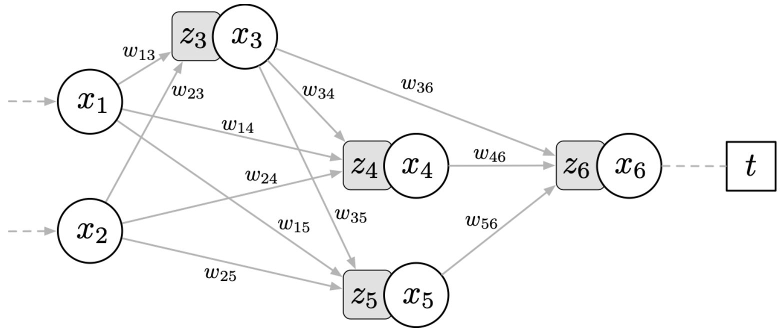
A3-Q1: Backprop



a) $L(x,t) = \frac{t}{x} - \ln\frac{t}{x} - 1$ and $\rho(z) = z^2$

$$\frac{\partial L}{\partial z_6} = \frac{\partial L}{\partial x_6}\frac{dx_6}{dz_6} = \frac{\partial L}{\partial x_6}\rho'(z_6)$$

Deriv.

What is $\frac{\partial L}{\partial x}$?

$$L(x,t) = \frac{t}{x} + \ln\frac{x}{t} - 1$$

$$\therefore \frac{\partial L}{\partial x} = \frac{-t}{x^2} + \frac{t}{x}\frac{1}{t} = \frac{-t}{x^2} + \frac{1}{x}$$

What is $\rho'(z)$? $\longrightarrow \rho'(z) = 2z$

$$\therefore \frac{\partial L}{\partial z_6} = \left(\frac{-t}{x_6^2} + \frac{1}{x_6}\right)2z_6$$

$$= \left(\frac{-t}{z_6^4} + \frac{1}{z_6^2}\right)2z_6$$

$$= \boxed{2z_6\left(\frac{x_6 - t}{x_6^2}\right)} \text{ or } \boxed{2\,\frac{x_6 - t}{z_6^3}} \text{ or } \boxed{2\,\frac{z_6^2 - t}{z_6^3}}$$

or any algebraically equivalent that is reasonably simplified.

b) Update for w_{46}.

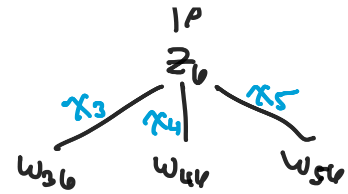
Grad. descent is $w_{46} \leftarrow w_{46} - K\frac{\partial L}{\partial w_{46}}$

What is $\frac{\partial L}{\ }$?

L
$|$
x_6

∂w_{46}

$$\frac{\partial L}{\partial w_{46}} = \underbrace{\frac{\partial L}{\partial x_6}\frac{dx_6}{dz_6}}_{\text{from (a)}} \boxed{\frac{\partial z_6}{\partial w_{46}}}$$



$z_6 = x_3 w_{36} + x_4 w_{46} + x_5 w_{56} + b_6$

$\therefore \frac{\partial z_6}{\partial w_{46}} = x_4$

For the rest of these questions, it is ok to use $\frac{\partial L}{\partial z_6}$ here. In general, they may reuse any quantity that they already defined.

$$= \left(2\,\frac{x_6 - t}{z_6^3}\right) x_4$$

$$\boxed{\therefore\ w_{46} \leftarrow w_{46} - k\left(2\,\frac{x_6 - t}{z_6^3}\right) x_4}$$

c) $b_6 \leftarrow b_6 - k\,\frac{\partial L}{\partial b_6}$

$$\frac{\partial L}{\partial b_6} = \frac{\partial L}{\partial z_6}\frac{\partial z_6}{\partial b_6}$$

$\frac{\partial z_6}{\partial b_6} = 1$

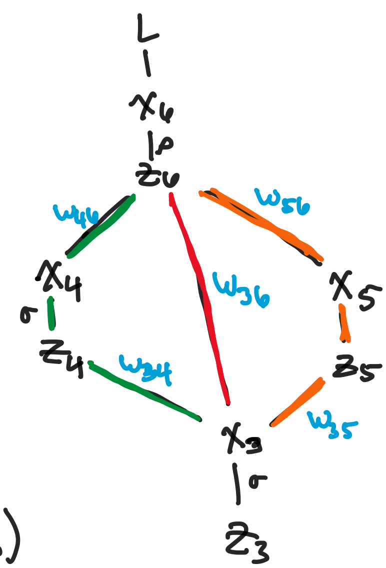
$$= 2\,\frac{x_6 - t}{z_6^3}$$

$$\boxed{\therefore\ b_6 \leftarrow b_6 - 2k\left(\frac{x_6 - t}{z_6^3}\right)}$$

d) We want $\frac{\partial L}{\partial z_3}$.

Did they include all 3 paths?

There are 3 paths to z_3.



$$\frac{\partial L}{\partial z_3} = \frac{\partial L}{\partial z_6}\left(\frac{\partial z_6}{\partial z_4}\frac{\partial z_4}{\partial x_3} + \frac{\partial z_6}{\partial z_5}\frac{\partial z_5}{\partial x_3} + \frac{\partial z_6}{\partial x_3}\right)\frac{dx_3}{dz_3}$$

$$= \frac{\partial L}{\partial z_6}\left(\frac{\partial z_6}{\partial x_4}\frac{dx_4}{dz_4}\frac{\partial z_4}{\partial x_3} + \frac{\partial z_6}{\partial x_5}\frac{dx_5}{dz_5}\frac{\partial z_5}{\partial x_3} + w_{36}\right)\sigma'(z_3)$$

$$= \frac{\partial L}{\partial z_6}\left(w_{46}\,\sigma'(z_4)\,w_{34} + w_{56}\,\sigma'(z_5)\,w_{35} + w_{36}\right)\sigma'(z_3)$$

Did they apply chain rule correctly?

$$\boxed{\therefore\ \frac{\partial L}{\partial z_3} = 2\,\frac{x_6 - t}{z_6^3}\left(w_{46}\,\sigma'(z_4)\,w_{34} + w_{56}\,\sigma'(z_5)\,w_{35} + w_{36}\right)\sigma'(z_3)}$$

The order of these terms doesn't matter.