

# hw2

---

id: 519021910861

name: huidong xu

---

## 1. 判断题

- a. 错误
- b. 正确
- c. 正确
- d. 错误
- e. 正确

我理解的适用是指在所有类型的深度神经网络中都可以运用 Dropout 和 Batch Normalization 来防止过拟合，但运用的效果好坏则因为模型的不同而有所差异。

## 2. 选择题

- a. C
- b. C
- c. D     我理解的 RGB 图片需要有三通道，所以实际输入大小为  $64 \times 64 \times 3$ .

## 3. 简答题

- a. 深度学习模型为何在训练中容易出现过拟合？试分析原因并给出如何在深度学习模型训练中环节过拟合？  
原因：1. 深度学习模型模型复杂度过高，参数过高。2. 训练集数据太少。3. 训练集测试集的数据分布不一致，样本里的噪声数据干扰过大。  
解决方案：1. 针对模型复杂度过高的问题，可以降低模型复杂度，例如减少层的数量或者减少神经元的个数来缩小网络的规模。2. 针对训练数据太少的问题，可以增加训练数据。3. 正则化。4. 数据增强。5.

dropout。神经网络在每一次迭代过程中随机地丢弃神经网络中的神经元。6. 早停。在训练过程中，如果训练误差继续减小，但是测试误差已经开始增加，此时可以停止训练。7. 集成学习。把多个模型集成在一起，降低单个模型过拟合风险。8. Batch Normalization。在 CNN 每层之间加上将神经元的权重调成标准正态分布的正则化层。

b. 深度学习模型训练过程中为何会出现梯度消失和梯度爆炸问题？有哪些方法可以解决梯度消失或梯度爆炸？

- 梯度消失

梯度消失是指在神经网络中，当前面隐藏层的学习速率高于后面隐藏层的学习速率，即随着隐藏层数目的增加，分类准确率反而下降。

梯度消失主要由隐藏层层数过多和采用了不合适的激活函数导致，根本原因在于神经网络的不稳定。

具体而言，从深层网络角度来讲，不同的层学习速率差异很大，表现为网络中靠近输出的层学习的情况很快，靠近输入的层学习的很慢，有时甚至训练了很久前几层的权值和刚开始随机初始化的值差不多。因此，梯度消失其根本原因在于反向传播训练法则，属于先天不足。

而且，当选用了不恰当的激活函数，在网络中对神经元的激活函数进行链式求导时，当层数越多时求导结果越小，会加剧梯度消失的问题。

我们可以使用诸如 ReLU 等替代函数作为激活函数，以及使用 Batch Normalization 等对数据进行处理，而 LSTM 中的结构设计也可以帮助改善 RNN 中的梯度消失问题。

- 梯度爆炸

梯度爆炸是指在神经网络中，当前面隐藏层的学习速率低于后面隐藏层的学习速率，导致随着隐藏层数目的增加分类准确率反而下降。

梯度爆炸主要由隐藏层的层数过多和权重的初始化值过大导致，根本原因在于神经网络的不稳定。

具体而言，在深层网络或循环神经网络中，误差梯度可在更新中累积，变成非常大的梯度，然后导致网络权重的大幅更新，并因此使得网络变得不稳定，在极端情况下，权重的值变得非常大，以至于溢出，导致 NaN 值。

我们可以通过重新设计网络模型（即减少隐藏层数和批尺寸），使用 ReLU 激活函数并选择合适的初始权重值，使用长短期记忆网络，使用梯度截断和使用权重正则化等方法来改善梯度爆炸问题。