

第二章测试

题量: 6 满分: 100

作答时间: 09-12 11:42 至 09-13 12:42

智能分析

100分

一. 单选题 (33.2分)

1

2

二. 多选题 (16.7分)

3

三. 判断题 (50.1分)

4

5

6

一. 单选题 (共2题, 33.2分)

1. (单选题)在机器学习中, 查全率是指什么?

- A. 正确预测为正例的样本占有所有实际为正例的样本的比例
- B. 正确预测为反例的样本占有所有实际为反例的样本的比例
- C. 正确预测为正例的样本占有所有预测为正例的样本的比例
- D. 正确预测为反例的样本占有所有预测为反例的样本的比例

我的答案: A 正确答案: A



16.6 分

答案解析: 查看课件中查全率的定义

2. (单选题)在机器学习中, 哪些因素可能导致模型过拟合?

- A. 模型太简单
- B. 训练数据太少
- C. 模型太复杂
- D. 所有以上

我的答案: C 正确答案: C



16.6 分

二. 多选题 (共1题, 16.7分)

3. (多选题)以下哪个是机器学习中的性能度量指标?

- A. 精度
- B. 召回率
- C. F1值
- D. 均方误差

我的答案: ABCD 正确答案: ABCD



16.7 分

答案解析: 第二章课件中有关于性能度量的部分, 针对回归和分类任务, 有不同的度量指标。

三. 判断题 (共3题, 50.1分)

4. (判断题)在机器学习中, 增加模型的复杂度总是能提高模型的性能。

- A. 对
- B. 错

作业详情

答案解析：错误。过度增加模型复杂度可能导致过拟合。

5. (判断题)
k折交叉验证法的折数k越大越好。

- A. 对
- B. 错

我的答案: 错 正确答案: 错  16.7 分

答案解析：错误。过多的折数会增加计算成本。

6. (判断题)在机器学习中，过拟合只能通过增加数据来解决。

- A. 对
- B. 错

我的答案: 错 正确答案: 错  16.7 分

答案解析：错误。还可以通过正则化方法来解决。

一. 单选题 (33.2分)

- 1
- 2

二. 多选题 (16.7分)

- 3

三. 判断题 (50.1分)

- 4
- 5
- 6

3-6章练习题

题量: 7 满分: 100

作答时间: 11-07 11:43 至 11-07 23:43

智能分析

100分

一. 单选题 (100分)

1

2

3

6

7

一. 单选题 (共7题, 100分)

1. (单选题)假设我们使用软间隔SVM作为优化目标函数, 我们需要做什么来保证得到的模型与硬间隔SVM作为优化目标函数得到的模型基本一样?

- A. 惩罚系数 $C=0$
- B. 惩罚系数 $C=1$
- C. 惩罚系数 C 正无穷大
- D. 惩罚系数 C 负无穷大

我的答案: C 正确答案: C



14.2 分

2. (单选题)对决策树进行剪枝处理的主要目的是什么?

- A. 避免欠拟合
- B. 避免过拟合, 提升泛化能力
- C. 提高对训练集的学习能力
- D. 避免过拟合, 降低泛化能力

我的答案: B 正确答案: B



14.3 分

3. (单选题)线性回归能完成的任务是 ()。

- A. 分类
- B. 预测离散值
- C. 预测连续值
- D. 聚类

我的答案: C 正确答案: C



14.3 分

4. (单选题)在SVM中, margin的含义是 ()。

- A. 损失误差
- B. 幅度
- C. 差额
- D. 间隔

作业详情

一. 单选题 (100分)

1

2

3

6

7

5. (单选题)神经网络算法有时会出现过拟合的情况，那么采取以下哪些方法解决过拟合更为可行（）。

- A. 设置一个正则化项
- B. 减少训练数据集中数据的数量
- C. 增大学习的步长
- D. 为参数选取多组初始值，分别训练，再选取一组作为最优值

我的答案: A 正确答案: A



14.3 分

6. (单选题)决策树学习的关键是（）。

- A. 初始结点选择
- B. 选择最优划分属性
- C. 分枝
- D. 剪枝

我的答案: B 正确答案: B



14.3 分

7. (单选题)关于决策树结点划分指标描述正确的是（）。

- A. 类别非纯度越大越好
- B. 信息增益越大越好
- C. 信息增益率越小越好
- D. 基尼指数越大越好

我的答案: B 正确答案: B



14.3 分

第1-4章测试-2024

题量： 19 满分： 100.0

智能分析

考试时间： 2024-12-09 15:30 至 2024-12-25 23:48

一. 判断题 (共 9 题, 45.0 分)

