**一、选择题**

1．NLP的主要研究任务和应用包括以下( )。

A.图像识别 **B. 语音识别**  **C.文本生成**  D.计算机视觉

2．下列属于句法分析的技术是( )。

A.依存分析 B.词袋模型 C.词嵌入模型 D.循环神经网络

3．下列衡量分类器性能的指标是( )。

A.精确率 B.正确率 C.召回率 **D. 以上都对**

4． 在自然语言处理中，BOW模型的一个主要特征是( )。

A. 词语顺序信息被保留

B. 仅记录词语是否出现或出现频次

C. 仅适用于短文本的表示

D. 将每个词表示为一个独立向量，但不记录频次

5．不属于未登录词的是（）。

A.网络热门词语

B.人名、地名和组织机构名

C.化学试剂的名称

D.经典文学作品

6．CBOW模型是通过( )预测( )。

A.中心词，上下文 B.上下文，中心词

C.中心词，周围词 D.周围词，中心词

7．以下公式用来计算文本相似度的是( )。

A.欧式距离 B.马氏距离  **C.余弦相似度**  D.曼哈顿距离

8．在N-gram模型中，以下类型最常用的是( )。

A. 1-gram

1. 2-gram
2. 3-gram
3. 以上都对

9．前馈神经网络传递信息的方式为( )。

**A. 逐层前向传递**

B. 随机传递

C. 逐层反向传递

D. 同时传递

10．中文NLP的基本流程由语料获取、( )、文本向量化、模型构建、模型训练和模型评价6部分组成。

A.语料预处理 B.中文分词 C.去停用词 D.词性标注

11. RNN模型主要用于处理的数据类型为( )。

**A. 时间序列数据** B.非时间序列数据 C.图像数据 D.无序数据

12. Transformer模型中，Encoder包含以下重要组件的为( )。

**A.多头自注意力** B.单层神经元 C.循环连接 D.卷积层

13. 以下关于BERT和GPT设计差异的描述，正确的为( )。

**A. BERT使用Encoder堆叠结构，而GPT使用Decoder堆叠结构**

B. BERT和GPT均采用双向Transformer架构

C. BERT仅使用单向注意力，而GPT使用双向注意力

D. BERT和GPT均不包含Transformer模块

14. 以下激活函数中可以减少梯度消失问题的为( )。

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

15. 在依存句法分析中，主要体现词语间的关系为( )。

A．依存关系 B．词袋模型 C．词向量表示 D．句子分词

16. 情感分析的基础性工作是( )。

A．文本信息抽取 B．文本的主客观分类

C．情感分类 D．情感极性判断

17. 问答系统流程由问题理解、( )、答案生成3个部分组成。

1. 词性标注 B.关键词提取 C.问题分类  **D.知识检索**

18. CRF模型思想主要来源于（ ）。

A.无向图模型 B.最大熵模型

**C.马尔科夫随机场** D.统计方法

19. 以下方法中将一个列表 [1, 2, 3] 中的元素逐一添加到新的列表 new\_list中的是（ ）。

1. new\_list = list([1, 2, 3])

