

2025 年上海市高等学校信息技术水平考试试卷

二三级 人工智能技术及应用 模拟卷

提示：题包 A 必做，题包 B~G 是 6 选 2 或 3 个，题包 H 和 I 是 2 选 1
(本试卷考试时间 150 分钟；本试卷实际是以自适应考试形式进行，因此每题分值仅供参考)

题包 A

一、单选题（本大题 10 道小题，每小题 1 分，共 10 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. 下列关于人工智能的叙述，错误的是_____。
A. 人工智能技术很新，与其他科学技术很少结合
B. 人工智能+是科学技术发展趋势之一
C. 人工智能技术它与其他科学技术相结合，极大地提高了应用技术的智能化水平
D. 人工智能有力地促进了社会的发展
2. 下列关于大模型（Large Language Model, LLM）的描述中，错误的是_____。
A. 大模型通过海量无标注数据进行预训练，具备强大的上下文理解能力
B. 参数规模超过 10 亿的模型可称为“大模型”，参数数量是其核心特征之一
C. 大模型仅能处理文本数据，无法直接对图像、音频等多模态信息进行联合建模
D. 大模型“涌现能力”指随参数规模增长，模型会展现出预训练阶段未明确设计的新能力
3. 2016 年 3 月，AlphaGo 围棋软件战胜韩国棋手李世石，2017 年 3 月又战胜我国棋手柯洁。AlphaGo 使用的搜索技术是_____。
A. 深度优先搜索
B. A*算法
C. A 算法
D. 蒙特卡洛树搜索
4. 线性回归中，总体误差平方和(TSS)、残差平方和(RSS)与回归误差平方和(ESS)三者的关系是_____。
A. $RSS=TSS+ESS$
B. $TSS=RSS+ESS$
C. $ESS=RSS-TSS$
D. $ESS=TSS+RSS$
5. 过拟合模型表现在训练集上的特点是_____。
A. 低方差和低偏差
B. 高方差和低偏差
C. 低方差和高偏差
D. 高方差和高偏差

6. 小明在家中使用智能音箱时，发现说“打开客厅灯光”时音箱经常误识别为“打开客厅电视”。下列_____措施最可能改善这种情况。

- A. 将音箱音量调大
- B. 对着音箱清晰缓慢地说话
- C. 在音箱旁边放置加湿器
- D. 给音箱更换更鲜艳的外壳

7. 物以类聚人以群分，当企业面对大量的消费者用户时，他不可能对所有的用户做到真正的 1 对 1 的营销，这个时候往往就需要通过聚类分析先对用户进行细分，然后针对细分人群制定针对性的营销策略。如果已经结合企业中的真实数据和业务场景，构建用户细分模型，让你对用户进行消费品推荐，你可以采用_____来实现。

- A. 协同过滤算法
- B. 优化算法
- C. 迁移学习算法
- D. 搜索算法

8. 在一个分类问题中，如果样本的特征有限，且都取离散值。最适合的分类模型为_____。

- A. 决策树
- B. 线性回归模型
- C. 生成对抗网络
- D. K 近邻分类器

9. 不适合使用 Logistic 回归的任务有_____。

- A. 电影票房预测
- B. 人脸检测
- C. 情感分类
- D. 垃圾邮件分类

10. 如果开发一个新闻推荐系统，下面技术中不需要的有_____。

- A. 用户画像
- B. 协同过滤
- C. 文本分类、聚类
- D. 目标检测

二、多选题（本大题 10 道小题，每小题 1 分，共 10 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

1. 下列_____场景直接体现了人工智能在“智能交通与智能医疗”领域的应用。

- A. 医院使用 AI 机器人为患者配送药品
- B. 车辆通过激光雷达与视觉算法实现自动泊车
- C. 音乐 APP 根据用户听歌历史生成个性化歌单
- D. 穿戴设备通过 AI 算法分析心电图识别异常

2. 在人工智能数据处理相关工作中，下列说法正确的有_____。
- A. NumPy 可用于构建多维数组，为多维数据处理提供基础支持
 - B. Pandas 能对表格数据进行高效处理，是表格数据处理的重要工具
 - C. Matplotlib 只能绘制简单的 2D 图表，无法实现 3D 图表绘制
 - D. 利用 Pandas 处理后的数据，可结合 Matplotlib 进行常见图表（如折线图、饼图等）的绘制
3. 梯度下降法的特性包括_____。
- A. 多步迭代
 - B. 步长会影响收敛性
 - C. 更新方向为梯度的反方向
 - D. 更新方向为梯度的同方向
4. 机器学习/深度学习项目所需的步骤主要有_____。
- A. 采集数据
 - B. 数据预处理与特征选择
 - C. 选择模型、训练模型、评估模型
 - D. 预测、上线运行
5. 卷积神经网络 CNN 的层级结构包括_____。
- A. 卷积层
 - B. 池化层（汇聚层）
 - C. 循环层
 - D. 全连接层
6. 典型计算机视觉任务的主要步骤包括_____。
- A. 图像预处理
 - B. 特征设计与提取
 - C. 特征汇聚与变换
 - D. 分类器或回归器设计
7. 如果项目目标是开发一个算法，能自动的对给定的每一篇中文文章进行频道分类，例如分为“体育类”、“财经类”、“旅游类”等，涉及到的技术会有_____。
- A. 中文分词
 - B. 文本特征表示和特征选择技术
 - C. Word2Vec 算法
 - D. 机器学习的分类算法
8. 生成对抗网络主要包含_____两部分。
- A. 生成网络（生成器）
 - B. 判别网络（判别器）
 - C. 预测网络
 - D. 价值网络

