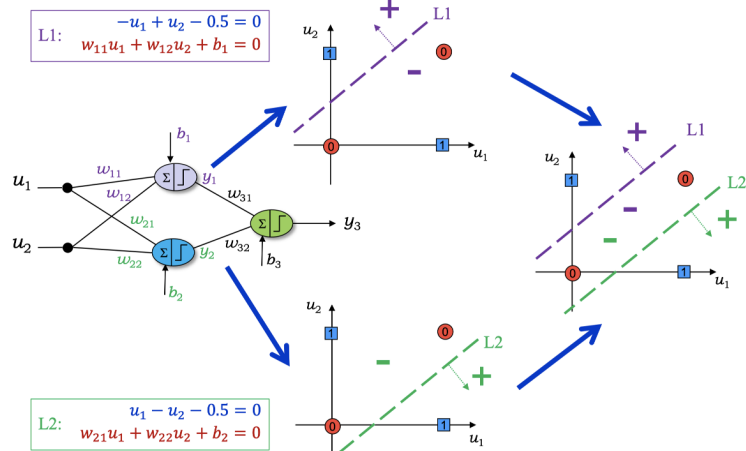


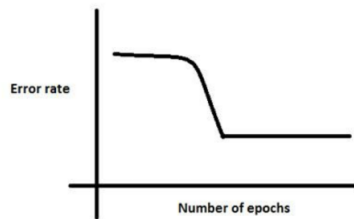
设网络  $f(x, \theta)$  输入  $\vec{x}_k$  有输出  $\hat{y}_k = f(\vec{x}_k, \theta)$   
 设损失函数为均方误差  $E_k =$

- 对于图 5.7 中的  $v_{ih}$ , 试推导出 BP 算法中的更新公式 (5.13) .
- 试述式 (5.6) 中学习率的取值对神经网络训练的影响.
- 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。

4. 根据图示分界线给出异或门神经网络的参数  $w_i, b_i$ .



- 在训练神经网络时, 损失函数 (Loss) 在最初的几个 epochs 时没有下降, 可能的原因是?



- 简述 Dropout 能够防止过拟合的原因。
- 假设 Dropout 概率为  $p$ , 为了保证期望值不变, 则在测试时, 该层模型权重  $w$  应该变为?
- 设模型初始参数  $\theta_0$  为  $[0.1, 0.2]$ , 学习率为 0.1, 动量更新权重为 1, 若模型在两次训练过程中的梯度依次为  $g_0 = [2, -1], g_1 = [1, -2]$ , 求两次训练后模型的参数  $\theta_2$ .