**特许全球金融科技师（CGFT）一级**

**《机器学习原理及应用》考试大纲**

（修订日期： 2022年09月18日）

[一、 考试性质 2](#_Toc114653581)

[二、 考查目标 2](#_Toc114653582)

[三、 考试形式和试卷结构 3](#_Toc114653583)

[1. 试卷满分 3](#_Toc114653584)

[2. 答题方式 3](#_Toc114653585)

[3. 试卷内容结构 3](#_Toc114653586)

[4. 试卷题型结构 3](#_Toc114653587)

[四、 考察内容与考点解析 4](#_Toc114653588)

[1. 章节内容考点清单 4](#_Toc114653589)

[五、 历年真题与考试样题 8](#_Toc114653590)

# 考试性质

特许全球金融科技师（CGFT）一级包含《机器学习原理及应用》科目，机器学习是金融科技从业关键技能。特许全球金融科技师（CGFT）一级考试是技能认证性考试，按照金融科技人才知识技能能力标准，以标准化考试的方式对考生对于机器学习基础算法理论和编程技能的掌握程度进行评价。

# 考查目标

1. **知识技能要求**

CGFT一级对考生机器学习相关知识技能要求如下：

* 掌握常见机器学习算法模型的基本概念和基本原理，包括但不限于：K近邻、多元线性回归、逻辑回归、决策树、SVM、PCA主成分分析、K-Means聚类分析等。
* 能熟练使用Scikit-Learn、XGBoost、LightGBM等机器学习建模工具进行机器学习建模，了解使用Tensorflow、PyTorch、Keras等深度学习建模工具进行深度学习建模。
* 掌握常见机器学习算法模型在金融风控和营销等业务场景中的应用。

1. **能力素质要求**

CGFT一级对考生机器学习相关能力素质要求如下：

* 理解常见机器学习算法模型的原理。
* 具备将各类机器学习算法模型在金融业务场景中的灵活应用能力，根据实际业务场景选择和设计模型。
* 对直接从事算法模型研究的考生，应具备一定的算法模型设计和创新研究能力。