9. 当处理非线性可分数据时，可以使用的方法有_____。

- A. 使用带核函数的支持向量机
- B. 使用神经网络
- C. 通过人工构建更多的特征
- D. 使用 K 近邻分类器

10. 如果开发一个智能门禁系统，可以使用的技术有_____。

- A. 人脸识别
- B. 声纹识别
- C. 指纹识别
- D. 虹膜识别

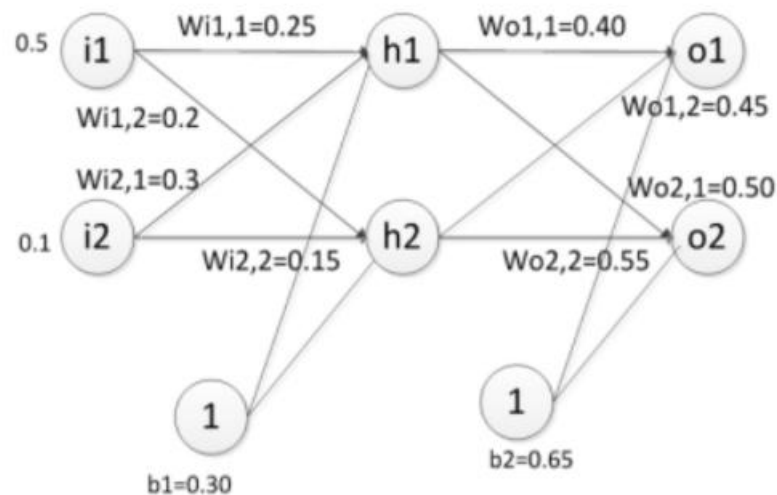
三、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

1. 从以下答案集合中为每空选择一个正确的答案，将其字母编号填入相应空格。

答案集合如下：

A. x	B. b2	C. z2
D. b1	E. a1	F. z1

实现如下神经网络：输入层包含两个神经元i1, i2和bias项b1；第二层隐含层包含h1,h2两个神经元和bias项b2；第三层是输出层包含o1,o2。每条线上标的w是层与层之间连接的权重，激活函数为sigmoid函数，假设两个输入是0.5和0.1。



```
import numpy as np
def sigmoid(x):
    return 1 / (1 + np.exp(-x))
if __name__ == '__main__':
    wi = [[0.25, 0.20], [0.30, 0.15]]          # Weight of input layer
```

```

wo = [[0.40, 0.45] , [0.50, 0.55]]          # Weight of output layer
b1 = 0.30
b2 = 0.65
x = [0.5, 0.1]
# 前向传播
z1 = np.dot (wi,__ (1) __) + b1          # 矩阵相乘
a1 = sigmoid (z1)
z2 = np.dot(wo,__ (2) __) + b2
a2 = sigmoid (__ (3) __)
print( " result:" + str(a2[0]) + " , result:" +str(a2[1]))

```

题包 B

一、单选题（本大题 10 道小题，每小题 1 分，共 10 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. 人工智能的目的是让机器能够_____，以实现某些脑力劳动的机械化。
A. 思考和创造
B. 和人一样工作
C. 完全代替人的大脑
D. 模拟、延伸和扩展人的智能
2. 下列关于机器学习的说法，正确的是_____。
A. 无监督学习只包含聚类，不包含特征降维
B. 有监督学习的“分类”任务输出的是连续数值
C. 数据预处理是机器学习流程中可有可无的步骤
D. 模型训练是机器学习实现从数据到预测能力的关键环节
3. 知识图谱采用_____知识表示方法，以结构化的形式描述客观世界中存在的概念、_____及其之间的复杂关系，是语义 WEB 技术在互联网应用的典型方法。
A. 本体、实体
B. 实体、本体
C. 主观、实体
D. 显示、主体
4. _____可以用来评估最终模型的泛化能力，但不能作为调参、选择特征等算法相关的选择的依据。
A. 训练集
B. 测试集
C. 验证集
D. 参数集
5. 在设计线性分类器的过程中，_____不是常用的损失函数。
A. Fisher 准则函数
B. 感知器准则函数
C. 最小平方误差函数
D. 高斯函数
6. 在机器学习中，损失函数（loss 函数）的作用是_____。
A. 牵引模型参数的更新
B. 增加网络层次
C. 修改模型参数
D. 防止过拟合
7. 深度学习中的“深度”是指_____。

