1. **对于图 5.7 中的v_{ih}，试推导出 BP 算法中的更新公式**（5.13）.

$\triangle v_{ih}=-\eta\frac{\partial E_k}{\partial v_{ih}}$，因$v_{ih}$只在计算$b_h$时用上，所以$\frac{\partial E_k}{\partial v_{ih}}=\frac{\partial E_k}{\partial b_h}\frac{\partial b_h}{\partial v_{ih}}$，其中$\frac{\partial b_h}{\partial v_{ih}}=\frac{\partial b_h}{\partial a_h}\frac{\partial a_h}{\partial v_{ih}}=\frac{\partial b_h}{\partial a_h}x_i$，所以$\frac{\partial E_k}{\partial v_{ih}}=\frac{\partial E_k}{\partial b_h}\frac{\partial b_h}{\partial a_h}x_i=-e_hx_i$，即得原书中5.13式。

2. 试述式（5.6）中学习率的取值对神经网络训练的影响.

学习率太高会导致误差函数来回震荡，无法收敛；而学习率太低则会收敛太慢，影响训练效率。

3. Minsky 与 Papert 指出：单层感知机因为是线性模型，所以不能表示复杂的函数，如异或（XOR）。验证单层感知机为什么不能表示异或。

假设感知机模型可以表示异或问题，即满足异或函数(XOR)输入与输出的情况（见**第1步**）。假设x向量只有两个维度x_1，x_2：

1. 根据$x_1=0,x_2=0,f(x)=-1$，则$w\cdot x+b<0$，可得$b<0$；
2. 根据$x_1=0,x_2=1,f(x)=1$，则$w_2+b>0$，结合$b<0$，可得$w_2>-b>0$；
3. 根据$x_1=1,x_2=0,f(x)=1$，则$w_1+b>0$，结合$b<0$，可得$w_1>-b>0$；
4. 根据$x_1=1,x_2=1$，并结合$w_1+b>0$、$w_2>0$，则$w_1+w_2+b>0$，可得$f(x)=1$，与异或条件中的$f(x)=-1$矛盾。

所以假设不成立，原命题成立，即感知机模型不能表示异或。

4. 正样本点是$x_1=(3,3)^T,x_2=(4,3)^T$，负样本点是$x_3=(1,1)^T$，试用梯度下降算法求解感知机模型，模型参数初值取 0。

表 2.1　例 2.1 求解的迭代过程

迭代次数	误分类点	w	b	$w\bullet x+b$
0		0	0	0
1	x_1	$(3,3)^{\mathrm{T}}$	1	$3x^{(1)}+3x^{(2)}+1$
2	x_3	$(2,2)^{\mathrm{T}}$	0	$2x^{(1)}+2x^{(2)}$
3	x_3	$(1,1)^{\mathrm{T}}$	-1	$x^{(1)}+x^{(2)}-1$
4	x_3	$(0,0)^{\mathrm{T}}$	-2	-2
5	x_1	$(3,3)^{\mathrm{T}}$	-1	$3x^{(1)}+3x^{(2)}-1$
6	x_3	$(2,2)^{\mathrm{T}}$	-2	$2x^{(1)}+2x^{(2)}-2$
7	x_3	$(1,1)^{\mathrm{T}}$	-3	$x^{(1)}+x^{(2)}-3$
8	0	$(1,1)^{\mathrm{T}}$	-3	$x^{(1)}+x^{(2)}-3$