

1. Page 5-6。请默写出多层感知神经网络的递归计算公式，即 h_l 和 s_l 的计算公式（Page 5）。请推导出对 s_l 和 h_{l-1} 的反向传播的一般形式，要有推导过程（见Page 6）。给大家降低难度，对 W_l 的梯度的公式就不要要求默写了，目的是让大家记住基本的运算规则。

$$h_l = f_l(s_l)$$

$$s_l = W_l h_{l-1} + b_l$$

$$\begin{aligned}\frac{\partial L}{\partial s_l^\top} &= \frac{\partial L}{\partial h_l^\top} \frac{\partial h_l}{\partial s_l^\top} \\ &= \frac{\partial L}{\partial h_l^\top} f'_l \\ \frac{\partial L}{\partial h_{l-1}^\top} &= \frac{\partial L}{\partial h_l^\top} \frac{\partial h_l}{\partial s_l^\top} \frac{\partial s_l}{\partial h_{l-1}^\top} \\ &= \frac{\partial L}{\partial h_l^\top} f'_l W_l\end{aligned}$$

2. Page 7，给出当 $f_l = \text{sigmoid}$ 函数，和 $f_l = \text{ReLU}$ 函数时，两种不同的 f'_l 所对应的梯度矩阵（是个对角矩阵）的解析式。需要把矩阵的对角线上每个单元的梯度的解析式写一下，而不是仅仅写出一个求导的形式。

sigmoid:

$$f'_l = \frac{\partial h_l}{\partial s_l^\top} = \begin{bmatrix} h_{l1}(1 - h_{l1}) & & \\ & \ddots & \\ & & h_{ln}(1 - h_{ln}) \end{bmatrix}$$

ReLU:

$$f'_l = \frac{\partial h_l}{\partial s_l^\top} = \begin{bmatrix} 1(h_{l1} > 0) & & \\ & \ddots & \\ & & 1(h_{ln} > 0) \end{bmatrix}$$