- A. 计算机理解的深度
 - B. 中间神经网络的层次很多
 - C. 计算机对问题的处理更加灵活
 - D. 计算机的求解更加精准
8. 下列属于自然语言处理基础中词法相关内容的是_____。
- A. 句子成分分析
 - B. 分词
 - C. RNN 句子建模
 - D. word2vec 词表示
9. 目前智能家居控制方式不包括_____。
- A. 意念控制
 - B. APP 控制
 - C. 无线遥控
 - D. 语音交互
10. 机器感知是人工智能研究的内容，下列属于机器感知领域的是_____。
- A. 使机器具有视觉、听觉、触觉、味觉、嗅觉等感知能力
 - B. 让机器具有理解文字的能力
 - C. 使机器具有能够获取新知识、学习新技巧的能力
 - D. 使机器具有听懂人类语言的能力

二、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

1. 从以下答案集合中为每空选择一个正确的答案，将其字母编号填入相应空格。
答案集合如下：

A. -	B. ^	C. ⊕
D. pyplot	E.	F. &

输入两个集合setA和setB，分别输出它们的交集、并集、差集。

```
setA = [1,2,3,4]
setB = [3,4,5,6]
print('交集:', setA __ (1) __ setB)
print('并集:', setA __ (2) __ setB)
print('差集:', setA __ (3) __ setB)
```

题包 C

一、单选题（本大题 7 道小题，每小题 1 分，共 7 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. 下列_____不属于人工智能的研究领域。

- A. 模式识别
- B. 机器学习
- C. 深度学习
- D. 编译原理

2. 用于调超参数，监控模型是否发生过拟合（以决定是否停止训练）的数据集是_____。

- A. 训练集
- B. 测试集
- C. 验证集
- D. 参数集

3. _____可用于描述和解决智能体在与环境的交互过程中通过学习策略以达成回报最大化或实现特定目标的问题的一种机器学习方法。

- A. 强化学习
- B. 迁移学习
- C. 监督学习
- D. 深度学习

4. 某机器学习模型对训练集的准确率很高，但对测试集则效果不佳，其原因可能是_____。

- A. 欠拟合
- B. 过拟合
- C. 参数过少
- D. 机器性能问题

5. 在决策树中，_____属于特征重要性评估的方法。

- A. 交叉熵
- B. 布朗运动
- C. 信息熵
- D. 核方法

6. Sigmoid 函数的定义为_____。

- A. $f(x)=1/(1+e^{(-x)})$
- B. $f(x)=e^{(-x)}/(1+e^{(-x)})$
- C. $f(x)=e^{(-x)}$
- D. $f(x)=1+e^{(-x)}$

7. 协同过滤算法可以解决推荐系统中的_____问题。

- A. 矩阵稀疏
- B. 用户相似性和物品相似性
- C. 冷启动

D. 数据高维

二、多选题（本大题 3 道小题，每小题 1 分，共 3 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

1. 属于启发式搜索的算法有_____。

- A. 蚁群算法
- B. 模拟退火算法
- C. 爬山算法
- D. A*算法

2. 文本处理的常见任务有_____。

- A. 机器翻译
- B. 文本分类
- C. 自动摘要
- D. 自动问答

3. 机器人判断物体的位置和形状通常需要_____信息。

- A. 行为
- B. 距离
- C. 明暗（光照）
- D. 颜色

三、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

1. 从以下答案集合中为每空选择一个正确的答案，将其字母编号填入相应空格。答案集合如下：

A. deque()	B. person
C. search_quene	D. graph["钱"]
E. graph["赵"]	F. graph[person]

在某社交网络中存在一位意见领袖，需要使用广度优先搜索算法找出这位意见领袖。社交网络通过散列表的形式保存，如用户“赵”关注了“钱”、“孙”和“李”，而“钱”关注了“周”和“吴”，则保存为：

```
graph["赵"] = ["钱", "孙", "李"]
```

```
graph["钱"] = ["周", "吴"]
```

程序使用一个预定义的函数person_is_leader(p)来判断用户是否为意见领袖。假设从用户“赵”开始搜索，需要确认他的关系网中是否有意见领袖。改正下列代码，使其能够得到正确的结果。

```
from collections import deque
# deque()函数用于创建一个双向队列
search_quene= deque()
search_quene += __(1)
while search_quene:
    person = __(2).popleft()
```

```
    if person_is_leader(person):
        print (person + " is the leader.")
        return True
    else:
        search_quene += __(3) #将person的关注者加入队列search_quene
return False
```

题包 D

一、单选题（本大题 9 道小题，每小题 1 分，共 9 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. 下列属于集成模型的是_____。
A. 逻辑回归
B. 决策树
C. 随机森林
D. 朴素贝叶斯
2. 以下属于无监督学习方法的是_____。
A. 逻辑回归
B. 决策树
C. 层次聚类
D. 朴素贝叶斯
3. 下列_____属于日常生活中常见的人工智能应用。
A. 用手机计算器计算乘法
B. 扫地机器人自动避开障碍物
C. 用 U 盘存储文件
D. 手动调节空调温度
4. 以下_____属于计算机视觉方法基础内容。
A. 图像分割与分类
B. 自然语言处理
C. 数据库设计
D. 网络协议
5. 在机器学习中，用于学习的经验数据集合称为_____。
A. 训练集
B. 测试集
C. 验证集
D. 标签集
6. 在深度学习网络训练时，进行正则化计算的目的是_____。
A. 防止欠拟合
B. 防止过拟合
C. 加快收敛速度
D. 减少 Loss 值
7. LENET5 的基本结构有输入层、C1-卷积层、S2 层-池化层（下采样层）、_____、S4 层-池化层（下采样层）、C5 层-卷积层、F6 层-全连接层和 Output 层-全连接层。

