

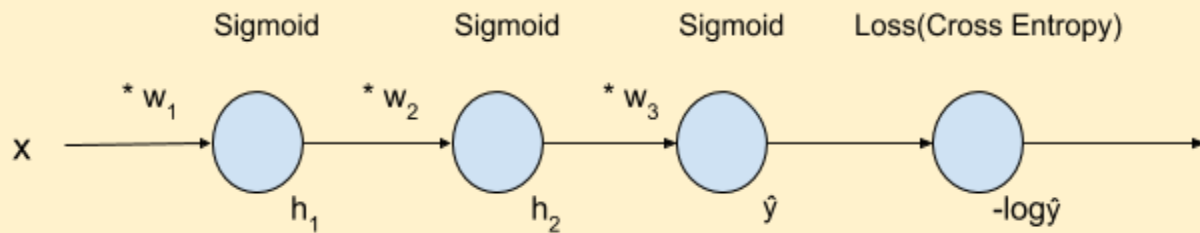
PadhAI: Backpropagation - the light math version

One Fourth Labs

Why do we care about the chain rule of derivatives

Importance of chain rule in Deep Learning

1. Let us look at a sample chain rule flow of a shallow neural network



2. Here, the output \hat{y} is a composite dependent on input x and all of the parameters w
3. *Loss function* : $L = f(x, w_1, w_2, w_3)$
4. Now, for the gradient, we want the derivative of the loss function with respect to the various weights $\frac{\partial L}{\partial w_i}$
5. If we want the derivative w.r.t w_2 then we do the following $\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial w_2}$
6. Here, computation happens from input layer to the output layer ie forward propagation
7. Derivative calculation happens backwards from the output layer to the input, ie back propagation