1.

$$\frac{\partial s_i}{\partial z_i} = e^{z_i}\left(\frac{1}{\sum_k e^{z_k}}\right)' + \frac{e^{z_i}}{\sum_k e^{z_k}}$$

$$= \frac{e^{z_i}}{\sum_k e^{z_k}} - \left(\frac{e^{z_i}}{\sum_k e^{z_k}}\right)^2 = s_i(1-s_i)$$

$$\frac{\partial s_i}{\partial z_j} = e^{z_i}\left(-\frac{e^{z_j}}{\left(\sum_k e^{z_k}\right)^2}\right) = -\frac{e^{z_i}e^{z_j}}{\left(\sum_k e^{z_k}\right)^2} = -s_i \cdot s_j$$

$$\therefore \quad \frac{\partial s_i}{\partial z_j} = \begin{cases} s_i(1-s_i), & \text{if } i=j \\ -s_i s_j, & \text{otherwise} \end{cases}$$

The Jacobian of $\frac{\partial s}{\partial z} = \begin{pmatrix} \frac{\partial s_1}{\partial z_1} & \frac{\partial s_1}{\partial z_2} & \cdots & \frac{\partial s_1}{\partial z_n} \\ \frac{\partial s_2}{\partial z_1} & & \ddots & \vdots \\ \vdots & & & \\ \frac{\partial s_n}{\partial z_1} & & \cdots & \frac{\partial s_n}{\partial z_n} \end{pmatrix}$

$$= \begin{pmatrix} s_1(1-s_1) & -s_1 \cdot s_2 & \cdots \\ -s_2 \cdot s_1 & s_2(1-s_2) & \\ \vdots & & \ddots \\ & & & s_n(1-s_n) \end{pmatrix}$$

2. $W_{AND} = (1, 1)$

$b_{AND} = -1.5$

$W_{OR} = (1, 1)$

$b_{OR} = -0.5$

3. If XOR can be represented using a linear model

$W_{XOR} = (w_1, w_2)$

$b_{XOR} = b$

We have

$$\begin{cases} b<0 & ① \\ w_1x_1 + b \geq 0 & ② \\ w_2x_2 + b \geq 0 & ③ \\ w_1x_1 + w_2x_2 + b < 0 & ④ \end{cases}$$

② + ③ − ④ : $b > 0$ contradicts with ①

$\therefore$ XOR can not be represented using a linear model.