PadhAI: Backpropagation - the full version

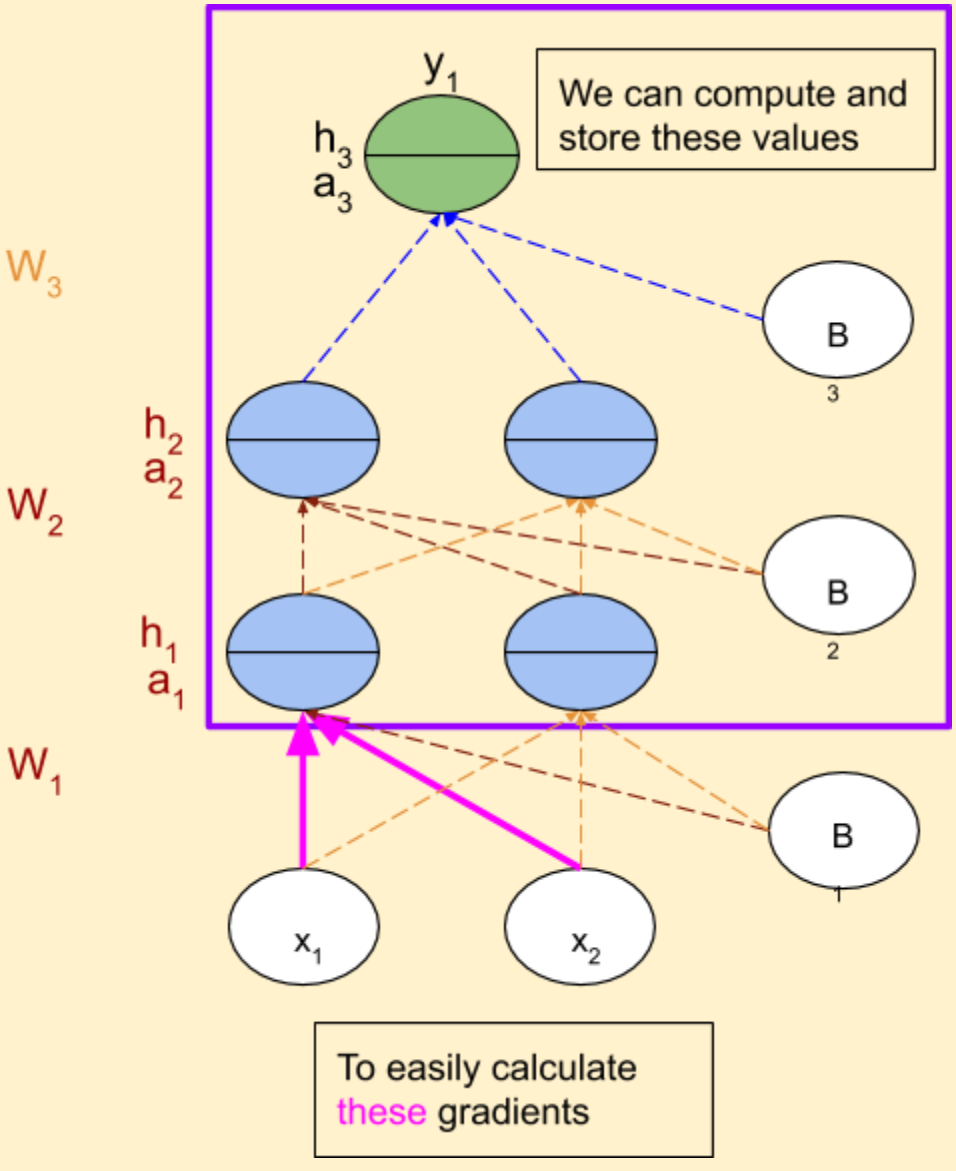
One Fourth Labs

Backpropagation (Math-heavy/Vectorized)

Setting the context

How does this differ from the previous section

1. We've looked at two different levels of complexity to the backpropagation algorithm so far
   a. No-math: A simple forward pass with no gradient calculation
   b. Light-math: Gradient calculation for each weight using chain rule
2. Now, with the Heavy-math version, our objective is to identify common calculations between different weights and re-use them to make our work simpler

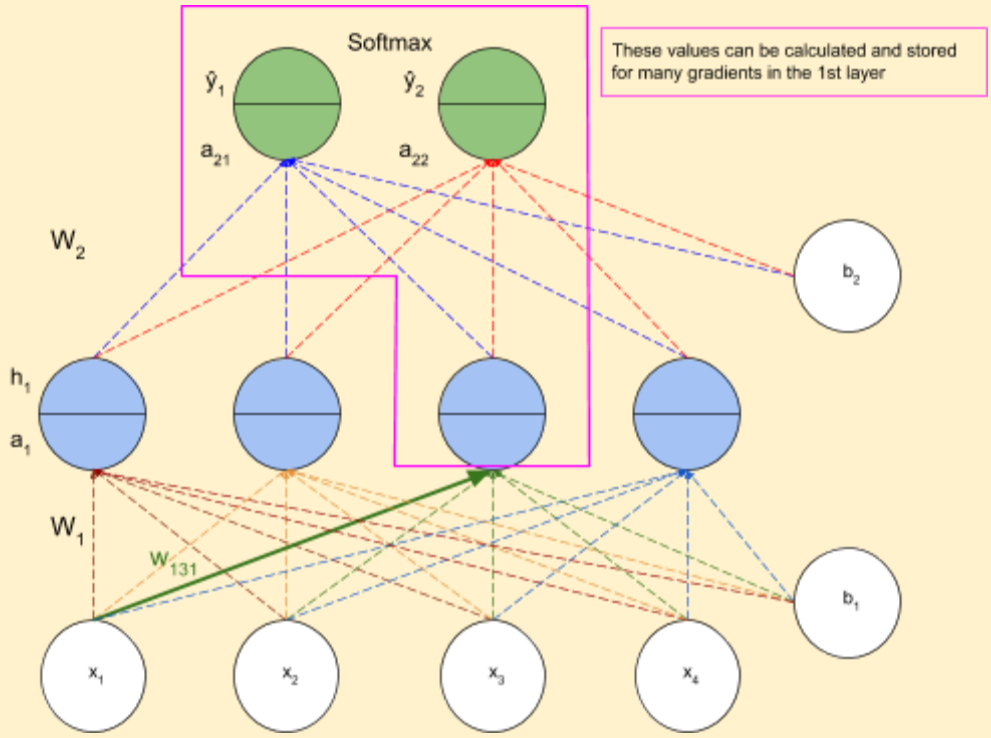


PadhAI: Backpropagation - the full version

One Fourth Labs

---

3. Let us consider the example from the light-math backpropagation chapter



4. Consider dw_{131} or $\frac{\partial L}{\partial w_{131}}$
   a. Here, we are certain about using the highlighted values for gradient computation. So we can pre-calculate and store them
   b. In the outermost layer

$$\frac{\partial L}{\partial a_2} = \begin{bmatrix} \frac{\partial L}{\partial a_{21}} \\ \\ \frac{\partial L}{\partial a_{22}} \end{bmatrix}$$

5. Similarly, storing the values of $\frac{\partial h_1}{\partial a_2}$ will prove useful down the line