**2020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）**

（本试卷考试时间 150 分钟）

**客观题部分**

# 题包A

1. 【单选题】人工智能在发展进程中的三个流派为\_\_\_\_\_、连接主义和行为主义。

A.符号主义

B.联合主义

C.对抗主义

D.传输主义

1. 【单选题】遗传算法的遗传操作主要有选择、交配和\_\_\_\_\_\_\_\_。

A.变异

B.继承

C.规划

D.降维

1. 【单选题】广度优先搜索是一种\_\_\_\_\_策略。

A.盲目搜索

B.启发式搜索

C.对抗搜索

D.博弈搜索

1. 【单选题】在手写数字识别问题中，如果忽略训练数据的标签，仅根据特征组将训练数据分类，这是一个\_\_\_\_\_\_\_\_学习过程。

A.无监督

B.有监督

C.弱监督

D.半监督

1. 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

1. CNN卷积神经网络结构中的隐藏层主要包括卷积层和\_\_\_\_\_。

A.池化层

B.连接层

C.输入层

D.传输层

1. 【多选题】产生式系统主要由\_\_\_\_\_\_\_\_组成。

A.规则库

B.综合数据库

C.推理机

D.行为库

1. 【多选题】文本处理的常见任务有\_\_\_\_\_。

A.文本分类

B.信息检索

C.自动问答

D.信息抽取

1. 【多选题】知识图谱的生命周期主要包括\_\_\_\_\_。

A.知识建模

B.知识获取

C.知识管理

D.知识赋能

1. 【程序单选题】请为程序中的每个空选择一个合适的选项

针对sklearn库中的鸢尾花数据集（4个特征：花瓣长、花瓣宽、花萼长、花萼宽，1个标签：0/1/2），随机抽取70%数据作为训练集，30%数据为测试集，采用Knn模型对训练集进行训练，然后对测试集预测分类结果。

#导入sklearn中的鸢尾花数据集

from sklearn.datasets import load\_iris

#导入sklearn中的训练数据与测试数据切分模块

from sklearn.model\_selection import train\_test\_split

#导入Knn模型

from sklearn.neighbors import KNeighborsClassifier

#载入鸢尾花数据集

data = load\_iris()

data\_x =\_\_\_\_(1)\_\_\_\_

data\_y = data.target

#切分70%训练集与30%测试集

X\_train, X\_test, y\_train, y\_test = train\_test\_split(data\_x, data\_y, \_\_\_\_(2)\_\_\_\_)

#模型学习

knn = KNeighborsClassifier()

knn.fit(X\_train,y\_train)

#对测试集进行预测

print(\_\_\_\_(3)\_\_\_\_)

|  |  |
| --- | --- |
| A. data.data | B.load\_data |
| C. test\_size=0.3 | D.test\_size=0.7 |
| E. knn.predict(X\_test) | F.knn.predict(data\_x) |

# 题包B

1. **【**单选题**】**下列说法正确的是？
2. 增加卷积核的尺寸，一定能提高卷积神经网络的性能。
3. 假如现在有个神经网络，激活函数是 ReLU，若使用线性激活函数代替 ReLU，那么该神经网络还能表征 XOR 函数。
4. 有 N 个样本，一般用于训练，一般用于测试。若增大 N 值，则训练误差和测试误差之间的差距会减少。
5. 监督学习需要可以处理任何数据。
6. **【**单选题**】**假设你在卷积神经网络的第一层中有 5 个卷积核，每个卷积核尺寸为 7×7，具有零填充且步幅为 1。该层的输入图片的维度是 224×224×3。那么该层输出的维度是多少？