# 考试形式和试卷结构

## 试卷满分

CGFT一级考试包含七门课程，其中《机器学习原理及应用》包含40题，该部分满分为100分。

## 答题方式

机考，闭卷，可以跨区答题

## 试卷内容结构

不对内容结构分布做明确限定。

## 试卷题型结构

以客观题方式进行考察，题型包括：判断题、单选题、多选题。

不对题型结构分布做明确限定。

# 考察内容与考点解析

## 章节内容考点清单

本表中考察要求的定义如下

* 基础识记：主要是事实知识，考察对事实性知识的记忆，“知道”即可。
* 一般理解：主要是概念定义，仅需一般理解。
* 深入理解：主要是核心原理，需要较深入理解并能灵活应用。
* 熟练掌握：主要是操作技能，需要熟悉基本操作流程。
* 不做要求：不作为考试内容，考生可以自主选择阅读。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 章 | 节 | 考点 | 考查要求 | 重要性 | 难度 |
| 1. 机器学习理论基础 | 1-1：机器学习的概念和定义 | * 机器学习的概念定义 * 机器学习的分类：监督学习、无监督学习、强化学习 * 理解有监督学习的概念定义，了解常见有监督学习算法模型 * 理解无监督学习的概念定义，了解常见无监督学习算法模型 * 机器学习和人工智能、数据科学的关系 | 一般理解 | 低 | 低 |
| 1-2：极大似然估计 | * 理解极大似然估计的概念和原理 * 理解极大对数似然估计的概念和原理 | 一般理解 | 高 | 高 |
| 1-3：贝叶斯原理和贝叶斯推断 | * 统计学中的贝叶斯定理 * 贝叶斯统计推断的概念 * 贝叶斯统计推断的计算步骤：先验、估计、后验 * 贝叶斯统计推断的优点 * 最大化后验概率 * 贝叶斯统计推断与机器学习的关系 | 深入理解 | 高 | 高 |
| 1-4：最优化理论基础 | * 最优化问题的定义：目标函数、约束条件 * 最优化问题的分类：线性优化、非线性优化 * 线性优化的概念定义 * 多元线性回归最小二乘法 * 凸优化的概念定义 * 非线性优化的概念定义 * 局部最优解与全局最优解，凸优化的局部最优等价于全局最优 * 梯度下降法基本原理 | 深入理解 | 高 | 高 |
| 1-5：偏差与方差 | * 了解偏差与方差的概念定义 * 了解偏差和方差分解 * 结合多项式回归的例子理解偏差和方差概念 | 一般理解 | 高 | 高 |
| 1-6：机器模型性能的衡量指标 | * TP、FP、FN、TN与混淆矩阵的概念定义 * 准确率Accuracy的概念定义 * 精确率Precision、召回率Recall、F1的概念定义 * 理解ROC曲线、AUC指标、KS指标 | 深入理解 | 高 | 高 |
| 1-7：模型验证和模型选择 | * 模型性能评价标准 * 理解Hold-out-sample验证的概念定义 * 理解交叉验证：K-Fold法、LOO留一法 | 一般理解 | 低 | 中 |
| 1-8：正则化 | * 为什么需要正则化 * 理解Lasso一阶正则化 * 理解Ridge二阶正则化 | 一般理解 | 低 | 中 |
| 1. 线性回归模型 | 2-1：线性回归模型 | * 线性回归模型的基本概念和定义形式 * 了解线性回归模型的业务应用 | 深入理解 | 高 | 中 |
| 2-2：均方误差最小化与最小二乘法 | * RMSE均方误差的定义 * 线性回归最小二乘法的概念和计算原理 | 深入理解 | 高 | 中 |
| 2-3：使用Scikit-Learn工具包实现线性回归 | * 使用Scikit-Learn实现线性回归 | 熟练操作 | 高 | 中 |
| 1. K-近邻算法 | 3-1：K近邻算法 | * K近邻算法基本原理 * K近邻算法计算步骤 * 参数K的选取原则和方法 * K近邻算法的优缺点 | 深入理解 | 高 | 中 |
| 3-2：使用Scikit-Learn工具包实现K近邻算法 | * 使用Scikit-Learn实现K近邻算法 | 熟练操作 | 高 | 中 |
| 1. 朴素贝叶斯(Naive Bayes) | 4-1：朴素贝叶斯分类 | * 贝叶斯分类的概念定义 * 朴素贝叶斯分类概念和原理 * 高斯朴素贝叶斯分类概念和原理 | 深入理解 | 高 | 中 |
| 4-2：朴素贝叶斯分类的优缺点 | * 朴素贝叶斯分类的优缺点 | 一般理解 | 低 | 中 |
| 4-3：使用Scikit-Learn工具包实现朴素贝叶斯分类 | * 使用Scikit-Learn工具包实现朴素贝叶斯分类 | 熟练操作 | 高 | 中 |
| 1. 逻辑回归 | 5-1：逻辑回归 | * 逻辑回归模型的定义 * Sigmoid函数的定义和性质 * 逻辑回归损失函数：交叉熵 * 逻辑回归优化求解：梯度下降法 | 深入理解 | 高 | 高 |
| 5-2：使用Scikit-Learn工具包实现逻辑回归 | * 使用Scikit-Learn工具包实现逻辑回归 | 熟练操作 | 高 | 高 |
| 1. 决策树 | 6-1：决策树算法模型 | * 决策树模型的基本概念 * 决策树模型的训练步骤 | 深入理解 | 高 | 高 |
