一. 单选题（共6题，30分）

1. **多头注意力中每个头的计算方式与什么相同？**
   - A. 残差连接
   - **B. 单个缩放点积注意力**
   - C. 全连接层
   - D. 池化操作
2. **在 Transformer 模型中，LayerNorm 的作用是？**
   - A. 计算注意力
   - **B. 进行层归一化以稳定训练**
   - C. 增强非线性
   - D. 提供位置信息
3. **在 Bahdanau 注意力机制中，注意力权重是如何计算的？**
   - A. 余弦相似度
   - B. 固定平均
   - C. softmax(Wx)
   - **D. 使用可学习的打分函数**
4. **Transformer 中引入位置编码的主要原因是？**
   - A. 避免梯度消失
   - **B. 因为模型不具备顺序信息**
   - C. 为了减少参数数量
   - D. 为了增强卷积能力
5. **多头注意力的主要优点是？**
   - **A. 能从不同子空间学习信息**
   - B. 降低了模型复杂度

- C. 使模型变得简单
- D. 提高了并行计算效率

6. **Transformer 中的前馈神经网络是？**
   - A. 生成位置编码
   - B. 多头注意力的一部分
   - **C. 在每层注意力机制后独立应用**
   - D. 用于计算注意力权重

二. 多选题（共10题，50分）

7. **以下哪些是 Bahdanau 注意力机制的组成部分？**
   - ✅ A. 解码器的隐藏状态
   - ✅ B. 上下文向量
   - ❌ C. 固定位置编码
   - ✅ D. 可学习的注意力函数
8. **位置编码的常见实现方式有哪些？**
   - ✅ A. 学习位置嵌入
   - ✅ B. 正余弦函数
   - ❌ C. 卷积编码
   - ❌ D. 位置掩码
9. **多头注意力的优势有哪些？**
   - ✅ A. 提供不同子空间的表示
   - ✅ B. 提高模型泛化能力
   - ❌ C. 降低训练复杂度
   - ✅ D. 实现并行处理

10. **Transformer中前馈神经网络(FFN)的特点是？**

- ✅ A. 跨时间步共享权重
- ✅ B. 使用ReLU激活
- ✅ C. 输入输出维度相同
- ✅ D. 单隐藏层全连接网络

11. **Transformer解码器与编码器的差异包括？**

- ✅ A. 解码器包含编码器-解码器注意力层
- ✅ B. 解码器使用掩码自注意力
- ✅ C. 编码器包含残差连接
- ❌ D. 解码器无前馈网络

12. **以下哪些机制有助于 Transformer 稳定训练？**

- ✅ A. Dropout
- ✅ B. 自注意力
- ✅ C. 残差连接
- ✅ D. 层归一化

13. **位置编码可直接相加到输入嵌入的原因包括？**

- ✅ A. 模型可自动分离信息
- ❌ B. 满足线性变换关系
- ✅ C. 输入嵌入与位置编码维度相同
- ❌ D. 位置编码无需训练

14. **缩放点积注意力的计算涉及哪些操作？**

- ✅ A. 向量点积

- ✅ B. 除以维度的平方根
- ❌ C. 残差连接
- ✅ D. Softmax

15. **自注意力机制的优势有哪些？**

- ✅ A. 参数少于 CNN
- ❌ B. 训练速度慢
- ✅ C. 易于并行处理
- ✅ D. 可以捕捉长距离依赖

16. **Transformer 中注意力机制包括哪些类型？**

- ✅ A. 多头注意力
- ✅ B. 编码器-解码器注意力
- ❌ C. Bahdanau 注意力（原用于 RNN）
- ✅ D. 自注意力

三. 简答题（共4题，20分）

17. **解释多头注意力机制的计算流程以及它相比单头注意力的优势。**

> 多头注意力将输入分别投影到多个查询、键、值子空间中，分别计算缩放点积注意力，最后将所有头的输出拼接并再线性变换。它相比单头注意力能并行捕捉不同子空间的特征信息，从而提升模型的表示能力。

18. **自注意力机制在序列建模中是如何捕捉长程依赖的?请简要说明其实现过程。**

> 自注意力通过对序列中所有位置之间进行相似度计算，使得每个位置都能“看到”整个序列的信息，不依赖于固定步长，因而能直接建模远距离依赖关系。

19. **描述 Transformer 模型的基本结构，并解释其训练时的主要优化技巧。**

> Transformer由编码器和解码器堆叠组成，每个编码器包含多头自注意力和前馈网络，解码器还包含编码器-解码器注意力。训练中采用残差连接、LayerNorm、Dropout等技术优化训练稳定性，使用Adam优化器和学习率调度策略加速收敛。

20. **详细解析Transformer解码器的掩码自注意力有何作用?为什么要是因果掩码遮蔽?**

> 掩码自注意力确保每个解码步骤只能访问当前位置及其前面的位置，防止“看到未来信息”；因果掩码实现了这一点，保证模型训练时符合自回归生成的顺序性。