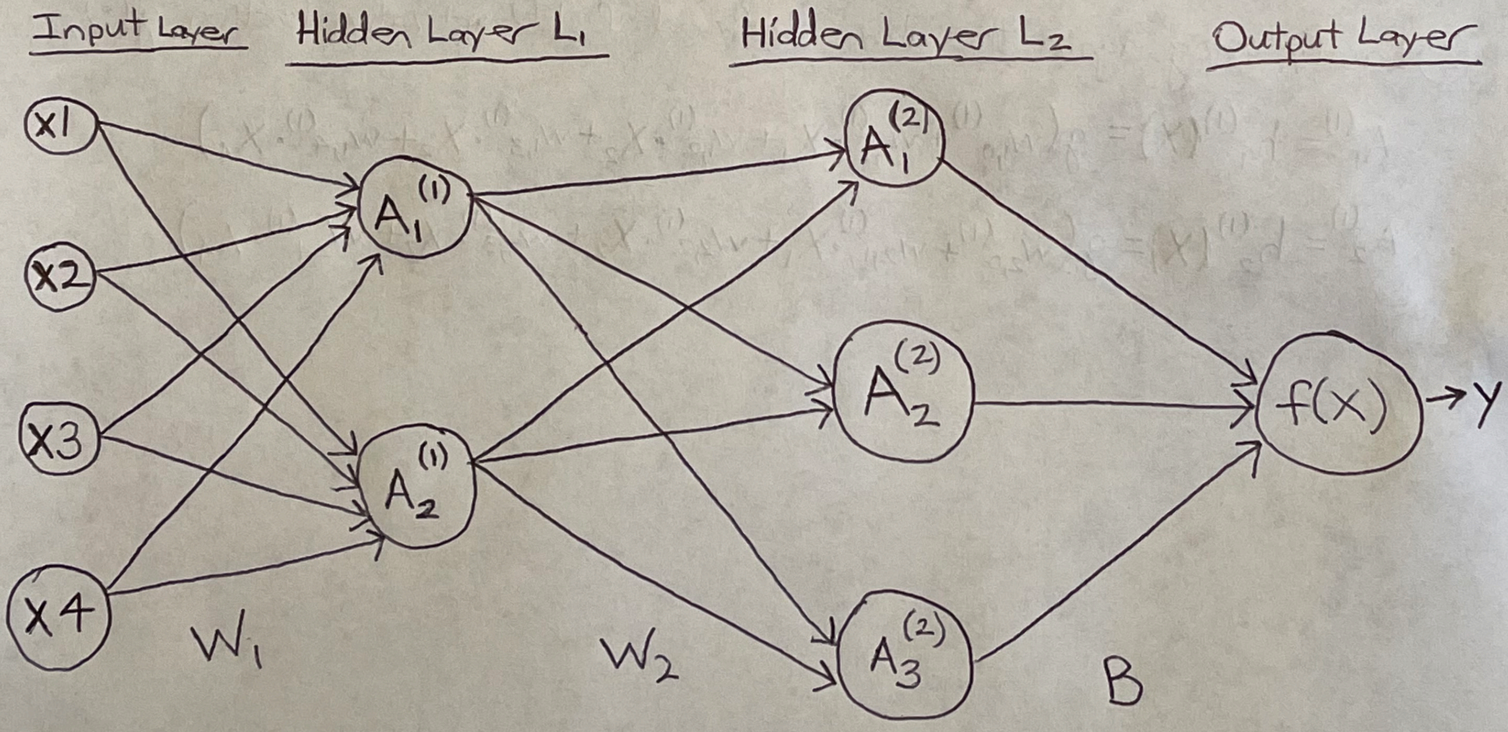
Chapter 10: Conceptual Exercises

1.) I assume:

(i) There is no activation from Hidden Layer L_2 to the Output Layer.

(ii) There is a bias term to be included in addition to both the $p=4$ inputs, another bias term in addition to the 2 Activations in Hidden Layer L_1, and a third bias term in addition to the 3 Activations in Hidden Layer L_2.

(a.) Picture



(b.) Assuming ReLU activation functions, below I write out (very explicitly) an expression for $f(X)$:

① $$f(X) = \beta_0 + \sum_{\ell=1}^{3} \beta_\ell A_\ell^{(2)} = \beta_0 + \beta_1 \cdot A_1^{(2)} + \beta_2 \cdot A_2^{(2)} + \beta_3 \cdot A_3^{(2)}$$

② Defining the ReLU activation as: $g(z) = (z)_+ = \begin{cases} 0 & \text{if } z<0 \\ z & \text{otherwise} \end{cases}$,

$$A_1^{(2)} = h_1^{(2)}(X) = g\left(w_{1,0}^{(2)} + w_{1,1}^{(2)} \cdot A_1^{(1)} + w_{1,2}^{(2)} \cdot A_2^{(1)}\right)$$

$$A_2^{(2)} = h_2^{(2)}(X) = g\left(w_{2,0}^{(2)} + w_{2,1}^{(2)} \cdot A_1^{(1)} + w_{2,2}^{(2)} \cdot A_2^{(1)}\right)$$

$$A_3^{(2)} = h_3^{(2)}(X) = g\left(w_{3,0}^{(2)} + w_{3,1}^{(2)} \cdot A_1^{(1)} + w_{3,2}^{(2)} \cdot A_2^{(1)}\right)$$

③ Lastly, using again the above-defined ReLU activation function $g(\cdot)$,

$$A_1^{(1)} = h_1^{(1)}(X) = g\left(w_{1,0}^{(1)} + w_{1,1}^{(1)} \cdot X_1 + w_{1,2}^{(1)} \cdot X_2 + w_{1,3}^{(1)} \cdot X_3 + w_{1,4}^{(1)} \cdot X_4\right)$$

$$A_2^{(1)} = h_2^{(1)}(X) = g\left(w_{2,0}^{(1)} + w_{2,1}^{(1)} \cdot X_1 + w_{2,2}^{(1)} \cdot X_2 + w_{2,3}^{(1)} \cdot X_3 + w_{2,4}^{(1)} \cdot X_4\right)$$

(c) Given the above in part (b), I plug in some values for the coefficients. I select:

$$B = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}; \quad W_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 4 & 7 \end{pmatrix}; \quad W_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 7 & 9 \end{pmatrix}$$

Given again the above definition of $g(X)$, we again write out the value for $f(X)$:

$$f(X) = 2 + 3 \cdot A_1^{(2)} + 4 \cdot A_2^{(2)} + 5 \cdot A_3^{(2)}$$

where:

$$A_1^{(2)} = g(1 + 2 \cdot A_1^{(1)} + 3 \cdot A_2^{(1)})$$
$$A_2^{(2)} = g(1 + 3 \cdot A_1^{(1)} + 5 \cdot A_2^{(1)})$$
$$A_3^{(2)} = g(1 + 4 \cdot A_1^{(1)} + 7 \cdot A_2^{(1)})$$

where:

$$A_1^{(1)} = g(1 + 2 \cdot X_1 + 3 \cdot X_2 + 4 \cdot X_3 + 5 \cdot X_4)$$
$$A_2^{(1)} = g(1 + 3 \cdot X_1 + 5 \cdot X_2 + 7 \cdot X_3 + 9 \cdot X_4)$$

(d.) There are a total of:

$$\underbrace{4}_{B's} + \underbrace{3(3)}_{W_2} + \underbrace{5(2)}_{W_1} = \boxed{23 \text{ parameters}}$$