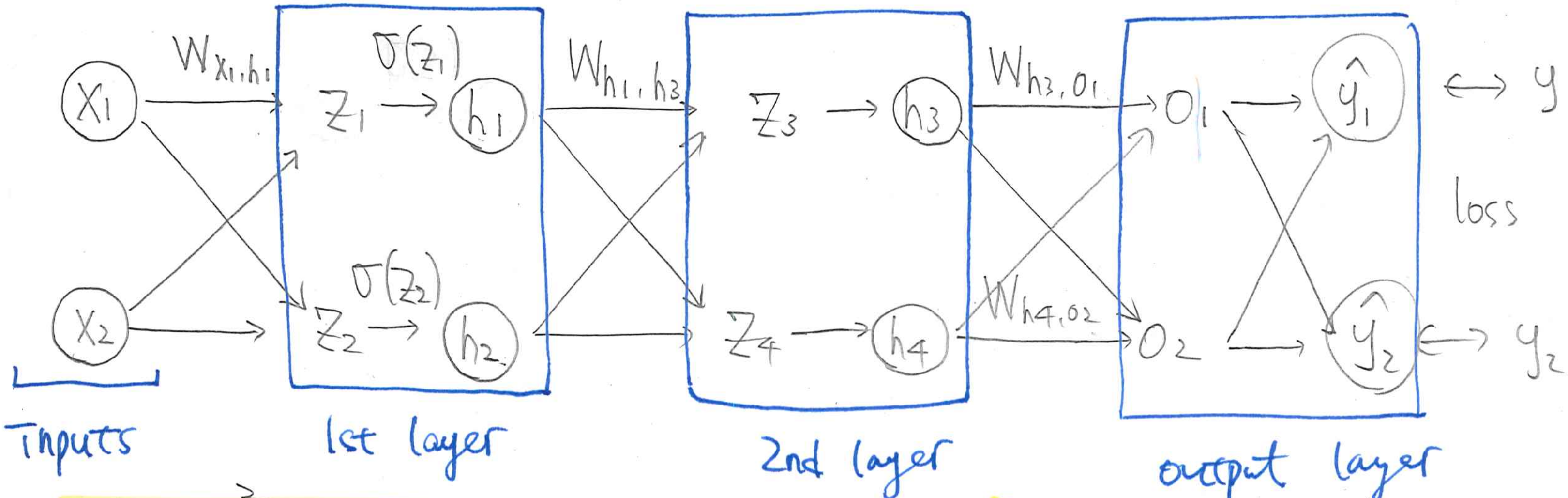


$$L(y, \hat{y}) = E_y(-\log \hat{y}) = -\sum_{i=1}^{2} y_i \log \hat{y}_i$$

$$z_1 = \sum_{i=1}^{2} x_i \cdot W_{x_i,h_1} = (x_1 \cdot W_{x_1,h_1} + x_2 \cdot W_{x_2,h_1})$$

$$h_1 = \sigma(z_1) = \frac{1}{1+e^{-z_1}}$$

$$o_1 = \sum_{i=3}^{4} h_i \cdot W_{h_i,o_1} = (h_3 \cdot W_{h_3,o_1} + h_4 \cdot W_{h_4,o_1})$$

$$\hat{y}_1 = \frac{e^{o_1}}{e^{o_1}+e^{o_2}} \qquad \hat{y}_2 = \frac{e^{o_2}}{e^{o_1}+e^{o_2}}$$

Goal: $$\frac{\partial L}{\partial W_{x_1,h_1}} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial W_{x_1,h_1}} = \frac{\partial L}{\partial z_1} \cdot x_1$$

obtain from forward pass

(similarly: $\frac{\partial z_3}{\partial W_{h_1,z_3}} = h_1$, and so on)

