

一. 判断题（每题 1 分，正确的打勾，错误的打叉，共 20 分）

1.机器学习是深度学习的子集，可以通过人工设计特征值从数据中获取知识（）

2.所有的多层感知机都属于深度学习模型（）

3.为了使神经网络具备非线性处理能力，激活函数就必须是一种非线性函数。（）

4.根据“没有免费的午餐”定律，ReLU 是目前最好的激活函数，因此被默认采用。（）

5.在误差反向传播算法里，网络的权重是根据实际值和网络计算值之间的误差自动进行网络权重的调整。（）

6.在 CNN 模型设计时，为了获取更多的图像特征数量，可以通过增大卷积核形状来实现（）

7.池化的原理是可以用某一区域图像的子块的统计信息来包含该子块的全局信息（）

8.在 CBOW 模型中，是通过周边词来计算中心词的词向量值。（）

9.Transformer 结构中，每个编码器之间是通过词嵌入来进行数据的传递，而且必须保证词向量的形状相同。（）

10.在用 transformer 结构进行模型设计时，必须保证模型同时包含编码器结构和解码器结构。（）

11.AIGC 的工作原理是通过用户输入的提示词，从海量的数据库中搜索出最接近输入提示词语义的内容，并展现给用户。（）

12.GAI 是通过人类已经产生的语料进行训练的人工智能模型，因此不

会造假。（）

13.当一种系统在复杂性增加到某一临界点时，会出现其子系统或较小规模版本中未曾存在的行为或特性，这种能力叫“涌现”。（）。

14.GAI 和 AIGC 是同一个概念，都是指的人工智能具备了自我生成内容的能力。（）

15.只要给大语言模型足够多的训练语料，就能保证模型每次生成的内容都相同。（）

16.在扩散模型的前向过程中，通过带参数的模型 U-Net 一步步加入噪声，经过 N 步，最终生成一副不再扩散的稳定图像。（）

17.多模态人工智能是通往 AGI 的必由之路。（）

18.Transformer 结构中，在进入编码器层前，应对输入文本数据进行编码，这个编码可以是独热码，也可以是 Word2Vec 训练好的词向量。（）

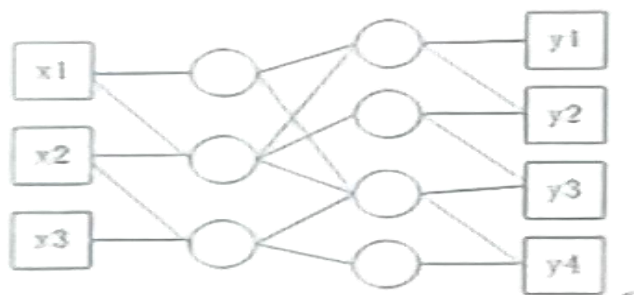
19.在用欧式距离计算文本相似度时，距离越短测文本越相似。（）

20.事实证明，transformer 结构是自然语言处理中最好的模型，所以其他 NLP 模型都可以淘汰不用了。（）

二．单选题（每小题 2 分，共 30 分）

1.对于下图所示的人工神经网络，一共有多少个权重参数？（）

A.12 B.14 C.19 D.144

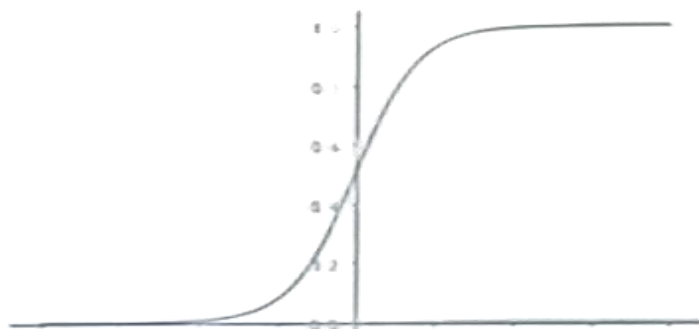


2.神经元细胞的哪种现象被称为“全或无”现象？（）

- A.只有当外来刺激有足够大的强度时，才能产生动作电位，且动作电位的幅度和传播范围恒定
- B.动作电位只能在神经元胞体内产生，不能在突起部分传播。
- C.动作电位的产生与刺激强度无关，但传播范围与刺激强度成正比，
- D.神经元的兴奋只能持续一段时间，然后自动恢复到静息状态，

3.下图所示是哪个激活函数（）[图中上下限为 1，0]

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



4.目前一副彩色图像的每个像素是用多大的内存保存的？（）

- A.1 个 bit B.1 个 byte C.3 个 bit D.3 个 byte

5.已知输入 X 为 12×12 的矩阵，卷积核大小为 5×5 ，步长为 2，边缘填充为 1 个像素则经过卷积运算后的矩阵形状是：（）

- A. 14×14 B. 4×4 C. 5×5 D. 8×8 .

