

练习3：缺陷数据集分析

(1) R介绍

(2) 下载xalan2.4数据集:

<https://zenodo.org/record/268436/files/xalan-2.4.csv>

(3) 编写一个R脚本，分析 wmc, dit, noc, cbo, rfc和 lcom的缺陷预测能力

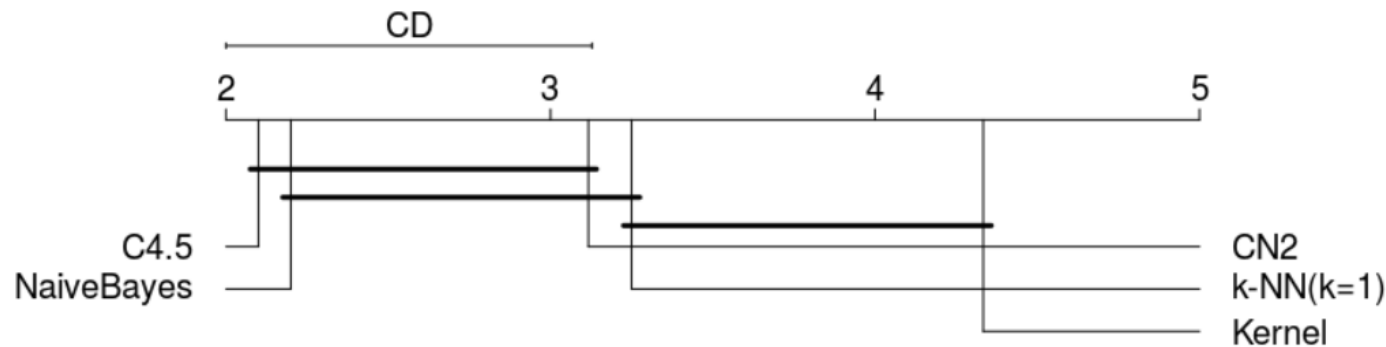
- a. 描述性统计：最小值、25%处值、中位值、75%处值、最大值、平均值、偏度(skewness)和峰度(kurtosis)
- b. 与bug数据的相关系数: Spearman和Pearson相关系数以及统计显著性



练习3：缺陷数据集分析

(3) 编写一个R脚本， ...

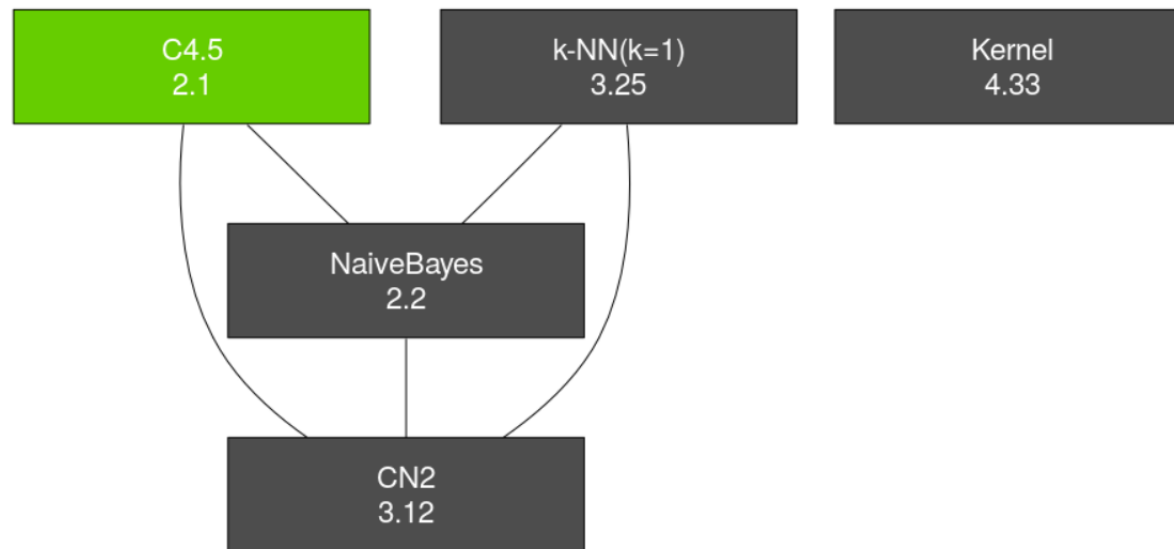
- c. 利用Naïve Bayes和logistic回归等10种机器学习方法建立多变量的缺陷预测模型，不需要进行特征选择
- d. 利用10x10交叉验证方法评价上述缺陷预测模型的性能，包括分类性能(评价指标为AUC)和排序性能(评价指标为CE)
- e. 利用CD图比较这10种模型在统计上的差别(plotCD)



练习3：缺陷数据集分析

(3) 编写一个R脚本， ...

f. 利用Algorithm图比较这10种模型在统计上的差别
(`drawAlgorithmGraph`)

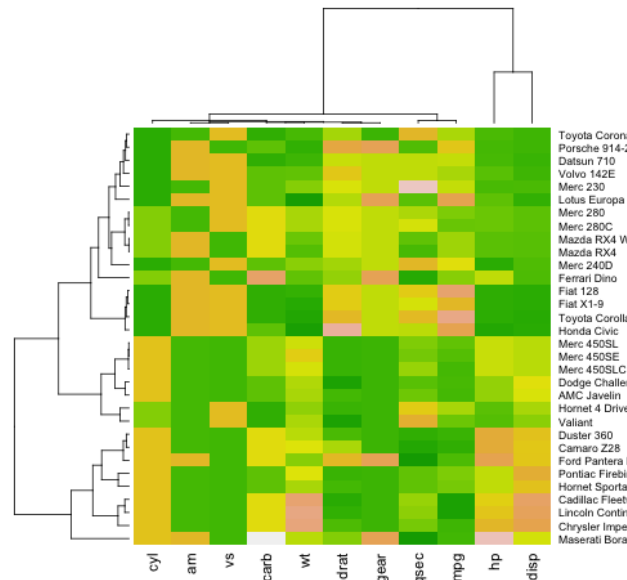


https://rdr.io/cran/scmamp/f/vignettes/Statistical_assessment_of_the_differences.Rmd

练习3：缺陷数据集分析

(3) 编写一个R脚本， ...

g. 用heatmap展示10个模型在100个测试集上的结果(行为模型，列为100个测试集上的结果)



在12月30日前提交：R脚本以及报告