- A. C3 层-卷积层
- B. C3 层-全连接层
- C. C3 层-循环层
- D. C3 层-池化层

8. 在图像识别中，最常用的神经网络是_____。

- A. 长短时记忆网络 LSTM
- B. 循环神经网络 RNN
- C. 卷积神经网络 CNN
- D. 自编码器 Auto-Encoder

9. 研究促进人与计算机之间通信和交互方法的人工智能领域是_____。

- A. 自然语言处理
- B. 符号处理
- C. 决策支持
- D. 机器人学

二、多选题（本大题 1 道小题，每小题 1 分，共 1 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

1. 机器人的触觉传感器主要包括_____。

- A. 接触觉
- B. 压力觉
- C. 接近觉
- D. 温度觉

三、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

1. 从以下答案集合中为每空选择一个正确的答案，将其字母编号填入相应空格。答案集合如下：

A. <code>s = Distances[i]</code>	B. <code>len(Distances)</code>
C. <code><=</code>	D. <code>>=</code>
E. <code>=</code>	F. <code>s += Distances[i]</code>

某辆车加满油后最多可以行驶600公里，同时高速公路上有若干加油站，相邻站点之间的距离均小于600公里，加油站之间的距离信息保存在一个列表Distances里。现在需要通过贪婪算法求出此车辆跑满全程所需要加油的最少次数。补全下述代码，使其能够求得正确的解。

```

num = i = s = 0
while i <= len(Distances):
    __ (1) __
    if s __ (2) __ 600:
        __ (3) __
        num += 1
    i += 1

```

```
print(num)
```

题包 E

一、单选题（本大题 4 道小题，每小题 1 分，共 4 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. _____不属于人工智能的“三驾马车”。
A. 算法
B. 算力
C. 数据
D. 深度学习
2. 下列关于“AI for Science (AI4Science)”的描述中，正确的是_____。
A. AI4Science 仅指用科学方法研究 AI 技术的发展
B. 利用机器学习预测蛋白质三维结构属于 AI4Science 的应用
C. AI4Science 与传统科学研究完全独立，不依赖领域知识
D. 电商平台的商品推荐系统属于 AI4Science 的典型场景
3. 专家系统是以_____为基础，以推理为核心的系统。
A. 知识
B. 计算
C. 符号
D. 概率
4. 决策树的核心是通过_____来逐步划分数据，形成分类或回归的规则。
A. 随机选择特征
B. 计算指标选择最优特征
C. 人工指定特征
D. 按特征出现顺序

二、多选题（本大题 6 道小题，每小题 1 分，共 6 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

1. 强化学习的核心是让智能体通过_____方式学习最优行为。
A. 与环境互动并获得奖励
B. 直接读取环境的全部规则
C. 试错尝试不同动作
D. 提前记住所有正确步骤
2. 机器学习的关键要素包括_____。
A. 数据
B. 模型
C. 损失函数
D. 优化算法

3. 神经网络的学习和优化是指调整神经网络的_____,使输入和输出完成预设需求的任务。

- A. 连接权值
- B. 结构
- C. 层数
- D. 激活函数

4. 长短时记忆网络（LSTM）的门（gate）结构包括_____。

- A. 记忆门
- B. 输入门
- C. 遗忘门
- D. 输出门

5. 语音合成的基本步骤包括_____。

- A. 文本分析
- B. 韵律处理
- C. 声学处理
- D. 文本生成

6. 下列属于经典神经网络模型的有_____。

- A. LeNet
- B. ResNet
- C. LSTM
- D. KNN

三、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

1. 从以下候选答案集中为每空选择一个正确的答案，将其字母编号填入相应空格。候选答案集合如下：

A. LogisticRegression()	B. LinearRegression()
C. 5	D. 2.5
E. m.score(x,y)	F. m.score(array)

现有学习驾驶技术投入的练习时间（月）与通过考试成绩（是否通过）的历史数据，现使用Logistic回归预测训练时间为两个半月的考试成绩。

```
import numpy as np
x=np.array([0.50,0.75,1.00,1.25,1.50,1.75,1.75,2.00,2.25,2.50,2.75,3.00,3.25,3.50,4.00,4.25,4.50,
            4.75,5.00,5.50]) #训练时间（月）
y=np.array([0,0,0,0,0,0,1,0,1,0,1,0,1,0,1,1,1,1,1,1]) #通过考试为1，不通过考试为0

from sklearn.linear_model import LogisticRegression
```

```
m=_____ (1) _____  
x=x.reshape(-1,1)  
y=y.reshape(-1,1)  
m.fit(x,y)  
p=m.predict(____ (2) ____)  
print(p)  
print(____ (3) ____)
```

