

序号	题型	题干	选项A	选项B	选项C	选项D
1	单选题	Python3 支持的数字类型不包括以下哪一项	int	float	complex	range
2	单选题	下面哪项不属于Python异常处理的关键字?	try	catch	raise	except
3	单选题	[1, 2, 3]+[4, 5, 6] 的结果是多少	[6][15]	[1, 2, 3][4, 5, 6]	[21]	[1, 2, 3, 4, 5, 6]
4	单选题	假设有这样的一个示例: types=['娱乐','体育','科技'], 我们在使用列表时, 以下哪个选项, 会引起索引错误?	types[-1]	types[-2]	types[0]	types[3]
5	单选题	range() 函数的作用, 描述正确的是?	可以将结果转换为列表	生成一系列的数字	可以解析列表	可以统计计算
6	单选题	列表 books=['半生缘','往事并不如烟','心是孤独的猎手'], 执行切片操作, 以下代码输出错误的是?	books[0:2], 输出: ['半生缘','往事并不如烟']	books[:2], 输出: ['半生缘','往事并不如烟']	books[1:], 输出: ['半生缘','往事并不如烟']	books[-2:], 输出: ['往事并不如烟','心是孤独的猎手']
7	单选题	pip 的作用?	对操作系统进行操作	字符加密功能	输出运行日志	Python 包安装器
8	单选题	关于 python 类继承, 下列描述错误的是?	定义子类的实例时, 可以通过 子类的 init() 方法, 给父类的所有属性赋值	对于继承而来的父类方法, 如果它不符合子类所期望的行为, 那么就必须建立新的类。	super() 是一个特殊函数, 它会把父类和子类关联起来。	子类除了拥有继承父类而来的属性和方法之外, 还可以自定义子类自己的属性和方法
9	单选题	关于 python 读取文件, 下列描述错误的是	关键字 with 会在程序不再需要访问文件或出现异常的情况下, 关闭文件	Python 对数据量没有大小限制, 需要读取的文件多大都可以	函数 open(), 入参如果只有文件名, 那么 Python 会在当前执行的 .py 文件的所在目录中, 查找文件	读取文本文件时, Python 会将文件中的所有文本都解释为字符串
10	单选题	关于注释, 下列描述错误的是?	注释用井号 ( # ) 标识	编写注释的主要目的是阐述代码要做什么, 以及是如何做的	写不写注释, 无所谓	必须编写有意义的注释
11	单选题	在大数据集上训练决策树。为了减少训练时间, 我们可以	增加树的深度	增加学习率 (learning rate)	减少树的深度	减少树的数量
12	单选题	哪种推荐系统技术从其他用户偏好数据中获益最多	基于条目的过滤	基于内容的过滤	基于用户的过滤	以上所有方法都不满足
13	单选题	聚类算法中, 关于簇的初始化, 以下那种说法不正确	选择彼此距离尽可能远的那些点	初始中心点的选择对迭代的快速停止和聚类结果质量很重要	初始中心点周围的数据点密度很高, 有可能更快地得到高质量的聚类结果	初始中心点的选择对聚类结果无关紧要
14	单选题	我们建立了一个有5000个特征, 100万数据的机器学习模型。以下哪些是有效地训练这些数据的方法	随机抽取一些样本, 在这些少量样本之上训练	可以试用在线机器学习算法	应用PCA算法降维, 减少特征数	以上所有
15	单选题	某超市研究销售纪录数据后发现, 买啤酒的人很大概率也会购买尿布, 这种属于数据挖掘的哪类问题	关联规则发现	聚类	分类	自然语言处理
16	单选题	以下哪项关于决策树的说法是错误的	冗余属性不会对决策树的准确率造成不利的影响	子树可能在决策树中重复多次	决策树算法对于噪声的干扰非常敏感	寻找最佳决策树是NP完全问题
17	单选题	以下两种描述分别对应哪两种对分类算法的评价标准? (a)警察抓小偷, 描述警察抓的人中有多少个是小偷的标准。 (b)描述有多少比例的小偷给警察抓了的标准	Precision, Recall	Recall, Precision	Precision, ROC	Recall, ROC