| 6-2：不纯度的度量指标：信息熵、GINI系数 | * 信息熵的概念定义 * GINI系数的概念定义 | 深入理解 | 高 | 高 |
| 6-3：决策树分类模型 | * 基于信息熵增益的ID3算法原理 * 基于信息熵增益率的C4.5算法原理 * 基于GINI系数增益的CART算法原理 | 深入理解 | 高 | 高 |
| 6-4：决策树回归模型 | * 分枝变量度量指标：方差 * 案例使用波士顿房价数据进行决策树回归 | 深入理解 | 高 | 高 |
| 6-5：使用Scikit-Learn工具包实现决策树模型 | * 使用Scikit-Learn工具包实现决策树模型，包括：模型参数、决策树可视化 | 熟练操作 | 高 | 中 |
| 1. 支持向量机SVM | 7-1：支持向量机：线性可分的情形 | * 线性可分问题 * SVM概念：超平面、支持向量SV、边距Margin * SVM模型的几何直观解释 * 高维空间的超平面 * SVM模型的数学定义，SVM优化目标是最大化决策边距 | 深入理解 | 高 | 高 |
| 7-2：支持向量机：线性不可分的情形 | * 线性不可分问题 * 理解软边际与松弛变量的概念和原理 * 核函数概念、使用核函数的好处、各类常见核函数 | 深入理解 | 高 | 高 |
| 7-3：使用Scikit-Learn工具包实现SVM模型 | 使用Scikit-Learn工具包实现SVM模型 | 熟练操作 | 高 | 中 |
| 1. 集成学习 | 8-1：集成学习 | * 集成学习基本概念 * 袋装法Bagging原理 * 提升法Boosting原理 | 一般理解 | 高 | 高 |
| 8-2：常见集成学习算法 | * 随机森林算法原理 * XGBoost算法原理 * LightGBM算法原理 | 一般理解 | 高 | 高 |
| 8-4：使用Scikit-Learn工具包实现各类集成学习算法 | * 使用Scikit-Learn工具包实现各类集成学习算法 | 熟练操作 | 高 | 高 |
| 1. 无监督学习：降维 | 9-1：无监督学习 | * 无监督学习的定义 * 为什么要使用无监督学习 * 无监督学习的类型：降维、聚类 | 深入理解 | 高 | 高 |
| 9-2：降维 | * 协方差矩阵、相关系数矩阵 * 主成分分析PCA的计算步骤，特征向量的计算步骤 * 降维的应用 | 深入理解 | 高 | 高 |
| 9-3：使用Scikit-Learn工具包实现PCA | * 使用Scikit-Learn工具包实现PCA | 熟练操作 | 高 | 中 |
| 1. 无监督学习：聚类 | 10-1：聚类 | * 聚类的基本概念 * K-Means聚类算法原理和步骤 | 深入理解 | 高 | 中 |
| 10-2：使用Scikit-Learn工具包实现K-Means聚类 | * 使用Scikit-Learn工具包实现K-Means聚类 | 熟练操作 | 高 | 中 |
| 1. 神经网络 | 11-1：人工神经网络基本概念 | * 理解人工神经网络基本概念 * 感知机Perceptron模型原理，Perceptron与逻辑回归的比较 * 多层感知机MLP模型原理 | 一般理解 | 高 | 中 |
| 11-2：人工神经网络算法原理 | * 激活函数：sigmoid、tanh、ReLu，掌握函数形式、函数图示以及导数，理解使用非线性激活函数的原因 * 目标损失函数：MSE、交叉熵 * 梯度下降法原理 * 神经网络的正向传播和反向传播 * 神经网络超参数 | 深入理解 | 高 | 高 |
| 11-3：使用Scikit-Learn工具包实现多层感知机模型 | * 基于Sklearn.neural\_network模块 * 基于MLPClassifier模块 | 熟练操作 | 高 | 中 |
| 11-4：使用Tensorflow和Keras建立神经网络模型 | * Tensorlow和Keras包的安装 * 常用数据集 * Keras包的类：datasets、models、layers、optimizers | 熟练操作 | 低 | 中 |
| 1. 深度学习基础入门 | 12-1：深度学习基本概念 | * 深度学习基本概念 * 深度学习的常用优化方法：SGD、Adam * Mini-Batch训练机制 | 不做要求 | 高 | 高 |
| 12-2：卷积神经网络CNN | * 理解图像数据 * 理解CNN模型原理：卷积、池化、非线性激活函数 * 经典CNN模型的网络结构：LeNet、AlexNet、VGGNet、ResNets * CNN模型在图像识别中的应用 | 不做要求 | 高 | 高 |
| 12-3：循环神经网络RNN | * 理解序列数据 * 理解RNN模型原理、理解GRU模型原理 * RNN模型的缺点 * RNN模型在时间序列预测中的应用 | 不做要求 | 高 | 高 |
| 12-4：长短时记忆网络LSTM | * LSTM模型原理与网络结构：门控、遗忘门、输入门 * LSTM在自然语言处理中的应用 | 不做要求 | 高 | 高 |

