1、解：(a) 由题：$W_1=\begin{pmatrix}0.1 & 0.1 & 0.3\\ 0.2 & 0.3 & 0.2\\ 0.1 & 0.2 & 0.1\\ 0.3 & 0.1 & 0.1\end{pmatrix}$ $W_2=W_1^T=\begin{pmatrix}0.1 & 0.2 & 0.1 & 0.3\\ 0.1 & 0.3 & 0.2 & 0.1\\ 0.3 & 0.2 & 0.1 & 0.1\end{pmatrix}$ $W_3=(0.2\ 0.2\ 0.4)$ $\begin{cases}b_1=0.5\\ b_2=0.5\\ b_3=0.2\end{cases}$

设 $X=(x_1,x_2,x_3)^T$，$h_1=(h_{11},h_{12},h_{13},h_{14})^T$，$h_2=(h_{21},h_{22},h_{23})^T$

已知：$X=(0.05,0.10,0.05)^T$，计算过程如下：

① $z_1=W_1X+(b_1\ b_1\ b_1\ b_1)^T=(0.53\ \ 0.55\ \ 0.53\ \ 0.53)^T$ $\quad h_1=\cos(z_1)=(0.8628,0.8525,0.8628,0.8628)^T$

② $z_2=W_2h_1+[b_2\ b_2\ b_2]^T=(1.1019\ \ 1.1009\ \ 1.1019)^T$ $\quad h_2=\cos(z_2)=(0.4519,0.4528,0.4519)^T$

③ $z_3=W_3h_2+b_3=0.5617$ $\quad \hat{y}=\frac{1}{1+e^{-z_3}}=0.6368$

(b) 损失函数：$L=(y-\hat{y})^2$

计算梯度：$\frac{dL}{d\hat{y}}=2(\hat{y}-y)$ $\quad \frac{d\hat{y}}{dz_3}=\hat{y}(1-\hat{y})$ $\quad \frac{\partial z_3}{\partial W_3}=h_2^T$

$\Rightarrow \frac{\partial L}{\partial W_3}=\frac{dL}{d\hat{y}}\frac{d\hat{y}}{dz_3}\frac{\partial z_3}{\partial W_3}=2(\hat{y}-y)\hat{y}(1-\hat{y})h_2^T=(-0.0655,-0.0656,-0.0655)$

(c) $W_3'=W_3-0.1\frac{\partial L}{\partial W_3}=(0.2065,0.2066,0.4065)$

4、解：输入 $X=\begin{pmatrix}x_{11} & x_{12} & x_{13}\\ x_{21} & x_{22} & x_{23}\\ x_{31} & x_{32} & x_{33}\end{pmatrix}=\begin{pmatrix}1 & 1 & 2\\ 1 & 1 & 2\\ 1 & 1 & 2\end{pmatrix}$，输出 $Y=\begin{pmatrix}y_{11} & y_{12}\\ y_{21} & y_{22}\end{pmatrix}$，设 $y=\begin{pmatrix}y_{11}\\ y_{12}\\ y_{21}\\ y_{22}\end{pmatrix}$，$H=\begin{pmatrix}H_1 & H_2\\ H_3 & H_4\end{pmatrix}$，设 $h=\begin{pmatrix}H_1\\ H_2\\ H_3\\ H_4\end{pmatrix}$

① Forward：$y=\begin{pmatrix}y_{11}\\ y_{12}\\ y_{21}\\ y_{22}\end{pmatrix}=\begin{pmatrix}H_1+H_2+H_3+H_4\\ H_1+2H_2+H_3+2H_4\\ H_1+H_2+H_3+H_4\\ H_1+2H_2+H_3+2H_4\end{pmatrix}=W\cdot h$，这里 $W=\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 2 & 1 & 2\\ 1 & 1 & 1 & 1\\ 1 & 2 & 1 & 2\end{pmatrix}$，$h=\begin{pmatrix}H_1\\ H_2\\ H_3\\ H_4\end{pmatrix}=\begin{pmatrix}0.2\\ 0.2\\ 0.3\\ 0.3\end{pmatrix}$，代入得 $y=\begin{pmatrix}1\\ 1.5\\ 1\\ 1.5\end{pmatrix}$ 即 $\begin{cases}y_{11}=y_{21}=1\\ y_{12}=y_{22}=1.5\end{cases}$