18	单选题	哪一种分类器假定各个特征之间是独立的	朴素贝叶斯	SVM	决策树	最近邻居法
19	单选题	通常来说, 哪个模型被认为易于解释?	SVM	Logistic Regression	Decision Tree	K-Mean
20	单选题	一次调查表明苏州三分之一的共享单车都是摩拜, 并且在那之中一半的车都是黄色的。 如果你在上海的街头看到一辆黄色的单车过来, 请问它是摩拜的可能性有多大	0. 5	0. 333333333	0. 166666667	0. 2
21	单选题	SVM (支持向量机) 与 LR (逻辑回归) 的数学本质上的区别是什么	损失函数	是否有核技巧	是否支持多分类	其余选项均不对
22	单选题	在回归模型中, 下列哪一项在权衡欠拟合 (under-fitting) 和过拟合 (over-fitting) 中影响最大?	回归函数阶数	更新权重 w 时, 使用的是矩阵求逆还是梯度下降	使用常数项	其余选项均不对
23	单选题	如果我们说“线性回归”模型完美地拟合了训练样本 (训练样本误差为零), 则下面哪个说法是正确的?	很可能训练模型过拟合	测试样本误差始终为零	测试样本误差不可能为零	其余选项均不对
24	单选题	为了观察测试 Y 与 X 之间的线性关系, X 是连续变量, 使用下列哪种图形比较适合	散点图	柱形图	直方图	其余选项均不对
25	单选题	一般来说, 下列哪种方法常用来预测连续独立变量	线性回归	逻辑回顾	线性回归和逻辑回归都行	其余选项均不对
26	单选题	在Numpy中构建一个全是0的数组一般使用	zeros()	ones()	arange()	linspace()
27	单选题	梯度为 () 的点, 就是损失函数的最小值点, 一般认为此时模型达到了收敛	-1	0	1	∞
28	单选题	创建一个 3 x 3 的数组, 下列方法错误的是	np. arange(0, 9). reshape(3, 3)	np. eye(3)	np. random. random([3, 3, 3])	np. mat("1, 2, 3;4, 5, 6;7, 8, 9")
29	单选题	关于数据集的标准化, 正确的描述是	数据集标准化有助于加快模型的收敛速度	标准化一定是归一化, 即数据集的取值分布在 [0, 1] 区间上	数据集的标准化一定是让标准差变为1	所有的模型建模之前, 必须进行数据集标准化
30	单选题	DL是 () 的简称	人工智能	机器学习	神经网络	深度学习
31	单选题	一般使用以下哪种方法求解线性回归问题	最小二乘法	最大似然估计	对数变换	其余选项均不对
32	单选题	以下哪个函数可以实现画布的创建	subplots()	add_subplot()	figure()	subplot2grid()
33	单选题	Numpy中统计数组元素个数的方法是	ndim	shape	size	itemsiz
34	单选题	决策树中的分类结果是最末端的节点, 这些节点称为	根节点	叶节点	父节点	子节点
35	单选题	ML是 () 的简称	人工智能	机器学习	神经网络	深度学习

序号	题型	题干	选项A	选项B	选项C	选项D
36	单选题	当数据特征不明显、数据量少的时候，采用下面哪个模型	线性回归	逻辑回归	支持向量机	神经网络
37	单选题	机器学习的流程包括:数据获取、特征提取、（）和模型验证这四个过程	数据清洗	数据分析	模型训练	数据开发
38	单选题	以下哪个软件包用于矩阵和向量的科学计算	Numpy	Pandas	Matplotlib	sk-learning
39	单选题	使下面哪个函数可以将线性回归线转为逻辑回归线	Sigmoid	高斯核函数	P(A)	H(x)
40	单选题	下面不属于人工神经网络的是	卷积神经网络	循环神经网络	网络森林	深度神经网络
41	多选题	不可变数据类型包括？	Set	String	Tuple	Number
