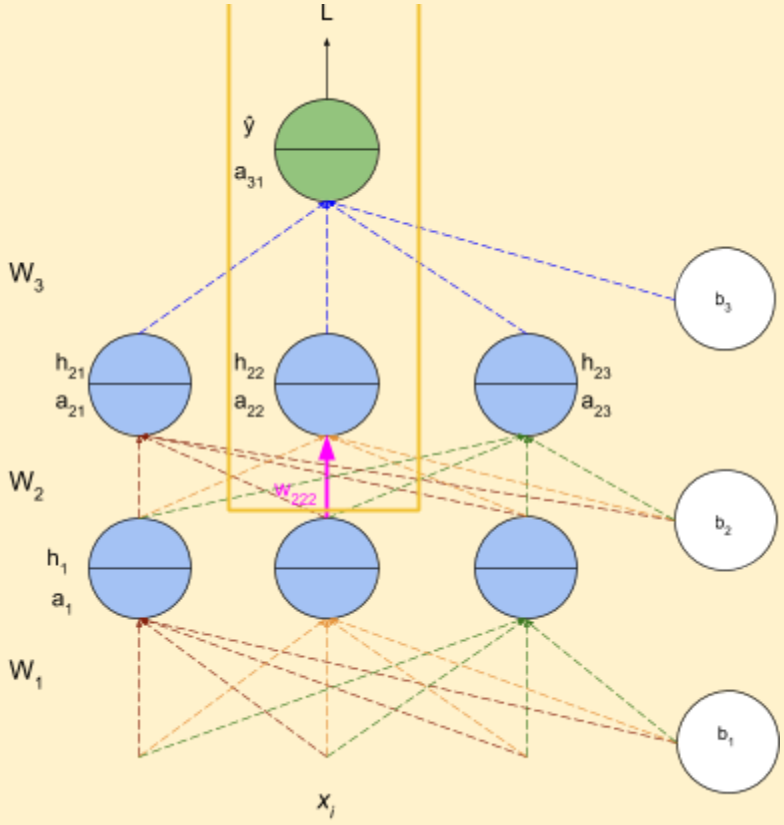
PadhAI: Backpropagation - the light math version

One Fourth Labs

Applying chain rule in a neural network

How many derivatives do we need to compute and how do we compute them?

1. Let's focus on the highlighted weight(w_{222}) of the following neural network



2. To learn this weight, we have to compute the partial derivative w.r.t loss function a

$$(w_{222})_{t+1} = (w_{222})_t - \eta * (\frac{\partial L}{\partial w_{222}})$$

3. We can calculate $\frac{\partial L}{\partial w_{222}}$ as follows
   a. $\frac{\partial L}{\partial w_{222}} = (\frac{\partial L}{\partial a_{22}}).(\frac{\partial a_{22}}{\partial w_{222}})$
   b. $\frac{\partial L}{\partial w_{222}} = (\frac{\partial L}{\partial h_{22}}).(\frac{\partial h_{22}}{\partial a_{22}}).(\frac{\partial a_{22}}{\partial w_{222}})$
   c. $\frac{\partial L}{\partial w_{222}} = (\frac{\partial L}{\partial a_{31}}).(\frac{\partial a_{31}}{\partial h_{22}}).(\frac{\partial h_{22}}{\partial a_{22}}).(\frac{\partial a_{22}}{\partial w_{222}})$
   d. $\frac{\partial L}{\partial w_{222}} = (\frac{\partial L}{\partial \hat{y}}).(\frac{\partial \hat{y}}{\partial a_{31}}).(\frac{\partial a_{31}}{\partial h_{22}}).(\frac{\partial h_{22}}{\partial a_{22}}).(\frac{\partial a_{22}}{\partial w_{222}})$
4. Thus, by breaking the partial derivative into all the subdivisions along that path and multiplying it, we will get the required solution.