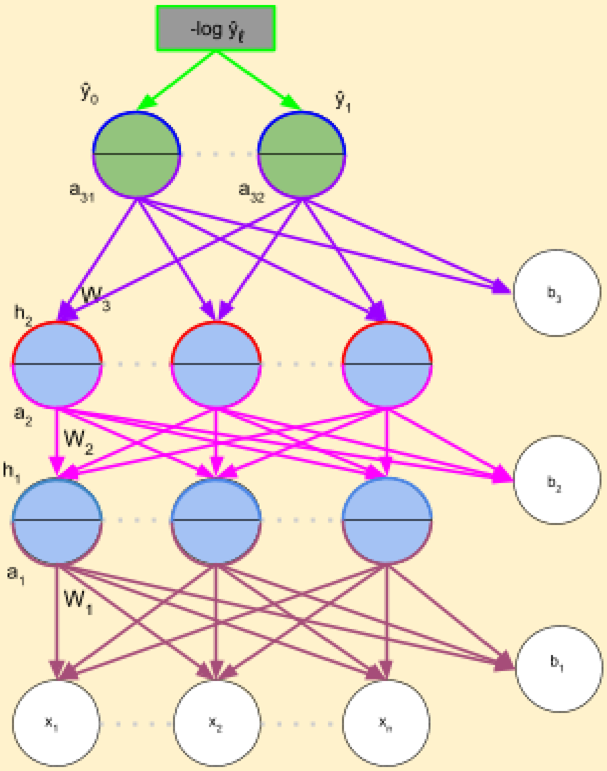
PadhAI: Backpropagation - the full version

One Fourth Labs

Understanding the dimensions of gradients

What are we interested in?

1. Consider the backpropagation illustration from the previous section



2. What we are interested in is a $\frac{\partial L(\theta)}{\partial a_{Li}} = \frac{\partial(-\log \hat{y}_l)}{\partial a_{Li}}$ (where true output y = 1, L = Layer number, l is the index of the correct class-label for the given input, and i is the neuron number)
3. We know that $\frac{\partial L(\theta)}{\partial a_{Li}}$ is dependent on a_{31} and a_{32}
4. Therefore, the derivative at the output layer

$$\nabla_{a_3} L(\theta) = \begin{bmatrix} \frac{\partial L}{\partial a_{31}} \\ \frac{\partial L}{\partial a_{32}} \end{bmatrix}$$

5. In the above gradient, L = 3 and i ∈ {1, 2}
6. Henceforth, we can use these notations in place of numbers to simplify gradient calculation for all possible gradients.