

史上最萌最认真的机器学习/深度学习/模式识别入门指导手册(一)

夕小瑶 夕小瑶的卖萌屋 2017-02-11

喵喵喵~大家元宵节快乐噢。有没有要陪小夕出去看烟花的吖...

小夕借此给热爱学习的喵喵们献上这篇拙文，希望不要嫌弃哦~

还有，小夕画的封面图是不是很棒呀(￣▽￣)

小夕发现现在想进军人工智能领域的程序yuan甚至少年少女喵好多呀，但是有几只向我吐槽过网上的资料很多很杂，入门指导基本是杂而不实，很容易让人满腔热情的开始，却又一脸懵逼的放弃。于是小夕来拯救世界啦!~

小夕的车辙

小夕从大二开始做人脸识别，后来发现很快遇到了瓶颈。于是开始学机器学习理论，又很快遇到了瓶颈，于是又开始补数学。补完数学又钻模式识别，又被虐了，但还是辛苦的啃完了《模式分类》。后来学深度学习。再后来不停的论文论文论文、代码代码代码...

由于小夕是从工程出发然后到理论，再回工程的道路，小夕觉得这样走的弯路挺大的。而且前期做工程的时候真的很懵逼，各种看不懂然后跳过。

所以小夕这里分享给大家的道路是从理论到工程的平坦上升的道路。当然啦，理论中肯定会穿插代码实践。小夕希望这是一条靠谱的、没有知识断层的深度学习/机器学习的入门之路，希望能在真正意义上帮到大家。

前言

本指导适合于真正有志于钻研机器学习(含深度学习)、模式识别及其相关应用领域的人。对于那种“1个月入门机器学习”的大忽悠学习模式，小夕这里没有噢~

理论上说，完全小夕的整个入门指导后，应对整个机器学习大框架、理论细节、工程能力都有了比较好的积累。

在此之后可以偏向工程，轻松玩TensorFlow、Caffe等DL框架，转CV、NLP等应用性更强的方向；也可以偏向理论，比较轻松的看看最新的paper，跟上学术界的最前沿。但是工程与理论不是绝对割裂的哦，偏工程也要看论文、补数学，偏理论也要敲代码、了解优秀框架哦。

首先，在开始之前，要确保有以下的基础：

- 语文能力最少达到初中水平。
- 英语能力最少要达到高中优秀水平，尽量在大学四级水平之上。
- 数学最少达到高中水平，微积分一定得学过。

尽量有以下的基础(这样可以边学边练，体会更深的喵，但是不会也没有关系的)

- 面向过程编程范式
- 常见数据结构
- 常见算法思想及算法分析能力

然后小夕为了避免文章过长，将计划截成了四五个阶段。本文为第一阶段。

阶段一

线性代数

- 前置课程
 - 中学代数
- 主参考资料
 - 《线性代数应该这样学》(英文叫《Linear Algebra done right》)
- 辅助参考资料(有先后顺序)
 1. 《Deep Learning》Bengio等，第二章(中译本勉强能看，链接<https://github.com/exacity/deeplearningbook-chinese> 呜呜好想赶紧开通原创功能插超链接)；
 2. Wiki百科(翻墙不用教吧...)；
 3. 《矩阵分析与应用》张贤达
- 重点内容(无先后顺序)：
 - 向量及向量空间
 - 内积与范数
 - 线性映射
 - 矩阵
 - 张成、线性相关、线性无关
 - 特征值与特征向量
 - 特征分解
- 高级内容(最起码要了解)：
 - 谱定理
 - 奇异值分解(SVD)
 - 矩阵的迹
 - 行列式
- 学习方法
 1. 适当参考小夕总结的重点内容，细细的品味《线性代数应该这样学》(这本书真的棒呆了)。
 2. 对于书中依然理解不了的部分，参考其他辅助资料哦。另外如果大家有哪方面难以理解，可以告诉小夕，小夕会尽量解答，若有必要的话直接写一篇小文章帮助大家理解哦。
- 主要意义

