

人工智能赛项考试-样例卷

(单选题) 1.如果你想冻结模型中的一部分层的参数,使它们不再更新,应该使用哪个操作?

A.param.requires_grad = False

B.torch.no_grad()

C.optimizer.step()

D.param.data.zero_()

(单选题)2.运行下列代码,结果中 A 的值是?

```
import pandas as pd
```

```
df = pd.DataFrame({'A' : [1,2,2,3,4,5,5]})
```

```
sum(df.drop_duplicates())
```

A.22

B.15

C.8

D.7

(单选题)3.在 PyTorch 中, torch.nn.ReLU() 是以下哪一种类型的函数?

A.优化器

B.激活函数

C.损失函数

D.数据加载器

(单选题) 4.如果你想为 GPU 上的张量创建一个与其相同形状但值全为零的张量, 应该使用哪个函数?

A.torch.zeros_like()

B.torch.ones_like()

C.torch.full()

D.torch.randn()

(单选题) 5.下面的代码是使用线性回归对 California Housing dataset 进行分类的代码, 请根据代码选择正确的选项:

...

1. 导入必要的库

```
import numpy as np # 用于数值计算
```

```
import pandas as pd # 用于数据处理
```

```
from sklearn.datasets import fetch_california_housing # 加载加州房价数据集
```

```
from sklearn.model_selection import train_test_split # 用于数据集划分
```

```
from sklearn.preprocessing import StandardScaler # 数据标准化
```

```
from sklearn.linear_model import LinearRegression # 线性回归模型
```

```
from sklearn.metrics import mean_squared_error, r2_score # 评估指标
```

2. 加载加州房价数据集

```
california = fetch_california_housing() # 加载加州房价数据集
```

```
X = california.data # 特征数据
```

```
y = california.target # 目标变量 (房价中位数)
```

将数据集转换为 DataFrame 方便查看

```
df = pd.DataFrame(X, columns=california.feature_names)
```

```
df['MedHouseValue'] = y
```

```
print(____ (题目 1) _____) # 打印前几行数据查看
```

3. 数据预处理

将数据集划分为训练集和测试集, 80%用于训练, 20%用于测试

```
X_train, X_test, y_train, y_test = ____ (题目 2) _____
```

数据标准化 (均值为 0, 方差为 1)

```
scaler = StandardScaler()
```

```
X_train = scaler.fit_transform(X_train) # 训练集标准化
```

```
X_test = scaler.transform(X_test) # 测试集标准化 (使用训练集的标准)
```

4. 选择模型

```
model = LinearRegression() # 创建线性回归模型
```

5. 模型训练

```
model.fit(X_train, y_train) # 使用训练数据训练模型
```

6. 模型评估

```
y_pred = model.__(题目 3)_____ # 使用模型对测试集进行预测
```

计算均方误差 (MSE) 和决定系数 (R^2)

```
mse = mean_squared_error(y_test, y_pred) # 计算均方误差
```

```
r2 = __(题目 4)_____ # 计算决定系数
```

```
print(f"均方误差 (MSE): {mse:.2f}") # 输出均方误差
```

```
print(f"决定系数 ( $R^2$ ): {r2:.2f}") # 输出决定系数
```

7. 预测 (可选)

假设我们有一个新的样本进行预测

新样本包含八个特征, 按照 'AveRooms', 'AveOccup', 'AveIncome' 等顺序

填入

```
new_data = np.array([[8.3252, 41.0, 6.984126, 1.023809, 322.0, 2.555556,  
37.88, -122.23]])
```

```
new_data = scaler.__(题目 5)_____ # 标准化新数据
```

```
predicted_price = model.predict(new_data) # 预测房价中位数
```

```
print(f"预测的房价中位数: {predicted_price[0]:.2f}") # 输出预测的房价  
'''
```

题目 (4) 处应为:

A.r2_score(y_train, y_pred)

B.r2_score(y_test, y_pred)

C.r2_score(y_test, y_test)

D.r2_score(y_test, y_train)

(单选题)6.代码 np.nonzero([1, 0, 2, 4, 3, 0]) 的返回值为:

A.[1, 2, 4, 3]

B.[0, 2, 3, 4]

C.[True, False, True, True, True, False]

D.[True, 0, True, True, True, True, 0]

(单选题)7.代码 df.sort_values(by=['age', 'visits'], ascending=[False, True])

的运行结果是:

A.将 DataFrame 中的数据先按照 visits 降序排列, 然后按照 age 升序排列

B.将 DataFrame 中的数据先按照 visits 升序排列, 然后按照 age 降序排列

C.将 DataFrame 中的数据先按照 age 降序排列, 然后按照 visits 升序排列

D.将 DataFrame 中的数据先按照 age 升序排列, 然后按照 visits 降序排列

(单选题) 8.在一个 DataFrame 中, 包含整数格式的 age 列, 代码 `df[df.age>3]`

运行的结果是:

- A.取出 age 的值大于 3 的行
- B.取出 age 列中所有大于 3 的行的标号
- C.返回一个布尔值的列表, 其中 age 列大于 3 的行对应的值为 True
- D.判断一行 age 的值是否大于 3

(单选题)9.关于 Pandas 中的 head 函数, 下列说法正确的是:

- A.可以默认打印 DataFrame 的表头数据
- B.可以默认打印 DataFrame 中前 5 条数据
- C.可以默认打印 DataFrame 中前 10 条数据
- D.可以默认打印 DataFrame 的描述信息

(单选题)10.下列哪一项可以实现 0-9 元素的逆序输出?

- A.`np.arange(10)[::-1]`
- B.`np.arange(10)[-1]`
- C.`np.arange(10)[1::-1]`
- D.`np.arange(10)[1:1:-1]`

(单选题) 11.在 Numpy 中, 哪个函数可以指定创建数组的形状, 并且可以用随机数填充这个指定形状的数组?

- A.`np.random.random`

B.np.random.randn

C.np.random.rand

D.np.random.randint

(单选题) 12.如何获取数组的形状、大小和维度?

A.arr.shape, arr.size, arr.dim

B.arr.shape, arr.size, arr.ndim

C.arr.shape, arr.length, arr.ndim

D.arr.size, arr.shape, arr.dim

(单选题)13.下列表达式中, 哪个的值一定为 True

A.0.3 == 3 * 0.1

B.0.3 equals 3 * 0.1

C.math.equals(0.1, 3 * 0.1)

D.math.isclose(0.3, 3 * 0.1)

(单选题)14.下列哪个属于可变数据类型:

A.str

B.tuple

C.list

D.int

(单选题) 15.给定两个变量 a、b, 要求输出逻辑是: 如果 a 大于 b, 则返回 a-b, 否则返回 a+b。请问下列哪项可以满足上述逻辑:

A.(a > b) if (a - b) else (a + b)

B.(a > b) if (a + b) else (a - b)

C.(a + b) if (a > b) else (a - b)

D.(a - b) if (a > b) else (a + b)

(单选题)16.以下哪一选项是训练静态词向量的方法?

A.Fasttext

B.ELMO

C.BERT

D.GPT

(单选题)17.1997 年, IBM 开发的一台超级电脑打败了前世界象棋冠军加里·卡斯帕罗夫, 以下哪项是这台计算机的名字?

A.Deep Mind

B.Deep Blue

C.Deep Green

D.Deep Thinking

(单选题)18.在数据较少的情况下,可使用迁移学习完成对图片进行图像识别的任务, 下列哪种方法是合适的?

- A.使用现有的数据重新训练整个模型
- B.冻结模型最后几层，只对前面的部分进行调参
- C.冻结模型前面的部分，只对模型最后几层进行调参
- D.对模型的每一层进行评估，只选择其中的少数几层

(单选题)19.下列哪一种架构包含有反馈连接?

- A.全连接神经网络
- B.卷积神经网络
- C.循环神经网络
- D.受限玻尔兹曼机

(单选题)20.卷积神经网络中 Dropout 层的作用是:

- A.加快收敛速度
- B.防止过拟合
- C.丰富训练样本
- D.增加正样本强度

(单选题) 21.分批归一化的好处都有啥?

- A.是一种非常有效的反向传播算法
- B.是一种非常有效的优化算法
- C.可以将权重分布调整为正态分布
- D.可以大致固定每一层输入的范围

(单选题)22.在神经网络中,在特定神经元得到的输出为-0.321,则此神经网络可能使用了哪种激活函数?

- A.ReLU
- B.Sigmoid
- C.tanh
- D.softmax

(单选题) 23.与传统机器学习方法相比,深度学习的优势在于:

- A.深度学习可以自动学习特征
- B.深度学习可以完全不需要数据预处理
- C.深度学习不需要调参
- D.以上三条都正确

(单选题)24.对于一个图像识别问题,下面那种神经网络可以更好地解决问题?

