b) $g(x) = x$

$$x = V_1 + (w_1 + w_3 x_1 + w_5 x_2) v_2 + (w_2 + w_4 x_1 + w_6 x_2) v_3$$

$$= V_1 + w_1 V_2 + x_1 w_2 V_2 + x_2 w_5 V_2 + w_2 V_3 + x_1 w_4 V_3 + x_2 w_6 V_3$$

$$= 1\underbrace{(V_1 + w_1 V_2 + w_2 V_3)}_{m} + x_1 \underbrace{(w_2 V_2 + w_4 V_3)}_{n} + x_2 \underbrace{(w_5 V_2 + w_6 V_3)}_{o}$$

$$g(x) = g((V_1 + w_1 V_2 + w_2 V_3) + x_1 (w_2 V_2 + w_4 V_3) + x_2 (w_5 V_2 + w_6 V_3))$$

c. Since MLP can be used and solved in a linear activation function it has better expressivity than it’s singular linear counterpart.

d.

d) $g(x) = \frac{1}{1+e^{-x}}$

$$g(V_1 + z_1 V_2 + z_2 V_3) = \frac{1}{1 + e^{-(V_1 + z_1 V_2 + z_2 V_3)}}$$

$$\frac{1}{g(x)} = 1 + e^{-(V_1 + z_1 V_2 + z_2 V_3)}$$

$$\ln\left(\frac{1}{g(x)} - 1\right) = \ln\left(e^{-(V_1 + z_1 V_2 + z_2 V_3)}\right)$$

$$\ln\left(\frac{1}{g(x)} - 1\right) = -(V_1 + z_1 V_2 + z_2 V_3)$$

Not solvable because you can not find a clear solution for g(x) with the exponential.

e. Because the MLP expression is not solvable, the expressive power is not as good as the single sigmoid perceptron.

f. The activation function must be differentiable so that you can go backwards within the model. It also must be nonlinear so that it can shape to curves.