42	多选题	关于变量的命名和使用，下列描述正确的是？	变量名可以包含字母、数字、下划线和特殊字符	变量名不能包含空格，但可使用下划线来分隔其中的单词	可以将Python关键字和函数名用作变量名	变量名应既简短又具有描述性
43	多选题	在列表中添加元素的方法有哪些	append()	insert()	tuple()	add()
44	多选题	关于元组的描述，以下说法正确的是	元组是使用圆括号来标识的	定义好元组后，可以访问，但不能修改	修改元组，Python 会返回类型错误	虽然不能修改元组中的元素，但我们可以，为原来存储元组的变量重新赋值
45	多选题	以下关于字典，描述正确的是	字典是一系列的键值对	与键相关联的值可以是任何 Python 对象，比如数字、字符串、列表甚至是字典	可以先使用一对空的花括号，定义一个空字典，然后再分行添加键值对	可以指定字典名、用方括号括起的键以及与该键相关联的新值，来修改字典值
46	多选题	以下哪种是用于降维的重要特征提取技术	PCA	LDA	RNN	SVM
47	多选题	以下那些属于深度学习算法	RNN	BP	CNN	SVM
48	多选题	SVM的算法性能取决于	核函数的选择	核函数的参数	软间隔参数C	权重系数
49	多选题	对于机器学习表述下列不正确的是	机器学习和人工智能是独立的两种技术	机器学习是人工智能的核心技术和重要分支	机器学习的目标是让机器设备像人类一样学习书本知识	机器学习是指一系列程序逻辑控制算法
50	多选题	下列选项中能创建Numpy数组的选项是	a = numpy.array([1,2,3])	a = numpy.array([1,[1,2],3])	a = numpy.array([[1,2,3],[4,5,6]])	a = numpy.array([[ ' xiao' , ' qian ' ],[ ' xiao' , ' feng' ]])
51	多选题	Pandas里提供的两种基本数据结构是什么	Dataframe	Series	array	func
52	多选题	下列哪种方法可以用来减小过拟合	更多的训练数据	L1 正则化	L2 正则化	减小模型的复杂度
53	多选题	训练误差会降低模型的准确率，产生欠拟合，此时如何提升模拟拟合度？	增加数据量	特征工程	减少正则化参数	提高模型复杂度
54	多选题	以下哪些学科属于人工智能	运维系统	图像识别	自然语言处理	语音识别
55	多选题	逻辑回归与线性回归分析有哪些不同之处	逻辑回归用来预测事件发生的概率	逻辑回归用来计算拟合优度指数	逻辑回归用来对回归系数进行估计	逻辑回归结果集是二元对立关系
56	多选题	如果一个 SVM 模型出现欠拟合，那么下列哪种方法能解决这一问题	增大惩罚参数 C 的值	减小惩罚参数 C 的值	减小核系数（gamma参数）	增大核系数（gamma参数）
57	多选题	下列哪些算法可以用来构造神经网络	kNN	svm	线性回归	逻辑回归
58	多选题	train_test_split(data)分割后，一般来说会有哪些返回数据集	x_train	y_train	x_test	y_test
59	多选题	在决策树中，当一个节点分叉的时候，以下关于“信息增益”正确的是	不纯的节点越少，越需要更多的信息来描述种群	信息增益可以用熵作为“1-Entropy”来推导	信息增益偏向于数值大的属性	信息增益是随机决策的
60	多选题	以下关于主成分分析哪些是正确的	在 PCA 前必须将数据标准化	我们应该选择说明最高方差的主成分	我们应该选择说明最低方差的主成分	我们可以用 PCA 来可视化低维数据
61	多选题	如何在监督的机器学习挑战使用聚类方法	可以先创建簇，然后在不同簇中分别使用监督机器学习算法	在使用监督机器学习算法之前可以把簇的id 作为特征空间中额外的特征	无法在使用监督机器学习算法之前创建簇	在使用监督机器学习算法之前不能把簇的id 作为特征空间中额外的特征