B. new\_list = [element for element in [1, 2, 3]]

C. new\_list = ([1, 2, 3])

D. new\_list = [[1, 2, 3]]

20. 如果要定义一个带有默认参数值的函数，以下选项中正确的是（ ）。

A. def func(a, b=2):

B. def func(a=2, b):

C. def func(a=2, b=2, c):

D. def func(a, b, c=2, d)

21．政府部门利用NLP技术分析人们对某一事件、政策法规或社会现象的评论，实时了解百姓的态度，这属于NLP研究内容的( )。

A.信息检索 B.文本分类 C.信息过滤 D.自动文摘

22．不属于NLP应用场景的是( )。

A.百度翻译 B.图灵机器人 **C.数据挖掘** D.微信语音转文字

23．在python程序中安装需要的程序包时使用的命令是( )。

A.pip package\_name B.pip uninstall package\_name

C.pip install package\_name D.pip upgrade package\_name

24．不属于基于规则的分词方法的是( )。

A.正向最大匹配法 B.逆向最大匹配法

C.反向最大匹配法 D.双向最大匹配法

25．不属于关键词提取算法的是( )。

A.TF-IDF算法 B.TextRank算法 C.主题模型算法 D.关联算法

26．独热(One-Hot)表示的缺点不包括( )。

**A.构造简单**  B.维数过高 C. 矩阵稀疏 D. 不可以保留语义

27．DM模型与CBOW模型的区别为( )。

A.DM模型的输入包括上下文

B.DM模型预测目标词出现的概率

C.DM模型输入不仅包括上下文，而且还包括相应的段落

D.CBOW模型输入包括上下文

28．RNN适用于处理视频、语音、文本等与**时序**相关的问题，其常见的应用领域不包括( )。

A.图像处理 CNN B.视频剪辑

B 视频本质上是一帧帧连续的图像序列，RNN 可以用于视频摘要、动作识别、帧间预测等时序分析任务。

C.语音识别 D.文本相似度计算

D 属于 NLP 序列对比任务，RNN／BiRNN、注意力机制等都是主流做法。

29．RNN**经典结构**的输入和输出的序列长度为( )。

**A.多对一**  B.一对多

C.等长的多对多 D.非等长的多对多

30.下列关于LSTM说法不正确的是( )。

A.通过改进使RNN具备避免梯度消失的特性

**B.LSTM只能够刻画出输入数据中的短距离的相关信息，不能够捕捉到具有较长时间间隔的依赖关系**

C.LSTM神经网络模型使用门结构实现了对序列数据中的遗忘与记忆

D.使用大量的文本序列数据对LSTM模型训练后，可以捕捉到文本词的依赖关系，训练好的模型就可以根据指定的文本生成后序的内容

31．以下选项中，( )不是 Transformer 的特点。

A.位置信息编码 **B.循环神经网络** C.残差连接 D.多头注意力机制

32．在 Transformer 中，残差连接的主要作用是( )。

A.增加计算复杂度 **B.改善训练过程中的梯度流动**

C.增加模型的容量 D.提高模型的可解释性

33．在 Python 中，缩进通常用于( )。

A.变量声明 B.定义函数和控制结构

C.注释 D.数据类型转换

34．以下选项中，可以创建一个包含数字的列表的是( )。

A.list = (1, 2, 3) B.list = {1, 2, 3}

C.list = [1, 2, 3] D.list = <1, 2, 3>

35．依存句法分析的核心是( )。

A.确定短语结构 B.分析语义角色

C.生成句法树 D.找出词之间的依存关系

36．句法树的每个节点代表( )。

A.一个短语或词类 B.一个单词

C.句子的情感 D.句子的长度

37．离散表示方法中，( )方法将词表示为一个二进制向量，其中只有一个元素为1，其余为0。

A.Bag of Words (BoW)  **B.One-hot Encoding** C.TF-IDF D.Word2Vec

