

重庆某双一流大学机器学习期末真题

考试时间: 2020 年 12 月 17 日 9:00-11:00

回忆整理: Vayne Duan

写在前面:

1. 试卷总体比较简单, 闭卷, 不能带计算器(也没有什么计算量).
2. 复习全面一点, 90 分没有什么问题.
3. 本回忆版真题于 2020 年 12 月 17 日上午 11:56 写成, 刚吃完饭, 有点撑.
4. 计院专业课的试卷似乎都不准老师发出来, 希望有学弟学妹们能将我“回忆试卷”的习惯传承下去, 为之后的学弟学妹们做一点微小的贡献 $O(\cap_\cap)O$
5. 其 余 专 业 课 的 回 忆 版 试 卷 也 许 可 以 在 **github.com/VayneDuan** 找到, 记得 **star & follow!**

简答题(10 分*4)

1. 监督学习 非监督学习的区别; 举例
回归和分类的区别
2. 支持向量机的原理 和 为什么核函数有用
2. 决策树连续值怎么处理
3. 随机森林的生成过程, 随机性体现在哪

计算题(10 分*3)

1. 给定神经网络的各个参数
(1) 画出神经网络拓扑图
(2) 计算隐层的输入, 输出 和 输出层的输入, 输出
2. 贝叶斯计算
给定: $P(\text{患病})$, $P(\text{阳性} | \text{患病})$, $P(\text{阳性} | \text{正常})$
求 $P(\text{患病} | \text{阳性})$
答案很小, 10/10111
3. 给定 4 个样本(每个样本有 x 和 y)和初始中心点, 进行聚类
很简单, 两轮就完了

思考题(15 分*2)

1. 过拟合 欠拟合概念; 结合偏差和方差, 解释造成过拟合欠拟合的原因; 神经网络中怎么解决过拟合, 给出至少 2 种方法
2. 一些数据有标记, 一些数据没有标记; 设计一种方法, 能够尽可能充分利用这些数据.

其实就是半监督, 可以去网上查一下相关算法