- A.感知器
- B.多层感知器
- C.循环神经网络
- D.卷积神经网络

(单选题)25.当在卷积神经网络中加入池化层时(pooling layer),变换的不变性会被保留吗?

- A.无法确定
- B.需视实际情况而定
- C.会保留
- D.不会保留

(单选题) 26.训练神经网络时, 我们可以通过哪种方法来优化网络, 最终达到最小化损失函数的目的?

- A.正向传播算法
- B.反向传播算法
- C.双向传播算法
- D.以上均不正确

(单选题)27.下列哪一项不属于数据清洗的范畴?

- A.将数据分批次
- B.填补缺失值
- C.删除离群点
- D.解决数据的不一致性

(单选题) 28.对机器学习性能的评估指的是:

- A.对算法学习能力的评估
- B.对算法学习方法的评估
- C.对算法学习效率的评估

D.对查准率和查全率的评估

(单选题) 29.在数据中只有一个特征的情况下，构建一个最简单的线性回归模型需要几个系数？

A.1 个

B.2 个

C.3 个

D.4 个

(单选题) 30.下列方法中，哪个无助于防止过拟合？

A.L1/L2 正则化

B.Dropout

C.Batch Normalization

D.SGD

(单选题)31.超市通过对顾客购物小票的分析，得到购买啤酒的人有很大的概率也会购买尿不湿，他们使用了哪种类型的算法？

A.分类算法

B.聚类算法

C.关联规则算法

D.自然语言处理技术

(单选题)32.假设 X 核 Y 都服从正态分布, 那么 $P(X < 5, Y < 0)$ 是一个:

- A.先验概率
- B.后验概率
- C.联合概率
- D.几何概率

(单选题) 33. “均方根误差” 的简写是:

- A.MAE
- B.RMAE
- C.MSE
- D.RMSE

(单选题) 34.下列哪种方法可能改善模型的欠拟合现象?

- A.从数据集中抽样进行训练
- B.对模型进行裁剪
- C.增加训练过程的迭代次数
- D.使用 L1 或者 L2 正则化

(单选题)35.对于多层感知器 (MLP), 输入层节点数为 10, 隐藏层节点数为 5,

从输入层道隐藏层的最大链接数是多少?

- A.小于 50
- B.等于 50

C.大于 50

D.无法确定

(单选题) 36.下列哪个函数不应作为神经网络的激活函数?

A. $y = \sin(x)$

B. $y = \tanh(x)$

C. $y = \max(0, x)$

D. $y = 2x + 3$

(单选题) 37.关于机器学习与深度学习的关系, 下列描述正确的是:

A.机器学习是深度学习的一个子集。

B.深度学习是机器学习的一个新的研究方向。

C.机器学习与深度学习是两种完全不同的技术。

D.深度学习的训练速度远高于机器学习的训练速度。

(单选题)38.下面的代码是使用决策树对鸢尾花数据进行分类的代码, 请根据代码

选择正确的选项:

'''

第一步: 导入必要的库

from sklearn.datasets import load_iris # 导入鸢尾花数据集

from sklearn.model_selection import train_test_split # 用于划分训练集

和测试集

```
from sklearn.tree import DecisionTreeClassifier # 导入决策树分类器

from sklearn.metrics import accuracy_score, confusion_matrix # 用于评
估模型性能

# 第二步：加载鸢尾花数据集

# 这里我们使用 scikit-learn 自带的鸢尾花数据集

iris = ____ (题目 1) _____

# 输出数据集的描述信息，帮助我们了解数据

print(iris.DESCR)

# 特征变量（花萼长度、花萼宽度、花瓣长度、花瓣宽度）

X = iris.data

# 目标变量（鸢尾花的类别: 0 -> Setosa, 1 -> Versicolor, 2 -> Virginica)

y = iris.target

# 打印特征和目标的基本信息

print("特征数据集的形状:", X.shape) # (150, 4) 表示 150 个样本，每个样本
有 4 个特征

print("目标数据集的形状:", y.shape) # (150,) 表示 150 个样本的类别标签

# 第三步：划分训练集和测试集
```

```
# 我们将数据集按 80%训练、20%测试的比例划分
```

```
X_train, X_test, y_train, y_test = ____ (题目 2) _____
```

```
# 打印训练集和测试集的大小
```

```
print("训练集的大小:", X_train.shape)
```

```
print("测试集的大小:", X_test.shape)
```

```
# 第四步：训练决策树分类器
```

```
# 使用 scikit-learn 中的 DecisionTreeClassifier 来构建模型
```

```
clf = DecisionTreeClassifier(____ (题目 3) _____)
```

```
# 使用训练数据训练模型
```

```
clf.fit(X_train, y_train)
```

```
# 第五步：模型预测
```

```
# 使用测试集进行预测
```

```
y_pred = ____ (题目 4) _____
```

```
# 打印测试集的真实标签和预测结果
```

```
print("测试集的真实标签:", y_test)
```

```
print("预测结果:", y_pred)
```



```
# 第六步：模型评估

# 计算模型的准确率

accuracy = ____ (题目 5) _____

print("模型的准确率: {:.2f}%".format(accuracy * 100))

# 计算并输出混淆矩阵

conf_matrix = confusion_matrix(y_test, y_pred)

print("混淆矩阵:\n", conf_matrix)

'''
```

