



$$\frac{dL}{dL} = 1$$

$$\frac{dL}{Z_1} = 1$$

$$\frac{dL}{Z_2} = 1$$

$$\frac{dL}{Z_3} = 1$$

$$\frac{dL}{Z_4} = 1$$

$$\frac{dL}{P_{Z_1}} = 0$$

$$\frac{dL}{P_{Z_2}} = 0$$

$$\frac{dL}{P_{Z_3}} = 1$$

$$\frac{dL}{P_{Z_4}} = 0$$

$$\frac{dL}{P_1} = 0$$

$$\frac{dL}{P_2} = 0$$

$$\frac{dL}{P_2} = \frac{dP_{Z_3}}{dP_3} \times \frac{dL}{P_{Z_3}}$$

$$\frac{dL}{P_4} = 0$$

$$= \frac{-1}{P_3} \times 1 = \frac{-1}{P_3}$$

$$\frac{dL}{c_1} = \frac{dP}{dc_1} \times \frac{dL}{D} + 0$$

$$\frac{dL}{c_2} = \frac{dP}{dc_2} \times \frac{dL}{D} + 0$$

$$\frac{dL}{c_4} = \frac{dP}{dc_2} \times \frac{dL}{D} + 0$$

$$\frac{dL}{D} = (c_3 \times -D^{-2}) \left(\frac{-1}{P_3} \right)$$

$$\frac{dL}{c_3} = \frac{1}{D} \left(\frac{-1}{P_3} \right)$$

$$\frac{dL}{c_3} = \frac{dD}{dc_3} \times \frac{dL}{D} + \frac{1}{D} \left(\frac{-1}{P_3} \right)$$

Will give positive
gradients, thus decrease
wrong logits

will give negative
gradients, thus
increase correct
logits