绪论

假设具有以下能力：

Python主要的库的基础，如Numpy, Pandas和Matplotlib

本科数学知识（微积分、线性代数、概率和统计学）

熟悉Jupyter

学习框架：

第一部分：ML基础

1. ML可以解决哪些问题？主要类别和基本概念
2. 典型ML项目的主要步骤
3. 学习用模型拟合数据
4. 优化代价函数
5. 处理、清洗和准备数据
6. 选择和特征工程
7. 选择模型并使用交叉验证调整超参数
8. ML的主要挑战，尤其是欠拟合和过拟合（偏差/方差权衡）
9. 数据降维
10. 最常见算法：

线性回归，多项式回归, LR, KNN, SVM, DT, RF, EM

第二部分：NN和DL

1. NN的优势
2. 用TF构建和训练NN
3. 最重要的NN结构：前向NN，卷积网络，循环网络，LSTM网络，自动编码器
4. 训练DNN的技术
5. 缩放大数据集NN
6. 强化学习

第一部分主要用Scikit-Learn，第二部分主要用TensorFlow.

假设拥有足够的数据，计算力和耐心，DL适合用于图像识别，语音识别和NLP.

第一章：

ML用于：

1. 需要大量手动调整或规则的问题
2. 传统方法得不到好的解决方案的问题
3. 波动环境下的问题
4. 复杂问题和海量数据问题

属性attribute：数据类型，特征feature：数据类型和数据值；通常不区分二者。

练习：

1. 如何定义ML?
2. 能说出四种问题吗？
3. 什么是带变迁的训练集？
4. 最常见的两个监督任务是什么？
5. 说出四个常见的无监督任务？
6. 答：增强学习
7. 什么算法需要将数据分为多个组别？ 答：在线学习
8. 垃圾邮件检测问题会作为有监督/无监督问题来解？答：有监督
9. 什么是在线学习系统？
10. 什么是包外学习（out-of-core）？
11. 什么算法依赖于相似度测量做预测？ 答：基于实例的算法
12. 模型参数和学习算法的超参数的区别？答：模型参数通过训练得到，超参数通过多次实验人工调整确定
13. 基于模型的学习算法在求什么？采用哪种最常见的策略达到目标？怎么做出预测？答：求模型的决策边界，常使用交叉验证来判断模型的好坏，通过学得的模型边界对新数据进行判断，在边界内为一类，否则在另一类
14. 说出ML中的四个挑战？
15. 如果训练出色，泛化很差。