**一、填空题 (每题2分，共40分)**

1. 自然语言处理的目标是实现\_\_\_\_\_\_\_\_\_。让计算机能够理解和生成自然语言
2. 在自然语言处理中，POS Tagging是指识别文本中每个单词的\_\_\_\_\_\_\_\_\_。词性（Part-of-Speech）
3. 机器翻译是自然语言处理的一个应用方向，其主要目标是将一种语言的文本自动翻译成另一种语言，目前常见的机器翻译方法主要包括统计机器翻译和\_\_\_\_\_\_\_\_\_机器翻译。神经（或神经网络）
4. 自然语言处理研究的核心任务包括 \_\_\_\_\_\_ 等（请列出4项任务）。分词、词性标注、命名实体识别、句法分析，文本分类，情感分析，信息抽取，问答系统，机器翻译
5. 20世纪60年代，第一个大规模的语料库是\_\_\_\_\_\_\_\_\_。布朗语料库（Brown Corpus）
6. 自然语言处理的发展经历了基于规则的方法、基于机器学习的方法、基于\_\_\_\_\_\_\_\_\_的方法等多个阶段。深度学习
7. 语言模型的目的是估计一个句子出现的\_\_\_\_\_\_\_\_\_。概率
8. 在n元语言模型中，计算词的条件概率时，我们假设词的出现仅与\_\_\_\_\_\_\_个词相关。前（n-1）
9. \_\_\_\_\_\_\_技术可以缓解n元语言模型中的数据稀疏问题。平滑
10. 在前馈神经网络中，输出层通常使用\_\_\_\_\_\_\_\_\_激活函数来处理多分类问题。softmax
11. 在神经网络中，损失函数的作用是量化模型的预测与\_\_\_\_\_\_\_\_\_之间的差异。实际标签（或真实值、ground truth）
12. 循环神经网络（RNN）适用于处理\_\_\_\_\_\_\_\_\_数据。序列（或时序）
13. RNN的输出不仅依赖于当前时刻的输入，还会依赖于\_\_\_\_\_\_\_\_\_。上一个时刻的状态（或隐藏状态）
14. 反向传播算法的核心是通过链式法则计算每一层的\_\_\_\_\_\_\_\_\_。梯度
15. \_\_\_\_\_\_\_\_模型是基于Transformer结构的编码式预训练语言模型。BERT
16. 预训练语言模型的一大优势是能够通过\_\_\_\_\_\_\_\_\_，适应多种下游任务。微调（fine-tuning）
17. GPT模型采用了\_\_\_\_\_\_\_\_\_训练策略，生成下一个单词的条件概率。自回归（或auto-regressive）
18. 在计算语言模型时，\_\_\_\_\_\_\_\_\_问题会影响到高阶n元模型的效果。数据稀疏
19. 基于神经网络的语言模型避免了传统n元模型中的\_\_\_\_\_\_\_\_\_问题。数据稀疏/数据存储
20. BERT预训练任务包括\_\_\_\_\_\_\_\_\_和\_\_\_\_\_\_\_\_\_。掩码语言建模（MLM）和下一句预测（NSP）

**二、判断题 (每题2分，共20分，正确填T，错误填F)**

1. (T ) 自然语言处理的研究内容包括语言的理解和语言的生成。
2. (T ) 神经网络模型的一个优势是能够在较长的文本中建模长距离的依赖关系。
3. (F ) 语言模型只能用于文本生成任务，不能用于文本分类。
4. (F ) BERT模型使用的是单向语言模型方法，而GPT使用的是双向语言模型方法。
5. (T ) 在神经网络语言模型中，词向量表示能有效减少数据稀疏问题。
6. (F ) ELMo模型是一种基于Transformer结构的语言模型。
7. (F ) 预训练语言模型需要大量的标注数据进行训练。
8. (F ) 在深度神经网络中，激活函数的选择对网络性能没有显著影响。
9. (F ) 在卷积神经网络中，每一层的神经元都与前一层的所有神经元相连接。
10. (F ) BART是一种纯解码结构的预训练语言模型。

**三、简答题 (每题10分，共40分)**

1. 简要介绍神经网络语言模型与传统n元语言模型的区别，并说明神经网络模型如何解决n元语言模型中的问题。

* **传统n元模型基于统计方法，假设当前词只与前n-1个词有关，存在数据稀疏问题。**
* **神经网络语言模型使用词向量将词映射到稠密空间，能捕捉词间相似性，缓解数据稀疏。**
* **神经网络能更好地泛化到未见过的组合，并通过非线性变换建模复杂关系。**

1. 解释循环神经网络（RNN）与前馈神经网络的区别，并说明RNN的主要优势。

* **前馈网络：无记忆，每个输入独立处理。**
* **RNN：具有记忆能力，输出依赖当前输入和之前状态，适用于处理序列。**
* **优势：可捕捉序列中的上下文信息，实现时间依赖建模。**

1. 请简述预训练语言模型的三种模式，并给出具体例子。

* **编码式：BERT**
* **编解码式：BART**
* **解码式：GPT**

1. 介绍Transformer的主要结构，比较Transformer编码器和解码器的差异。

* **核心结构包括：多头自注意力机制、前馈网络、残差连接与LayerNorm。**
* **编码器：使用自注意力，处理输入序列，输出上下文表示。**
* **解码器：包括自注意力和编码器-解码器注意力，逐步生成目标序列。**