

24机器学习-试题回忆

编者	Egopposer (line2345)
日期	2024/12/13

0、前言及试题概览

为帮助软院同学正确考量考试难度，合理规划复习时间，预防应付考试复习冗余或者复习不足的情况，依照24年考试内容，制定本参考版试题回忆，帮助同学更好的制定复习计划。

斜体为记忆模糊的题目，不保证可信度，**黑体**为我认为需要关注的点。

试题概览：（主观评价较多，仅供娱乐）

题量	特大（取决于计算器质量）
难易度（软院专业课中）	8+1（计算量）/10
送分题占比	20%
背诵记忆占比	5%（几乎没有硬背内容，都是需要理解才能写出来的）
21-23均分	
21-23平均满绩率	

1、选择题（2*15=30）

质量很高的选择题，完全不是选词填空的硬背就行的情况。对问题的理解和各章节的综合理解有要求，也有送分题

忘得差不多了.....

- 下列算法中不能直接对图像数据进行分类的是（）
 - CART
 - SVM
 - logistic
 - LDA
- 对下列线性样本使用“留一法”进行数据集划分，得到的均方误差为（）

x	y
1	0
3	1
2	2

- 预测股市价格走向属于下列哪类任务 ()
 - 回归
 - 分类
 - 表征
- 给定软间隔支持向量机定义，当约束变量 ξ_i 大于零小于一时，该样本点位于 ()
- 当 (且仅当?) 一个向量为支持向量时，下列选项成立的是 () (选项仅供参考，可能就没有正确的，正确的我记得好像是约束生效的形式变了一下? 忘了)
 - $1 - y_i(\omega x_i + b) = 0$
 - $\lambda_i \neq 0$
 - $\lambda_i y_i x_i = 0$
 - $x_i(\hat{y}_i - y_i) \geq 0$
- 给定PCA的结果图像，指出其中的第一主成分 ()
- 好的聚类结果具有的特征是 ()
 - 簇内相似度高，簇间相似度低
 - 簇内相似度低，簇间相似度低
 - 簇内相似度高，簇间相似度高
 - 簇内相似度低，簇间相似度高

2、判断题 (2*10=20)

和选择一样，30%左右的题对理解有一定要求

- 逻辑回归算法只能做分类任务
- 机器学习中的经验 (E) 指的是数据
- 信息增益率比信息增益更不偏好样本类别少的数据
- 剪枝是解决决策树过拟合的有效手段
- 深度学习不能解决回归问题
- 反向传播算法中应用了链式求导法则

3、简答题 (4+4+6+6=20)

和软院其他简答题大不相同了，不需要怎么背

- 画出M-P神经元的工作图，并简述其工作流程 (4分)
- 关于优化目标 $\max f(x)$ 和约束条件 $g(x) - n(x + k) \geq C$,乱七八糟的一坨 \leq 另一坨.写出其拉格朗日函数 (6分)
- (填空) Logistic的推导过程 (6分)
 - $p_0 = P(y_i = 0) =$
 - $p_1 = P(y_i = 1) =$
 - 化简到最简，优化目标为 \min
- 简述PCA算法的工作流程 (4分)

4、综合题 (6+7+8+10)

卡西欧991cnx拼尽全力无法战胜

- 给定下列数据集，初始化3和5作为聚类的初始中心点，使用kmeans算法，给出Lloyd算法前两次迭代的中心点和聚类结果 (6分)

编号	花萼长度	花萼宽度	花瓣长度	花瓣宽度	种类
1					A
2					A
3					A
4					A
5					B
6					B
7					B
8					B

初见端倪， 8×4 的样本，还非整数
你不会指望我把数据集给你背下来吧(˘•ω•˘)

- 使用上一题的数据集，计算其LDA中的类间散度矩阵和类内散度矩阵（7分）

昨天 21:21

我 ~~算了~~ 机器学习考试按了50分钟计算器

老师上来就说“同学们写快一点哈，后面计算量可能有点大”

往后面一看我 ~~算了~~ 8×8 的非整数据集算四个协方差矩阵
 4 $4+4+2$



• 给出下列二分类结果，给出其混淆矩阵，精度，召回率，F1值，TP_r，FP_r。（8分）

样本属性	□	○	□	□	□	□	○	○	○	○
预测结果	○	□	○	□	□	○	□	○	□	○

- 对于正样例 (x_1, y_1) , 负样例 (x_2, y_2) , (x_3, y_3) , 使用SVM计算其分类超平面 (10分) **(课件原题微调, 并且直接给出了优化目标)**