62	多选题	在应用线性回归时我们需要遵循哪些假设	检查异常值	要求所有变量都遵循正态分布	假设数据中很少或不存在多重共线性	可以适用任何情况
63	多选题	以下关于ndarray(数组)的属性说明正确的是	ndim 表示数组的维数	shape 表示数组的尺寸	size 表示数组的尺寸	dtype 表示数组中元素的类型
64	多选题	下列关于数据合并的说法正确的是	concat() 可实现任意两张表的横向堆叠或者纵向堆叠	append() 可实现任意两张表的横向堆叠或者纵向堆叠	concat() 函数中的参数axis=0时表示纵向操作，axis=1时表示横向操作	concat() 函数中的参数join表示其他轴向上的索引是按交集还是并集合并
65	多选题	以下步骤属于数据清洗有哪些	去重	删除缺失值	数据合并	异常值检测
66	多选题	以下属于分类模型的评价指标的是	精确率	召回率	均方误差	ROC曲线
67	多选题	以下回归方法属于线性回归的是	分段回归	一元线性回归	多元线性回归	多个因变量与多个自变量的回归
68	多选题	以下方法属于回归模型评价指标的是	平均绝对误差	精确度	均方误差	中值绝对误差
71	判断题	string、list、tuple都属于序列（sequence）				
72	判断题	“我有” + 10 + “个苹果。”，这个表达式会报错				
73	判断题	函数可以减少重复的代码，实现程序复用。从而让程序更短，并且更容易阅读与修改				
74	判断题	全局作用域指的是在整个程序中都有效；而局部作用域，只在代码块中有效，比如函数内部				

序号	题型	题干	选项A	选项B	选项C	选项D
75	判断题	把可能导致异常的代码放在 try 代码块中；发生异常时要执行的代码放在 catch 代码块中。				
76	判断题	保存在字典中的项是无序的，而列表中的项是有序的				
77	判断题	RMSE (Root Mean Square Error) 均方根误差可用于评估推荐系统和线性回归模型				
78	判断题	假如你在训练一个线性回归模型，如果数据量较少，容易发生过拟合				
79	判断题	假如你在训练一个线性回归模型，如果数据维度较少，容易发生欠拟合				
80	判断题	朴素贝叶斯分类器的特点是，假设样本各维属性独立				
81	判断题	SVM的原理的简单描述，可概括为最大间隔分类				
82	判断题	随机森林属于Bagging方法				
83	判断题	机器学习的随着数据量的增大上限大于深度学习				
84	判断题	K-means聚类是有监督学习算法				
85	判断题	集成学习中，软投票的意思是分类器“一人一票”				
86	判断题	使用Pandas库设置索引使用set_index()方法				
87	判断题	matplotlib中xlabel()方法是设置y轴的名称				
88	判断题	支持向量机的简称是ANN				
89	判断题	如果学习率过大，模型需要更长的训练时间才能收敛到最优值				
90	判断题	a=np.random.randn(4,3) b=np.random.randn(3,2) c=a*b，c的结果是.c.shape=(4,3)				
91	判断题	对没有标签(label)的数据进行分类的问题属于机器学习中聚类问题，无label即无监督学习				
92	判断题	SVM 只能应用于线性分类				
93	判断题	增加数据，能够有效减小过拟合，减小训练样本误差和测试样本误差之间的差距				
94	判断题	无监督算法只处理“特征”，不处理“标签”				
95	判断题	降维算法不属于无监督学习				
96	判断题	K-meas算法和SVM算法都属于无监督学习				
97	判断题	深度学习的概念源于神经网络的研究，含有多个隐层的神经网络算法就是一种深度学习算法				
98	判断题	逻辑回归只能做2值分类，不能直接做多值分类				
99	判断题	卷积神经网络属于深度学习的一种，单层神经网络也是深度学习的一种				
100	判断题	逻辑回归要求自变量和目标变量是线性关系				