题目 (1) 处应为:

- A.load_iris_dataset()
- B.get_iris_data()
- C.load_data_iris()
- D.load_iris()

(单选题)39.关于下列代码运行结果描述中, 正确的是:

```
import numpy as np

data_01 = np.arange(6).reshape(3,2,1,1).squeeze()

data_02 = np.arange(6).reshape(3,1,2,1).squeeze()
```

- A.data_01 和 data_02 的形状和数据完全一样。
- B.data_01 和 data_02 的形状一样但数据不一样。
- C.data_01 和 data_02 的形状不一样和数据一样。
- D.data_01 和 data_02 的形状和数据均不一样。

(单选题)40.以下代码的运行结果是：

```
for i in range(1,3):  
    print(i)  
for i in range(2):  
    pring(i)
```

- A.1 3 2
- B.1 2 0 1
- C.1 3 0 1
- D.1 3 0 2

(单选题) 41.已有数据: data = [1,2,3,4,5,4,3,2,1], 下列哪行代码可以去除 data 中的重复数据?

- A.result = [d for d in list(data)]
- B.result = [d for d in set(data)]
- C.result = [d for d in dict(data)]

D.以上都不对

(单选题)42.下列程序的运行结果是:

```
data = [ i for i in range(1,7)]  
  
data[-1:1:-1]
```

A.python [-1,1,-1]

B.python [6,5,4,3]

C.python [1,2,3,4]

D.程序报错

(单选题)43.下面的代码是使用决策树对鸢尾花数据进行分类的代码,请根据代码

选择正确的选项:

...

第一步: 导入必要的库

```
from sklearn.datasets import load_iris # 导入鸢尾花数据集
```

```
from sklearn.model_selection import train_test_split # 用于划分训练集  
和测试集
```

```
from sklearn.tree import DecisionTreeClassifier # 导入决策树分类器
```

```
from sklearn.metrics import accuracy_score, confusion_matrix # 用于评  
估模型性能
```

```
# 第二步：加载鸢尾花数据集

# 这里我们使用 scikit-learn 自带的鸢尾花数据集

iris = ____ (题目 1) _____

# 输出数据集的描述信息，帮助我们了解数据

print(iris.DESCR)

# 特征变量（花萼长度、花萼宽度、花瓣长度、花瓣宽度）

X = iris.data

# 目标变量（鸢尾花的类别: 0 -> Setosa, 1 -> Versicolor, 2 -> Virginica)

y = iris.target

# 打印特征和目标的基本信息

print("特征数据集的形状:", X.shape) # (150, 4) 表示 150 个样本，每个样本
有 4 个特征

print("目标数据集的形状:", y.shape) # (150,) 表示 150 个样本的类别标签

# 第三步：划分训练集和测试集

# 我们将数据集按 80%训练、20%测试的比例划分

X_train, X_test, y_train, y_test = ____ (题目 2) _____

# 打印训练集和测试集的大小
```

```
print("训练集的大小:", X_train.shape)

print("测试集的大小:", X_test.shape)

# 第四步：训练决策树分类器

# 使用 scikit-learn 中的 DecisionTreeClassifier 来构建模型
clf = DecisionTreeClassifier(____ (题目 3) _____)

# 使用训练数据训练模型
clf.fit(X_train, y_train)

# 第五步：模型预测

# 使用测试集进行预测
y_pred = ____ (题目 4) _____

# 打印测试集的真实标签和预测结果
print("测试集的真实标签:", y_test)

print("预测结果:", y_pred)

# 第六步：模型评估

# 计算模型的准确率
accuracy = ____ (题目 5) _____

print("模型的准确率: {:.2f}%".format(accuracy * 100))
```

```
# 计算并输出混淆矩阵

conf_matrix = confusion_matrix(y_test, y_pred)

print("混淆矩阵:\n", conf_matrix)

...