A. 217 x 217 x 3

B. 217 x 217 x 8

C. 218 x 218 x 5

D. 220 x 220 x 7

1. **【**单选题**】**机器学习训练时，Mini-Batch 的大小优选为2个的幂，如 256 或 512。它背后的原因是什么？

A. Mini-Batch 为偶数的时候，梯度下降算法训练的更快

B. Mini-Batch 设为 2 的 幂，是为了符合 CPU、GPU 的内存要求，利于并行化处理

C. 不使用偶数时，损失函数是不稳定的

D. 以上说法都不对

1. **【**多选题**】**关于搜索与求解，描述正确的是（）

A.搜索是为了达到某一目标而多次进行某种操作、运算、推理或计算的过程

B.所有的智能活动过程，都可以看作或者抽象为一个基于搜索的问题求解

C.搜索是人在求解问题时不知现成解法的情况下所采取的一种普遍方法

D.搜索可以看作人类和其他生物所具有的一种元知识

1. **【**多选题**】**按用途分类，专家系统可分为（）  
   A.诊断型解释型

B预测型决策型

C.设计型规划型

D控制型调度型

1. **【**多选题**】**联想存储的特点是（）

A.可以存储许多相关（激励，响应）模式对

B.以分布、稳健的方式存储信息

C.即使输入激励模式完全失真时，仍然可以产生正确的响应模式

D.可在原存储中加入新的存储模式

1. **【**多选题**】** 下列哪种方法可以用来减小过拟合？

A. 更多的训练数据

B. L1 正则化

C. L2 正则化

D. 减小模型的复杂度

1. **【**多选题**】**下列说法错误的是？

A. 当目标函数是凸函数时，梯度下降算法的解一般就是全局最优解

B. 进行 PCA 降维时，不需要计算协方差矩阵

C. 沿负梯度的方向一定是最优的方向

D. 利用拉格朗日函数能解带约束的优化问题

1. **【**多选题**】**文本语料库的可能用到的特征有

A. 文本中词计数

B. 词的向量标注

C. 词性标注（Part of Speech Tag）

D. 基本依存语法

1. 【程序单选题】选择适当的选项，补全如下一段程序。

|  |  |  |
| --- | --- | --- |
| A. range(10) | C. str(np.random.randint(0,2)) | E. entity |
| B. range(9) | D. str(randint(0,2)) | F. entity+1 |

一个遗传算法，种群规模为90000（216<90000<217)，采用二进制编码。

程序生成初始群体population, 其初始个体数为10（每个个体为一个17位的二进制字符串），生成完成后打印这些个体的编码。

import numpy as np

import math

population=[ ]

for i in ①

entity = '' ''

for j in range(17)

entity=entity + ②

population. append( ③ )

pirnt(entity+"\n")

# 题包C

1. **【**单选题**】**遗传算法中，以下选项中**不是**常用算子的是（ ）。

A. 选择 B. 优化 C. 突变 D. 交叉

1. **【**单选题**】**协同过滤算法**无法**解决推荐系统中的（ ）问题。

A. 矩阵稀疏 B. 冷启动 C. 用户相似性 D. 物品相似性

1. **【**单选题**】**下列算法中属于有监督学习的是（ ）。

A. 谱聚类 B. K-Means聚类 C. 主成分分析 D. 支持向量机

1. 【单选题】KNN (k-nearest neighbors algorithm)是常用的\_\_\_\_\_机器学习算法。

A.分类

B.聚类

C.正则

D.降维

1. **【**单选题**】**搜索分为启发式搜索和（ ）。

A.大数据搜索 B.精确搜索 C.盲目搜索 D.模糊搜索

1. **【**单选题**】**产生式系统中，推理机主要完成推理、（ ）、执行规则、检查推理终止条件任务。

A.冲突消解 B.条件匹配 C.规则选取 D.规则合并

1. **【**单选题**】**对于时序数据的提取，（ ）方法不属于基于变换的特征提取。

A.统计分析 B.小波变换 C.主成分分析 D.快速傅里叶变换

1. **【**单选题**】**语义网络的组成部分为（ ）

A.框架和弧线 B.状态和算符 C.节点和链 D.槽和值

1. **【**多选题**】**网络过拟合可以通过（ ）来缓解。

A. 增加数据 B. 正则化 C.增加网络容量 D.增加网络路径

1. 选择适当的选项，补全如下一段程序。

|  |  |  |
| --- | --- | --- |
| A. arr[i][j] = arr[i - 1][j – 1] | C. remain -= w[j] | E. arr[i] [remain] |
| B. arr[i][j] = arr[i ][j] | D. remain -= w[i] | F. arr[i - 1][remain] |