# 历年真题与考试样题

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题面 | 选项列 | | | | 正确答案 |
| A | B | C | D |
| 机器学习模型的计算是事先定好的，而统程传序的计算是随着数据而逐渐适应改变的 | 对 | 错 |  |  | B |
| K-最近邻算法是一种“懒惰学习者”算法 | 对 | 错 |  |  | A |
| 当目标变量是类型的时侯使用逻辑回归模型，而当目标变量是连续的时侯使用逻辑分类模型，分类和回归都属于监督学习 | 对 | 错 |  |  | B |
| SVM训练完成后，模型的分类能力只和支持向量有关，而和所有非支持向量无关。 | 对 | 错 |  |  | A |
| 人工神经网络（Artificial Neural Network， ANN）是受构成动物大脑的生物神经网络启发而来的计算系统 | 对 | 错 |  |  | A |
| 机器学习可以分成 | 分类学习和混合学习 | 监管学习和自动学习 | 监督学习和无监督学习 | 预测学习和分析学习 | C |
| 关于贝叶斯理论，如下哪个陈述是不对的： | 贝叶斯理论是联系不同条件概率的一个理论 | 贝叶斯理论通常是结合先验知识(prior knowledge)，来计算相关事件的概率 | 贝叶斯理论使用一些我们已经拥有的知识或信念，通常称为后验概率，来帮助我们计算相关事件的概率 | 如果我们想要计算在事件B这个条件下，A发生的概率，我们只要知道了，在条件A下B发生的概率，再结合事件A和事件B各自的边际概率就行了 | C |
| 以下哪个陈述是对的： | 优化实际上就是对约束函数进行最小化或最大化，同时需要满足一系列的目标要求 | 优化实际上就是对约束函数进行最小化或最大化，或者需要满足一系列的目标要求 | 优化实际上就是对目标函数进行最小化或最大化，同时需要满足一系列的约束要求 | 优化实际上就是对目标函数进行最大化，同时最小化约束函数 | C |
| 如下哪个是正确的： | 一般非线性规划问题总能找到解析解 | 凸优化是非线性优化的一种特殊情况 | 凸优化是线性优化的一种特殊情况 | 当目标函数是非线性的但约束函数是线性的时候，就是凸优化 | B |
| 把数据分成两部分， | 一部分用来做训练也就是样本内数据，另外一部分用来测试模型的好坏，叫样本外数据。 | 一部分用来做训练也就是样本外数据，另外一部分用来测试模型的好坏，叫样本内数据。 | 一部分用来做测试也就是测试数据，另外一部分用来判断模型的好坏，叫判断数据。 | 一部分用来做判断也就是判断数据，另外一部分用来测试模型的好坏，叫测试数据。 | A |
| 用来预测连续变量的常用方法是： | 逻辑回归 | 线性回归 | 逻辑回归和线性回归都可以 | 以上都不是 | B |
| 以下哪个陈述是对的： | 因为 Lasso 正则化方法可以使得某些模型参数缩小到0，所以可以用来做特征选择 | 因为 L2 正则化方法可以使得某些模型参数缩小到0，所以可以用来做特征选择 | 因为 Lasso 正则化方法可以使得某些模型参数缩小到0，所以不能用来做特征选择 | 因为 L2 正则化方法可以使得某些模型参数缩小到0，所以不能用来做特征选择 | A |
| 以下哪个陈述是对的： | 图像的表现形式主要是几何，图像处理是提取其中的形状，对图像进行分类识别。 | 图像的表现形式主要是坐标，图像处理是提取支持向量，对图像进行分割。 | 图像的表现形式主要是像素，图像处理是提取其中色彩，对图像进行灰度分析。 | 图像的表现形式主要是像素，图像处理是提取其中的模式和结构，对图像进行分类识别。 | D |
| 以下哪个陈述是对的： | 决策树叶节点也叫根节点 | 决策树叶节点是各个分叉后的节点 | 决策树分叉最后的节点叫做叶节点 | 决策树分叉最后的节点叫根节点 | C |
| 以下哪个陈述是对的： | 基尼不纯度和熵的概念有一些类似之处，基尼系数越高，信息内容含量越多，也就代表越纯。 | 基尼不纯度和熵的概念有一些类似之处，基尼系数越高，信息内容含量越少，也就代表越纯。 | 熵的概念和基尼不纯度有一些类似之处，熵越高，信息内容含量越多，也就代表越不纯。 | 熵的概念和基尼不纯度有一些类似之处，熵越高，信息内容含量越少，也就代表越纯。 | C |
| 支持向量机 (Support Vector Machine, SVM) 算法的目标是 | 在n 维空间（n-特征数）中找到一个对数据点进行清晰分类的超平面 | 在n 维空间（n-特征数）中找到对所有数据点的平均距离最近的超平面 | 在n-1 维空间（n-特征数）中找到一个对数据点进行清晰分类的超平面 | 在2维空间中找到一个对数据点进行清晰分类的超平面 | A |
| 以下哪个陈述是对的： | 支持向量机硬边际分类意味着选择一个超平面，可以允许一些数据点停留在超平面的不正确的一边，但只能在超平面的边际之外。 | 支持向量机软边际分类意味着选择一个超平面，可以允许一些数据点停留在超平面的不正确的一边，以及在超平面的边际之间。 | 支持向量机软边际分类意味着选择一个超平面，可以允许一些数据点停留在超平面的不正确的一边，但不能在超平面的边际之外。 | 支持向量机硬边际分类意味着选择一个超平面，可以允许一些数据点停留在超平面的不正确的一边，但只能在超平面的边际之内。 | B |
