

一、选择题（每题3分，共30分）

1. 下列哪一项不属于监督学习？
 - A. 分类
 - B. 回归
 - C. 聚类
 - D. 识别
2. 在模型评估中，如果正负样本比例不平衡，以下哪个指标可能不是一个好的评估方法？
 - A. 精确率 (Precision)
 - B. 召回率 (Recall)
 - C. F1-Score
 - D. 准确率 (Accuracy)
3. 在神经网络中，哪个激活函数在输入为正数时导数恒为1，有效避免了梯度消失问题？
 - A. Sigmoid
 - B. Tanh
 - C. ReLU
 - D. Softmax
4. 为了解决简单循环神经网络（RNN）在处理长序列时可能出现的梯度消失问题，引入了哪种模型？
 - A. 前馈神经网络 (Feed-forward Neural Network)
 - B. 卷积神经网络 (Convolutional Neural Network)
 - C. 长短时记忆模型 (LSTM)
 - D. 感知机 (Perceptron)
5. [cite_start]在卷积神经网络（CNN）中，哪种操作通过共享权重参数来减少模型参数数量，并能在一定程度上防止过拟合？ [cite: 14]
 - A. 池化 (Pooling)
 - B. 卷积 (Convolution)
 - C. 批归一化 (Batch Normalization)
 - D. 填充 (Padding)
6. 在循环神经网络（RNN）的应用模式中，常用于文本情感分类的是哪一种？
 - A. 多对多
 - B. 多对一
 - C. 一对多
 - D. 一对一
7. 在自注意力（Self-Attention）机制中，计算注意力权重时，每个单词的查询向量（Query）需要和什么进行点积运算？
 - A. 其他所有单词的值向量 (Value)
 - B. 其他所有单词的键向量 (Key)

- C. 其他所有单词的查询向量 (Query)
 - D. 其他所有单词的词向量 (Word Embedding)
8. 下列哪种正则化方法是在训练神经网络时，以一定概率随机“丢掉”一部分神经元？
- A. L1 正则化
 - B. L2 正则化
 - C. 批归一化 (Batch Normalization)
 - D. Dropout
9. 在逻辑回归 (Logistic Regression) 模型中，其输出值 y 的范围是？
- A. (0, 1)
 - B. (-1, 1)
 - C. [0, 1]
 - D. 任何实数
10. 误差反向传播算法解决了神经网络中的哪一个核心难题？
- A. 网络结构设计
 - B. 参数优化
 - C. 激活函数选择
 - D. 数据预处理

二、判断题（每题2分，共10分）

- 1. 训练集、验证集和测试集的数据可以有交叉，以提高模型的训练效果。（错）
- 2. 经验风险越小，模型在测试集上的表现就一定越好。（错）
- 3. 单层感知机可以模拟“逻辑异或”这一线性不可分的功能。（错）
- 4. 在循环神经网络中，参数 W_x 、 W_o 和 W_h 在不同时间步是共享（复用）的。（对）
- 5. Transformer 模型通过引入自注意力机制，未对输入的方向、距离信息进行编码。（对）

三、填空题（每空2分，共20分）

- 1. 机器学习根据数据利用方式可分为 **监督学习**、**无监督学习** 和半监督学习。
- 2. F1-Score 是 **精确率 (Precision)** 和 **召回率 (Recall)** 的调和平均数。
- 3. 循环神经网络 (RNN) 是一类用于处理 **序列数据** 的网络结构。
- 4. LSTM 模型通过引入三种门结构来控制信息流动，这三种门分别是 **输入门**、**遗忘门** 和输出门。
- 5. 梯度下降算法中，梯度的反方向是函数值 **下降最快** 的方向。
- 6. 为了防止过学习，可以在损失函数中加入 **正则化项（或惩罚项）** 来降低模型复杂度。

7. 批归一化 (Batch Normalization) 通过规范化手段, 将神经网络每层神经元的输入值分布改变成均值为 0、方差为1的标准正态分布。

四、简答题 (每题10分, 共30分)