② Backward：$d=(1,1,2,1)^T$，$L=\frac{1}{2}(y-d)^T(y-d)$，$y=W\cdot h$

梯度：$\frac{\partial L}{\partial h}=\frac{\partial L}{\partial y}\frac{\partial y}{\partial h}=(y-d)^TW=(0,0.5,-1,0.5)\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 2 & 1 & 2\\ 1 & 1 & 1 & 1\\ 1 & 2 & 1 & 2\end{pmatrix}=(0,1,0,1)$

更新：$h'=h-\alpha\cdot\left(\frac{\partial L}{\partial h}\right)^T=(0.2\ \ 0.2\ \ 0.3\ \ 0.3)^T-(0,1,0,1)^T=(0.2\ \ -0.8\ \ \ 0.3\ \ -0.7)^T$，即更新后 $H=\begin{pmatrix}0.2 & -0.8\\ 0.3 & -0.7\end{pmatrix}$

2、解: (a) 若使用线性激活函数，则多层神经网络仅相当于单个线性层，网络加深无法提高模型容量，徒增计算复杂度

证: 考虑 $n+1$ 隐层神经网络，其输入 $x\in\mathbb{R}^p$，输出 $y\in\mathbb{R}^q$，第 i 层节点数为 d_i ($d_0=p, d_n=q$)，参数为 W_i (维数 $d_{i-1}\times d_i$) 与 $b_i\in\mathbb{R}^{d_i}$

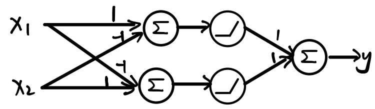
$$\text{Forward:}\begin{cases} h_0 = x \\ h_i = \sigma_i(W_i h_{i-1} + b_i) \quad , i=1,2,\cdots,n \\ y = h_n \end{cases}$$

式中 $\sigma_i(\cdot)$ 为第 i 层的激活函数 ($i=1,2,\cdots,n$)，假设 $\sigma_i(\cdot)$ 为线性函数，表示为 $\sigma_i(x) = W_i' x + b_i'$，这里 W_i' 维数为 $d_i\times d_i$，$b_i'\in\mathbb{R}^{d_i}$，则:

$$y = W_n'\Big(W_n\big(\cdots(W_0'(W_0 x + b_0) + b_0')\cdots\big) + b_n\Big) + b_n' = \tilde{W}x + \tilde{b}$$，这里 $\tilde{W} = W_n' W_n W_{n-1}' W_{n-1}\cdots W_0' W_0$，$\tilde{b} = W_n'\Big(W_n\big(\cdots(W_1(W_0' b_0 + b_0') + b_1)\cdots\big) + b_n\Big) + b_n'$

可见该 $n+1$ 隐层神经网络仅相当于单个线性层，网络加深但其表征能力没有提升，徒增计算复杂度

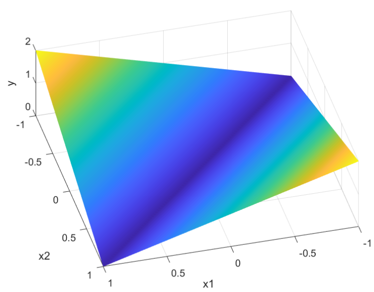
(b)



图中 Σ 表示相加，表示 ReLU，Forward 公式: $y = \mathrm{ReLU}(x_1 - x_2) + \mathrm{ReLU}(x_2 - x_1)$

验证: $x_1=1, x_2=0 \Rightarrow y=1$; $x_1=0, x_2=1 \Rightarrow y=1$; $x_1=x_2=1 \Rightarrow y=0$; $x_1=x_2=0 \Rightarrow y=0$ (即 XOR)

图像:



3、解: (a) 设卷积核宽度为 k，则 $\left\lfloor \frac{100-k}{2} + 1 \right\rfloor = 48$，得 $k=6$ 或 5

∴ 卷积层输入通道 3，输出通道 4，卷积核宽度为 6 或 5

(b) 最少可用 4 个卷积核:

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix} \quad \begin{bmatrix}0&1&0\\0&0&0\\0&0&0\end{bmatrix} \quad \begin{bmatrix}0&0&0\\1&0&0\\0&0&0\end{bmatrix} \quad \begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}$$