线性代数是机器学习的不能更基础的数学基础。不仅仅是因为矩阵是机器学习中运算的基本单位，而且一些线性代数中的高级理论也被借鉴吸收到了机器学习算法中，比如用SVD(奇异值分解)来对特征降维，迹运算可以加深对PCA及某些聚类算法本质的理解等。

Matlab

- 前置课程
 - 线性代数
 - 面向过程编程范式
- 主参考资料：
 - Coursera平台斯坦福大学Andrew Ng的“机器学习”课程的Matlab章节的视频及补充资料
 - Matlab内嵌帮助文档
- 辅助参考资料

1. Google
 2. CSDN
 3. StackOverFlow
- 重点内容：
 - Matlab基本语法
 - 矩阵运算常用的内置API(即用户编程接口/函数接口)
 - 学习方法
 1. 参考Coursera平台“机器学习”课程的Matlab教学视频及阅读材料来掌握Matlab基础知识。
 2. 养成使用Matlab内嵌帮助文档来学习新API的方法的习惯。例如在解释器中可以用“help +命令”或“doc +命令”快速掌握某API的用法等。
 3. 多多总结积累常用的矩阵运算API。
 - 主要意义

利用Matlab入门和理解机器学习算法,可以很大限度的忽略编程语言语法特性对算法实现的影响,也就是说语法对数学表示的还原度很高~这样可以很轻松的打通数学到计算机算法的桥梁,对于以后深入学习机器学习的理论和代码实现都大有裨益。

机器学习-上

- 前置课程
 - 线性代数
 - 微积分
 - Matlab
- 主参考资料：
 - Coursera平台斯坦福大学Andrew Ng的“机器学习”课程
 - 《数据挖掘导论》
- 重点内容：
 - 机器学习基本概念及应用领域
 - 回归与分类的概念
 - 线性回归模型(Linear regression model)
 - 逻辑回归模型(Logistic regression model)
 - 浅层神经网络(Neural Network)
 - 支持向量机(Support Vector Machine)
 - 交叉验证思想及用途(Cross Validation)
 - 聚类的概念
 - K-Means模型
- 学习方法
 1. 乖乖的跟着Coursera课程计划来就好,要认真完成课后习题和编程题哦。若课程中的有些内容实在没有理解透,强烈建议参考《数据挖掘导论》中的相关章节呐(这本书对于机器学习部分的讲解很好喵)。
小夕注: Matlab代码实现中多多体会和建立“用矩阵运算”解决问题的思想,努力摆脱“循环套循环”的low做法。
 2. Coursera结束后,仔细阅读《数据挖掘导论》中的第4、5章。先跳过其中的决策树和贝叶斯分类器那一节(以后的计划中会安排的哦),另外第五章的组合方法能看懂就看,看不懂也没关系,放在以后还会深入学习。
- 主要意义

Coursera上的“机器学习”课程中几乎没有涉及到概率统计的知识,自然也没有贝叶斯分类器这个经典统计模型。同时也没有深入讲解最优化的知识。

不过这样极大的降低了课程的难度,非常适合拿来快速接触机器学习又不会受很大打击。对于提升兴趣,以及后续理论深入的学习真的很有帮助呀,小夕当年就是看了这个视频然后爱上了可爱的Andrew Ng,然后

爱上了这一领域\(/▽/)\。

而《数据挖掘导论》也非常适合新手入门，讲解的十分浅显易懂，在Coursera课程结束后再根据这个复习一遍，效果会非常好的。

下阶段预告

1. 概率论与数理统计
2. 最优化算法-上
3. 机器学习-下

本文重点

记得推荐给身边有需要的人哦，ta与小夕都会很感激您的~

...要哭哭了，最近都快没人打赏了...对小夕的文章不够满意，记得在订阅号主页给小夕提意见哦~小夕真的很用心哦(。ゝ。)



微信支付



Transfer to 夕小瑶

没有吃的了..>_<..



文章已于修改

声明：pdf仅供学习使用，一切版权归原创公众号所有；建议持续关注原创公众号获取最新文章，学习愉快！