| 以下哪个陈述是不对的： | 当分类问题线性不可分时，可以用多项式回归模型达到非线性效果 | 当分类问题线性不可分时，可以利用原始特征设计非线性特征。 | 当分类问题线性不可分时，可以将数据投影到一个更高维的空间，在那里线性分类器的性能会更好 | 当分类问题线性不可分时，需要用到非线性边界 | A |
| SVM中，边距（Margin) 的长度等于支持向量到超平面距离的 | 0.5倍 | 1倍 | 2倍 | 4倍 | C |
| 下面不属于无监督学习模型的是： | 主成分分析 | 奇异值分解 | K-均值聚类 | K-近邻 | D |
| 神经网络: | 是用来处理简单数据输入并协同工作的不同机器学习算法的复杂框架 | 通过考虑训练样本数据来“学习”怎样执行计算任务 | 需要使用特定任务规则的计算程序 | 神经网络是逻辑回归模型和聚类模型的结合 | B |
| sigmoid函数， | 也可以作为神经网络中的激活函数 | 数学表达式是：1-exp(-z) | 数学表达式是：1/(1+exp(z)) | 数学表达式是：1/(1-exp(z)) | A |
| 关于损失函数： | 回归问题和分类问题都可以用RELU损失函数 | 回归问题用平方误差损失函数，分类问题用交叉熵损失函数 | 回归问题可以用平方误差损失函数，也可以用交叉熵损失函数 | 分类问题可以用平方误差损失函数，也可以用交叉熵损失函数 | B |
| 可以对手写数字 MNIST数据集做分类的模型不包括： | 逻辑回归 | 多层感知器 | 线性回归 | 支持向量机 | C |
| keras.datasets 的数据包不包括: | CIFAR10小图像分类 | IMDB影评情感分类 | MNIST手写数字数据库 | SP500股票数据 | D |
| K-近邻方法属于 | 勤快的预测方法 | 懒惰的预测方法 | 懒惰的学习方法 | 勤快的学习方法 | C |
| 朴素贝叶斯(Naive Bayes)中的”朴素“是因为 | 假定各个样本之间的概率是条件独立的 | 假定各个特征之间的概率是条件独立的 | 假定各个样本之间的概率是无条件独立的 | 假定各个特征之间的概率是无条件独立的 | B |
| sklearn的决策树分类器有一个函数参数是随机状态 random\_state， | 为了在拟合过程中获得随机行为，random\_state必须固定为一个确定的整数。 | 为了在拟合过程中获得确定性行为，random\_state必须固定为一个确定的整数。 | 为了在预测过程中获得随机行为，random\_state必须固定为一个确定的整数 | 为了在预测过程中获得确定性行为，random\_state必须固定为一个确定的整数 | B |
| 以下哪个陈述是对的： | IRIS数据集有4个特征，可以通过数据降维把4个特征转换成两个特征，这两个特征是原来数据的最大的两个主成分 | IRIS数据集有4个特征，可以通过数据降维把4个特征转换成两个特征，这两个特征是原来数据的最小的两个主成分。 | IRIS数据集有2个特征，可以通过数据升维把两个特征转换成4个特征，这4个特征是原来数据的最大的4个主成分。 | IRIS数据集有2个特征，可以通过数据升维把两个特征转换成4个特征，这4个特征是原来数据的最小的4个主成分。 | A |
| 以下哪个陈述是对的： | 神经网络是一个简单的函数关系，由数百万个复杂的激活函数组成 | 训练神经网络模型，实质上就是最大化它的方差函数 | 神经网络激活函数的输出值是模型好坏的一个度量 | 神经网络是对某个问题的一个数学解 | D |
| 人工智能包括机器学习 | 对 | 错 |  |  | A |
| 贝叶斯推理就是利用贝叶斯定理从数据中推导一个总样本的性质或概率分布的过程。 | 对 | 错 |  |  | A |
| 逻辑回归模型的拟合需要最小化成本函数，通常用梯度下降法 | 对 | 错 |  |  | A |
| 支持向量机 (SVM) 中，一个数据点越是靠近决策边界，我们对预测越有信心。 | 对 | 错 |  |  | B |
| 神经网络是用来处理复杂数据输入并协同工作的不同机器学习算法的一个框架 | 对 | 错 |  |  | A |
| 监督学习就是对 | 无标签的特征数据建立特征和标签之间关系模型的过程 | 有标签的特征数据建立特征和标签之间关系模型的过程 | 有特征的标签数据建立标和签之间关系模型的过程 | 无标签的特征数据建立特征和特征之间关系模型的过程 | B |
| 贝叶斯统计推断三个步骤的顺序是： | 似然，先验，后验 | 先验，似然，后验 | 先验，后验，似然 | 后验，似然，先验 | B |
| 以下哪个陈述是对的： | 最大化某个函数，实际上就是最小化它的绝对值 | 最小化某个函数，实际上就是最小化它的负函数 | 最小化某个函数，实际上就是最小化它的绝对值 | 最大化某个函数，实际上就是最小化它的负函数 | D |
| 以下哪个陈述是对的： | 全局优化方法的最坏情况复杂度会随问题规模n和m呈指数增长，这里的n指变量的个数。 n指数据的样本数 | 全局优化方法的最坏情况复杂度会随问题规模n和m呈线性增长，这里的n指变量的个数。 n指数据的样本数 | 局部优化方法的最好情况复杂度会随问题规模n和m呈指数增长，这里的n指变量的个数。 n指数据的样本数 | 局部优化方法的最好情况复杂度会随问题规模n和m呈对数增长，这里的n指变量的个数。 n指数据的样本数 | A |
