① $\dfrac{\partial L}{\partial w_1} = \dfrac{\partial L}{\partial h_3} f_3' w_3 f_1' x \qquad \dfrac{\partial L}{\partial w_2} = \dfrac{\partial L}{\partial h_3} f_3' w_4 f_2' x$

$\dfrac{\partial L}{\partial w_3} = \dfrac{\partial L}{\partial h_3} f_3' h_1 \qquad \dfrac{\partial L}{\partial w_4} = \dfrac{\partial L}{\partial h_3} f_3' h_2$

$\dfrac{\partial L}{\partial w_4} = \dfrac{\partial L}{\partial h_3} f_3' h_2 \qquad \dfrac{\partial L}{\partial b_1} = \dfrac{\partial L}{\partial h_3} f_3' w_3 f_1' \qquad \dfrac{\partial L}{\partial b_2} = \dfrac{\partial L}{\partial h_3} f_3' w_4 f_2'$

$\dfrac{\partial L}{\partial b_3} = \dfrac{\partial L}{\partial h_3} f_3' \qquad \dfrac{\partial L}{\partial x} = \dfrac{\partial L}{\partial h_3} f_3' (w_3 f_1' w_1 + w_4 f_2')$

② 1. $x \to h = 2x + 1$

$h \to h^2$

$x^2 + h^2 \to z^2$

$x \to x^2$

$h \to e^{th} = 1 + e^{-h} = y = \dfrac{1}{1 + e^{-h}}$

2. $\dfrac{\partial y}{\partial x} = 2y(1-y)$

③ 1. MLP without ReLU simplifies to a linear transformation $y = wx$, as a stack of linear layers simplifies

2. A CNN without ReLU reduces to the same linear operation as convolution is a stacking of linear operations

④ We need to normalize the inputs as they differ on the scale they use.

⑤ 1. $y \geq 0 = $ Softplus(z)

$y \leq 0 = $ Softplus(z) $\cdot -1$

$a \leq y \leq b = a + (b-a) \, \delta(z)$

⑥ 1. MLP with ReLU is picewise linear as each layer is linear

2. 1D CNN with ReLU is also piecewise linear for the same reason,

its composed of exclusively linear operations.