

一. 单选题（共10题，47.2分）

1. 注意力机制的核心思想是：

- A. 对所有输入数据赋予相同权重
- **B. 动态分配权重以聚焦关键信息**
- C. 仅处理序列的最后一个时间步
- D. 完全忽略历史信息

2. Nadaraya-Watson核回归中，注意力权重通过以下哪种方式计算？

- A. 查询与键的余弦相似度
- B. 查询与键的欧式距离
- **C. 核函数(如高斯核)**
- D. 随机初始化

3. 注意力评分函数的作用是：

- **A. 计算查询与键的相似度**
- B. 直接生成输出序列
- C. 替代反向传播
- D. 初始化模型参数

4. 加性注意力的数学形式包含：

- A. 查询与键的点积
- **B. 查询与键的串联后通过全连接层**
- C. 随机采样
- D. 仅使用键的范数

5. 点积注意力的优势是：

- **A. 计算复杂度低**
- B. 必须使用高斯核

- C. 仅适用于短序列
- D. 需要更多参数

6. 注意力权重通常通过什么函数归一化？

- A. Sigmoid
- B. ReLU
- **C. Softmax**
- D. Tanh

7. 在机器翻译中，注意力机制可以缓解什么问题？

- A. 梯度消失
- **B. 长序列信息丢失**
- C. 过拟合
- D. 计算量过大

8. 注意力机制最早受什么启发？

- A. 图像处理
- **B. 人类视觉系统**
- C. 语音识别
- D. 强化学习

9. Nadaraya-Watson核回归是注意力机制的：

- A. 神经网络实现
- **B. 非参数化实例**
- C. 反向传播算法
- D. 优化器

10. Bahdanau注意力中，上下文向量的计算方式是：

- **A. 编码器隐藏状态的加权和**
- B. 解码器隐藏状态的均值
- C. 查询与键的拼接
- D. 随机采样

二. 多选题（共5题，24分）

11. 注意力评分函数的常见类型包括：

- ☒ A. 加性注意力
- ☒ B. 点积注意力
- ☒ C. 缩放点积注意力
- ☐ D. 随机注意力

12. 注意力机制可用于哪些任务？

- ☒ A. 机器翻译
- ☒ B. 图像分类
- ☒ C. 文本摘要
- ☒ D. 语音识别

13. 关于注意力权重的正确描述是：

- ☒ A. 和为1
- ☒ B. 非负
- ☒ C. 可动态变化
- ☐ D. 仅由查询决定

14. 注意力机制缓解了传统Seq2Seq的哪些问题？

- ☒ A. 信息瓶颈（固定长度上下文向量）
- ☒ B. 长距离依赖
- ☒ C. 梯度消失
- ☐ D. 计算量过大

15. 以下哪些是注意力机制的必要组件？

- ☒ A. 查询（Query）
- ☒ B. 键（Key）

- ☒ C. 值 (Value)
- ☐ D. 卷积核

三. 简答题（共3题，14.4分）

16. 简述注意力机制的基本工作原理。

注意力机制通过计算**查询（Query）与键（Key）之间的相似度**，得到注意力权重，并将其用于加权求和值（Value），从而动态聚焦于输入序列中与当前任务最相关的信息，提高模型对长距离依赖的建模能力。

17. 比较加性注意力和点积注意力的优缺点。

加性注意力使用一个前馈神经网络学习相似度，表达能力强但计算开销较大；点积注意力通过计算查询与键的内积，相对更高效，尤其在矩阵并行处理时性能更优，但在维度较大时需使用缩放处理避免梯度过小。

18. 为什么注意力权重需要归一化？常用什么函数？

因为注意力机制本质上是对不同输入的加权求和，归一化确保所有权重非负且总和为1，使其具备“概率解释”；常用归一化函数是 **Softmax**。