| 以下哪个陈述是对的： | 偏差描述的是预测值的最大值与真实值之间的差距 | 偏差描述的是预测值的最小值与真实值之间的差距 | 偏差描述的是预测值的平均与真实值之间的差距 | 偏差描述的是预测值的最大值与最小值之间的差距 | C |
| K-近邻方法是“懒惰学习者”算法的一个例子， | 即在训练时做好计算，预测时不做计算直接用事先算好的值 | 即在执行对数据集的查询前，它不会通过对训练数据集进行计算来构建模型 | 即在预测时不做任何运算，直接进行查询 | 即在训练时只计算一个回合就停止更新模型参数 | B |
| 朴素贝叶斯的一个假定是： | 各类别的先验概率相等 | P(X|L)是高斯分布（X是特征变量, L是类别标签） | 分布的均值是0， 方差为1 | 特征各维度是类别条件独立的 | D |
| 如下哪个陈述是不对的： | 逻辑 (logistic) 回归的输出 (标签) 为0或1，阈值通常为0.5 | 逻辑 (logistic) 回归的输出大于阈值的为1，小于阈值的视为0 | 逻辑回归只能用于分类，不能用来做回归 | 逻辑回归(Logistic regression)用于解决回归问题而不是分类问题 | D |
| 以下哪个陈述是对的： | 特征选择是选出比较有用和重要的特征，除了用人工的判断哪些特征重要，可以借助 Lasso 正则化来达到类似的效果。 | 特征选择是选出比较有用和重要的标签，除了用人工的判断哪些标签重要，可以借助 Lasso 正则化来达到类似的效果。 | 特征选择是选出比较有用和重要的特征，除了用人工的判断哪些特征重要，可以借助 L2 正则化来达到类似的效果。 | 特征选择是选出比较有用和重要的标签，除了用人工的判断哪些标签重要，可以借助 L2 正则化来达到类似的效果。 | A |
| 如下哪个陈述是不对的： | 图像的表现形式主要是像素 | 图像处理是对像素提取其中的模式和结构 | 对图像做回归是一个简单的图像识别问题 | 一个图像中有多个物体目标时，可以先进行图像分割，再做分类 | C |
| 为了度量决策树的不纯度或不确定性， | 需要共同使用基尼系数和信息增益 | 可以使用基尼系数或信息增益 | 可以使用基尼系数，但不能使用信息增益 | 可以使用信息增益，但不能使用基尼系数 | B |
| 以下哪个陈述是对的： | 熵因为要做指数计算，计算复杂性略为增加。基尼系数是把概率求平方，计算相对简单。 | 基尼系数因为要做指数计算，计算复杂性略为增加。熵是把概率求平方，计算相对简单。 | 熵因为要做对数计算，计算复杂性略为增加。基尼系数是把概率求平方，计算相对简单。 | 基尼系数因为要做对数计算，计算复杂性略为增加。熵是把概率求平方，计算相对简单。 | C |
| Python scikit-learn 包里的决策树模型，random\_state 为了在拟合过程中获得确定性行为，random\_state必须固定为： | 0 | 1 | 一个确定的整数 | 任意浮点数 | C |
| 以下哪个陈述是对的： | 最接近超平面的数据点称为支持向量 | 最远离超平面的数据点称为支持向量 | 支持向量是边距(margin)最大的向量 | 支持向量是边距(margin)最小的向量 | A |
| 以下哪个陈述是对的： | 模型的边际违反 (margin violation) 有一定的惩罚量，惩罚强度由支持向量机 (SVM) 的参数C控制 | 模型的边际违反 (margin violation) 有一定的惩罚量，惩罚强度由软边际的距离决定。 | 模型的边际违反 (margin violation)没有惩罚量，因为这是软边际允许的 | 模型的边际违反 (margin violation) 有一定的惩罚量，惩罚强度由超平面的维度决定 | A |
| 以下哪个陈述是对的： | 支持向量机的目标函数是一种将高维输入空间转化为低维空间的函数 | 支持向量机的目标函数是一种将低维输入空间转化为高维空间的函数 | 支持向量机的核函数是一种将低维输入空间转化为高维空间的函数 | 支持向量机的核函数是一种将高维输入空间转化为低维空间的函数 | C |
| 以下哪个陈述是对的： | 超平面的法向量𝐰 （垂直于超平面）和超平面上任意点 𝐱 的点积等于1 | 如果向量 𝐱 \_xDC31\_和 正交 (垂直)，和 𝐲 的点积等于0，到 𝐲的正交投影是0 | 如果向量 𝐱 \_xDC31\_和 正交 (垂直)，和 𝐲 的点积等于1，到 𝐲的正交投影是1 | 如果向量 𝐱 \_x005f\_xdc31\_和 𝐲 平行，和 𝐲 的点积等于1，到 𝐲的正交投影是0 | B |
| 监督机器学习的应用不包括： | 垃圾电子邮件检测 | 对高维数据降维 | 情绪分析-挖掘文档在情感上是积极的、还是消极的 | 医疗诊断 | B |
| 以下哪个陈述是对的： | 用训练数据找到PCA的变换矩阵，然后同样对训练数据和测试数据做变换，这样才能保证训练数据做的变换和测试数据做的变换是吻合的。 | 用训练数据找到PCA的变换矩阵，然后对训练数据做变换，另外用测试数据找到PCA的变换矩阵并对测试数据做变换，这样才能保证训练数据做的变换和测试数据做的变换是类似的。 | 用测试数据找到PCA的变换矩阵，然后同样对训练数据和测试数据做变换，这样才能保证训练数据做的变换和测试数据做的变换是吻合的。 | 用测试数据找到PCA的变换矩阵，然后对训练数据做变换，另外用训练数据找到PCA的变换矩阵并对测试数据做变换，这样才能保证训练数据做的变换和测试数据做的变换是对称的。 | A |
