

张凯

教育经历

2019.09-2020.08 康奈尔大学 生物统计与数据科学 硕士

- 成绩排名: Overall GPA: 3.98/4.0
- 核心课程: 机器学习、生存分析、因果推断、层次模型与纵向数据分析、药学统计

2015.09-2019.06 香港中文大学(深圳) 统计学 本科一等荣誉学位

- 成绩排名: Overall GPA: 3.40/4.0 (专业 7/71) ; Major GPA: 3.66/4.0 (专业 3/71)
- 奖学金: 2018-2019 学年本科生科研津贴; 2017-2018 学年理工学院校长嘉许名单 (学院 15%学生获此殊荣) ; 2017-2018 学年逸夫书院学业奖学金 (书院 10%学生获此殊荣)
- 核心课程: 回归分析、数据挖掘、数据分析导论、数据管理、统计推断、时间序列、随机过程、生存分析、最优化、金融数据建模、计量经济学、非参数统计、调查抽样等
- 线上课程: 深度学习系列 (Coursera 与 DataCamp) 、结构化语言系列 (DataCamp)



实习与研究经历

威尔康奈尔医学研究院-研究生研究助理

2020.01-至今

基于老鼠视网膜单细胞 RNA 序列的聚类分析与细胞类型识别 (指导老师: Kathy Zhou 教授)

- 对液滴测序后的单细胞 RNA 数据进行预处理, 具体包括利用基于 Dirichlet-multinomial 模型的方法筛选细胞, 利用离群值检测过滤低质量细胞, 运用去卷积正则化方法排除非生物学差异, 运用基于泊松分布的方法进行特征选择, 并通过主成分分析 (PCA) 进行降维。
- 运用多种算法对细胞进行聚类分析并比较其差异, 其中包括 K-means、分层聚类、DBSCAN、SNN 图聚类、K-means 与图聚类的结合算法、基于重采样的序列集成聚类算法。
- 通过配对 t 检验与 BH 校正发现标记基因, 并通过分类算法识别细胞类别。

基于老年人衰老因素的中介效应分析 (指导老师: Arindam RoyChoudhury 教授)

- 运用基于累计比数模型 (Proportional Odds) 与传统 Baron-Kenny 四步过程的 Sobel 检验, 探究身体成分、体适能和力量等衰老因素之间的中介效应。

深圳市大数据研究院-实习数据分析师

2018.09-2019.06

校园范围的轨迹网络分析研究项目 (指导老师: 贾建民教授, 周健君博士)

- 运用多线程方法, 查询 MongoDB 中的原始 WiFi 连接数据并计算港中深学生的个人轨迹网络, 输出轨迹网络矩阵
- 运用 NetworkX 工具包可视化学生的轨迹网络并计算相关网络属性, 如度中心性、簇系数与时间熵。
- 通过 Lasso 回归与 XGBoost 发现影响学生 GPA 的网络属性与对应的行为属性。

基于行为数据的 GPA 预测模型研究项目 (指导老师: 周健君博士)

- 从原始 WiFi 连接数据中提取 1710 名港中深本科生的行为属性, 包括在校园各建筑 (如图书馆、寝室、运动娱乐设施等) 的月停留时间, 与基于时空数据重合度的交际圈估计。
- 基于行为数据, 运用决策树模型、随机森林算法和 XGBoost 算法预测潜在的 GPA 下降的学生, 同时运用决策树模型回溯分类结果以备后续优化。最终模型敏感度达到 81.8%同时总准确率达到 96.3%。

长江证券风险管理部-信用风险实习生

2018.07-2018.08

- 对 2014-2017 年的上市制造业公司的财务数据进行数据预处理, 包括缺失值填充, 基于条件推断的连续变量分离化, 基于随机森林算法的特征选择, 和 PCA 降维。
- 运用 WOE 转换与逻辑回归方法搭建信用风险评分卡模型。

奖项证书与技能兴趣

- 奖项证书: 香港中文大学(深圳) 校长嘉许名单 (2018-2019) 、理工学院奖学金 (2018-2019)
- 技能与语言能力: R (tidyverse, ggplot2, shiny)、Python (pandas, scikit-learn, keras)、SQL
- 英语 (GRE 324 V: 154/Q: 174)