

# 《计算机应用数学第一次作业-2024 秋》

截止时间：2025. 10. 31

## 计算题（50 分）

1、已知某工厂某批次水泥重量服从正态分布，总体方差为 2.65 公斤，从该工厂随机抽取 18 袋水泥，其平均重量为 24.9 公斤，试求该工厂水泥平均重量的 95%和 99%的置信区间。

（10 分）

2、从某学校抽取 20 名高一学生，经测量，这 20 名学生的平均身高为 155cm，标准差为 10cm，假设平均身高服从正态分布，试求该学校高一学生总平均身高的 95%和 99%的置信区间。（10 分）

3、1.2 节 P19，证明被抽样总体分布未知情况下的无偏估计。（10 分）

4、证明 Lemma 1.3.1。（20 分）

## 编程题（50 分，每题 25 分）

说明：

- 建议使用开源工具包，例如 scikit-learn 中有朴素贝叶斯等函数实现；
- 相关可视结果用 PDF 提交，代码可以放在 PDF 中也可以作为附件上传；
- **完整结果文件作为附件上传。所有除 PDF 外的文件都打成一个压缩包。**

### 1、PageRank

数据集：PageRank\_Dataset.csv。

数据描述：数据集中每一行是一条边<起点 ID，终点 ID>。

任务描述：使用PageRank算法计算每个节点的PageRank值，参数设置可以参考谷歌PageRank算法。

要求输出：1）结果统计，例如PageRank分数最高的20个node；2）csv文件，每一行是一个节点ID和它对应的PageRank值。

Optional：探索不同参数（ $\beta$ ）对结果的影响。

### 2、朴素贝叶斯分类器（Naive Bayes Classifier）

数据集：Bayesian\_Dataset\_train.csv, Bayesian\_Dataset\_test.csv。

数据描述：列名分别为“年纪、工作性质、家庭收入、学位、工作类型、婚姻状况、族裔、性别、工作地点”，最后一列是标签，即收入是否大于 50k 每年。

任务描述：使用朴素贝叶斯（Naïve Bayesian）预测一个人的收入是否高于 50K 每年。

要求输出：1）结果统计，例如 precision、recall、F1 score 等；2）csv 文件，在 test 文件最后增加一列，填入模型预测的收入标签（ $\leq 50K$  或  $> 50K$ ）

Optional：探索不同参数对结果的影响。