#逻辑回归模型定义
#如果只有一个特征，重塑为一列数组

#模型学习
#预测

#评估模型准确率

题包 F

一、单选题（本大题 4 道小题，每小题 1 分，共 4 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. 梯度爆炸的解决办法是_____。
A. 批量归一化 Batch Normalization
B. 梯度截断
C. 正则化
D. 修改损失函数
2. 下列属于数据增强技术相关内容的是_____。
A. 扩散模型基础
B. 逻辑回归
C. 决策树
D. 朴素贝叶斯
3. 多模态技术的基础不涉及以下_____。
A. 多模态感知
B. 多模态数据
C. 单一模态的独立分析
D. 多模态数据的融合方法
4. 在深度学习模型的训练过程中，丢弃法（dropout）操作的目的是_____。
A. 防止过拟合
B. 加快收敛
C. 减少震荡
D. 降低学习率

二、多选题（本大题 6 道小题，每小题 1 分，共 6 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

1. 人工智能发展的三个阶段包括_____。
A. 计算智能
B. 感知智能
C. 认知智能
D. 知识智能
2. 神经网络模型的常用激活函数_____。
A. Sigmoid
B. Relu
C. Tanh
D. Sin
3. 基本的知识表示方法有_____。

- A. 产生式表示法
 - B. 框架表示法
 - C. 状态空间表示法
 - D. 谓词公式表示法
4. 使用深度学习方法对图像进行目标检测的基本步骤包括_____。
- A. 采用 sobel 算子提取目标物体的边缘
 - B. 使用无监督算法确定图像中目标物体的可能位置
 - C. 确定候选区域并缩放到相同大小，输入卷积神经网络提取图像特征
 - D. 采用分类器对每个候选区域内的特征进行分类
5. Transformer 模型的核心机制不包括以下_____。
- A. 自注意力（Self-Attention）机制
 - B. 卷积层（Convolutional Layer）
 - C. 编码器 - 解码器（Encoder-Decoder）结构
 - D. 位置编码（Positional Encoding）
6. 常用的神经网络优化算法有_____。
- A. 随机梯度下降法
 - B. 小批量梯度下降法
 - C. 动量法
 - D. 遗传算法

三、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

1. 从以下候选答案集合中为每空选择一个正确的答案，将其字母编号填入相应空格。候选答案集合如下：

A. fitness_sum[i] +=sum(value)	B. fitness_sum[i]/=sum(value)
C. population_new.append(population[j])	D. population_new.append(rand)
E. rand<=fitness_sum[j]	F. rand< fitness_sum[j]

用Python编程语言实现的基本遗传算法的程序如下：

```

selection.py
from __future__ import division
import numpy as np
def selection(population,value):
    #轮盘赌选择
    fitness_sum=[ ]
    for i in range(len(value)):
        if i ==0:
            fitness_sum.append(value[i])
        else:
            fitness_sum.append(fitness_sum[i-1]+value[i])
    for i in range(len(fitness_sum)):

```

```

        _____(1)_____
#选择新种群
population_new=[ ]
for i in range(len(value)):
    rand=np.random.uniform(0,1)
    for j in range(len(value)):
        if j==0:
            if 0<rand and rand<=fitness_sum[j]:
                _____(2)_____
        else:
            if fitness_sum[j-1]<rand and _____(3)_____:
                population_new.append(population[j])
return population_new

```

题包 G

一、单选题（本大题 6 道小题，每小题 1 分，共 6 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择一个正确答案。

1. 以下属于可信人工智能关键特性的是_____。

- A. 可解释性
- B. 仅追求算法高性能
- C. 不考虑安全性
- D. 无需关注公正性

2. 以下属于计算机视觉方法与应用的是_____。

- A. 图像分割与分类
- B. 文本情感分析
- C. 音频信号降噪
- D. 数据库索引优化

3. 下列属于预训练模型核心优势的是_____。

- A. 随机初始化参数后直接训练
- B. 利用大规模无标注数据学习通用特征
- C. 仅适用于图像分类任务
- D. 完全不需要微调即可解决所有下游任务

4. 欠拟合的直观表现是_____。

- A. 在训练集上精度较低
- B. 在训练集上精度较高
- C. 在验证集上精度较低
- D. 在验证集上精度较高

5. 以下属于循环神经网络模型的是_____。

- A. 长短时记忆网络 LSTM
- B. 深度信念网络 DBN
- C. 残差网络 ResNet
- D. 生成对抗网络 GAN

6. 自然语言处理可分为_____两个子领域。

- A. 符号和数字
- B. 时间和空间
- C. 算法和启发式
- D. 理解和生成

二、多选题（本大题 4 道小题，每小题 1 分，共 4 分），从下面题目给出的 A、B、C、D 四个可供选择的答案中选择所有正确答案。

- 下列场景中，最能体现“智能教育”核心价值的是_____。
 - 老师使用 PPT 向全班学生讲解统一的数学知识点
 - 学生通过智能学习平台获取根据自身薄弱点推荐的专项练习题
 - 智能教学系统自动批改学生作业并生成个性化错题集
 - 家长手动记录孩子的课后作业完成情况
- 与传统模式识别方法相比，神经网络方法的特点包括_____。
 - 智能化程度高
 - 自适应学习能力强
 - 并行处理能力强
 - 对计算资源要求低
- 生成对抗网络（GAN）的主要组成部分包括_____。
 - 生成器（负责生成模拟数据）
 - 判别器（负责判断数据真假）
 - 存储器（存储所有训练数据）
 - 转换器（转换数据格式）
- 机器人自动避碰的基本过程包括_____。
 - 确定机器人的静态参数（机器人本体大小、负载）
 - 确定机器人的动态参数（机器人速度、方向等）
 - 确定机器人本体与障碍之间的相对位置参数
 - 根据障碍物参数分析机器人本体的运动态势