| 人工神经网络（Artificial Neural Network， ANN）是 | 根据计算机电子线路标准化后得到的网络 | 无机物生长形成的网络 | 有机晶体形成的网络 | 受构成动物大脑的生物神经网络启发而来的计算系统。 | D |
| 以下哪个陈述是对的： | 对于隐藏单元，tanh 激活通常比 sigmoid 激活函数更有效，因为它的输出平均值接近于零，因此它可以更好地为下一层将数据居中化 | 对于输出单元，tanh 激活通常比 sigmoid 激活函数更有效，因为它的输出最大值是1，因此可以作为分类的概率 | 对于隐藏单元， sigmoid 激活通常比 tanh 激活函数更有效，因为它的输出接近于分类的概率 | 对于输入单元，tanh 激活通常比 sigmoid 激活函数更有效, 因为它的中间值等于零, 正好和做了数据标准化的输入数据吻合 | A |
| 多层感知器的目标损失函数选择: | 回归问题和分类问题都可以用平方误差损失函数 | 回归问题和分类问题都可以用交叉熵损失函数 | 回归问题可以用 tanh 函数, 分类问题可以用 RELU 函数 | 回归问题可以用平方误差损失函数, 分类问题可以用交叉熵损失函数 | D |
| 在神经网络模型的训练过程中， | 损失函数的数值逐渐减增加，可以通过画损失曲线来直观的感受。 | 损失函数的数值逐渐减小，可以通过画损失曲线来直观的感受。 | 激活函数的数值逐渐减增加，可以通过画激活曲线来直观的感受。 | 目标函数的数值逐渐减小，可以通过画梯度曲线来直观的感受。 | B |
| 不包括在keras.datasets数据包里的数据集是： | CIFAR10小图像分类数据集 | IMDB影评情绪分类数据集 | MNIST手写数字数据集 | 纽约时报新闻热线主题分类数据集 | D |
| 以下哪个陈述是不正确的： | 神经网络中有大量的非线性变换，通过梯度下降法最小化损失函数，可以收敛到许多不同的局部极小值 | 利用随机扰动避开局部极小值和鞍点，同时试图收敛到全局极小值 (或尽量接近全局极小值 ) | 一般情况下，梯度下降中的损失函数是通过对训练集单个样本的损失求导数得到的 | 用损失函数在单个 (或少部分) 训练样本上的值做梯度下降可以逃离局部极小点 | C |
| 只有一个观察结果时不能用贝叶斯推理，因为更新先验需要大量数据 | 对 | 错 |  |  | B |
| 函数f\_0(𝐱)的最大化可以通过最小化函数的负值来实现：maximize{f\_0(𝐱)} = minimize{−f\_0(𝐱)} | 对 | 错 |  |  | A |
| 文本处理是从文本数据中提取可以计算的信息 | 对 | 错 |  |  | A |
| 在 𝑛 维空间中，超平面的数学表达式是一个线性表达式 | 对 | 错 |  |  | A |
| 深度学习是机器学习的一种，“深度”的意思来自于人工神经网络可以有多个隐含层。 | 对 | 错 |  |  | A |
| 以下哪个陈述是对的： | 机器学习将训练数据特征和标签之间的关系泛化到了更普遍的情况 | 机器学习将训练数据特征和标签之间的关系简化到了更特殊的情况 | 机器学习将测试数据特征和标签之间的关系泛化到了更特殊的情况 | 机器学习将测试数据特征和标签之间的关系简化到了更特殊的情况 | A |
| 贝叶斯统计推断按顺序排三个步骤是： | 似然、先验、后验 | 先验、似然、后验 | 先验、后验、似然 | 后验、似然、先验 | B |
| 以下哪个陈述是对的： | 没有约束的优化问题就叫做非线性优化，非线性优化通常比线性优化要难很多 | 没有约束的优化问题就叫做线性优化，线性优化通常比非线性优化要难很多 | 没有约束的优化问题就叫做无约束优化，无约束优化通常比有约束优化要容易很多 | 没有约束的优化问题就叫做无约束优化，无约束优化通常比有约束优化要难很多 | C |
| 如下哪个陈述是不对的： | 凸优化的局部最优不一定是全局最优 | 凸优化的局部最优就是全局最优 | 用一个更宽松但凸的约束代替每个非凸约束，松弛后的问题的最优值为非凸问题的全局最优值提供了一个下界 | 找到原始非凸目标的凸近似公式，通过求解该近似问题，可以得到一个比较好的初始起点 | A |
| 以下哪个陈述是对的： | 机器学习模型的预测误差中的方差描述的是预测值的协方差 | 机器学习模型的预测误差中的方差描述的是预测值的发散程度 | 机器学习模型的预测误差与模型只和方差有关 | 机器学习模型的预测误差与模型的偏差和方差都无关 | B |
| 如下哪个陈述是不对的： | K-近邻是一个非常简单的监督学习算法，它原数据点都有它对应的标签。 | K-近邻在执行对数据集的查询前，需要事先通过对训练数据集进行训练来构建模型 | K-近邻根据它的邻居的数据点的组别来决定新数据点是属于哪个组，这样就可以做分类预测 | K-近邻可以做回归问题，根据它属于那个组的平均值来决定它这个新数据点的数值 | B |
| 高斯朴素贝叶斯模型拟合时, 只需找到 | 每个标签内样本点的特征平均值和标准差 | 每个标签内样本点的特征平均值、标准差、相关系数 | 每个标签内样本点的特征平均值、标准差、偏度（skewness) | 每个标签内样本点的特征平均值、标准差、峰度（kurtosis) | A |
| 如下哪个陈述是不对的： | 逻辑回归问题中，我们的对象标签只有类，没有一个数值可以让我们计算均方误差 | 逻辑回归模型把特征线性加权组合后，权重是模型参数，通过一个sigmoid函数转换成一个类似于概率的数值 | 逻辑回归输出概率大于阈值的可以认为倾向于一个肯定的结果，如果小于阈值的，是一个倾向于否定的结果 | 逻辑回归的对象标签是一个概率，所以要用sigmoid函数产生一个类似于概率的输出数值 | D |