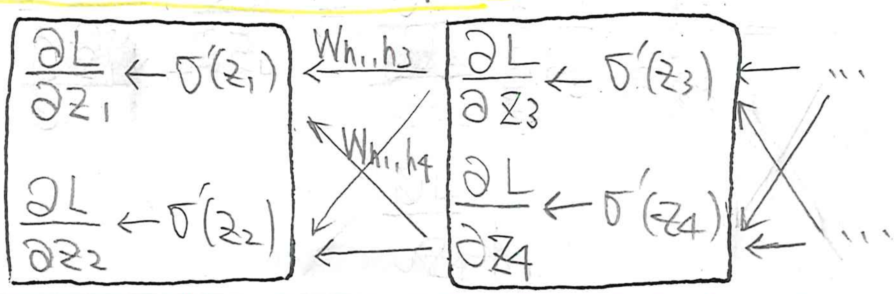
(try $\frac{\partial L}{\partial W_{x_2,h_1}}$ yourself)

backward pass

need to calculate

$$\frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} = \left(\frac{\partial L}{\partial z_3} \cdot \underbrace{\frac{\partial z_3}{\partial h_1}}_{= W_{h_1,h_3}} + \frac{\partial L}{\partial z_4} \cdot \underbrace{\frac{\partial z_4}{\partial h_1}}_{= W_{h_1,h_4}}\right) \cdot \underbrace{\frac{\partial h_1}{\partial z_1}}_{= \sigma'(z_1)}$$

So let's illustrate backward pass:



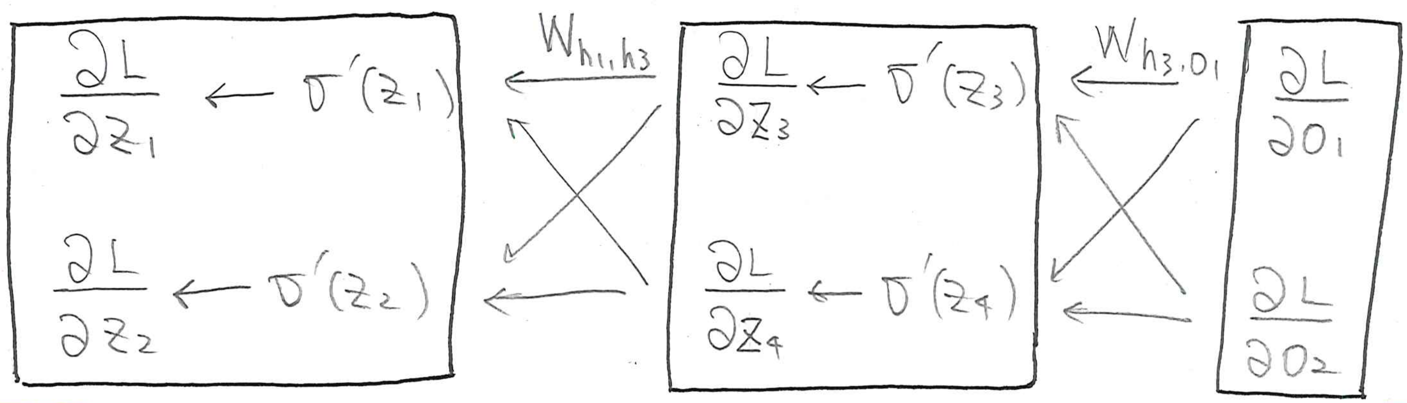
Continue on backward pass: need to calculate

$$\frac{\partial L}{\partial z_3} = \frac{\partial L}{\partial h_3}\cdot\frac{\partial h_3}{\partial z_3} = \left(\frac{\partial L}{\partial o_1}\cdot\underbrace{\frac{\partial o_1}{\partial h_3}}_{=W_{h_3,o_1}} + \frac{\partial L}{\partial o_2}\cdot\underbrace{\frac{\partial o_2}{\partial h_3}}_{=W_{h_3,o_2}}\right)\cdot\underbrace{\frac{\partial h_3}{\partial z_3}}_{=\sigma'(z_3)}$$

$$\frac{\partial L}{\partial o_1} = \frac{\partial L}{\partial \hat{y}_1}\cdot\frac{\partial \hat{y}_1}{\partial o_1} + \frac{\partial L}{\partial \hat{y}_2}\cdot\frac{\partial \hat{y}_2}{\partial o_1} = -\frac{y_1}{\hat{y}_1}\cdot\left[\frac{e^{o_1}}{e^{o_1}+e^{o_2}} - \left(\frac{e^{o_1}}{e^{o_1}+e^{o_2}}\right)^2\right] - \frac{y_2}{\hat{y}_2}\left[\cdots\right]$$

So complete backward pass:

$\longleftarrow$ from output to input



Summary

backward propagation has two phases

- forward pass: calculate $\boxed{\frac{\partial z_k}{\partial W_{x_i,h_j}}}$
- backward pass: calculate $\boxed{\frac{\partial L}{\partial z_k}}$