1. (判断题, 5.0 分)

实际应用中, “一个模型肯定比另一个模型具有更强的泛化能力”的这种情况是不存在的。

- A. 对
- B. 错

我的答案: 对 正确答案: 对

知识点:

2. (判断题, 5.0 分)

奥卡姆剃刀原则: 即“若有多个假设与观察一致, 选最简单的一个”。

- A. 对
- B. 错

我的答案: 对 正确答案: 对

知识点:

3. (判断题, 5.0 分)

基尼值 $Gini(D)$ 越小, 当前样本集 D 纯度越高。

- A. 对
- B. 错

我的答案: 对 正确答案: 对

知识点:

4. (判断题, 5.0 分)

线性判别分析的目标是使得同类样本尽可能分散和异类样本尽可能接近。

- A. 对
- B. 错

我的答案: 错 正确答案: 错

知识点:

5. (判断题, 5.0 分)

决策树学习算法中, 属性 a 的信息增益越大, 则使用该属性进行划分所获得的“纯度提升”越小。

- A. 对

一. 判断题 (45.0 分)

- 1 2 3
- 6 7 8

二. 单选题 (20.0 分)

- 10 11 12

三. 多选题 (10.0 分)

- 14

四. 填空题 (25.0 分)

- 15 16 17

返回

考试详情

一. 判断题 (45.0 分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 单选题 (20.0 分)

- 10
- 11
- 12

三. 多选题 (10.0 分)

- 14

四. 填空题 (25.0 分)

- 15
- 16
- 17

我的答案: 错 正确答案: 错
知识点:

6. (判断题, 5.0 分)
线性模型学得的参数 ω 直观地表达了各属性在预测中的重要性，因此，该模型具有较好的可解释性。

A. 对

B. 错

我的答案: 对 正确答案: 对
知识点:

7. (判断题, 5.0 分) 多变量决策树的学习过程中，是试图建每个非叶节点是一个形如 $w_i \cdot a_i = t$ 的线性分类器。

A. 对

B. 错

我的答案: 错 正确答案: 错
知识点:

8. (判断题, 5.0 分) 预剪枝训练时间的开销比后剪枝训练时间开销大很多。

A. 对

B. 错

我的答案: 错 正确答案: 错
知识点:

9. (判断题, 5.0 分) 训练数据较少时更容易发生欠拟合

A. 对

B. 错

我的答案: 错 正确答案: 错
知识点:

二. 单选题 (共 4 题，20.0 分)

10. (单选题, 5.0 分)
对西瓜的成熟度进行预测得到结果为0.72，这属于（）学习任务。

A. 分类

B. 回归

C. 聚类

D. 以上都不是

我的答案:B 正确答案: B

返回

考试详情

一. 判断题 (45.0 分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 单选题 (20.0 分)

- 10
- 11
- 12

三. 多选题 (10.0 分)

- 14

四. 填空题 (25.0 分)

- 15
- 16
- 17

11. (单选题, 5.0 分)
决策树基于树结构来进行预测，根结点对应了 ()

- A. 样本全集
- B. 一个属性测试
- C. 一个决策结果
- D. 部分样本全集

我的答案:A 正确答案: A
知识点:

12. (单选题, 5.0 分) 模型的泛化能力是指()。

- A. 适用于测试集样本的能力
- B. 适用于验证集样本的能力
- C. 适用于训练集样本的能力
- D. 适用于新样本的能力

我的答案:D 正确答案: D
知识点:

13. (单选题, 5.0 分) 决策树算法的泛化性能提升,则验证集精度()

- A. 降低
- B. 提高
- C. 降为零
- D. 不变

我的答案:B 正确答案: B
知识点:

三. 多选题 (共 1 题, 10.0 分)

14. (多选题, 10.0 分) 【多选题】解决类别不平衡的方法包括()

- A. 欠采样
- B. 过采样
- C. 去除正例样本
- D. 阈值移动

我的答案:ABD 正确答案: ABD
知识点:

四. 填空题 (共 5 题, 25.0 分)

返回

考试详情

一. 判断题 (45.0 分)

- 1
- 2
- 3
- 6
- 7
- 8

二. 单选题 (20.0 分)

- 10
- 11
- 12

三. 多选题 (10.0 分)

- 14

四. 填空题 (25.0 分)

- 15
- 16
- 17

我的答案:

(1) 泛化误差

正确答案:

(1) 泛化误差

知识点:

16. (填空题, 5.0 分) 保留类别比例的采样方式通常称为“分层采样”，例如，对D进行分层采样而获得含70%样本的训练集S和含30%样本的测试集T，若D中包含400个正例，300个反例，则分层采样得到测试集T应包含反例__ (1) __个。

我的答案:

(1) 90

正确答案:

(1) 90

知识点:

17. (填空题, 5.0 分) 基于均方误差最小化来进行模型求解的方法，称为：__ (1) __，就是试图找到一条直线，使得所有样本到直线上的欧式距离之和最小。

我的答案:

(1) 最小二乘法

正确答案:

(1) 最小二乘法

知识点:

18. (填空题, 5.0 分) 剪枝是解决__ (1) __拟合的主要手段。

我的答案:

(1) 过

正确答案:

(1) 过

知识点:

19. (填空题, 5.0 分) (__ 1 __)准则对可取值数目较多的属性有所偏好。

我的答案:

(1) 信息增益

正确答案:

(1) 信息增益

知识点:

第5-7章测试-2024

题量： 16 满分： 100.0

考试时间： 2024-12-09 15:30 至 2024-12-25 23:55

智能分析

待批阅

一. 填空题 (共 1 题, 10.0 分)

1. (填空题, 10.0 分) 理想激活函数是阶跃函数, 0表示抑制神经元, 1表示激活神经元, 但具有____、____等不好的性质。

我的答案:

- (1) 不连续
(2) 不光滑

正确答案:

- (1) 不连续;不光滑
(2) 不光滑;不连续

知识点:

二. 判断题 (共 6 题, 30.0 分)

2. (判断题, 5.0 分) 估计类条件概率 $P(x|C)$ 的常用策略先假定其具有某种确定的概率分布形式, 再基于训练样本对概率分布参数估计。

- A. 对
B. 错

我的答案: 对

正确答案: 对

知识点:

3. (判断题, 5.0 分) 对属性条件独立性假设进行一定程度的放松, 属于半朴素贝叶斯分类器。

- A. 对
B. 错

我的答案: 对

正确答案: 对

知识点:

4. (判断题, 5.0 分)
Boosting算法中, 学习器是顺序生成。

- A. 对
B. 错

我的答案: 对

正确答案: 对

知识点:

一. 填空题 (10.0 分)

1

二. 判断题 (30.0 分)

2

3

4

7

三. 单选题 (25.0 分)

8

9

10

四. 多选题 (15.0 分)

13

14

15

五. 计算题 (20.0 分)

16

返回

考试详情

- A. 对
- B. 错

我的答案: 对 正确答案: 对

答案解析: 课本, P148, 欲使用贝叶斯判定准则来最小化决策风险, 首先要获得后验概率 $P(C|x)$, 即对每个样本必, 选择能使后验概率 $P(c|x)$ 最大的类别标记

于是, 最小化分类错误率的贝叶斯最优分类器为

$$h^*(x) = \arg \max_{c \in \mathcal{Y}} P(c | x), \quad (7.6)$$

即对每个样本 x , 选择能使后验概率 $P(c | x)$ 最大的类别标记.

知识点:

6. (判断题, 5.0 分) Sigmoid函数常用于神经网络中的输出层, 用于将输出映射到0到1之间。

- A. 对
- B. 错

我的答案: 对 正确答案: 对

答案解析: 看函数图像。

知识点:

7. (判断题, 5.0 分) 在多层前馈神经网络中, 同一层神经元之间存在全互联。

- A. 对
- B. 错

我的答案: 错 正确答案: 错

知识点:

三. 单选题 (共 5 题, 25.0 分)

8. (单选题, 5.0 分)

BP神经网络由于其强大的表示能力, 经常遭遇 () 问题, 即训练误差持续下降, 但测试误差却可能上升。

- A. 不收敛
- B. 梯度消失
- C. 过拟合
- D. 欠拟合

我的答案:C 正确答案: C

知识点:

9. (单选题, 5.0 分) 贝叶斯最优分类器是指对每个样本 x , 选择能使后验概率 $P(C|x)$ () 的类别标记。

一. 填空题 (10.0 分)

1

二. 判断题 (30.0 分)

2

3

4

7

三. 单选题 (25.0 分)

8

9

10

四. 多选题 (15.0 分)

13

14

15

五. 计算题 (20.0 分)

16

 返回

考试详情

- B. 与先验概率相同
- C. 小于先验概率
- D. 最大

我的答案:D

正确答案: D

知识点:

10. (单选题, 5.0 分) 用有向无环图刻画属性间的依赖关系的是 () 。

- A. 半朴素贝叶斯分类器
- B. 朴素贝叶斯分类器
- C. 贝叶斯网
- D. AODE

我的答案:C

正确答案: C

知识点:

11. (单选题, 5.0 分)
BP算法的主要目标是:

- A.
最小化训练时间
- B.
最大化输出层神经元的输出
- C.
最小化训练集误差
- D.
增加网络层数

我的答案:C

正确答案: C

知识点:

12. (单选题, 5.0 分)
下面哪个不是BP算法的核心步骤？

- A.
前向传播计算输出
- B.
根据误差反向传播更新权
- C.

