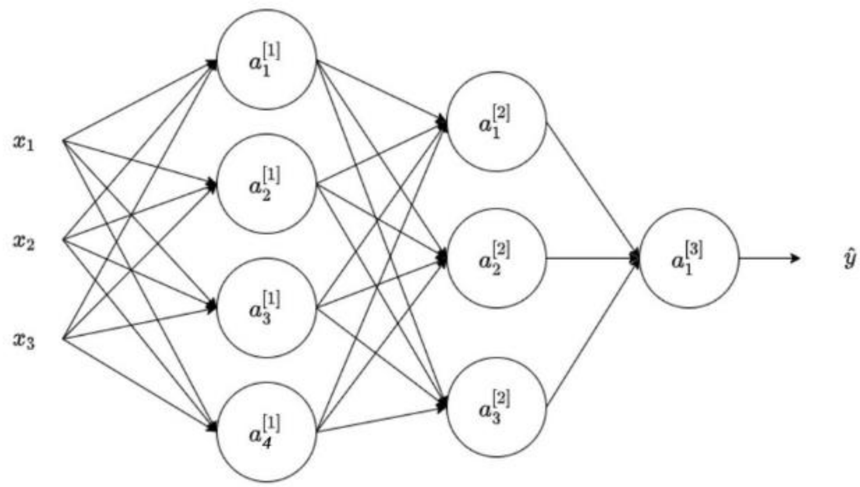
9. Consider the following 2 hidden layers neural network: 1 point



We are given a **3-layer neural network**:

- Input: x_1, x_2, x_3 → so input layer size = **3**
- Hidden Layer 1: 4 units → $a_1^{[1]}, a_2^{[1]}, a_3^{[1]}, a_4^{[1]}$
- Hidden Layer 2: 3 units → $a_1^{[2]}, a_2^{[2]}, a_3^{[2]}$
- Output Layer: 1 unit → $a_1^{[3]} = \hat{y}$

Formula for weight and bias shapes:

For layer l:

$$W^{[l]} \in \mathbb{R}^{(n^{[l]}, n^{[l-1]})}, \quad b^{[l]} \in \mathbb{R}^{(n^{[l]}, 1)}$$

where

- $n^{[l]}$ = number of units in layer l.
- $n^{[l-1]}$ = number of units in the previous layer.

Step 1: First hidden layer ($l = 1$)

- Previous layer: input size = 3 → $n^{[0]} = 3$
- Current layer: $n^{[1]} = 4$

So:

- $W^{[1]} \in \mathbb{R}^{(4,3)}$ ✅
- $b^{[1]} \in \mathbb{R}^{(4,1)}$ ✅

Step 2: Second hidden layer ($l = 2$)

- Previous layer: $n^{[1]} = 4$
- Current layer: $n^{[2]} = 3$

So:

- $W^{[2]} \in \mathbb{R}^{(3,4)}$ ✅
- $b^{[2]} \in \mathbb{R}^{(3,1)}$ ✅

Step 3: Output layer ($l = 3$)

- Previous layer: $n^{[2]} = 3$
- Current layer: $n^{[3]} = 1$

So:

- $W^{[3]} \in \mathbb{R}^{(1,3)}$ ✅
- $b^{[3]} \in \mathbb{R}^{(1,1)}$ ✅

✅ Correct Statements from the options:

- $W^{[1]}$ will have shape **(4, 3)**
- $b^{[1]}$ will have shape **(4, 1)**
- $W^{[2]}$ will have shape **(3, 4)**
- $b^{[2]}$ will have shape **(3, 1)**
- $W^{[3]}$ will have shape **(1, 3)**

10. Whereas the previous question used a specific network, in the general case what is the dimension of $b^{[l]}$, the bias vector associated with layer l?

- ○ $b^{[l]}$ has shape $(1, n^{[l-1]})$
- ○ $b^{[l]}$ has shape $(1, n^{[l]})$
- ○ $b^{[l]}$ has shape $(n^{[l]}, 1)$
- ○ $b^{[l]}$ has shape $(n^{[l+1]}, 1)$

What is the general dimension of $b^{[l]}$, the bias vector of layer l?

Step 1: Recall the general formulas

For a given layer l:

- Number of units in layer l = $n^{[l]}$.
- **Weights:**

$$W^{[l]} \in \mathbb{R}^{(n^{[l]}, n^{[l-1]})}$$

- **Biases:**

$$b^{[l]} \in \mathbb{R}^{(n^{[l]}, 1)}$$

- **Activations:**

$$A^{[l]} \in \mathbb{R}^{(n^{[l]}, m)}$$

  where m = number of training examples.

Step 2: Why $b^{[l]} \in (n^{[l]}, 1)$?

- Each unit in layer l has **its own bias.**
- So, if there are $n^{[l]}$ units, then $b^{[l]}$ must have **one bias per unit.**
- Since biases are stored as a column vector, the shape is:

$$(n^{[l]}, 1)$$

Good question! Let's sort them out carefully:

Hyperparameters (set by you, not learned during training):

- ✅ **Number of iterations** (you decide how long to train)
- ✅ **Size of the hidden layers (n^{[l]})** (you design the architecture)
- ✅ **Number of layers (L)** (you choose how deep the network is)
- ✅ **Learning rate (\alpha)** (you set the step size for updates)

Not Hyperparameters (these are learned values or results of computation):

- ❌ **Bias vectors (b^{[l]})** → learned during training
- ❌ **Weight matrices (W^{[l]})** → learned during training
- ❌ **Activation values (a^{[l]})** → computed outputs, not set manually

👉 So the **hyperparameters** are:
Number of iterations, Size of hidden layers, Number of layers, Learning rate.