6.将“大马高人骑人高马大”这句话按照单个汉字进行分词，则词汇

表有 () 个 token，句子有 () 个 token： ()

A.5 和 5 B.9 和 9 C.5 和 9 D.9 和 5

7.关于余弦相似度，则以下哪个描述是正确的： ()

A.计算结果的范围在 0~1 之间

B.如果计算结果为 0，则两个向量完全相同，

C.如果计算结果为 1，则两个向量完全相同，

D.如果计算结果为 1，则两个向量完全不同，

8.GPT 通过自回归生成技术预测了“问君能有几多愁”的下一句是“恰似一江春水向东流”，则该模型需要通过几次推理（按单个字分词）？

()

A.1 B.9 C.2 D.10

9.以下哪个模型是多模态的？ ()

A.DALL-E1 B.AARON C.GAN D.DeepDream

10.有一向量矩阵 $A_{20 \times 5000}$ ，要对其进行降维变换成 $B_{20 \times 5}$ ，可以采用矩阵 C 对 A 进行线性变换，那么 C 的形状应该是： ()

A.20X5 B.5X20 C.5X5000 D.5000X5.

11.对一个 6×6 的特征矩阵，进行平均池化计算，池化窗口为 2×2 ，步长=2，则池化后的特征矩阵形状是： ()

A. 5×5 B. 6×6 C. 3×3 D. 4×4

12.在多分类任务中，输出层的激活函数常采用 Softmax，原因是 ()

A.防止过拟合 B.获取全局最优点

C.加快训练速度 D.输出结果可以代表分类概率

13.下面哪种任务比较适合使用 RNN? ()

- A.图像分类
- B.文本生成
- C.图像分割
- D.时间不相关的静态任务,

14.开源的大语言模型 Llama3 具有了非常好的语义理解和输出能力, 为了使其快速的应用到专业的领域, 可以对该模型采用什么样的技术手段? ()

- A.用专业领域的数据进行微调训练
- B.无需进行任何修改, 可以直接部署应用,
- C.用专业领域的数据重新进行训练
- D.不管采取何种技术手段, 都无法用到专业领域,

15.人工智能技术将在今后进入我们生活工作的方方面面, 给我们带来很大的便利, 以下说法正确的是: ()

- A.可以用大语言模型帮我们写研究论文
- B.可以用大语言模型帮我们快速总结文献,
- C.AI 生成的新闻有图有真相, 不会有假
- D.可以用文生视频技术任意创作并发布到网上。

三. 多选题 (每小题 2 分, 共 20 分, 多选少选均不给分)

1.以下哪些模型属于深度学习的模型? ()

- A.卷积神经网络
- B.循环神经网络
- C.扩散模型
- D.Transformer.

2.MLP 一般由以下哪些网络层组成? ()

- A.输入层
- B.池化层
- C.输出层
- D.隐含层

3.下列说法正确的有? ()

- A.感知机模型可以解决二分类问题
- B.感知机模型可以解决线性回归问题
- C.感知机模型可以解决异或问题
- D.感知机模型可以解决多分类问题
- 4.在实际应用中，对数据进行哪种处理可以使其落入 0 到 1 区间？（）
- A.ReLU B.Softmax C.Sigmoid D.添加高斯噪声。
- 5.BP 算法具有如下的局限性（）
- A.易陷入局部最小值
- B.学习新样本时可能遗忘旧样本，
- C.缺乏严谨的推导过程，
- D.隐节点选取缺乏理论指导
- 6.为了预测“天行健君（X）以自强不息”这句话的 X 时，假设使用单向 RNN 来预测，则预测 X 时可能用到的信息来自以下哪些字（）
- A.天 B.自 C.君 D.不
7. Transformer 是一种序列到序列模型，可包含如下哪些结构：（）
- A.编码器 B.解码器 C.注意力机制 D.跨层和归一化
- 8.LLM 中的“大”包含如下几个含义（）
- A.训练数据庞大 B.参数规模巨大 C.训练耗资巨大 D.应用范围大
- 9.AGI 在模型训练时有以下几个要求（）
- A.任务无关 B.领域无关
- C.必须使用 GPU D.必须使用监督学习，
- 10.多模态大语言模型（MLLM）可能对以下哪些模态数据进行处理：

()

A.视频 B.声音 C.文本 D.气味.

