

1.不完全可以，梯度消失通常发生在激活函数引入的梯度饱和区域，当输入远离中心时，导数趋近于0，导致反向传播时梯度快速衰减。在某些情况下，较大的学习率可能使梯度在初期保持一定规模，从而一定程度上延缓梯度消失，增大学习率可能会导致训练不稳定，尤其是当梯度突然变大时，会引发梯度爆炸或模型无法收敛。因此，仅仅通过增加学习率并不能从根本上解决梯度消失问题。

2.输入是一张 $100 * 100$ 的 RGB 图像，那么总的输入特征数量为： $100 * 100 * 3 = 30000$

全连接层有 100 个神经元，每个神经元需要接收 30,000 个输入特征，

则权重参数总数为： $30000 * 100 = 3000000$

每个神经元都有一个偏置参数，所以偏置参数为100个

总可学习参数数量为权重参数加上偏置参数为 $3000000 + 100 = 3000100$ 个

3.输入是 256×256 的 RGB 图像：

输入通道数为3

卷积核为 $3 * 3 * 3 = 27$

总权重参数为 $27 * 100 = 2700$

偏置参数为100

总参数数量： $2700 + 100 = 2800$

输入是 256×256 的灰度图像：

输入通道数为1

卷积核为 $3 * 3 * 1 = 9$

总权重参数为 $9 * 100 = 900$

偏置参数为100

总参数数量： $900 + 100 = 1000$

4.根据公式可知 $63 * 63 * 16 = (63 * 63 * 16 + 2 * P - 7 * 7) / 1 + 1$

所以 $P = 3$ ，即填充应设置为3

5.(1)

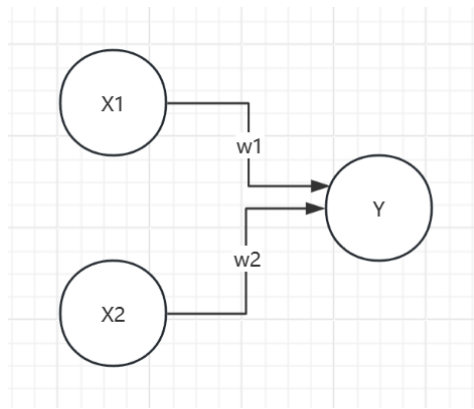
$$Y = w_5 h_1 + w_6 h_2 = w_5(w_1 X_1 + w_3 X_2) + w_6(w_2 X_1 + w_4 X_2)$$

$$Y = (w_5 w_1 + w_6 w_2) X_1 + (w_5 w_3 + w_6 w_4) X_2$$

$$W_1 = w_5 w_1 + w_6 w_2$$

$$W_2 = w_5 w_3 + w_6 w_4$$

$$Y = W_1 X_1 + W_2 X_2$$



(2)

$$w1 = 1, w2 = -1, w3 = 1, w4 = -1, w5 = 1, w6 = 1$$