

## 作业二

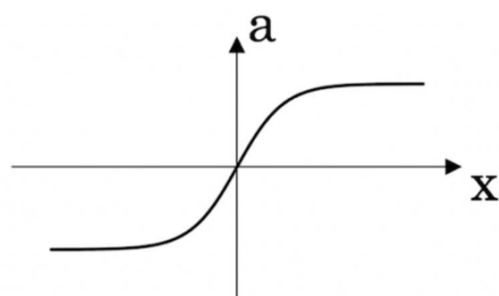
### 一. 判断题

1. 神经网络随着网络层数的增加表达能力变得更强, 因此网络中的激活函数可以去掉, 只需要加深网络就能够训练足够强的神经网络。✗
2. Long short-term Memory (LSTM)网络具有记忆和遗忘功能, 适用于序列建模。✓
3. 注意力机制(attention)相比 LSTM 更高效, 因为在每个 time step 计算时都可以读到全局信息, 而不需要像 LSTM 那样串行计算。✗
4. 训练卷积神经网络 (CNN) 时, 如果对训练样本通过平移、旋转和缩放等操作额外生成一些补充样本, 会从整体上降低训练样本的质量, 影响网络提取特征, 从而导致预测准确率下降。✗
5. 在网络中加入 Dropout 和 Batch Normalization 都是深度学习中常见的防止过拟合的手段, 在所有类型的深度神经网络中都适用。✗

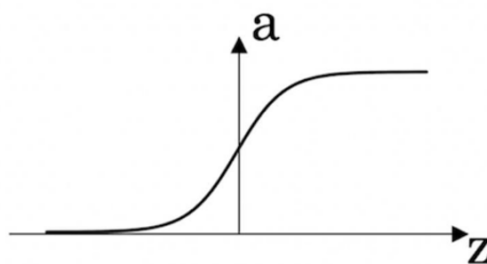
### 二. 选择题

1. 下图中哪一个表述 ReLU 激活函数? C

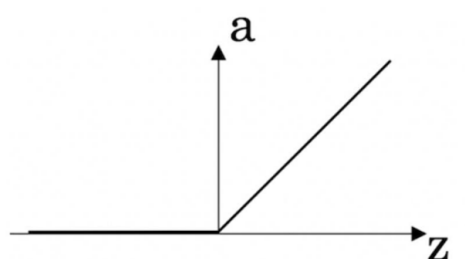
A.



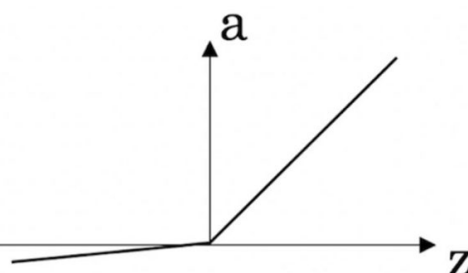
B.



C.



D.



2. 输入为  $64 \times 64$  的 RGB 图片, 使用 32 个  $3 \times 3$  的卷积核提取特征, 步长为 1, 不使用 padding 填充, 则输出的大小为: C  
(a)  $64 \times 64 \times 64$  (b)  $64 \times 32 \times 32$  (c)  $32 \times 62 \times 62$  (d)  $32 \times 64 \times 64$
3. 上一题中, 该卷积层的参数有多少个? D  
(a) 64 (b) 32 (c)  $32 \times 3 \times 3$  (d)  $32 \times 3 \times 3 \times 3$

### 三. 简答题

1. 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缓解过拟合?

原因: ① 数据量太小. ② 训练集和测试集分布不一致. ③ 模型复杂度过高

方法: ① 增加更多数据 ② 降低模型复杂度 ③ 数据增强 ④ 正则化

2. 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?

原因: ① 使用了深层网络. 当层数越深时, 梯度将以指数形式传播. 如果对激活函数进行求导的部分大小  $> 1$ , 梯度更新将以指数形式增加, 即发生梯度爆炸; 若  $< 1$  则梯度更新信息以指数形式衰减, 即发生梯度消失.

② 采用了不合适的损失函数. 如果激活函数选择不合适, 如 sigmoid 函数, 则很容易发生梯度消失的情况

解决方案: ① 预训练 + 微调

② 梯度剪切. 正则

③ 选择 relu, leakrelu, elu 等激活函数

④ batch normalization

⑤ 使用残差网络

⑥ 使用 LSTM.