1. 网络爬虫原理及实例(抓取新闻及小说)
2. TF, TF-IDF 文本向量化
3. PageRank 网页排序算法**(TextRank 关键词提取)**
4. Scikit-learn 及xgboost 机器学习工具包使用
5. 参数估计（**最小二乘估计、极大似然估计**）
6. 分类回归算法

Knn

贝叶斯分类

LR 逻辑回归**(梯度下降，随机梯度下降)**

决策树(回归树)

SVM

1. 聚类算法

K-means 聚类

DBSCAN 聚类

谱聚类

1. 组合学习**(bagging, boosting)**

Adaboost 算法

随机森林

GBDT

1. 主题模型

PLSA 及EM

LDA

1. CTR 广告点击率预测
2. 新闻标签提取**（word2vec）**
3. 相似新闻文章挖掘
4. 气象降水量预测**(天池竞赛项目)**
5. Netflix 电影推荐系统