| 标准化缩放是通过 | 去除最小值并缩放到单位方差来标准化特征 | 去除最大值并缩放到单位方差来标准化特征 | 去除平均值并缩放到单位方差来标准化特征 | 去除单位方差并缩放到平均值来标准化特征 | C |
| 图像识别的应用不包括： | 指纹识别 | 瞳孔身份识别 | 声纹识别 | 自动驾驶中对行人的识别 | C |
| 构造决策树模型的关键点是决定怎么根据 | 分杈来进行特征选取，分杈的目的是使得特征比以前更好地标准化。 | 分杈来进行特征选取，分杈的目的是使得特征比以前更少。 | 特征来进行分杈，分杈的目的是使得样本点分杈以后比以前更好地分成了不同的类。 | 特征来进行分杈，分杈的目的是使得样本点分杈以前比以后更好地分成了不同的类。 | C |
| 总共有三类，概率（频率）分别是0、0.5、0.5, 计算熵的结果是： | 负无穷 | 1 | -1 | 正无穷 | B |
| 回归树中目标值具有连续性， | 因此基尼指数、信息增益都可以作为拆分的标准 | 因此基尼指数不是合适的拆分标准、只有信息增益合适 | 因此基尼指数、信息增益都不是合适的拆分标准，需要用方差来拆分 | 因此基尼指数、信息增益都不是合适的拆分标准，需要用交叉熵来拆分 | C |
| 以下哪个陈述是对的： | 为了分离两类数据点，可以选择许多可能的超平面，支持向量机是找其中距离数据点最近的平面 | 逻辑回归不关心实例是否接近决策边界, 因此它选择的决策边界可能不是最优的 | 支持向量机不关心实例是否接近决策边界，它只关心离边界最远的的数据点 | 为了分离两类数据点, 逻辑回归和支持向量机用的都是同样的分隔边界条件 | B |
| 以下哪个陈述是对的： | 逻辑回归和支持向量机的成本（损失）函数具有相同的极限行为，区别发生在中间区域 | 逻辑回归和支持向量机的成本（损失）函数具有不同的极限行为，中间区域的表现类似 | 逻辑回归和支持向量机的成本（损失）函数没有可比性，因为一个是回归，一个是分类 | 逻辑回归和支持向量机的成本（损失）函数类似，因为够可以做回归模型 | A |
| 关于SVM中的核函数，以下哪个陈述时对的： | 用核函数的点积定义一个核函数，相当于隐式地定义了非线性映射函数𝜙(𝐱) | 用核函数的点积定义一个核函数，相当于显式地定义了非线性映射函数𝜙(𝐱) | 用核函数的线性叠加定义一个核函数，相当于隐式地定义了非线性映射函数𝜙(𝐱) | 用核函数的线性叠加定义一个核函数，相当于显式地定义了非线性映射函数𝜙(𝐱) | A |
| 训练集总共有两个样本点(1, 1)和(-1,-1), 一个线性SVM分类器的最优分割超平面方程是： | y+x=0 | y+x = 1 | y-x=0 | y-x = -1 | A |
| 以下哪个陈述是对的： | 数据压缩利用的是核函数的降维功能 | K-均值聚类也被称为降维方法 | 主成分分析（PCA) 和奇异值分解（SVD）依赖于线性代数中的概念，将矩阵分解为更易于理解和更具信息的部分 | 数据压缩假设大量数据是冗余的，它把数据中显示大部分方差的部分舍弃掉 | C |
| 以下哪个陈述是不对的： | SVM模型的训练结果只由支持向量决定，使得它有较好的鲁棒性 | SVM模型利用核函数把非线性问题转换成高位空间的线性问题 | 当数据量比较大时 (特征数大，样本数大）SVM算法比较耗时, 解决方法是先降维再做SVM学习 | 当数据量比较大时 (特征数大，样本数大）SVM算法比较耗时, 解决方法是先做SVM学习再降维 | D |
| 以下哪个不属于神经网络的节点： | 输入节点 | 输出节点 | 神经单元 | 关键节点 | D |
| 以下哪个陈述是不对的： | 激活函数Sigmoid的函数值范围是0到1 | Tanh的函数值范围是-1到1 | tanh⁡(𝑧) = 2×𝑠𝑖𝑔𝑚𝑜𝑑(2𝑧) −1 | Tanh的函数值范围是0到正无穷 | D |
| 关于神经网络的损失函数，以下哪个陈述是错的： | 回归问题可以平方误差作为损失函数 | 分类问题可以用交叉熵作为损失函数 | 单样本的平方误差 = (y - yb)(y - yb)，其中y是模型输出值，yb是实际值 | 单样本的交叉熵= y log(yb)， 其中y是模型输出值，yb是实际标签 | D |
| 以下哪个陈述是对的： | 损失函数值下降的速度太快说明学习速率可能太小。 | 损失函数值的逐渐变平时说明模型已经确定得差不多了，可以停止训练。 | 可以看损失函数值下降的速度来大致判断神经网络参数 W 和 b 是不是太大或太小。 | 可以从损失函数值的逐渐变平的趋势中看出是不是模型已经优化不动了，需要重新开始训练。 | B |
| 用神经网络做的分类模型，需要用softmax激活函数来转换成概率的层是： | 输入层 | 输出层 | 隐含层 | dropout层 | B |
| 以下哪个陈述是不正确的： | 神经网络中有大量的非线性变换，通过梯度下降法最小化损失函数，可以收敛到许多不同的局部极小值 | 利用随机扰动避开局部极小值和鞍点，同时试图收敛到全局极小值 (或尽量接近全局极小值 ) | 一般情况下，梯度下降中的损失函数是通过对训练集单个样本的损失求导数得到的 | 用损失函数在单个 (或少部分) 训练样本上的值做梯度下降可以逃离局部极小点 | C |