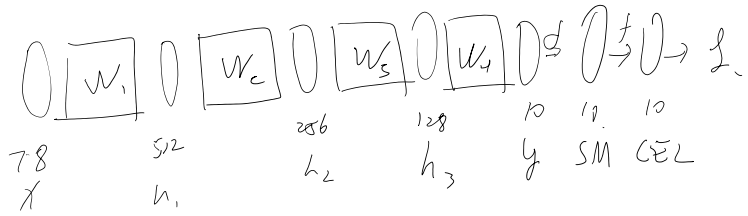

Crossentropy Error Function and Softmax

Effect on Gradient Descent

- $E_x = -\sum_k [t_k^x \log(o_k^x) + (1 - t_k^x) \log(1 - o_k^x)]$
- Crossentropy derivative: $\frac{\partial E_x}{\partial o_j^x} = -\frac{t_j^x}{o_j^x} + \frac{1-t_j^x}{1-o_j^x}$
- Softmax derivative: $\frac{\partial o_j^x}{\partial net_j^x} = o_j^x(1 - o_j^x)$
- $\frac{\partial E_x}{\partial net_j^x} = \left(-\frac{t_j^x}{o_j^x} + \frac{1-t_j^x}{1-o_j^x}\right)(o_j^x(1 - o_j^x)) = -t_j^x + t_j^x o_j^x + o_j^x - t_j^x o_j^x$
- $\frac{\partial E_x}{\partial net_j^x} = -t_j^x + o_j^x \quad \Longrightarrow \quad \delta_j = t_j^x - o_j^x$

2 / 3

$$\min_W L_{CEL} \rightarrow \frac{\partial L_{CEL}}{\partial SM} \cdot \frac{\partial SM}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W_4}$$

$$\frac{\partial L}{\partial W_4} = y - O(y)$$

$$\frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial W_4} \cdot \frac{\partial W_4}{\partial W_3}$$

h_3