1. a) 激活函数可以引入非线性，将神经网络中的线性变换转换为非线性变换，使神经网络可以学习和表示复杂的非线性关系。如果没有激活函数，不管网络有多少层，最终的输出都是输入的线性组合，无法处理复杂的任务

b) 当输入为 $X=(x_1\ x_2\ x_3)^T=(0.05, 0.10, 0.05)^T$ 时，

$$Z_1=(z_{11}\ z_{12}\ z_{13})^T=W_1X=\begin{bmatrix}0.1 & 0.2 & 0.3\\ 0.4 & 0.5 & 0.6\\ 0.5 & 0.4 & 0.3\end{bmatrix}\begin{bmatrix}0.05\\ 0.10\\ 0.05\end{bmatrix}=\begin{bmatrix}0.04\\ 0.1\\ 0.08\end{bmatrix}$$

$$h_1=(h_{11}\ h_{12}\ h_{13})^T=\sin(Z_1)=(0.04, 0.1, 0.08)^T$$

$$Z_2=(z_{21}\ z_{22}\ z_{23}\ z_{24})^T=W_2\begin{bmatrix}b_1\\ h_{11}\\ h_{12}\\ h_{13}\end{bmatrix}=\begin{bmatrix}-0.1 & 0.2 & -0.1 & 0.2\\ 0.2 & -0.1 & 0.2 & -0.1\\ 0.1 & -0.2 & 0.1 & -0.2\\ -0.2 & 0.1 & -0.2 & 0.1\end{bmatrix}\begin{bmatrix}0.1\\ 0.04\\ 0.1\\ 0.08\end{bmatrix}=\begin{bmatrix}0.004\\ 0.028\\ -0.004\\ -0.028\end{bmatrix}$$

$$h_2=(h_{21}\ h_{22}\ h_{23}\ h_{24})^T=\sin(Z_2)=(0.004, 0.028, -0.004, -0.028)^T$$

$$z_3=W_3(b_2\ h_{21}\ h_{22}\ h_{23}\ h_{24})^T=0.024$$

$$\therefore \hat{y}=\frac{1}{1+e^{-z_3}}=0.5060$$

c) 采用最小化均方误差作为优化准则，损失函数 $L=\frac{1}{2}(y-\hat{y})^2$

根据BP算法，$\frac{\partial L}{\partial W_3}=\frac{\partial L}{\partial \hat{y}}\cdot\frac{\partial \hat{y}}{\partial z_3}\cdot\frac{\partial z_3}{\partial W_3}$

$$=-(y-\hat{y})\cdot\hat{y}(1-\hat{y})\cdot(b_2\ h_{21}\ h_{22}\ h_{23}\ h_{24})$$

$$=(-0.0222, -0.00045, -0.0031, 0.00045, 0.0031)^T$$

d) $\because W_3'=W_3-\alpha\frac{\partial L}{\partial W_3}$ 且 $\alpha=0.1$

$\therefore$ 更新后的 $W_3'=[0.10222, 0.200045, 0.30031, 0.499955, 0.09969]$

2. a) $\because N=224$，$m=11$ 且步长为4

$\therefore$ 输出尺寸 $=\left\lfloor\frac{224-11}{4}\right\rfloor+1=54$

$\therefore$ 第一个卷积层输出的宽度和高度是54

b) $\because$ 第一个卷积层共有96个卷积核 $\therefore$ 输出通道数=卷积核个数=96

$\therefore$ 第一个卷积层输出的维度为 $54\times54\times96$

$\therefore$ 被激活的神经元数量 $=54\times54\times96=279936$ 个

c) $\because$ 卷积层参数个数=(卷积核高×卷积核宽×通道数+1)×卷积核个数

$\therefore$ 该卷积层中可被训练的参数数量 $=(11\times11\times3+1)\times96=34944$ 个

d) 输入特征数量 $=224\times224\times3=150528$

$\because$ 每个神经元与所有输入特征连接，且共有512个神经元

$\therefore$ 权重参数 $=150528\times512=77070336$ 个

$\because$ 每个神经元有一个偏置 $\therefore$ 偏置参数=512个

$\therefore$ 总参数个数 $=77070336+512=77070848$ 个

e) W_1：卷积核中间一行的值与上下两行的值存在显著差异，可用于提取图像中的水平边缘特征

W_2：中心像素与其周围像素的值有显著差异，可用于提取图片中的点特征或显著特征

W_3：卷积核为对角线对称结构，从左上到右下在对角线方向上有显著的值变化，可用于提取图像中的对角线边缘特征

3. $\because H_0=\begin{bmatrix}-0.2 & 0.4\\ 0.4 & 0.2\end{bmatrix}$ 输入值 $=\begin{bmatrix}1 & 1 & 2\\ 1 & 1.5 & 2\\ 1 & 2 & 2\end{bmatrix}$

$\therefore$ 网络输出值 $=\begin{bmatrix}0.9 & 1.6\\ 1.2 & 1.7\end{bmatrix}$

损失函数 $L=\frac{1}{2}(y-d)^2=\sum_{i=1}^{4}\frac{1}{2}(y_i-d_i)^2$

$\frac{\partial L}{\partial H_1}=\sum_{i=1}^{4}\frac{\partial L}{\partial y_i}\cdot\frac{\partial y_i}{\partial H_1}=\sum_{i=1}^{4}(y_i-d_i)\cdot\frac{\partial y_i}{\partial H_1}=1.75$；　$\frac{\partial L}{\partial H_2}=\sum_{i=1}^{4}\frac{\partial L}{\partial y_i}\cdot\frac{\partial y_i}{\partial H_2}=\sum_{i=1}^{4}(y_i-d_i)\cdot\frac{\partial y_i}{\partial H_2}=2.8$；

$\frac{\partial L}{\partial H_3}=\sum_{i=1}^{4}\frac{\partial L}{\partial y_i}\cdot\frac{\partial y_i}{\partial H_3}=\sum_{i=1}^{4}(y_i-d_i)\cdot\frac{\partial y_i}{\partial H_3}=2.4$；　$\frac{\partial L}{\partial H_4}=\sum_{i=1}^{4}\frac{\partial L}{\partial y_i}\cdot\frac{\partial y_i}{\partial H_4}=\sum_{i=1}^{4}(y_i-d_i)\cdot\frac{\partial y_i}{\partial H_4}=2.85$；

$\therefore$ 更新权值参数为 $\begin{bmatrix}H_1' & H_2'\\ H_3' & H_4'\end{bmatrix}=\begin{bmatrix}H_1-\alpha\frac{\partial L}{\partial H_1} & H_2-\alpha\frac{\partial L}{\partial H_2}\\ H_3-\alpha\frac{\partial L}{\partial H_3} & H_4-\alpha\frac{\partial L}{\partial H_4}\end{bmatrix}=\begin{bmatrix}-0.375 & 0.12\\ 0.16 & -0.085\end{bmatrix}$