38．在Bag of Words (BoW)模型中，文本表示为( )。

A.词的顺序  **B.词的频率**  C.词的语义 D.词的长度

39．在Word2Vec模型中，CBOW（Continuous Bag of Words）的主要目标是( )。

**A.从上下文预测目标词**  B.从目标词生成上下文

C.生成固定长度的向量 D.计算语义相似度

40．N-gram模型中，“N”表示( )。

A.单词的数量  **~~B.文本的长度~~**

C.上下文窗口的大小 D.词的类别

41．依存句法分析的主要目的是( )。

A.确定词汇的顺序 B.理解词与词之间的关系

C.生成句子的语法树 D.进行文本分类

42．在句法分析中，线型图（Dependency Graph）通常用于表示( )。

A.词频 B. 主题模型算法 C. 介词 D.TextRank算法

43．在依存句法中，通常被视为“根”节点的词是( )。

A.主语 B.谓语动词 C.形容词 D.矩阵稀疏

44．神经网络中的隐藏层的主要作用是（）

A.输入数据的预处理 B.生成最终输出

**C.学习输入数据的特征**  D.保存模型参数

45．在神经网络中，梯度下降法的主要目的是( )。

**A.最小化损失函数**  B.最大化输出值

C.选择激活函数 D.增加隐藏层的数量

46．以下算法中，属于梯度下降法的一种变种的是( )。

A支持向量机（SVM） **B.随机梯度下降（SGD）**

C.决策树 D. K-means

47. CNN中，卷积操作的目的是 ( )。

A.输入数据降维 **B.提取局部特征**

C.增加计算复杂度 D.处理序列数据

48. LSTM（长短期记忆网络）的主要优势是（ ）。

A.处理静态数据  **B.解决梯度消失问题** C.提高训练速度 D.生成图像

49. 在Encoder/Decoder架构中，掩码（masking）的主要作用是（ ）。

A.提高模型的计算效率  **B.防止模型访问未来的信息**

C.加速训练过程 D.简化模型结构

50．以下方法中，用于句法分析中的句法树生成的是( )。

A.词向量 B.递归神经网络

C.句法分析器 D.朴素贝叶斯分类器

51．在神经网络训练中，以下不是激活函数的为( )。

**A. Sin**  B. ReLU C. Tanh D. Softmax

52．神经网络中的激活函数主要作用是( )。

A.数据归一化 B.中文分词 C.去停用词  **D.引入非线性特性**

53．LSTM通过( )相关的门机制来解决RNN的梯度消失问题。

A.激活门、选择门 B.正向传播门、反向传播门

C.线性门、非线性门 **D.输入门、遗忘门、输出门**

54．循环神经网络最主要的优势是( )。

A.能够处理图像数据 **B.能够处理时间序列数据和具有上下文依赖关系的序列数据**

C.计算速度非常快 D.可以并行计算每个时间步的输出

55．在 Transformer 中，残差连接的主要作用是( )。

A.增加计算复杂度  **B.改善训练过程中的梯度流动**

C. 增加模型的容量 D.提高模型的可解释性

56．自注意力机制中的Query与Key的点积结果经过以下处理后形成注意力权重的为( )。

A.正则化 B.归一化

**C. Softmax** D.激活函数

57．Transformer中的多头注意力机制的主要作用是( )。

A.提高模型的训练速度

**B.提高模型对不同特征的捕捉能力**

C.减少模型的参数数量

D.增加模型的鲁棒性

58．在 Python 中，使用 insert() 方法时，以下说法正确的是( )。

A.insert() 将元素添加到列表的末尾 B.insert() 需要两个参数：索引位置和要插入的元素

C.insert() 只能添加一个元素 D.insert() 不能插入重复元素

59．在 Python 中，如果要从列表中删除一个指定位置的元素，并返回被该元素，应使用的方法是( )。

A.del 语句 B.remove() C.pop() D. clear()