一. 填空题 (10.0 分)

1

二. 判断题 (30.0 分)

2

3

4

7

三. 单选题 (25.0 分)

8

9

10

四. 多选题 (15.0 分)

13

14

15

五. 计算题 (20.0 分)

16

返回

考试详情

D. 根据梯度下降调整权重

我的答案:C

正确答案: C

知识点:

四. 多选题 (共 3 题, 15.0 分)

13. (多选题, 5.0 分) 具有最大间隔的划分超平面的优点是 ()

- A. 唯一
- B. 易构造
- C. 鲁棒
- D. 泛化能力强

我的答案:CD

正确答案: CD

知识点:

14. (多选题, 5.0 分) CNN包括下面几层:

- A. 输入层
- B. 卷积层
- C. 采样层
- D. 连接层

我的答案:BCD

正确答案: BCD

知识点:

15. (多选题, 5.0 分)
以下哪些是支持向量机的优点?

- A. 高维空间中依然有效
- B. 稀疏解, 有效减少模型复杂度
- C. 训练时间较快
- D. 能处理非线性问题

我的答案:ABD

正确答案: ABD

知识点:

一. 填空题 (10.0 分)

1

二. 判断题 (30.0 分)

2 3 4 7

三. 单选题 (25.0 分)

8 9 10

四. 多选题 (15.0 分)

13 14 15

五. 计算题 (20.0 分)

16

返回

考试详情

16. (计算题, 20.0 分)

贝叶斯决策计算题:

识别班级中是男生c1还是女生c2。训练样本集中有100个学生, 男生5人 (类别为c1), 女生95人 (类别为c2)。

现有一待识别的学生, 长发的特征指标为x, 从类条件概率表上查得: $P(x|c1) = 0.2$, $P(x|c2) = 0.8$ 。

(1) 试给出最小化分类错误率的贝叶斯最优分类器表达式和朴素贝叶斯最优分类器的表达式。

(2) 使用贝叶斯决策对该学生进行男女性别分类。

我的答案:

$$h^*(x) = \arg \max_{c \in Y} P(c|x), \quad h_{nb}(x) = \arg \max_{c \in Y} P(c) \prod_{i=1}^d P(x_i|c)$$

先验概率: $P(c_1) = 0.05, P(c_2) = 0.95$

全概率公式: $P(x) = P(c_1) \times P(x|c_1) + P(c_2) \times P(x|c_2) = 0.01 + 0.76 = 0.77$

贝叶斯公式: $P(c_1|x) = \frac{P(c_1)P(x|c_1)}{P(x)} = \frac{0.05 \times 0.2}{0.77} = 0.013$

后验概率: $P(c_2|x) = \frac{P(c_2)P(x|c_2)}{P(x)} = \frac{0.95 \times 0.8}{0.77} = 0.987$

正确答案:

(1) $h^*(x) = \arg \max_{c \in Y} P(c|x)$ 、 $h_{nb}(x) = \arg \max_{c \in Y} P(c) \prod_{i=1}^d P(x_i|c)$

(2) 先验概率: 男生的先验概率为 $P(c1) = 0.05$, 女生的先验概率为 $P(c2) = 0.95$

全概率公式: $P(x) = P(c1) \times P(x|c1) + P(c2) \times P(x|c2) = 0.01 + 0.76 = 0.77$

贝叶斯公式: $P(c_i|x) = \frac{P(c_i)P(x|c_i)}{P(x)}$

后验概率: $P(c1|x) = P(c1) \times P(x|c1) / P(x) = 0.05 \times 0.2 / 0.77 = 0.013$

$P(c2|x) = P(c2) \times P(x|c2) / P(x) = 0.95 \times 0.8 = 0.987$

问题分析: 在已知长发的特征指标情况下, 若学生留长发, x为男生的概率1.3%, x为女生的概率98.7%

知识点:

一. 填空题 (10.0 分)

1

二. 判断题 (30.0 分)

2

3

4

7

三. 单选题 (25.0 分)

8

9

10

四. 多选题 (15.0 分)

13

14

15

五. 计算题 (20.0 分)

16