

兰州大学 2022 至 2023 学年第二学期 统计学习与数据挖掘

作业二

请根据老师提供的模板完成下列任务.

1. 在使用决策树进行个人信用风险评估中,我们采用了一个 UCI 数据集”德国贷款数据”.对这个数据集,我们对数据源及索要解决的问题进行了描述.并对数据进行探索和预处理,使数据能够达到程序处理的目的.然后我们对数据进行划分,分为训练集和测试集.然后进行模型训练,训练完毕再进行模型性能评估.我们采用的是使用 Scikit-learn 中的 DecisionTreeClassifier 算法来训练决策树模型.发现模型的违约贷款预测准确率并不高,于是采用了通过创建一个代价矩阵定义模型犯不同错误时的代价的方法,提升了模型的性能,违约贷款的识别提高到了 85%.

试分析另外一个 UCI 数据集,并写出实验报告.这里采用的模型请同学们自己定,例如可以采用逻辑斯蒂回归,朴素贝叶斯,各种支持向量机(线性可分,线性,非线性等等,也可以都采用,比较这些方法).

阅读报告模板请从下列网址下载:

[实验报告模板](#)

注意: 同学们请将实验报告上传到超星学习平台.