60．在Python的列表操作中，以下表达式中可以对列表lst的所有元素进行降序排序是（ ）。

A. lst.sort(reverse=True) B. sorted(lst)

C. sorted(lst, reverse=True) D. lst.sort()

61．NLP的主要研究任务和应用包括以下( )。

A.图像识别  **B. 语音识别 C.文本生成** D.计算机视觉

62．TextRank算法属于( )的方法。

A.监督学习 B.无监督学习 C.半监督学习 D.强化学习

63．CBOW模型与Skip-gram模型的主要区别在于( )。

A.**CBOW使用上下文词预测中心词，Skip-gram使用中心词预测上下文词**

B.CBOW不使用任何上下文信息

C.Skip-gram模型使用固定窗口的上下文，CBOW模型不使用窗口

D.Skip-gram模型计算速度更快

64．CRF 模型主要用于解决的问题为( )。

A.句子生成 B.序列标注 C.文本分类 D.句法分析

65．在问答系统中，"检索式问答" 与 "生成式问答" 的主要区别是( )。

A.检索式问答直接从数据库中生成答案，而生成式问答基于上下文生成答案

**B.检索式问答从外部资源或数据库中检索并返回现有答案，而生成式问答生成全新的答案**

C.检索式问答依赖深度学习模型，生成式问答不需要任何模型

D.检索式问答只能处理结构化数据，生成式问答处理非结构化数据

66．在基于神经网络的分词方法中，常见的模型包括( )。

A.支持向量机（SVM） B.长短时记忆网络（LSTM）

**~~C.词袋模型（BoW）~~**  D.朴素贝叶斯模型

67．在中文分词中，未登录词通常会导致( )。

A.词性标注错误 B.分词效率提高

C.无法识别该词的情感极性 D.分词算法识别错误，并将其误分为已有的词

68. GloVe词嵌入方法通过优化目标获得词向量的为( )。

A.词频统计 B.共现矩阵的对数误差 C.相似性函数 D.最大似然估计

69. Skip-gram模型优化的目标是( )。

**A.最大化上下文词的概率**

B.最小化中心词的概率