三、程序填空题（本大题 1 道小题，每空 2 分，共 6 分）。

- 从以下候选答案集合中为每空选择一个正确的答案，将其字母编号填入相应空格。候选答案集合如下：

A. KMeans(n_clusters=3)	B. KMeans(n_clusters=4)
C. kmeans.labels	D. kmeans.tags
E. drawing	F. plotting

现给定iris.data鸢尾花测量数据集，包括花瓣的长度和宽度、花萼的长度和宽度4类特征。鸢尾花有三个品种：setosa, versicolor, virginica。使用K-Means算法对数据集样本进行品种聚类。

```
import pandas as pd
from sklearn.cluster import KMeans

filename = 'iris.data'
data = pd.read_csv(filename, header = None)
#生成k-means模型
X = data.iloc[:,0:4].values.astype(float)      #提取数据集中的四个特征数据
```

```
kmeans = _____ (1) _____ #模型定义
kmeans.fit(X) #模型学习
print('means.labels_:\n', _____ (2) _____) #输出聚类结果
pd. _____ (3) _____.scatter_matrix(data, c=kmeans.labels_) #生成散点矩阵图观察聚类效果
plt.show()
```

题包 H

二、操作题

提示：操作题包含编程题和论述题三种题型。

（一）编程题

1. 城市数字化转型中，智能技术赋能数字生活，采用推荐系统促进消费升级，通过在线用户的特征预测出用户购买偏好。如下表所示，其中“性别”属性中，0表示女，1表示男；“行业”属性中，0表示信息产业，1表示交通运输业，2表示金融业；“购买偏好”属性中，1表示购买智能音箱，2表示购买扫地机器人。根据用户的特征（“年龄”、“性别”和“行业”）可以预测出用户要购买智能音箱或购买扫地机器人的偏好。以下程序采用K近邻（KNN）算法对下表中的样本构建一个分类模型，并预测用户8（测试样本）的购买偏好。请为以下程序中的每个空，选择一个合适的选项。

	用户序号	年龄（岁）	性别	行业	购买偏好
训练样本	1	21	1	0	1
	2	19	0	0	1
	3	44	1	1	2
	4	32	1	2	1
	5	49	0	1	2
	6	28	0	1	2
	7	25	1	2	1
测试样本	8	38	1	2	?

```
import numpy as np
from sklearn import neighbors
x=np.array([[21,1,0],[19,0,0],[44,1,1],[32,1,2],[49,0,1],[28,0,1],[25,1,2]])#建立训练数据集
_____（1）_____#建立训练数据集标签
knn=neighbors.KNeighborsClassifier(_____（2）_____）
_____（3）_____#模型学习
_____（4）_____#使用 knn 模型对测试样本进行预测
print(p)
```

- | | |
|--------------------------------|--------------------------------|
| A. y=np.array([1,1,2,1,2,2,1]) | B. y=np.array([1,1,2,0,2,2,1]) |
| C. n_neighbors=5 | D. n_neighbor=5 |
| E. knn.fit(x,y) | F. knn.fit(y,x) |
| G. p=knn.predict([[38,1,1]]) | H. p=knn.predict([[38,1,2]]) |

2. 人工智能技术在智慧医疗和康养产业领域得到了广泛应用，是数字城市建设的重要内容。例如，根据用户的体征数据（年龄、BMI、血压、皮质厚度）预测血糖水平。如下表所示，将1-5号用户的体征数据作为训练样本。以下程序采用线性回归模型，预测6号用户的血糖水平结果。请为该程序中的每个空，选择一个合适的选项。

	用户序号	年龄（岁）	BMI	血压	皮质厚度（毫米）	结果（血糖水平）
训练样本	1	50	33.6	72	35	148
	2	23	28.1	66	23	89
	3	30	30.8	74	25.6	116
	4	49	33.5	77	41	167
	5	51	36.6	94	33	143
测试样本	6	43	32.1	74	33	?

```
import ____ (1) ____ as np
x=np.array([[50, 33.6, 72, 35], [23, 28.1, 66, 23], [30, 30.8, 74, 25.6], [49, 33.5, 77, 41], [
51, 36.6, 94, 33]]) #建立训练数据集
y = np.array([148, 89, 116, 167, 143]) #建立训练结果标签
print(x, y)
____ (2) ____ =np.array([[43, 32.1, 74, 33]])#赋值测试样本的数据
print("6 号用户的体征数据：", new_X)
from sklearn.linear_model import LinearRegression
reg_model=LinearRegression()#初始化回归模型
reg_model.fit(____ (3) ____ )
pre_Y=reg_model.predict(____ (4) ____ )#使用线性回归模型预测样本
print(pre_Y)
```

- | | |
|-----------------------|-------------|
| A. numpy | B. np.array |
| C. new_Y | D. new_X |
| E. array[x], array[y] | F. x, y |
| G. x | H. y |