1. 什么是过拟合? 请解释导致期望风险增加的原因, 并列举至少两种防止过拟合的方法。
 - **过拟合**: 指模型在训练集上表现很好 (经验风险小), 但在未见过的数据 (如测试集) 上表现很差 (期望风险大) 的现象, 即模型的泛化能力很弱。
 - **原因**: 当模型过于复杂, 反复学习训练数据后, 虽然经验风险会持续降低, 但模型的泛化误差 (err) 会增加, 导致期望风险 (真实风险) 随之增加。
 - **防止方法** (任选两种):
 - **结构风险最小化**: 引入正则化项 (如L1或L2正则化) 来惩罚模型的复杂度。
 - **Dropout**: 在训练过程中随机丢弃一部分神经元, 降低网络复杂度。
 - **批归一化 (Batch Normalization)**: 对网络中间层的输入进行归一化, 有助于稳定训练过程, 也有一定的正则化效果。
2. 请简述循环神经网络 (RNN) 为什么会出现梯度消失问题, 并说明长短时记忆模型 (LSTM) 是如何缓解这个问题的。
 - **RNN梯度消失原因**: RNN在反向传播计算梯度时, 需要用到链式求导法则。在长序列中, 梯度的计算涉及到多个激活函数 (如tanh) 导数的连乘。由于tanh函数的导数取值范围是(0, 1], 多个小于1的数连乘会导致最终的梯度值趋近于0, 这就是梯度消失问题。
 - **LSTM如何缓解**: LSTM引入了内部记忆单元 (Cell State) 和门控机制。内部记忆单元的信息更新是通过遗忘门和输入门控制的, 其状态更新公式为 $c_t = f_t c_{t-1} + i_t \tanh(\cdot)$ 。在对 c_{t-1} 求偏导时, 会包含一项遗忘门 f_t 的输出。如果遗忘门选择保留旧信息 (f_t 接近1), 梯度就可以在时间步之间很好地传递而不会消失, 从而缓解了梯度消失问题。
3. 请对比卷积神经网络 (CNN) 中的卷积操作与Transformer中的自注意力机制在感受野上的主要区别。
 - **CNN的感受野**: CNN中的卷积操作使用**固定大小的卷积核**在输入数据 (如图像) 上滑动, 每个输出点的取值仅依赖于输入中一个局部区域内的点。这个局部区域就是感受野, 其大小和形状是**预先固定**的, 并且通过参数共享在整个输入上重复使用。
 - **Transformer的自注意力机制**: Transformer中的自注意力机制的感受野是**动态的、全局的**。它通过计算查询向量 (Query) 与所有键向量 (Key) 的相似度来得到注意力权重, 然后用这些权重去加权所有的值向量 (Value)。这意味着模型可以自适应地决定感受野的大小和形状, 关注输入序列中任意位置的信息, 而不仅仅是局部邻域。

五、 计算题（10分）

假设一个卷积神经网络的输入图像大小为 32×32 像素，使用一个大小为 5×5 的卷积核进行卷积操作，步长（Stride）为1，边缘填充（Padding）像素数为2。请根据公式计算并写出卷积操作后得到的特征图（Feature Map）的分辨率（大小）。

解：

根据卷积结果分辨率的计算公式：

$$\text{输出尺寸} = \frac{W+2P-F}{S} + 1$$

其中：

- W = 输入图像大小 = 32
- F = 卷积核大小 = 5
- P = 边缘填充像素数 = 2
- S = 步长 = 1

将数值代入公式：

$$\text{输出尺寸} = \frac{32+2 \times 2-5}{1} + 1$$

$$\text{输出尺寸} = \frac{32+4-5}{1} + 1$$

$$\text{输出尺寸} = \frac{31}{1} + 1$$

$$\text{输出尺寸} = 31 + 1 = 32$$

答： 经过卷积操作后，得到的特征图分辨率为 32×32 。

选择题解析

1. 下列哪一项不属于监督学习？

- 正确答案：C. 聚类

• 解析：

- [cite_start]监督学习 (Supervised Learning) 的目标是利用**带有标签信息**的训练数据 $\{(x_i, y_i)\}_{i=1}^n$ 来学习一个最优的映射函数 f [cite: 1][cite_start]。分类、识别和回归都属于监督学习的范畴，因为它们都需要有标签的数据（例如，图片和它的类别，或房屋特征和它的价格）来进行训练 [cite: 1, 4]。
- [cite_start]无监督学习 (Unsupervised Learning) 则是直接从**无标签数据** $\mathcal{D} = \{x_i, i = 1, \dots, n\}$ 出发学习映射函数 [cite: 1]。聚类（Clustering）是无监督学习的典型例子，其目的是在没有预先定义类别的情况下，将相似的数据点分组。

2. 在模型评估中，如果正负样本比例不平衡，以下哪个指标可能不是一个好的评估方法？

- 正确答案：D. 准确率 (Accuracy)
- 解析：
 - [cite_start]材料中明确指出：“如果正负样例比例不平衡, 准确率不是一个好的方法” [cite: 3]。
 - [cite_start]原因：准确率计算的是正确预测的样本数占总样本数的比例 ($ACC = \frac{TP+TN}{P+N}$) [cite: 3]。在一个类别极不平衡的数据集（例如99%的样本是负例，1%是正例）中，模型即使将所有样本都预测为负例，也能获得99%的准确率，但这显然是一个没有区分能力的无效模型。而精确率、召回率和F1-Score更能反映模型对少数类别的识别能力。