C.提高词向量的维度

D.优化模型的运行时间

70．以下哪个模型基于序列到序列架构的是( )。

A. BERT B. GPT  **C. Transformer** D. LSTM

71. NLP 技术常用于以下场景的为( )。

A.病毒检测  **B.智能客服**

C.语音压缩 D.图像识别

72．在文本摘要中，生成式方法的缺点为( )。

A.无法灵活生成新内容

B.存在语法错误的可能性

C.计算效率较低

D.无法处理长文档

73. 在词嵌入模型中，常用来评估词向量质量的方法是( )。

**A.词类相似性测试**

B.模型收敛速度

C.文本摘要效果

D.模型复杂度

74．在问答系统中，BERT模型主要用途是( )。

A.提取候选答案 B.生成答案 C.数据预处理 **D.编码问题和上下文**

75．依存句法分析主要关注句子中词语之间的( )。

A.语义关系 B.句法关系 C.词性关系 D.顺序关系

76．对于词嵌入模型中的CBOW方法，其目标是( )。

**A.根据上下文预测中心词**

B.根据中心词预测上下文

C.优化词与上下文的相似度

D.构建词频分布

77．在文本分类任务中，传统分类方法不包括的是( )。

A.朴素贝叶斯

B. SVM

**C. TF-IDF**

D.决策树

78. 在深度学习框架中，支持动态图机制的为( )。

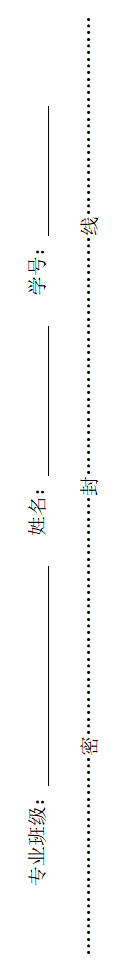
A.TensorFlow **B.PyTorch** C.Caffe D.PaddlePaddle

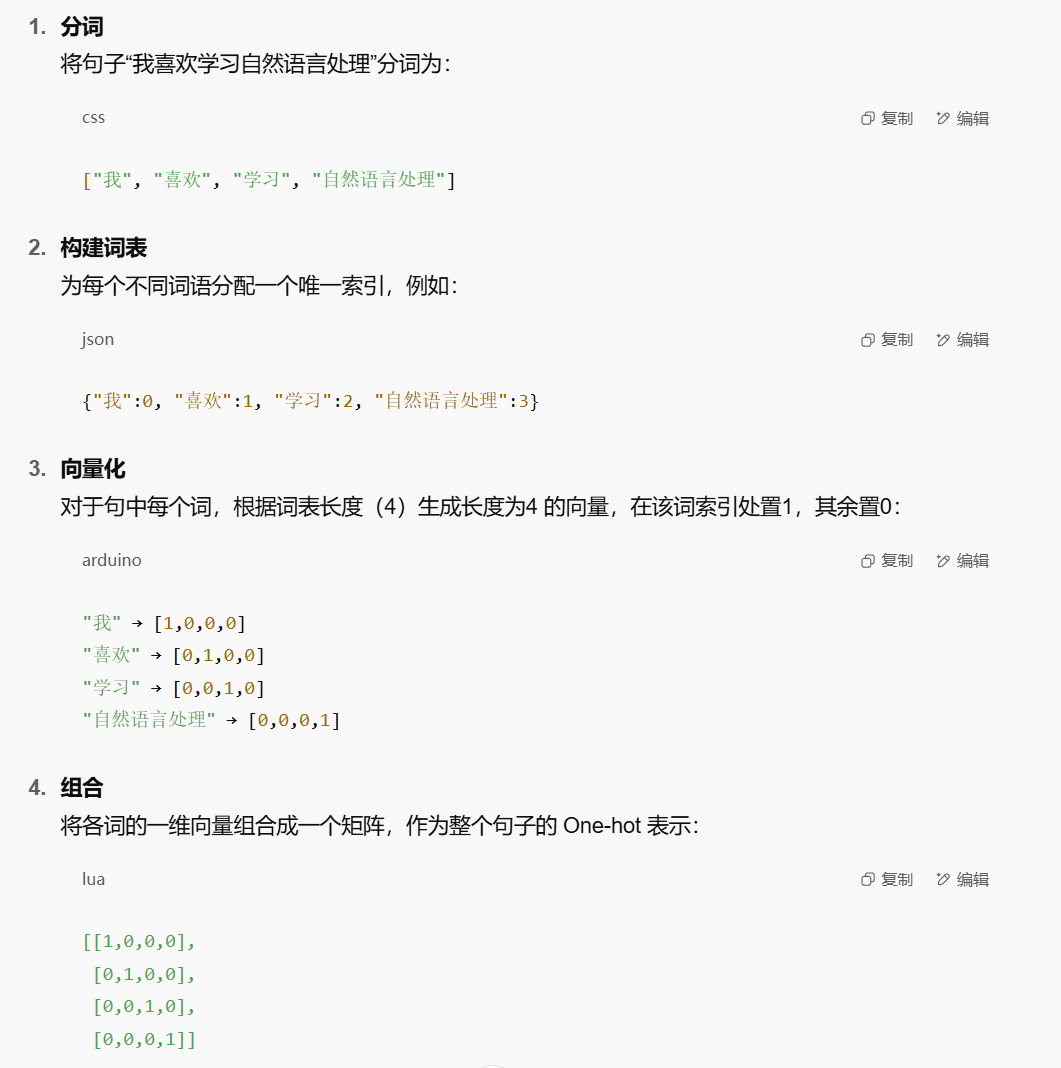
1. 判断题
2. ~~BERT模型是一种生成式模型。( )~~
3. ~~LSTM是一种特殊的CNN模型。（ ）~~  **~~RNN~~**
4. Skip-gram模型通过中心词来预测上下文。( )
5. ~~反向传播和梯度下降是两种不同的优化方法。( )~~
6. ~~CNN模型广泛用于处理非图像类数据。( )~~
7. 自注意力机制能够减轻RNN的梯度消失问题。（ √ ）

