

深度学习测验

总计问题：20

工作表时间：21分钟

姓名

班级

日期

1. 在机器学习的定义中，与传统编程相比，机器学习的核心特征是什么？

- a) 输入数据和结果（标签），计算机输出规则（模型）
- c) 不需要任何数据输入

- b) 输入规则和数据，计算机输出结果
- d) 只能处理图像数据

2.

下列哪种学习范式被称为“带标签的老师教学”？

- a) 有监督学习
- c) 强化学习

- b) 深度学习
- d) 无监督学习

3. 在有监督学习中，如果我们要预测房价（连续数值），这属于什么任务？

- a) 生成任务
- c) 分类任务

- b) 回归任务
- d) 聚类任务

4. 在神经网络（MLP）中，激活函数（如Sigmoid）的主要作用是什么？

- a) 计算加权求和
- c) 减少参数数量

- b) 计算损失值
- d) 引入非线性能力

5. 神经网络训练的“四步法”中，用来分析错误原因并计算梯度的是哪一步？

- a) 参数更新 (Adjust)
- c) 误差计算 (Check Answers)

- b) 反向传播 (Find Cause)
- d) 前向传播 (Guess)

6. 关于损失函数均方误差 (MSE)，计算公式 $E = \frac{1}{2} (y^{\text{hat}} - y)^2$ 中的平方项主要作用是什么？
- a) 为了让计算结果变为负数
 - b) 为了消除正负号影响并放大较大误差
 - c) 为了减小误差数值
 - d) 没有任何数学意义，只是习惯

7. 梯度 (Gradient) 在数学上的几何意义是指向什么方向？
- a) 函数值为0的方向
 - b) 函数值减小最快的方向
 - c) 函数值增长最快的方向
 - d) 切线水平的方向

8. 在参数更新公式 $\theta_{\text{new}} = \theta_{\text{old}} - \alpha \cdot \nabla E$ 中， α 代表什么？
- a) 偏置 (Bias)
 - b) 激活函数
 - c) 损失值 (Loss)
 - d) 学习率 (Learning Rate)

9. 下列哪种梯度下降方式会使用整个数据集来计算一次梯度？
- a) 小批量梯度下降 (Mini-batch GD)
 - b) 随机梯度下降 (SGD)
 - c) 批量梯度下降 (Batch GD)
 - d) 自动梯度下降

10. 在数据预处理中，为什么要对图像像素进行“归一化”（如缩放到0-1之间）？
- a) 为了增加图片的清晰度
 - b) 为了减小图片文件的存储大小
 - c) 让数据在合理范围内，加速训练收敛
 - d) 为了把彩色图变成黑白图

11. 在深度学习中，一张 RGB 彩色图片的通常被表示为什么维度的张量？

- a) 2维张量 (矩阵)
- b) 2维张量 (矩阵)
- c) 3维张量 (高度, 宽度, 通道数)
- d) 0维张量 (标量)

12. 为什么全连接网络 (MLP/FCNN) 不适合直接处理大尺寸图像？

- a) MLP 训练速度太快
- b) MLP 无法处理数值数据
- c) 破坏空间结构且参数量爆炸
- d) MLP 只能做回归不能做分类

13. CNN 中“权值共享”是指什么？

- a) 卷积层和全连接层共享权重
- b) 同一个卷积核在图像所有位置滑动时使用同一套权重
- c) 训练集和测试集共享权重
- d) 所有层使用相同的权重

14. 在卷积层中，Padding (填充) 的主要作用是什么？

- a) 改变图像的颜色
- b) 防止边缘信息丢失并控制输出尺寸
- c) 增加计算量
- d) 去除图像噪声

15. 最大池化 (Max Pooling) 层的主要功能不包括以下哪项？

- a) 减少后续计算量
- b) 增加参数数量
- c) 提供一定的平移不变性
- d) 减小特征图尺寸

16. CNN在最后展平 (Flatten) 时，与MLP一开始的展平有什么本质区别？

- a) 没有区别
- b) CNN展平不需要全连接层处理
- c) CNN展平的是“高阶特征”，保留了语义信息
- d) CNN展平后数据量变大了

17. VGG 网络架构的一个显著特点是什么?
- a) 使用了非常大的卷积核 (如 11×11)
 - b) 去掉了全连接层
 - c) 全部使用小卷积核 (3×3) 堆叠
 - d) 没有使用池化层
18. 在一个神经网络中, 如果损失函数 L 是关于所有参数 θ 的函数, 为什么我们说这是一个“高维”函数? 它的维度取决于什么?
- a) 取决于网络中所有可训练参数 (权重 W 和偏置 b) 的总数量
 - b) 取决于输出标签的类别数量
 - c) 取决于输入数据的特征维度 (Feature Dimensions)
 - d) 取决于训练样本的数量 (Batch Size)
19. 在 CNN 中, 关于“一个卷积核”的维度, 以下描述最准确的是?
- a) 尺寸大小 \times 输出通道数 ($K \times K \times C_{out}$)
 - b) 由输出通道数决定
 - c) 永远是二维的 ($K \times K$)
 - d) 尺寸大小 \times 输入通道数 ($K \times K \times C_{in}$)
20. 一张大小为 $100 \times 100 \times 3$ 的图片, 经历一层 $3 \times 3 \times 3 \times 16$ 卷积核、 $stride=1$ 、 $padding=0$ 的卷积运算后, 又经历 2×2 尺寸、 $stride=2$ 的最大池化层计算, 得到的特征图大小为
- a) $97 \times 97 \times 16$
 - b) $1 \times 98 \times 98$
 - c) $49 \times 49 \times 16$
 - d) $3 \times 48 \times 48$

答案要点

1. a) 输入数据和结果 (标签), 计算机输出规则 (模型)
2. a) 有监督学习
3. b) 回归任务
4. d) 引入非线性能力
5. b) 反向传播 (Find Cause)
6. b) 为了消除正负号影响并放大较大误差
7. c) 函数值增长最快的方向
8. d) 学习率 (Learning Rate)
9. c) 批量梯度下降 (Batch GD)
10. c) 让数据在合理范围内, 加速训练收敛
11. c) 3维张量 (高度, 宽度, 通道数)
12. c) 破坏空间结构且参数量爆炸
13. b) 同一个卷积核在图像所有位置滑动时使用同一套权重
14. b) 防止边缘信息丢失并控制输出尺寸
15. b) 增加参数数量
16. c) CNN展平的是“高阶特征”, 保留了语义信息
17. c) 全部使用小卷积核 (3x3) 堆叠
18. a) 取决于网络中所有可训练参数 (权重W和偏置b) 的总数量
19. d) 尺寸大小 x 输入通道数 ($K \times K \times C_{in}$)
20. c) 49x49x16