

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(三)

夕小瑶 夕小瑶的卖萌屋 2017-02-20

能够坚持走到阶段三的喵一定是很有毅力的喵!也是很合格的机器学习/人工智能研究者啦~但是走到这里,也意味着您不满足于使用看似孤立的机器学习模型来生硬的完成一些任务,那么,大一统的理论体系就在这一阶段建立吧~

前言

完成这一阶段的喵,在很多人眼里已经是小牛了。其实完成这一阶段后确实可以无压力的阅读大部分机器学习/模式识别/深度学习及相关领域的论文了,甚至可以动动小想法发表小论文啦(只要英语别太差)。但是在小夕看来,开始扔掉大牛的书籍,开始啃论文才是研究的开始哦~

发布上个阶段二时,有可爱的读者听到小夕说有4个阶段后表示压力很大,好像小夕这个入门指导太残忍了一些,2333~其实真的没有啦,静下心来,坚实的走好每一步,会发现路边都是风景哦。

还有人问小夕,小夕进行到阶段几了。小夕说过早就完成入门了呀,2333~加油哦,小夕在介里等你哦

阶段三

python

- 前置课程
 - 线性代数
 - 最优化算法
- 主参考资料
 - 《a byte of Python》
 - 《NumPy Tutorial》
 - 《SciPy Tutorial》
- 辅助参考资料
 - 《Python核心编程》
 - CSDN等各种中文技术博客
- 重点内容:
 - python基本语法
 - numpy
 - scipy
 - 线性代数(linear algrbra)
 - 优化(optimization)
- 高级内容(学嗨了就顺便学了,没学嗨留着以后现学现用)
 - scipy其他内容,比如统计、图像处理等
 - matplotlib

- scikit-learn
- libsvm
- 学习方法
 1. 关于基础语法：
 - a. 对于已经有C/C++/Java中的一种或几种编程语言基础的老手,相信不用小夕多啰嗦啦~一本《a byte of python》就足够掌握python基础语法啦,记得只有100页左右,一两天就可以学完。对了,这本书有中文版哦,翻译的还可以,喵喵一搜就能搜到啦。
 - b. 对于编程基础很薄弱的同学,可以忽略《a byte of python》,建议根着《python核心编程》的前十四章来,将基础语法掌握牢固呐。
 - c. 再啰嗦一句,其实入门最快的方式还是多看多写代码,《a byte of python》有附带源代码,看完《a byte of python》或者《python核心编程》后,再将这些源代码看一遍一定会有不少收获咯~
 2. 关于各种扩展库
 - a. 对于英语还不错的同学,强烈建议看官方各种库的tutorial,讲的很清晰,又简洁。
 - b. 对于英语实在不忍直视的同学,只能找各种中文博客啦,一搜xx库入门指导,肯定出来很多打小牛的技术博客。。虽然质量参差不齐,但是拿来入门勉强可以的。
- 主要意义

python在科学计算上的优势使得其无论在学术界还是工业界都非常流行,虽然一般不会用python开发最终的产品,但是对于新算法、机器学习系统的快速实现是任何编译型语言都比不来的。机器学习的主流编程语言不得不学哦。

机器学习工程实践

- 前置课程
 - python
 - 机器学习
- 主参考资料
 - 《机器学习实战》
- 重点内容：
 - k近邻
 - 决策树
 - 朴素贝叶斯
 - logistic回归
 - svm
 - adaboost
 - 线性回归
 - k-means
- 高级内容（学嗨了顺便学了，这些轮子没有前面的轮子重要）
 - PCA
 - SVD

- 学习方法

在保证机器学习理论已经掌握的情况下，参考《机器学习实战》，对各个典型的机器学习模型进行代码实现。最低要求建立代码实现与数学理论的联系，代码能力强者可以尝试独立复现主流机器学习模型及其在应用场景的优化。

- 主要意义

小夕经常看到有人对这本书的书评是“重复造轮子的无意义书籍”。小夕对于这种说法好不想解释呀，很多“经验丰富”的程序员转行做机器学习，却认为机器学习跟学安卓/iOS开发一样，把API用熟练就算入门了，这种想法大错特错呀。机器学习是一门理论与工程仍旧在不断突破的学科，而不是一件成熟的工业化产品。哪怕是在工业界，不懂轮子构造，只会用别人造好的轮子来实现机器学习系统的话，那这样的工程师可能还不如xx语言工程师有价值。等AI浪潮一突破或者一冷静，这些类库小王子必然是首当其冲的。所以用python深刻的造轮子是长久发展的不可或缺的一环哦。

模式识别与深度学习-上

- 前置课程

- 线性代数
- 概率与统计
- 最优化算法
- 机器学习

- 主要参考资料

- 《模式分类》
- 《DEEP LEARNING》

- 重点内容：

1. 贝叶斯决策
2. 参数估计
3. 非参数方法
4. 线性判别函数
5. 浅层神经网络
 - delta方法
 - BP算法及其优化
 - RBF网络
6. 深度神经网络
 - Hopfield了解
 - 玻尔兹曼机（了解
 - RBM了解
 - DBN了解
 - DBM了解
 - CNN
 - Autoencoder
 - RNN
 - LSTM
7. 聚类
 - 高斯混合密度

- K-means
- 层次聚类
- 8. 决策树与随机森林
- 9. 特征提取与特征选择
- 学习方法
 1. 这门课学习时一定要一定一定要记得自己多总结，梳理框架。不仅要深刻理解公式细节，更重要的是从全局出发，从更上层的视角来将机器学习模型串到一个大框架下。
 2. 重点内容的1-5和7-9根据《模式分类》来即可。
 3. 重点内容5中的SVM在模式分类中讲解很糙，在之前的机器学习中已经掌握SVM的可以跳过，否则建议再理解一下。
 4. 重点内容的6根据《Deep Learning》对应章节来即可。
 5. 重点内容的8、9也讲的很糙（我们的ppt讲的很细，但是老师不让分享ppt呀QAQ），如果看书觉得不够透彻的话，可以求助wiki或者小夕。
- 主要意义

这门课是小夕整个计划的最重点的一门课。这么课的重点和规划也是按照我们学校的《模式识别》课程来的，学完之后应该很扎实的具备了机器学习/模式识别的理论啦。

虽然网上经常把《模式分类》的难度吹的上天，其实只要按小夕说的，前置课程掌握好的话，这本书读起来非常享受~真的那么难的话，怎么可能作为我们的教材呢。不过话说这本书的小错误真的很多！小夕以后写一篇文章来把《模式分类》中发现的小错误与大家分享一下。



下阶段预告

1. 最优化算法-下
2. 模式识别与深度学习-下

加油哦，快来陪小夕并肩作战吧。

蟹蟹你o(≧v≦)o



微信支付



Transfer to 夕小瑶

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！