**第 i 个位置的表示计算中，会直接用一个注意力权重矩阵去加权所有其他位置的表示，这就像给每对位置都搭了一条“快速通道”，梯度可以沿这条通道直接回溯，而不是沿着 RNN 那条可能长达上百步的时间链。**

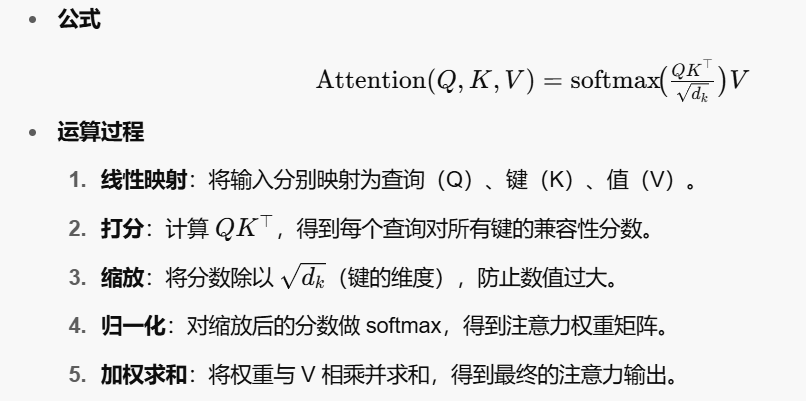
1. 停用词在自然语言处理中通常被忽略，因为它们对文本的意义贡献不大。（ ）
2. TF-IDF算法中，字词的重要性随着它在不同文档中出现次数的增加而上升。（ ）
3. ~~在文本生成任务中，模型生成的文本质量与输入的上下文信息无关，任何输入都能产生高质量的输出。（ ）~~
4. 在Python中，list.append()方法会返回一个新的列表，其中包含了原列表中的所有元素以及新添加的元素。( )
5. ~~词袋模型（Bag of Words）在处理文本时，考虑了单词之间的顺序信息。( )~~
6. 词嵌入（Word Embedding）技术如Word2Vec可以捕捉单词之间的语义关系。（ ）
7. One-hot编码中，随着语料库的增加，词向量的维数会线性增加，这可能导致维数灾难。( )
8. ~~自然语言处理中的文本分类任务只涉及二分类问题。( )~~
9. 多头注意力机制允许模型在同一时间关注多个不同的表示子空间。( )
10. ~~Transformer网络架构中，Encoder 和 Decoder 的结构是完全相同的。（ ）~~
11. ~~单层感知器可以处理任何线性和非线性类型的分类问题。（ ）~~
12. ~~CNN 神经网络只能用于图像数据处理。（ ）~~
13. 在 n-gram 模型中，随着 n 的增加，模型的复杂性也会相应增加。（ ）
14. 模型训练过程中要注意两个问题，分别是过拟合问题和欠拟合问题。( )
15. ~~Word2Vec的Skip-gram模型是从上下文词汇生成目标词汇的。( )~~
16. 词嵌入（Word Embedding）技术如Word2Vec可以捕捉单词之间的语义关系。（ ）
17. 线型图只能表示简单的句子结构，无法处理复杂句子。( )
18. ~~梯度下降法只能应用于线性模型，不能用于非线性模型。( )~~
19. RNN通过其隐藏状态在时间步之间传递信息。（ ）
20. LSTM网络在训练时比标准RNN更容易收敛。（ ）
21. ~~文本摘要技术只适用于长文本，短文本不需要摘要处理。（ ）~~
22. 填空题
23. jieba分词支持精确模式、全模式、\_\_\_\_\_\_\_\_模式。
24. LSTM包含输入门、遗忘门和\_\_\_\_\_\_\_\_\_这三个机制。
25. Doc2Vec技术从\_\_\_\_\_\_\_\_技术扩展而来。
26. 文本向量化的方法主要分为**离散（独热）表示和\_\_ 分布式\_\_\_\_\_\_表示**。
27. \_\_\_\_\_\_\_\_是自然语言中能够独立运用的最小单位，是语言信息处理的基本单位。
28. 如今常用的深度学习框架主要包括\_\_\_\_\_\_\_\_、Tensorflow、Keras、PaddlePaddle。
29. BiLSTM的独特之处在于，它采用双向结构，即包含正向和反向的两个独立\_\_\_\_\_\_\_\_子网络。
30. \_\_\_\_\_\_\_\_\_\_\_\_生态系统分为库家族和Hub平台，库家族是一组面向自然语言处理任务的开源Python库，包括Transformer、Tokenizers等，Hub平台是一个模型共享平台
31. 在自然语言处理领域，方法的发展通常经历了这几个阶段：基于规则的方法、基于统计的方法、机器学习方法、\_\_\_\_\_\_\_\_方法、预训练和迁移学习。
32. Transformer模型中在每个编码器中的每个子层的周围都有一个\_\_\_\_\_\_\_\_，并且都跟随着一个层归一化步骤。
33. 词嵌入的常见方法有基于词共现的\_\_\_\_\_\_, 基于上下文预测的Word2Vec，以及基于全局文本特征的FastText, 这些方法用于学习词的向量表示。
34. BERT模型属于Transformer的预训练模型之一，它使用自注意力层来编码句子，\_\_\_\_\_\_层来生成不同的注意力头，使得模型能够捕捉句子的复杂关系。
35. 在命名实体识别任务中，常见的实体类型包括\_\_\_\_\_\_\_\_、地名和组织名等，它们用于标记文本中具有特定含义的词语。
36. 在LSTM模型中，**\_\_细胞\_\_\_\_\_\_状态 cell** 用于保存长期记忆。门控机制决定哪些信息得以长期保存或者

