0813

1. 微积分和线性代数
2. 基本的Python能力（廖雪峰的Python教程，到面向对象编程一节即可）
3. 基本的Python数据处理能力（pandas和numpy）

pandas官方文档

Github上的pandas练习项目：<https://www.kesci.com>/home/project/5a8afe517f2d695222327e14

numpy官方文档

numpy菜鸟教程

常用机器学习算法：

* 线性模型原理和实现
* 集成学习模型（随机森林、GBDT、xgboost、lightgbm）原理和实现
* 深度学习模型、Tensorflow的使用
* 特征选择、模型调参与模型融合
* 不包含支持向量机（算法复杂度较高，在比较大的数据集上很难运行）

比赛平台：天池、Kaggle

腾讯广告赛、京东赛 – 数据量非常大、结果比较真实。

微信公众号：麻婆豆腐AI

Github：https://iphysresearch.github.io/DataSciComp?sub=PF,AC,DM

比赛解决方案汇总Github：<https://github.com/Smilexuhc/Data-Competition-TopSolution>

机器学习的目标：以机器算法替代人工劳动，以自动地完成预测或决策。

机器学习：从大量的数据中学习规律，需要依赖于统计学工具。大部分机器学习问题都可以转化为预测问题。

计算机程序的逻辑仍然由人类编写和实现，但机器学习中的规律是由算法从大量数据中自动学习的。

线性回归：性能上的优越性、良好的可解释性。

系数b1：在其他条件相同的条件下，x1每增加1，y每增加b1。

练习1、使用sklearn估计孩子身高和父母身高之间的线性回归关系。

男女分开预测、正则化后再预测

计算机程序设计课程：https://cn.udacity.com/course/design-of-computer-programs--cs212