(2)

Given data

input nodes : N

hidden nodes : k

output nodes = N

Activation function as follows:

$z = W_1 x + b_1$ → (1)

$h = S(z)$ → (2)

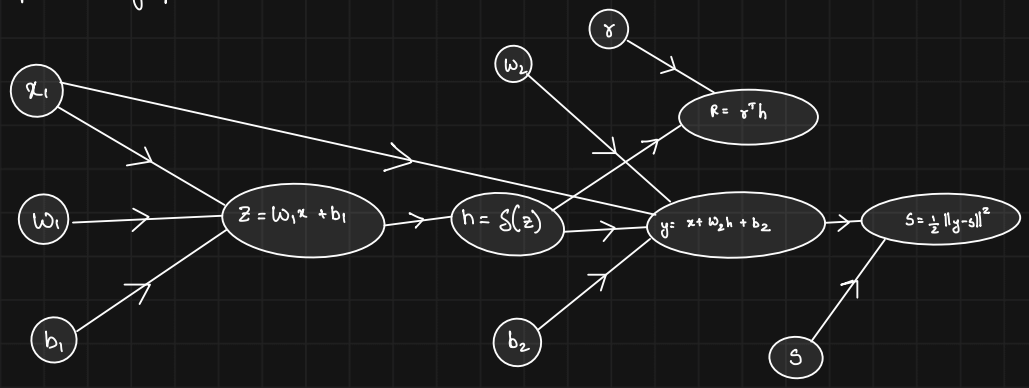
$y = x + W_2 h + b_2$ → (3)

The cost function $\boxed{E = R + S}$

where $R = \gamma^T h$ → (4)

$S = \frac{1}{2}\|y - s\|^2$ → (5)

(a) Computational graph



(b) $\frac{\partial E}{\partial x} = \frac{\partial R}{\partial x} + \frac{\partial S}{\partial x}$

$= \frac{\partial R}{\partial h}\frac{\partial h}{\partial x} + \frac{\partial S}{\partial y}\cdot\frac{dy}{dx}$

$= \frac{\partial R}{\partial h}\cdot\frac{\partial h}{\partial z}\cdot\frac{\partial z}{\partial x} + \frac{\partial S}{\partial y}\left(\frac{\partial h}{\partial x} + 1\right)$

$= \frac{\partial R}{\partial h}\cdot\frac{\partial h}{\partial z}\cdot\frac{\partial z}{\partial x} + \frac{\partial S}{\partial y}\left(\frac{\partial h}{\partial z}\cdot\frac{\partial z}{\partial x} + 1\right)$

$= \boxed{\gamma^T \cdot S'(z) \cdot W_1 + (y - s)\left(S'(z) \cdot W_1 + 1\right)}$