3. 下面的函数完成了决策树算法中的香农熵计算过程，请将代码补充完整。

```
def calcShannonEnt(dataSet):
    numEntries = len(dataSet)
    _____ 填空 1 _____ #the number of unique elements and their occurrence
    for featVec in dataSet:
        currentLabel = featVec[-1]
        if currentLabel not in labelCounts.keys():
            labelCounts[currentLabel] = 0
        else _____ 填空 2 _____ #假设 labelCounts 数组中已经有当前类标签
    shannonEnt = 0.0
    for key in labelCounts:
        prob = float(labelCounts[key])/numEntries
        shannonEnt -= _____ 填空 3 _____ #log base 2, 计算香农熵的公式
    return shannonEnt
```

(二) 论述题

请按如下方式补充下表，以说明 SVM 分类算法和逻辑回归方法的相同点和不同：

1) 用“是”或“否”填写表格第二列

2) 在下述选项中选择合适选项填入表格第二列和第三列。

选项：

- A. 非线性场景
- B. 线性场景
- C. 间隔最大化方程
- D. 似然概率最大化
- E. 最佳分割平面
- F. 分类结果
- G. 概率值

	是否相同	SVM	逻辑回归
目标			
优化目标			
输出结果			
适合场景			

题包 I

二、操作题

提示：操作题包含编程题、方案设计题和论述题三种题型。

(一) 编程题

1. 数字孪生是超大规模城市建设和管理的一种重要信息技术手段，针对高清城市地图、摩天大楼建筑内景图等，提出了图像压缩和实时处理的需求。以下程序段采用 K-means 算法的基本思想，对数字城市建设所需的高清地图进行图像压缩处理，请为以下程序中的每个空，选择一个合适的选项。

```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import sys
#读取文件
pic = mpimg.imread('\素材\pic.png') #需要指明高清虚拟地图的图像路径
print("原图片大小为：",pic.size)
print("原图片占用内存为：",sys.getsizeof(pic))
#压缩图片
image = pic[:, :4, ::4]
X = ____ (1) ____ #线性化处理
#模型构建
model = KMeans(n_clusters = 64)
#模型训练并预测
labels = model.fit_predict(X)
#聚类中心
colors = ____ (2) ____ #获取每个类别的颜色
new_image = colors[labels].reshape(image.shape)
print("压缩后图片大小为：",new_image.size)
print("压缩后图片占用内存为：",sys.getsizeof(new_image))
plt. ____ (3) ____ #展示压缩前的图片
plt.axis('off')
plt.show()
plt. ____ (4) ____ #展示压缩后的图片
plt.axis('off')
plt.show()
```

- | | |
|---------------------------|--------------------------|
| A. image.reshape(64, 64) | B. image.reshape(-1, 4) |
| C. model.cluster_centers_ | D. model.cluster_centers |
| E. imshow() | F. imshow(pic) |
| G. imshow(new_image) | H. imshow(image) |

2. 某城市拟定制一批公园景观桥,采用智能化手段对桥梁结构设计公司的桥梁作品进行合格性检验。其中公开征集到的某批次桥梁设计作品的合格情况如下表中训练样本所示。表中的“桥梁类型”字段,1表示混凝土桥梁,2表示实木桥梁,3表示全钢结构桥梁。“是否合格”字段,1表示作品合格,0表示作品不合格。假设新提交的13号征集作品(测试样本)的参数为(1, 3.3, 3205),以下程序采用支持向量机模型(SVM)预测13号桥梁作品是否合格。请为以下程序中的每个空,选择一个合适的选项。

	序号	桥梁类型	跨度(米)	载重(千克)	是否合格
训练样本	1	1	7.0	3210	0
	2	2	5.5	1000	1
	3	3	3.5	3500	1
	4	1	3.2	3250	1
	5	2	5.6	900	0
	6	3	5.5	3200	0
	7	1	3.5	3180	1
	8	2	7.5	1100	0
	9	3	3.6	3890	0
	10	1	3.4	3750	0
	11	2	5.3	1050	1
	12	3	3.2	3450	1
测试样本	13	1	3.3	3205	?

```

import numpy as np
#建立桥梁作品合格情况的训练数据集
X=np.array([[1, 7.0, 32100], [2, 5.5, 1000], [3, 3.5, 3500], [1, 3.2, 3250], [2,
5.6, 900], [3, 5.5, 3200], [1, 3.5, 3180], [2, 7.5, 1100], [3, 3.6, 3890], [1, 3.4,
3750], [2, 5.3, 1050], [3, 3.2, 3450]]) # 输入训练样本的桥梁类型、跨度、载重
Y= np.array([0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1]) # 输出分类结果: 作品是
否合格
print(X, Y)
#建立测试样本(13号作品的数据参数)
New_X=np.array([[1, 3.3, 3205]])
print("13号作品的检验数据为: ", New_X)
#基于sklearn引入支持向量机分类模型
from sklearn import svm
# 模型初始化
_____ (1)
# 模型学习
_____ (2)
#预测13号作品是否合格
pre_Y= _____ (3)
#根据预测数值输出检验结果

