简答题：(40分)

1. 数据预处理四个流程+邻近性度量方法
2. 过拟合、欠拟合+分类评估指标有哪些
3. 决策树与基于规则分类的区别与联系
4. 聚类、回归、分类的区别和联系

计算题：(60分)

1. Apriori算法+找出所有规则中正相关的规则(lift>1)
2. K-means++算法
3. 朴素贝叶斯分类+决策树
4. 最小二乘法线性回归(里面考了计算协方差)