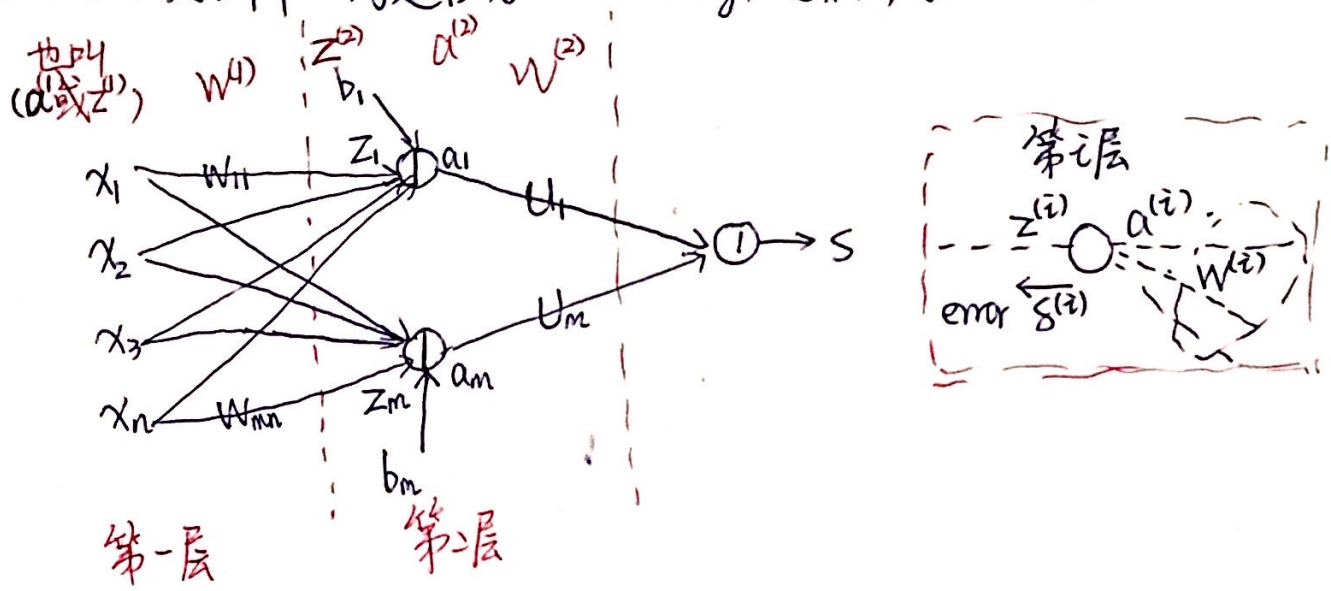
Backward Propagation 反向传播

NER task：判断中心词是否为 Named Entity，是输出高分数。



① Elemental Back-Prop:

$\theta^{(t+1)} = \theta^{(t)} - \alpha \nabla_{\theta^{(t)}} J$

($w_{ij}^{(i)}$ 中 i 是对应 (i+1) 层，j 是 i 层)

$$\frac{\partial s}{\partial w_{ij}^{(1)}} = \frac{\partial w^{(2)} a^{(2)}}{\partial w_{ij}^{(1)}} = \frac{\partial w_i^{(2)} a_i^{(2)}}{\partial w_{ij}^{(1)}} = w_i^{(2)} \frac{\partial a_i^{(2)}}{\partial w_{ij}^{(1)}} = w_i^{(2)} \frac{\partial a_i^{(2)}}{\partial z_i^{(2)}} \frac{\partial z_i^{(2)}}{\partial w_{ij}^{(1)}} = \underbrace{w_i^{(2)} f'(z_i^{(2)})}_{\delta_i^{(2)}} a_j^{(1)}$$

bias update (share error/gradients until $\delta_i^{(2)}$)

$$\frac{\partial s}{\partial b_i^{(1)}} = \delta_i^{(2)} \frac{\partial z_i^{(2)}}{\partial b_i^{(1)}} = \delta_i^{(2)} \cdot 1 = \delta_i^{(2)}$$

偏导数矩阵 Jacobian Matrix 雅可比矩阵 { input-dim: n; output-dim: m; Jab-dim: m×n }

② vectorized Back-Prop:

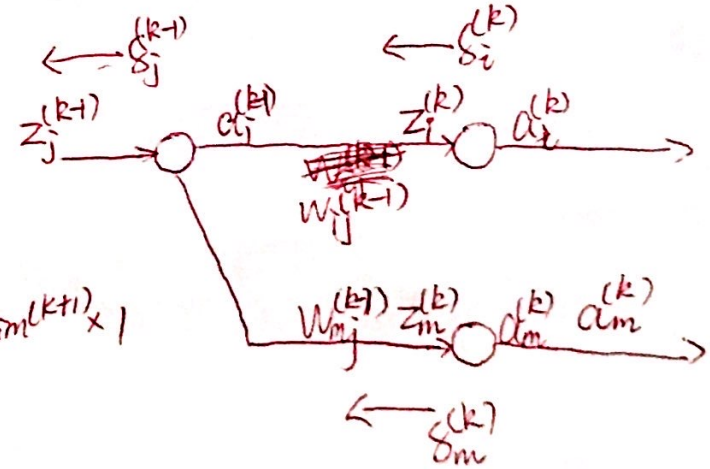
$$\nabla w^{(k)} = \begin{bmatrix} \delta_1^{(k+1)} a_1^{(k)} & \delta_1^{(k+1)} a_2^{(k)} & \cdots \\ \delta_2^{(k+1)} a_1^{(k)} & \delta_2^{(k+1)} a_2^{(k)} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \delta^{(k+1)} a^{(k)T}$$

$$\nabla b^{(k)} = \delta^{(k+1)} = \left[\delta_1^{(k+1)}, \delta_2^{(k+1)}, \cdots\right]^T$$

$$\delta_j^{(k)} = f'(z_j^{(k)}) \cdot \sum_i \delta_i^{(k+1)} w_{ij}^{(k)}$$

加法的导数也是加法

chain rule



$$\delta^{(k)} = f'(z^{(k)}) \circ (w^{(k)T} \delta^{(k+1)}) \quad \#$$

$\dim^{(k)} \times 1$; $\dim^{(k+1)} \times \dim^{(k)}$; $\dim^{(k+1)} \times 1$

$\circ$: element-wise dot product