```

题目 (3) 处应为:

- A.pre_train=True
- B.pretrain=True
- C.random_state=42
- D.get_state=42

(单选题)44.下面哪个方法用于将模型从训练模式切换到评估模式?

- A.optimizer.step()
- B.model.train()
- C.model.eval()
- D.torch.save()

(单选题)45.下列程序的运行结果是:

```
data = {'name': '张三'}

new_data = {'gender': '男'}
```

```
data.update(new_data)
```

```
print(data)
```

- A.{'gender': '男'}
- B.{'name': '张三'}
- C.{'name': '张三','gender': '男'}
- D.程序报错

(单选题) 46.下列哪种算法不属于分类算法:

- A.决策树算法
- B.逻辑回归算法
- C.支持向量机算法
- D.K 均值算法

(单选题)47.主成分分析 (PCA) 属于那种学习方式?

- A.有监督学习
- B.无监督学习
- C.半监督习
- D.强化学习

(单选题) 48.下面哪一个方法是用于保存 PyTorch 模型的参数?

- A.torch.load()

B.torch.save()

C.torch.manual_seed()

D.torch.optim.Adam()

(单选题) 49.下列哪条语句可以生成由 0-8 组成的 3x3 矩阵?

A.np.arange(9).reshape(3, 1)

B.np.arange(9).reshape(3, -1)

C.np.arange(8).reshape(3, 1)

D.np.arange(8).reshape(3, -1)

(单选题)50.关于生成式对抗网络中的两个模型之间的关系,可以用下列哪一对来对比?

A.组长与组员

B.警察与小偷

C.教师与学生

D.医生与护士

(多选题)51.关于树结构模型数据归一化问题, 下列说法不正确的事:

A.当样本不同特征之间取值范围相差过大的时候, 应当对数据进行归一化处理

B.需要根据实际情况判断是否对数据归一化

C.树结构模型是阶跃的, 阶跃点是不可导的, 因此无需对数据归一化

D.如果算法发生不收敛的情况, 即使是树结构模型也需要对数据进行归一化

(多选题) 52.关于集成学习，下列说法正确的是：

- A.集成学习中的并行学习方法是指算法的个体学习器之间存在强依赖关系
- B.集成学习中的串行学习方法是指算法的个体学习器之间不存在强依赖关系
- C.Boosting 属于串行集成学习方法
- D.随机森林属于并行集成学习方法

(多选题) 53.关于 K-Means 算法，下列说法正确的是

- A.K—Means 算法对大数据集有较高的效率并且有可伸缩行
- B.K-Means 算法是一种无监督学习算法
- C.K-Means 算法中的 K 值需要人工设定，且初始聚类中心随机选择
- D.K—Means 算法中初始聚类中心的选择对聚类结果影响不大

(多选题)54.关于 BP 算法，下列说法正确的是：

- A.BP 算法不能用于处理非线性问题
- B.通常情况下， BP 算法所需的训练时间较长
- C.BP 算法容易陷入局部最优解
- D.使用 BP 算法进行优化时有可能由于权重调整过大使得激活函数失活

(多选题) 55.下列场景中，对哪些场景的分析使用了正确的方法？

- A.根据商家最近几年的经营及服务数据，使用聚类算法判断出淘宝商家在各自主营类目下所属上架的层级

B.根据商家最近几年的成交数据,使用聚类算法预测出用户未来一个月内可能的消费金额

C.销售笔记本电脑的商家,使用关联规则方法预测可能购买打印机的顾客

D.根据用户最近的购买信息,使用决策树算法识别出买家的性别

(多选题) 56.下列哪些函数可以作为线性不可分支持向量机的核函数?

A.线性核函数

B.多项式核函数

C.径向基核函数

D.sigmoid 核函数

(多选题)57.下列属于降维算法的有:

A.主成分分析

B.奇异值分解

C.数据离散化

D.高斯核函数

(多选题)58.主成分分析获取超平面应具有的性质有:

A.最大可分性

B.局部极小性

C.信息增益最大

D.最近重构性

(多选题)59.下列哪些 python 语句是合法的?