以下是实现背包问题的动态规划算法。

def knapsack(p, w, v):

n = len(p)

lists,arr = [],[[0] \* (v + 1) for \_ in range(n + 1)]

for i in range(1, n + 1):

for j in range(1, v + 1):

if w[i - 1] <= j: # 如果当前物品的体积不超过背包的容量，p[i-1]当前物品的价值，w[i-1]当前物品的体积

arr[i][j] = max(arr[i - 1][j], p[i - 1] + arr[i - 1][j - w[i - 1]])

else: #如果当前物品的体积超过背包的容量

④

remain = v

for i in range(n, 0, -1):

if arr[i][remain] > ⑤ :

lists.append(i - 1) # (i-1)为当前物品的编号

⑥ # 容积减去已经找到的物品

return arr[-1][-1], lists

if \_\_name\_\_ == '\_\_main\_\_':

p = [700 ,500, 800 ,600 ,520] # 物品的价值

w = [3, 2, 9, 4, 3] # 物品占的体积

v = 8 # 背包的容量

print(knapsack(p, w, v))

# 题包D

1. **【**单选题**】**尽管人工智能学术界出现“百家争鸣”的局面，但是当前国际人工智能的主流派仍属于：（ ）

(A) 连接主义 (B) 符号主义 (C) 行为主义 (D) 经验主义

1. **【**单选题**】**监督学习与无监督学习的区别主要在于监督学习必须要（ ），使用标记数据牵引训练。

(A) 特征选择 (B) 调整参数 (C) 评估模型 (D) 标注

1. **【**单选题**】**在深度优先搜索策略中，open表是（ ）的数据结构

(A) 先进先出 (B) 先进后出

(C) 根据估价函数值重排 (D) 随机出

1. **【**单选题**】**网络训练时需要正则化的目的是（ ）。

(A) 评估模型 (B) 防止网络欠拟合

(C) 防止网络过拟合 (D) 调整参数

1. **【**单选题**】**深度学习中的损失函数（loss函数）的作用有（ ）。

(A) 估量模型的预测值 (B) 提高模型的鲁棒性

(C) 实现模型梯度下降 (D) 牵引网络的更新

1. **【**单选题**】**深度学习是利用（ ）提取特征进行学习。

(A) 机器学习 (B) 深度神经网络

(C) 损失函数 (D) 决策树

1. **【**多选题**】**网络过拟合不能通过（ ）缓解。

(A) 增加数据 (B) 加正则化

(C) 增加网络容量 (D) 减少数据

1. **【**多选题**】**机器学习/深度学习项目中所需的步骤包括（ ）。

(A) 采集数据、预处理与特征选择 (B) 选择模型、训练模型

(C) 评估模型、诊断模型、调整参数 (D) 预测、上线运行

1. **【**多选题**】**机器学习的技术包括（ ）。

(A) 深度学习 (B) 支持向量机(SVM)

(C) 决策树(Decision Tree) (D) K-近邻算法

1. 【程序单选题】请在以下选项中选择正确函数填入相应程序空格内。

|  |  |  |
| --- | --- | --- |
| A. random.choice() | B. getAge() | C. range() |
| D. getSex() | E. random.randint() | F. write() |

编写程序，生成20个人的模拟信息，包括性别、年龄并把生成的信息写入文本文件。程序如下：

def getSex():

return (1) (('男', '女'))

def getAge():

return str(random.randint(18,100))

def main(filename):

with open(filename, 'w', encoding='utf-8') as fp:

# 写入表头

fp.write('Sex,Age\n')

(1)

# 生成20 个人的随机信息

for i in range(20):

sex = (2)

age = (3)

line = ','.join([sex,age])+'\n'

fp.write(line)

**主观题部分**

# 1：程序排序

完成数据前处理任务：根据表1和表2中内容分别创建DataFrame对象df1和df2，将df2中内容与df1内容合并，并都存放在df1中；有数据项缺失（有空值）的行删除；统计df3中工业总产值和企业数量相关系数。