Hidden state 隐藏状态  **主要保存当前时刻的短期记忆和上下文信息，供下一时刻的计算和输出使用。**

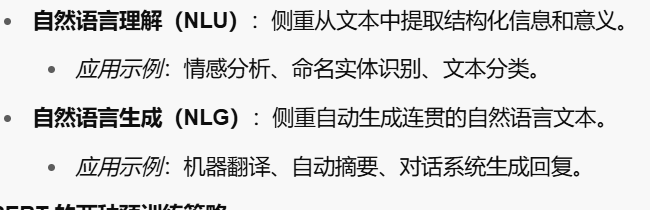
1. 文本分类任务的类型按照分类结果的不同可分为二分类、多分类。
2. 在Python中，如果需要将多个字符串合并成一个字符串，可以使用字符串的\_\_\_\_\_\_\_\_方法将列表中的字符串以指定分隔符连接起来。
3. 在情感分析任务中，常用的标签包括\_\_\_\_\_\_\_、中性和负面。
4. \_\_\_\_\_\_\_\_是自然语言生成的一项重要任务，它的目标是从长文本中生成简洁的摘要文本。
5. 在构建词向量的过程中，\_\_\_\_\_\_\_\_方法使用窗口大小和**上下文词的频率**来**学习**词语的**分布式表示**。
6. 生成模型和判别模型是两种常见的机器学习模型类型，\_\_\_\_\_\_\_\_模型的目标是直接生成数据分布，例如GAN和VAE。
7. 使用 numpy 库进行两个矩阵点乘的方法是\_\_\_\_\_\_\_\_\_\_\_\_。
8. \_\_\_\_\_\_\_\_\_\_\_\_是一种递归神经网络，被广发用于处理序列数据，如文本、语音和时间序列数据。 它具有自循环的特性，能够捕获先前时刻的信息并传递到当前时刻，从而对序列进行建模。
9. 在 n-gram 模型中，若 n=1 ,该模型被称为单元模型，若 n=2 ，则该模型被称为\_\_\_\_\_\_\_\_\_\_\_\_， 若 n=3 , 该模型被称为三元模型。
10. 使用词嵌入向量，可以计算两个单词之间的\_\_\_\_\_\_\_\_\_\_\_\_或者欧几里得距离等度量方式，比较 它们之间的语义相似度。
11. Transformer 使用\_\_\_\_\_\_\_\_\_\_\_\_来编码输入序列的位置信息。
12. LSTM 是 RNN 的一种变形，是为了克服 RNN 无法很好处理\_\_\_**长期依赖\_**\_而提出的。
13. \_\_\_\_\_\_\_\_\_\_\_\_算法的主要思想是词的重要性随着它在文档中出现次数的增加而上升，随着它的语料库中出现频率的升高而下降。
14. 在 Transformer 模型中，每层 Encoder 由\_\_自注意力机制\_\_和前馈神经网络组成。
15. \_\_\_\_\_\_\_\_\_\_\_\_ 编码是一种将每个单词或符号表示为一个高维向量的方法，其中每个维度对应于词汇表中的一个词或符号，且每个向量中只有一个元素为 1，其余元素均为 0。该方法常用于自然语言处理中的文本分类和词向量表示。
16. 在梯度下降算法中，常用的方法有\_\_\_ 随机梯度下降\_\_\_\_\_\_\_，通常这种方法通过计算目标函数的梯度来更新模型参数，从而逐步减少误差并使模型收敛。
17. 转换器网络中典型的编码器架构包括\_\_\_\_\_\_\_\_\_\_\_\_、RoBERTa 、ALBERT 和 DistilBERT等。
18. \_\_\_\_\_\_\_\_\_\_\_\_是通过将输入向量分别映射到多个子空间，并在每个子空间中执行独立的自注意力计算来增强模型的表现能力。
19. \_\_\_\_\_\_\_\_\_\_\_\_生态系统分为库家族和Hub平台，库家族是一组面向自然语言处理任务的开源Python库，包括Transformer、Tokenizers等，Hub平台是一个模型共享平台。
20. 机器学习模型的评价指标有\_\_\_\_\_\_\_\_\_\_\_\_率、准确率、召回率、F1值、ROC曲线、AUC曲线。
21. 自注意力机制中的Attention函数的计算公式是Attention(Qi，Ki,Vi) = \_\_\_\_\_\_\_。
22. 常用的深度学习框架包括 \_\_\_\_\_\_\_\_\_\_\_\_、Pytorch、Keras、PaddlePaddle，它们有不同的特点和应用场景。
23. 在随机梯度下降（SGD）中，每次更新参数时使用 \_\_\_\_\_\_\_\_\_\_\_个样本来计算梯度。
24. RNN在处理序列数据时，通过\_\_\_**\_记忆\_\_\_**\_\_\_\_机制将之前的信息传递到后续的时间步。
25. 在 n-gram 模型中，若n=1,该模型被称为单元模型，若 n=2，则该模型被称为二元模型，若 n=3, 该模型被称为\_\_\_\_\_\_\_\_\_\_\_。
26. LSTM和RNN的区别是LSTM具有更复杂的\_**\_门控\_\_\_\_\_结构**，能够更好地记忆和遗忘信息。
27. 余弦相似度是一种常用的计算\_\_\_\_\_\_\_\_\_\_\_的指标，用于评估两个向量之间的相似性。
28. Word2Vec中的CBOW模型是通过上下文预测\_\_\_\_\_\_\_\_\_\_\_，而Skip-gram模型则是通过目标词预测上下文。
29. 转换器网络（Transformer）中的典型的编码器-解码器架构模型包括T5、\_\_\_\_\_\_\_等。
30. Huggingface生态系统分为库家族和\_\_\_\_\_\_\_\_\_\_\_，前者是一组面向自然语言处理任务的开源Python库，包括Transformer、Tokenizers等，后者是一个模型共享平台。
31. 文本分类任务的类型按照分类结果的不同可分为二分类、\_\_\_\_\_\_\_\_\_\_\_和多标签分类。
32. jieba是一个广泛使用的中文分词库。根据不同的应用需求，jieba提供了三种主要的工作模式：\_\_\_\_\_\_\_\_\_\_、全模式、搜索引擎模式。
33. 机器学习模型的评价指标主要有精确率、准确率、\_\_\_\_\_\_\_\_\_\_\_、F1值、ROC曲线、AUC曲线等
34. 简答题
35. One-hot编码‌，又称为一位有效编码，是一种将分类变量转换为二进制向量的表示方法。接下来运用‌‌One-hot编码按照四个步骤对“我喜欢学习自然语言处理”这一句话进行编码。