A.x = y = z = 1

B.x, y = y, x

C.x = (y = z = 1)

D.x += y

(多选题) 60.在神经网络中, 下面哪种方法可以用来处理过拟合现象?

A.分批归一化

B.正则化

C.Dropout

D.最小二乘法

(多选题) 61.下列哪些情况可能会导致神经网络训练失败?

A.梯度消失

B.梯度下降

C.鞍点

D.激活函数死区

(多选题)62.可用于数据特征降维的方法有哪些?

A.SVD

B.PCA

C.AutoEncoder

D.LDA

(多选题)63.在决策树的构建过程中，下列哪些函数可以用于节点划分的衡量指标？

A.信息增益

B.信息增益率

C.基尼系数

D.核函数

(多选题)64.关于数据降维，以下说法正确的是：

A.降维是将训练样本的数据从高维空间转换到低维空间

B.降维不会对数据产生损伤

C.通过将为可以更加有效地发觉有意义的数据结构

D.降维有助于实现数据可视化

(多选题)65.卷积神经网络中，关于卷积输入层的通道数、卷积输出层的通道数以及卷积核的通道数，下列说法正确的是：

A.卷积输入层的通道数与卷积输出层的通道数相等

B.卷积输入层的通道数与卷积核的通道数相等

C.卷积输出层的通道数与卷积核的通道数相等

D.卷积输出层的通道数与卷积核的个数相等

(多选题) 66.关于逻辑回归 (LR) 和支持向量机 (SVM), 下列说法正确的是:

- A.两种算法的损失函数不同, 但是目的都是增加对分类影响较大的数据点的权重, 减少与分类关系较小的数据点的权重。
- B.两种算法解决的问题不同, 逻辑回归是解决回归算法, 支持向量机是分类算法
- C.支持向量机可以解决的问题逻辑回归也可以解决, 反之不一定
- D.逻辑回归可以解决的问题支持向量机也可以解决, 反之不一定

(多选题)67.关于 seq2seq 模型, 下列说法正确的是:

- A.seq2seq 包含编码器和解码器两部分
- B.seq2seq 可用于文本摘要任务
- C.可以将 seq2seq 理解为一种 NxM 模型
- D.可以使用 attention 机制对 seq2seq 进行优化

(多选题)68.常见的人工智能加速硬件包括:

- A.CPU
- B.GPU
- C.NPU
- D.TPU

(多选题) 69.关于标准正态分布, 正确的是:

- A.均值为 1

B.均值为 0

C.方差为 1

D.方差为 0

(多选题)70.数据预处理中清除数据集中的“脏”数据，这里的“脏”数据是指：

A.存储数据的设备受到某些污染物的影响

B.数据包含错误的记录或者异常点

C.数据中存在矛盾、有差异的记录

D.一些数据缺失某些特征或者包含一些缺失的值

(多选题)71.假设 `data = (1, 3, 3, 7, 9, 11)`，下列哪些操作是合法的？

A.`data[1 : -1]`

B.`data[1 : 7]`

C.`list(data`

D.`data * 3`

(多选题) 72.在 python 中, 哪下列哪个表达式可以创建一个包含 10 个 0 的一维数组?

A.`np.zeros(10)`

B.`np.zeros(10,)`

C.`np.zeros((10))`

D.`np.array([0.] * 10)`

(多选题)73.与机器学习相比，深度学习的主要特点在于：

- A.深度学习对算力的需求小于机器学习。
- B.深度学习可以自动从数据中提取特征。
- C.深度学习可以模拟更加复杂的函数。
- D.通过深度学习算法训练得到的模型更容易被人类所理解。

(多选题)74.使用 OneHot（独热码）编码来表示文本，下列说法正确的是：

- A.可以用较低维数来表示高维稀疏数据
- B.会浪费存储空间
- C.不同词的向量正交，无法衡量不同词之间的关系
- D.可以根据两个词在文本中出现的先后顺序进行语义分析

(多选题)75.下列关于 pandas 数据读/写的说法中，正确的是：

- A.read_csv 函数可以读取文本格式的结构化数据
- B.read_sql 函数可以读取数据库的数据
- C.to_csv 函数可以将结构化数据写入 csv文件
- D.to_excel 函数可以将结构化数据写入 Excel文件