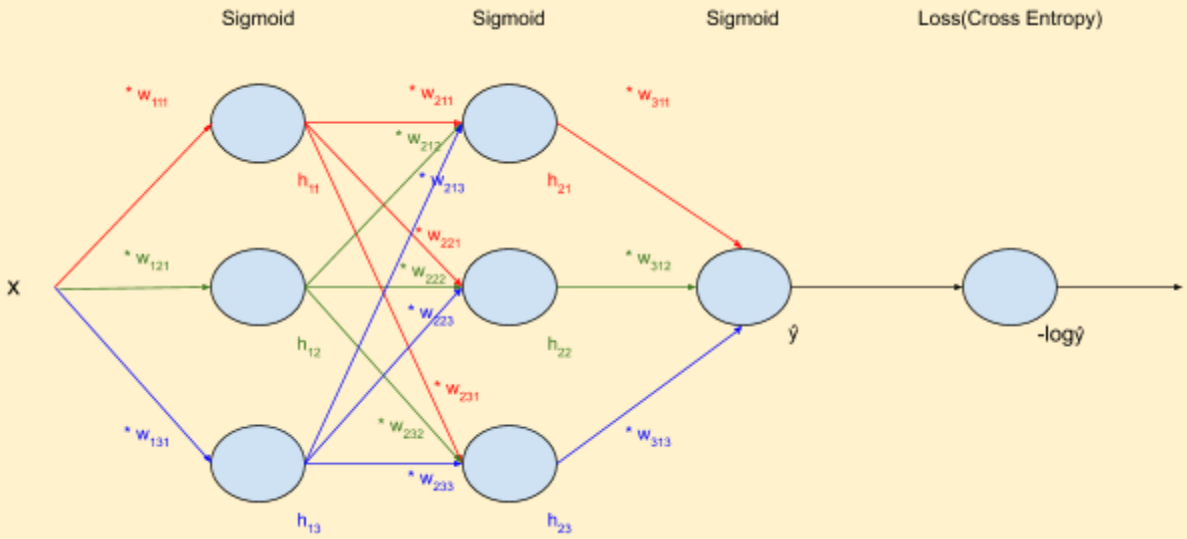
PadhAI: Backpropagation - the light math version

One Fourth Labs

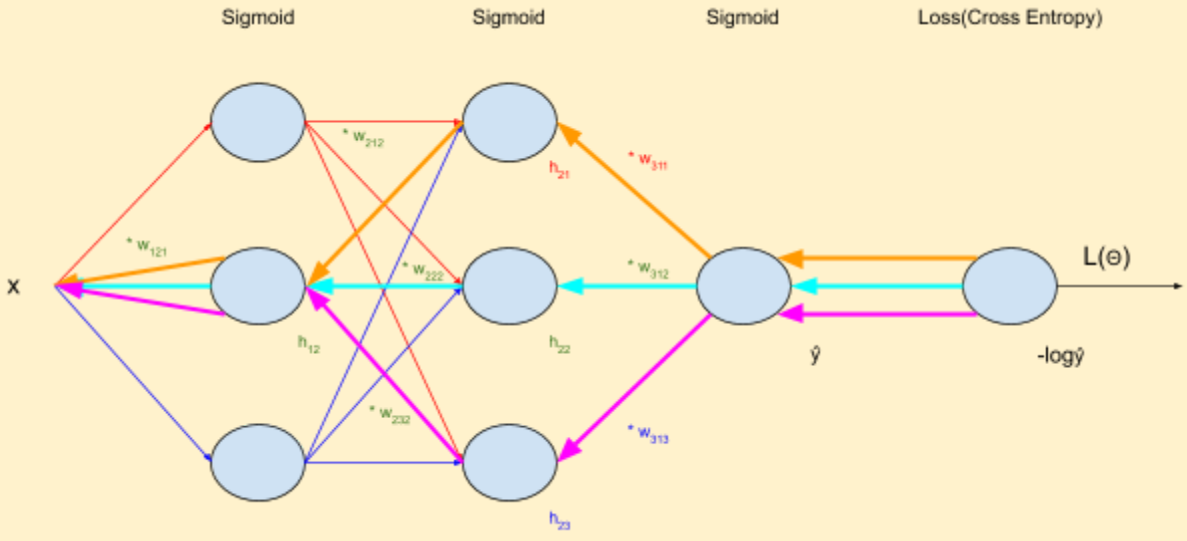
Applying chain rule across multiple paths

Importance of chain rule in deep learning

1. Let us look at a more complex neural network



2. In the shallow Neural Network from the previous example, we apply the chain rule along a straight path. However, in a more practical Neural Network as shown above, the chain rule needs to be applied across multiple parallel paths in order to find a particular gradient
3. For example, to calculate $\frac{\partial L}{\partial w_{121}}$ we need to operate along 3 different paths



4. Summing up the derivatives across the three paths (cyan, orange and pink) will give us the required derivative $\frac{\partial L}{\partial w_{121}}$
5. This scales across as many paths as there are in the neural network.
6. Here, these are not regular derivatives but partial derivatives.