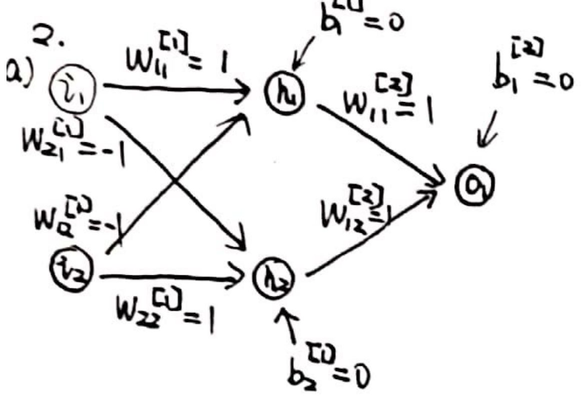
2.

a)



i_1	i_2	$z_1^{[1]}, h_1$	$z_2^{[1]}, h_2$	$z_1^{[2]}, o_1$
1	1	0, 0	0, 0	0, 0
1	0	1, 1	-1, 0	1, 1
0	1	-1, 0	1, 1	1, 1
0	0	0, 0	0, 0	0, 0

∴ 当且仅当 $i_1=1, i_2=0$ 或 $i_1=0, i_2=1$ 时 o_1 为1，即实现了 $o_1 = i_1 \text{ XOR } i_2$

b) 若网络输入限制为两输入 (x_1, x_2)，且激活函数为线性函数，则整个神经网络等价为一个线性函数：

$$y = a_0 + a_1 x_1 + a_2 x_2$$

当 $x_1=0, x_2=0$，则 $y_0 = a_0 = 0$

当 $x_1=1, x_2=0$，则 $y_1 = a_0 + a_1 = 1$

当 $x_1=0, x_2=1$，则 $y_2 = a_0 + a_2 = 1$

当 $x_1=1, x_2=1$，则 $y_3 = a_0 + a_1 + a_2 = 0$

无解（即使把 $y>0$ 定义为1也无法解决，因为 $y_0+y_3=y_1+y_2$）

∴ 无法解决异或问题

4.

a) $z_1^{[1]} = w_1 i_1 + w_2 i_2 + b_1 = 0.15\times0.05 + 0.2\times0.1 + 0.35 = 0.3775$

$z_2^{[1]} = w_3 i_1 + w_4 i_2 + b_1 = 0.25\times0.05 + 0.3\times0.1 + 0.35 = 0.3925$

$h_1 = \frac{1}{1+e^{-z_1^{[1]}}} \doteq 0.5933$ $\quad h_2 = \frac{1}{1+e^{-z_2^{[1]}}} \doteq 0.5969$

$z_1^{[2]} = w_5 h_1 + w_7 h_2 + b_2$

$= 0.3\times0.5933 + 0.45\times0.5969 + 0.6 = 1.047$

$z_2^{[2]} = w_6 h_1 + w_8 h_2 + b_2$

$= 0.5\times0.5933 + 0.55\times0.5969 + 0.6 = 1.225$

$\hat{o}_1 = \frac{1}{1+e^{-z_1^{[2]}}} \doteq 0.740$

$\hat{o}_2 = \frac{1}{1+e^{-z_2^{[2]}}} \doteq 0.773$

b) $J = \frac{1}{2}(\hat{o}_1 - o_1)^2 + \frac{1}{2}(\hat{o}_2 - o_2)^2$

$$\frac{\partial J}{\partial w_5} = \frac{\partial J}{\partial \hat{o}_1}\frac{\partial \hat{o}_1}{\partial z_1^{[2]}}\frac{\partial z_1^{[2]}}{\partial w_5} = (\hat{o}_1 - o_1)\hat{o}_1(1-\hat{o}_1)h_1$$

$= (0.740 - 0.12)\times0.740\times(1-0.740)\times0.5933 \approx 0.0708$

$$\frac{\partial J}{\partial w_6} = \frac{\partial J}{\partial \hat{o}_2}\frac{\partial \hat{o}_2}{\partial z_2^{[2]}}\frac{\partial z_2^{[2]}}{\partial w_6} = (\hat{o}_2 - o_2)\times\hat{o}_2(1-\hat{o}_2)h_1$$

$= (0.773 - 0.95)\times0.773\times(1-0.773)\times0.5933 \approx -0.0184$

c) $w_5 = w_5 - \alpha\frac{\partial J}{\partial w_5}$

≈ 0.2929

$w_6 = w_6 - \alpha\frac{\partial J}{\partial w_6}$

≈ 0.5018

3. a) 通过正则化，我们希望降低模型的复杂度，提高泛化能力。而偏置项b与模型的复杂度无关，这是因为模型的复杂度主要反映在输入的微小改变引起输出的变化程度，而b是对输入值

不敏感的，换句话说，b对输入是一视同仁的，
不会产生不一样的加权，因为偏置b并不影响
模型的复杂度，不需要把b加入惩罚项。

b) $\frac{\partial C}{\partial w_i} = \frac{\partial C_0}{\partial w_i} + 2\lambda w_i$

∴ 更新表达式：

$\forall i \quad w_i := w_i - \frac{\partial C_0}{\partial w_i} - 2\lambda w_i$

即 $w_i := (1-2\lambda) w_i - \frac{\partial C_0}{\partial w_i}$

因为 $1-2\lambda < 1$，而且一般来说 $1-2\lambda > 0$

在这种情况下，相比于原来的 $w_i := w_i - \frac{\partial C_0}{\partial w_i}$

在每一步权值更新时会产生权值的衰减，即
使普通情况的w变小，当w变小以后，
由于w是对x的加权，对于x的微小变化会
变得不那么敏感，可以一定程度上防止过拟合，
降低了模型的复杂度，提高了泛化能力。

c). 可以通过试凑法选取正则化参数。

① 首先，令λ=0，进行训练后观察train_acc和test_acc，

② 如果trainacc比test_acc大很多，则模型发生了过拟合，
逐步增大λ，用λ=0.001 λ=0.01 …等值进行试探，
确定λ的数量级。

③ 再对λ进行数量级内微调，使得模型的准确率和泛化能力都达到较好的即可。