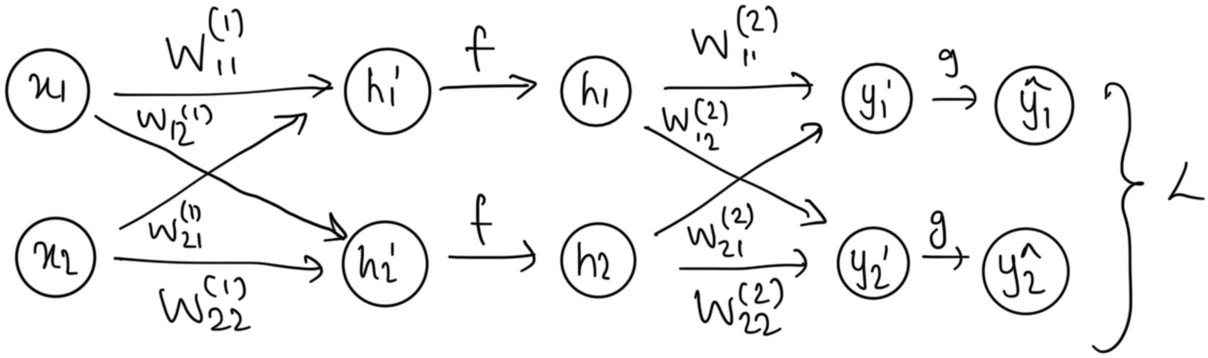

$$h_1' = W_{11}^{(1)} x_1 + W_{21}^{(1)} x_1 \qquad h_1 = f(h_1')$$
$$h_2' = W_{12}^{(1)} x_1 + W_{22}^{(1)} x_2 \qquad h_2 = f(h_2')$$

$$y_1' = W_{11}^{(2)} h_1 + W_{21}^{(2)} h_2 \qquad \hat{y}_1 = g(y_1')$$
$$y_2' = W_{12}^{(2)} h_1 + W_{22}^{(2)} h_2 \qquad \hat{y}_2 = g(y_2')$$

Loss function: $L(y, \hat{y}_1, \hat{y}_2)$

$$\frac{\partial L}{\partial W_{jK}^{(2)}} = \frac{\partial L}{\partial \hat{y}_K} \frac{\partial \hat{y}_K}{\partial y_K'} \frac{\partial y_K'}{\partial W_{jK}^{(2)}} = \frac{\partial L}{\partial \hat{y}_K} g'(y_K') h_j$$

Example

$$\frac{\partial L}{\partial W_{11}^{(2)}} = \frac{\partial L}{\partial \hat{y}_1} \frac{\partial \hat{y}_1}{\partial W_{11}^{(2)}} + \frac{\partial L}{\partial \hat{y}_2} \underbrace{\frac{\partial \hat{y}_2}{\partial W_{11}^{(2)}}}_{=0} \quad \hat{y}_2 = g(W_{12}^{(2)} h_1 + W_{12}^{(2)} h_2)$$

$$= \frac{\partial L}{\partial \hat{y}_1} \frac{\partial}{\partial W_{11}^{(2)}} \left[g(W_{11}^{(2)} h_1 + W_{21}^{(2)} h_2) \right]$$

$$= \frac{\partial L}{\partial \hat{y}_1} \cdot g'(W_{11}^{(2)} h_1 + W_{21}^{(2)} h_2) \cdot h_1$$

$$= \frac{\partial L}{\partial \hat{y}_1} \cdot g'(y_1') \cdot h_1$$

Similarly, $\dfrac{\partial L}{\partial W_{12}^{(2)}} = \dfrac{\partial L}{\partial \hat{y_1}} \underbrace{\dfrac{\partial \hat{y_1}}{\partial W_{12}^{(2)}}}_{=0} + \dfrac{\partial L}{\partial \hat{y_2}} \dfrac{\partial \hat{y_2}}{\partial W_{12}^{(2)}}$

since $\hat{y_1} = g\left(W_{11}^{(2)} h_1 + W_{21}^{(2)} h_2\right)$

$$= \frac{\partial L}{\partial \hat{y_2}} \frac{\partial}{\partial W_{12}^{(2)}} \left[g\left(W_{12}^{(2)} h_1 + W_{22}^{(2)} h_2\right) \right]$$

$$= \frac{\partial L}{\partial \hat{y_2}} g'\left(W_{12}^{(2)} h_1 + W_{22}^{(2)} h_2\right) h_1$$

$$= \frac{\partial L}{\partial \hat{y_2}} g'(y_2') h_1$$

$$\frac{\partial L}{\partial W_{jk}^{(1)}} = \sum_i \frac{\partial L}{\partial \hat{y_i}} \frac{\partial \hat{y_i}}{\partial y_i'} \frac{\partial y_i'}{\partial h_k} \frac{\partial h_k}{\partial h_k'} \frac{\partial h_k'}{\partial W_{jk}^{(1)}}$$

$$= \sum_i \frac{\partial L}{\partial \hat{y_i}} g'(y_i') W_{ki}^{(2)} f'(h_k') x_j$$

Example

$$\frac{\partial L}{\partial W_{12}^{(1)}} = \frac{\partial L}{\partial \hat{y_1}} \frac{\partial \hat{y_1}}{\partial W_{12}^{(1)}} + \frac{\partial L}{\partial \hat{y_2}} \frac{\partial \hat{y_2}}{\partial W_{12}^{(1)}}$$

Here, both $\hat{y_1}$ and $\hat{y_2}$ are affected by $W_{12}^{(1)}$ as follows:

$$\hat{y_1} = g(y_1')$$

$$= g\left(W_{11}^{(2)} h_1 + W_{21}^{(2)} h_2\right)$$

$$= g\left(W_{11}^{(2)} f(h_1') + W_{21}^{(2)} f(h_2')\right)$$

$$= g\left(W_{11}^{(2)} f\left(W_{11}^{(1)} x_1 + W_{21}^{(1)} x_2\right) + W_{21}^{(2)} f\left(W_{12}^{(1)} x_1 + W_{22}^{(1)} x_2\right)\right)$$

Similarly

$$\hat{y_2} = g\left(W_{12}^{(2)} f\left(W_{11}^{(1)} x_1 + W_{21}^{(1)} x_2\right) + W_{22}^{(2)} f\left(W_{12}^{(1)} x_1 + W_{22}^{(1)} x_2\right)\right)$$

$$\Rightarrow \frac{\partial \hat{y_1}}{\partial W_{12}^{(1)}} = g'\left(W_{11}^{(2)} f(W_{11}^{(1)} x_1 + W_{21}^{(1)} x_2) + W_{21}^{(2)} f(W_{12}^{(1)} x_1 + W_{22}^{(1)} x_2)\right) \times W_{21}^{(2)} \times f'(W_{12}^{(1)} x_1 + W_{22}^{(1)} x_2) \times x_1$$

$$= g'(y_1')\, W_{21}^{(2)}\, f'(h_2')\, x_1$$

Similarly $\dfrac{\partial \hat{y_2}}{\partial w_{12}^{(1)}} = g'(y_2')\, W_{22}^{(2)}\, f'(h_2')\, x_1$

$$\therefore \frac{\partial L}{\partial W_{12}^{(1)}} = \frac{\partial L}{\partial \hat{y_1}}\, g'(y_1')\, W_{21}^{(2)}\, f'(h_2')\, x_1 + \frac{\partial L}{\partial \hat{y_2}}\, g'(y_2')\, W_{22}^{(2)}\, f'(h_2')\, x_1$$

If you wish to find gradient wrt bias, the same formula holds, only for bias, the i/p becomes 1