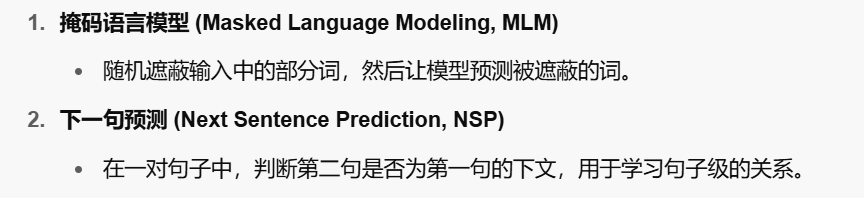
1. 简述自注意力机制的公式及运算过程。



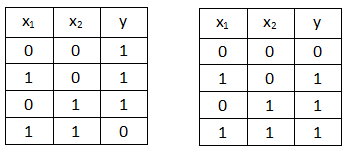
1. 简述自然语言处理的两个方向并至少列举每一方向的一项相关应用。



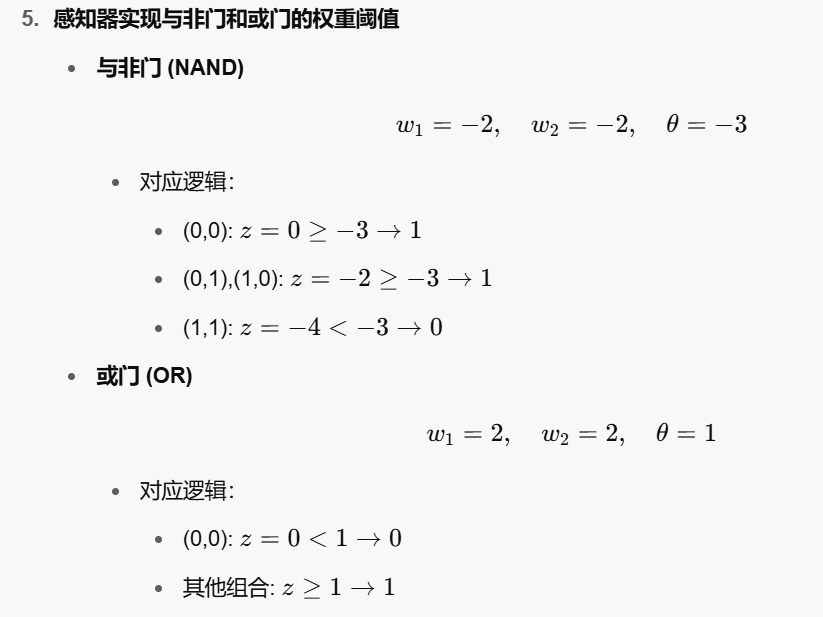
1. 简述BERT模型中采用的两种预训练策略。



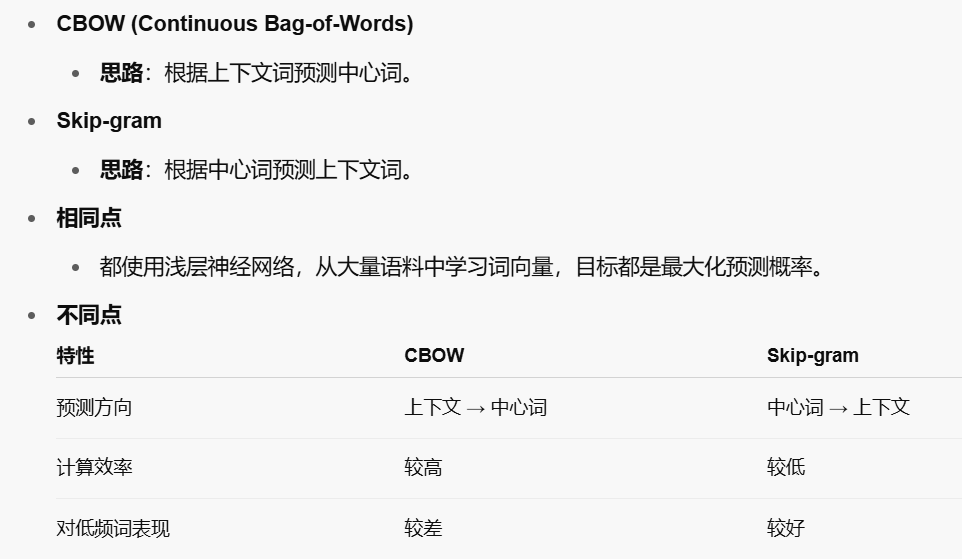
1. 感知器是一种简单形式的前馈式人工神经网络，接收一组输入，对它们进行加权处理，然后生成一个输出。用感知器来模拟逻辑电路中的与非门和或门，其对应的真值表如下图所示，输入信号（x1,x2）和输出信号(y)对应表称为真值表，对应输入与权重的线性组合方程为z=w1x1+w2x2，其阈值为θ。请分别给出一组能够满足与非门和或门的真值表对应的（w1,w2,θ）。



（a）与非门真值表 （b）或门真值表



1. Word2Vec是一种用于将词语映射到向量空间的技术，它通过捕捉词语之间的上下文关系来生成 高质量的词向量。请列举出Word2Vec 的2种主要的训练算法，并说明它们的相同点和不同



1. LSTM（长短期记忆网络）是一种用于处理和预测序列数据的递归神经网络（RNN）架构，旨在解决传统RNN在长序列学习中面临的梯度消失和爆炸问题。请列举LSTM结构中包含的三个门，并说明每个门作用。

**输入门：当前输入多少信息写入细胞状态**

**遗忘门：上一时刻细胞状态多少被保留/遗忘**

**输出门：决定当前细胞状态有多少信息输出到隐藏状态中**

1. jieba是一个广泛使用的中文分词库。根据不同的应用需求，jieba提供了三种主要的工作模式。请列举对应的三种模式，并给出对应模型的调用方法。
2. 请说明TF-IDF 算法的工作原理。
3. 请列举出神经网络中4种常见的激活函数，并分析对应函数的优缺点。