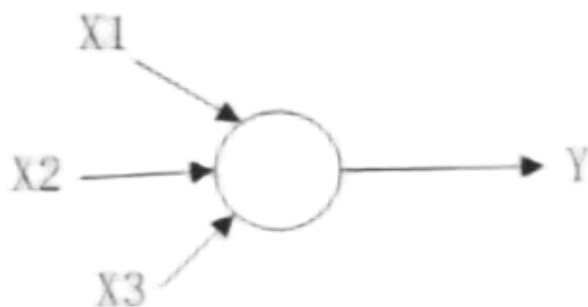
四. 计算题 (每小题 5 分, 共 15 分)

1.对于输入 $x=[0.5, 0.5, 0.3, 1]$, 计算经过 Softmax 激活后的输出结果 Y , 保留小数点 2 位。(5 分) ($e^{0.5}=1.65$, $e^{0.3}=1.35$, $e^1=2.72$)

2. (个人补全版, 学会计算就行) 给定一个 4×4 的输入矩阵 (Input Matrix) 和一个 2×2 的卷积核 (Kernel Matrix), 步长 (Stride) 为 1, 无补齐 (No Padding)。请计算并写出输出特征矩阵 (Output Matrix)

$$Input = \begin{pmatrix} 8 & 1 & 5 & 6 \\ 0 & 7 & 2 & 9 \\ 4 & 3 & 1 & 7 \\ 6 & 2 & 0 & 8 \end{pmatrix} \quad Kernel = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

3. 对于以下的感知机模型, 计算其输出 Y , 其中 $x=[2, 5, 3]$, $W=[1, 3, 5]$, 激活函数 $f(x) = x$, 偏置值 $\alpha=5$ 。(5 分)



五. 问答题（个人补全版）

一：

```
import math
```

```
# 定义两个句子
```

```
sentence_A = "天行健君子以自强不息"
```

```
sentence_B = "地势坤君子以厚德载物"
```

```
def calculate_jaccard_similarity(text1, text2):
```

```
    # a. 将句子转换为字符集合，以进行集合运算
```

```
    set1 = set(text1)
```

```
    set2 = set(text2)
```

```
    # 计算交集和并集
```

```
    intersection = set1.intersection(set2)
```

```
    union = set1.union(set2)
```

```
    # 计算 Jaccard 相似度
```

```
    similarity = len(intersection) / len(union)
```

```
    return similarity, intersection, union
```

```
jaccard_sim,          intersect_chars,          union_chars          =
```



```
calculate_jaccard_similarity(sentence_A, sentence_B)
```

```
print(f"句子的 Jaccard 相似度为: {jaccard_sim:.4f}")
```

问题：a. 第一行代码 `import math` 的作用是什么？

b. `def calculate_jaccard_similarity(text1, text2):` 这部分代码的作用是什么？

c. `jaccard_sim, intersect_chars, union_chars = calculate_jaccard_similarity(sentence_A, sentence_B)`

这行代码的作用是什么？

d. 请计算并写出该程序的最终输出结果。

二：

```
import tensorflow as tf
```

```
from tensorflow.keras.models import Sequential
```

```
from tensorflow.keras.layers import Dense
```

```
import numpy as np
```

```
# 假设已加载 MNIST 数据集，并将其分为训练集和测试集
```

```
X_train = np.random.random((60000, 784))
```

```
y_train = np.random.randint(10, size=(60000,))
```

```
X_train = X_train / 255.0
```

```
y_train = tf.keras.utils.to_categorical(y_train, 10)
```

```
model = Sequential([  
    Dense(64, activation='relu', input_shape=(20,)),  
    Dense(32, activation='relu'),  
    Dense(16, activation='relu'),  
    Dense(8, activation='relu'),  
    Dense(10, activation='softmax')  
])
```

c. 编译模型，为训练做准备

```
model.compile(optimizer='adam',  
              loss='???', # 请选择合适的损失函数  
              metrics=['accuracy'])
```

问题：

a. 在处理像 MNIST 这样的图像数据时，通常会有一行代码 `x = mtData.data / 255.0`。请问这行代码操作的名称是什么？其主要目的是什么？

b. 该模型最后一层（输出层）使用的激活函数是什么？

c. 该模型在“???”指定的损失函数（loss function）选什么？

A.MSE B.二元交叉熵 (binary_crossentropy)

C.MAE D.多分类交叉熵 (categorical_crossentropy)

d. 该 MLP 模型包含多少个隐藏层 (Hidden Layer)? 其中第三个隐藏层有多少个神经元?

三、 观察下图展示的一种深度学习网络训练模式，并回答这代表了什么过程