```

```

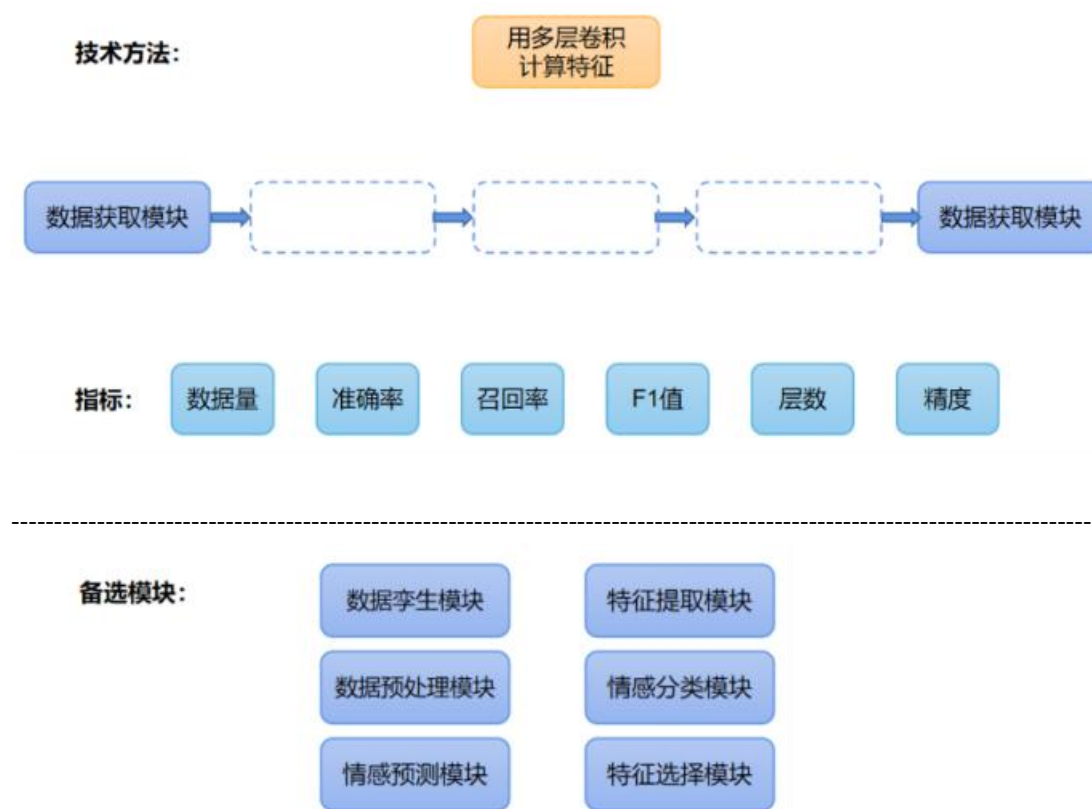
if _____(4)_____:
    print("13号作品的检验结果为：合格")
else:
    print("13号作品的检验结果为：不合格")

```

A. `clf = svm.SVC()` B. `clf = svm.SVC(X, Y)`
 C. `clf.fit(X, Y)` D. `clf.fit()`
 E. `clf.predict(New_X)` F. `clf.predict()`
 G. `pre_Y[0]==0` H. `pre_Y[0]==1`

（二） 方案设计题

上海是我国实施“垃圾分类”的先行城市，经过近几年的实践证明，“垃圾分类”切实增强了城市综合治理能力，提高城市居民的文明与素质。回顾“垃圾分类”实施初期，大量网民在多个网络平台自由评论了“垃圾分类”措施的优劣。基于卷积神经网络 CNN 原理和技术，已设计了一个面向文本评论的智能舆情分析系统的方案，包括 5 个模块及其对应的技术方法和指标。请填上缺少的三个模块，并将各技术方法和指标与相对应的模块连线。



（三） 论述题

假设某高校对人工智能有兴趣的本科生比例为 15%，而对人工智能有兴趣的研究生占 23%。如果五分之一的大学生是研究生，其余的是本科生，那么对人工智能有兴趣的学生是研究生的概率是多少？

请在以下解题过程中(6)选择，其余(1)-(5)、(7)填空：

由题可知：

$P(\text{对 AI 有兴趣} | \text{本科生}) = \underline{(1)} \%$

$P(\text{对 AI 有兴趣} | \text{研究生}) = \underline{(2)} \%$

$P(\text{研究生}) = \underline{(3)}$ (请填小数)

$P(\text{本科生}) = \underline{(4)}$ (请填小数)

求： $P(\text{研究生} | \text{对 AI 有兴趣}) = ?$

先算， $P(\text{对 AI 有兴趣}) = \underline{(5)} \%$

$\frac{(6)}{P(\text{对 AI 有兴趣})}$

根据朴素贝叶斯公式：

$P(\text{研究生} | \text{对 AI 有兴趣}) = \underline{(7)} \%$ (请保留 2 位小数)

(6) 备选答案：

$P(\text{对 AI 有兴趣} | \text{研究生}) * N(\text{研究生})$

$N(\text{对 AI 有兴趣 and 研究生})$

$P(\text{对 AI 有兴趣} | \text{研究生}) * P(\text{研究生})$

$P(\text{对 AI 有兴趣} | \text{本科生}) * P(\text{研究生})$