3. 在神经网络中，哪个激活函数在输入为正数时导数恒为1，有效避免了梯度消失问题？

- 正确答案：C. ReLU
- 解析：
 - [cite_start]ReLU (Rectified Linear Unit) 函数的定义为 $f(x) = \max(0, x)$ [cite: 7]。
 - [cite_start]其导数特性是：当输入 $x \geq 0$ 时，导数恒为1 [cite: 7][cite_start]。这一特性使得梯度在反向传播过程中能够保持其大小，不会因为多层连乘而趋向于0，从而有效避免了梯度消失问题 [cite: 7]。
 - [cite_start]Sigmoid 函数的导数小于1，在深度网络中容易导致梯度消失 [cite: 7][cite_start]。Softmax 通常用于输出层进行多分类，而不是作为隐藏层的激活函数来解决梯度问题 [cite: 8]。

4. 为了解决简单循环神经网络（RNN）在处理长序列时可能出现的梯度消失问题，引入了哪种模型？

- 正确答案：C. 长短时记忆模型 (LSTM)
- 解析：
 - [cite_start]材料中提到，简单的循环神经网络由于激活函数（如tanh）的导数小于1，在处理长序列时，多个小数连乘会导致梯度求导结果很小，从而引发梯度消失问题 [cite: 17]。
 - [cite_start]紧接着明确指出：“为了缓解这个问题，**长短时记忆模型（Long Short-Term Memory, LSTM）**被提出” [cite: 17][cite_start]。LSTM通过引入内部记忆单元和门控结构，使得梯度能够更好地在长序列中传递 [cite: 17, 23]。

5. 在卷积神经网络（CNN）中，哪种操作通过共享权重参数来减少模型参数数量，并能在一定程度上防止过拟合？

- 正确答案：B. 卷积 (Convolution)
- 解析：
 - [cite_start]材料在描述卷积算子的特点时提到：“局部感知、**参数共享**：卷积操作的权重参数可学习、可被重复使用，**减少了参数总数，一定程度防止过拟合**” [cite: 14]。
 - 参数共享意味着一个卷积核（一组权重）会滑过整个输入图像，用同一套参数去检测图像不同位置的特征，这极大地减少了模型需要学习的参数量，与每个像素都连接一个独立权重的前馈网络形成鲜明对比。

6. 在循环神经网络（RNN）的应用模式中，常用于文本情感分类的是哪一种？

- 正确答案：B. 多对一
- 解析：
 - [cite_start]**多对一 (Many-to-one)** 模式的特点是输入一个序列数据（多个单元），而输出只有一个单元 [cite: 20]。
 - [cite_start]文本情感分类任务正是如此：输入是一个由多个单词组成的句子（多输入），输出是整个句子的情感类别（如“积极”或“消极”，单输出）。材料中明确将“多对一”模式与“**情感分类**”应用相对应 [cite: 20]。

7. 在自注意力（Self-Attention）机制中，计算注意力权重时，每个单词的查询向量（Query）需要和什么进行点积运算？

- 正确答案：B. 其他所有单词的键向量 (Key)
- 解析：
 - [cite_start]自注意力机制的核心是计算一个词与句子中其他所有词的关联程度。这个关联度（或称注意力分数）是通过该词的**查询向量 (Query)** 与其他所有词的**键向量 (Key)** 进行点积得到的 [cite: 26]。
 - [cite_start]材料中的图示和公式清晰地展示了这一点：例如，为了计算单词 w_3 对其他单词的注意力，需要用 q_3 分别与 k_1, k_2, k_3, k_4 进行点积运算，得到原始的注意力分数 $\alpha_{3j} = q_3 \cdot k_j$ [cite: 26]。

8. 下列哪种正则化方法是在训练神经网络时，以一定概率随机“丢掉”一部分神经元？

- 正确答案：D. Dropout
- 解析：
 - [cite_start]Dropout 的定义就是“指在训练神经网络的过程中随机丢掉一部分神经元来降低神经网络的复杂度，从而防止过拟合” [cite: 23]。
 - [cite_start]L1和L2正则化是通过在损失函数中添加参数的范数作为惩罚项来实现的 [cite: 24] [cite_start]。批归一化是一种规范化手段，主要目的是加速收敛和克服梯度消失 [cite: 24, 31]。

9. 在逻辑回归（Logistic Regression）模型中，其输出值 y 的范围是？

- 正确答案：A. (0, 1)
- 解析：
 - [cite_start]逻辑回归使用了Sigmoid函数（或称Logistic函数） $y = \frac{1}{1+e^{-z}}$ 来将线性回归的输出映射到一个概率空间 [cite: 4]。
 - [cite_start]Sigmoid函数的特性决定了其输出值永远在0和1之间，但不会等于0或1。因此，其取值范围是开区间 (0, 1) [cite: 4]。这个输出值通常被解释为某个事件发生的概率。

10. 误差反向传播算法解决了神经网络中的哪一个核心难题？

- 正确答案：B. 参数优化
- 解析：
 - [cite_start]在神经网络的历史发展部分，材料提到误差反向传播算法“解决了多层感知机中参数优化这一难题” [cite: 8]。
 - [cite_start]在参数优化章节中也再次说明，模型会利用反向传播算法将损失误差由后向前传递，然后通过梯度下降算法对神经网络中的参数进行更新，这个过程就是参数优化 [cite: 12]。