表1:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 企业数量 | 工业总产值 | 从业人数 |
| 北京 | 34293 | 1665.43 | 241.40 |
| 天津 | 34798 | NaN | 240.14 |

表2：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 企业数量 | 工业总产值 | 从业人数 |
| 河北 | 229715 | 3688.16 | 662.13 |
| 山西 | 163163 | 1620.06 | 455.29 |

from pandas import DataFrame

from numpy import nan as NA

import pandas as pd

|  |  |
| --- | --- |
| （1） | data1 = [[34293,1665.43,241.40],[34798,NA, 240.14]] |
| （2） | df1 = pd.concat([df1,df2])  print ('df1=:\n', df1) |
| （3） | df1=DataFrame(data1,index=['北京','天津'],columns=['企业数量','工业总产值','从业人数']) |
| （4） | df2 = DataFrame(data2,index=['河北','山西'],columns=['企业数量','工业总产值','从业人数']) |
| （5） | data2 = [[229715, 3688.16, 662.13],[163163, 1620.06, 455.29]] |
| （6） | print ('工业总产值和企业数量的相关系数为：\n', df3['工业总产值']. corr(df3['企业数量'])) |
| （7） | df3 = df1.dropna()  print ('df3 =:\n',df3) |

# 2：程序开发

如下程序采用Sklearn绘制散点图，并呈现在右上方。请补全程序中缺失的部分。

import xlrd

import matplotlib.pyplot as plt

import numpy as np

from sklearn import model\_selection

from sklearn.linear\_model import LogisticRegression

from sklearn import metrics

data = xlrd.open\_workbook(sandianshuju.xlsx')

sheet = data.sheet\_by\_index(0)

Density = sheet.col\_values(6)

Sugar = sheet.col\_values(7)

Res = sheet.col\_values(8)

# 读取原始数据

X = np.array( [Density, Sugar] )

# y的尺寸为(17)

y = np.array(Res)

X = X.reshape(17,2)

# 绘制分类数据

f1 = plt.figure(1)

plt.title('watermelon\_3a')

plt.xlabel('density')

plt.ylabel('ratio\_sugar')

# 绘制散点图（x轴为密度，y轴为含糖率，呈现在右上方）

plt.scatter(X[y == 0,0], X[y == 0,1], marker = 'o', color = 'k', s=100, label = 'bad')

plt.scatter(X[y == 1,0], X[y == 1,1], marker = 'o', color = 'g', s=100, label = 'good')

plt.legend( ① )

plt.show( )

# 从原始数据中选取一半数据进行训练，另一半数据进行测试

X\_train, X\_test, y\_train, y\_test = model\_selection.train\_test\_split(X, y, test\_size=0.5, random\_state=0)

# 逻辑回归模型

log\_model = LogisticRegression( )

# 训练逻辑回归模型

log\_model.fit( X\_train, X\_test)

# 预测y的值

y\_pred = log\_model.predict( ② )

# 查看测试结果

print(metrics.confusion\_matrix(y\_test, y\_pred))

print(metrics.classification\_report(y\_test, y\_pred))

# 3：方案设计题：

如何为自动驾驶汽车提供训练数据？如何进行训练？

# 4：分析论述题

鸟能飞的实例见下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Instance | No. of Wings | Broken Wings | Living Status | Wing Area/Weight | Fly |
| 1 | 2 | 0 | alive | 2.5 | T |
| 2 | 2 | 1 | alive | 2.5 | F |
| 3 | 2 | 2 | alive | 2.6 | F |
| 4 | 2 | 0 | dead | 3.0 | F |
| 5 | 2 | 0 | dead | 3.2 | F |
| 6 | 0 | 0 | alive | 0 | F |
| 7 | 1 | 0 | alive | 0 | F |
| 8 | 2 | 0 | alive | 3.4 | T |
| 9 | 2 | 0 | alive | 2.0 | F |

问：（1）请画出鸟飞的决策树；

（2）鸟飞的规则是什么？