

第一题

为了打破对称性。如果所有权重和偏置初始化为0，那么网络中的每个神经元在隐藏层中都会计算出相同的输出。在反向传播时，所有神经元都会收到相同的梯度，从而以相同的方式更新权重。这会导致所有神经元学习到相同的特征，使得网络无法学习多样化的特征，从而大大降低模型的表达能力和学习效率。

第二题

- 正则化：通过在损失函数中添加权重的范数惩罚项（如L1或L2正则化），限制模型复杂度，防止权重过大。
- 丢弃法（Dropout）：在训练过程中随机丢弃一部分神经元，防止神经元之间的过度依赖，提高模型的泛化能力。
- 早停（Early Stopping）：在训练过程中监控验证集性能，当验证集性能不再提升时停止训练，避免过拟合训练数据。
- 数据增强：通过旋转、缩放、翻转等方式增加训练数据量，提高模型对变化的鲁棒性。
- 使用更简单的模型：减少网络层数或神经元数量，降低模型复杂度。

第三题

- 输入维度： $100 \times 100 \times 3 = 30000$
 - 权重： $30000 \times 100 = 3000000$
 - 偏置： 100
- 总参数量： $3000000 + 100 = 3000100$

第四题

- 每个卷积核权重： $3 \times 3 \times (in_channels = 3) = 27$
 - 每个卷积核偏置： 1
- 总参数量： $100 \times (27 + 1) = 2800$

第五题

- $OH = \frac{63+2\times1-5}{2} + 1 = 31$
- $OW = \frac{63+2\times1-5}{2} + 1 = 31$

- 通道数=卷积核个数=32
输出维度: $31 \times 31 \times 32$

第六题

$$\frac{63+2 \times padding - 5}{s=1} + 1 = 63 \\ \Rightarrow padding = 2$$

第七题

(1)

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

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

因此, 用**单层线性**模型 ($Y = \tilde{w}_1 X_1 + \tilde{w}_2 X_2$) 即可, 其中

$$\boxed{\tilde{w}_1 = w_5 w_1 + w_6 w_2, \quad \tilde{w}_2 = w_5 w_3 + w_6 w_4}$$

(2)

取 $w_1 = 1, w_2 = 1, w_3 = 3, w_4 = 3, w_5 = -5, w_6 = 4$

隐藏层 $\sigma(x) = \frac{1}{(1+e^{-x})}$, 输出层 $t(x) = 1[x > 0]$

四个输入点计算 (保留4位小数):

- (0,0): $h_1 = \sigma(0) = 0.5, h_2 = \sigma(0) = 0.5$
 $z = w_5 \cdot h_1 + w_6 \cdot h_2 = -5 \cdot 0.5 + 4 \cdot 0.5 = -0.5 \Rightarrow Y = 0$
- (1,0): $h_1 = \sigma(1) = 0.7311, h_2 = \sigma(3) = 0.9526$
 $z = -5 \cdot 0.7311 + 4 \cdot 0.9526 = 0.1550 > 0 \Rightarrow Y = 1$
- (0,1): $h_1 = \sigma(1) = 0.7311, h_2 = \sigma(3) = 0.9526$
 $z = -5 \cdot 0.7311 + 4 \cdot 0.9526 = 0.1550 > 0 \Rightarrow Y = 1$
- (1,1): $h_1 = \sigma(2) = 0.8808, h_2 = \sigma(6) = 0.9975$
 $z = -5 \cdot 0.8808 + 4 \cdot 0.9975 = -0.4139 < 0 \Rightarrow Y = 0$