

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(二)

夕小瑶 夕小瑶的卖萌屋 2017-02-14

谢谢你们。小夕昨晚又熬夜被发现了，谢谢对小夕的关心...
有你们陪着，小夕很幸运。

路人乙：“小夕，别人都说上完coursera上的那门机器学习课就已经入门机器学习了，可你又不这样认为。那你认为上完那门课，或者说学完上一篇的阶段一，处于什么水平呢？”

小夕：“大忽悠水平。”

前言

2333,小夕没有开玩笑哦,coursera那门课是为了帮助大家从整体上了解一下机器学习大框架,并培养兴趣。而仅仅完成阶段一的话,其实对企业而言几乎是毫无价值的,啃论文也会寸步难行的呢。

那么完成阶段二以后是处于什么水平呢?当然是真正的机器学习入门水平啦~

只要您认真学、认真总结了,相信此时配合您的编程能力,就可以尝试自己实现机器学习模型啦。调用别人写好的API,也会胸有成竹啦,而不是将其看作黑匣子了。

对于应用场景来说,完成该阶段就可以转行去做数据挖掘(DM)了,但要做计算机视觉(CV)、自然语言处理(NLP)中的模式识别问题的话,还不够哦~

所以说,小夕将东西放在这里咯,剩下的就靠各聪明的你们用探索欲去完成啦~

还有哦,在正式开始阶段二之前,小夕强烈建议您将阶段一中的matlab代码实现的机器学习算法再重新看一遍呐。不要光看你写的!更重要的是看老师给写好的大框架!甚至可以打开某个matlab内置的库函数感受一下!

阶段二

概率与统计

- 前置课程
 - 微积分
- 主参考资料
 - 《概率论与数理统计》陈希孺（注意不是浙大的那本！）
- 重点内容：
 - 整本书！
- 学习方法

这本书写的超棒！虽然学校的概率统计用的浙大那本教材，但是学完也有好多地方似懂非懂。直到在图书馆无意间遇到了这本书。。。所以认真读咯，是不是一想到小夕也读过这本书，就迫不及待想开始了呢【捂脸】

- 主要意义

这门课程不需要谈意义了吧 (ノ ◡ ノ) 这门课都没有掌握，那只能处于计划一的大忽悠水平咯~

最优化算法-上

- 前置课程

- 微积分（高等数学）
- 线性代数

- 主参考资料

- 《Deep Learning》第四章（中文版链接见手册（一））
- 《Numerical Optimization》Jorge Nocedal等

- 辅助参考资料

《最优化理论与方法》袁亚湘，孙文瑜（这本书已绝版，但某宝有卖复刻版；在学校同学可去图书馆借，没收藏这本书的大学应该可以取消数学和计算机专业了吧）

- 重点内容：

- 一阶无约束优化算法
 - 梯度下降法（简单了解步长的确定方法）
- 二阶无约束优化算法
 - 牛顿法
- 约束优化算法
 - 线性规划概念与应用
 - 二次规划概念与应用
 - 拉格朗日乘子法的简单认识

- 高级内容（依照自身数学基础，尽可能深的理解）

- 一阶无约束优化算法
 - 梯度下降法（仅掌握线搜索法，学嗨了可以看信赖域法）
- 二阶无约束优化算法
 - 共轭梯度法
 - 拟牛顿法
- 约束优化算法
 - 线性规划（仅掌握单纯形法，学嗨了可以看内点法）
 - 二次规划（仅掌握对偶法，学嗨了可以看积极集法）

- 学习方法

1. 小夕考虑到最优化算法对机器学习而言虽然至关重要，但是对数学基础要求很高。因此在本计划中采用个性化定制的方式：
 - a. 如果您的数学基础很好，强烈建议您尽可能的完成高级内容，这对后面机器学习算法的透彻理解极其重要。
 - b. 如果您的数学基础不够，只需完成重点内容即可。但是希望在业余再加深一下对微积分、线性代数等知识的理解哦~方便以后突破瓶颈呐。
2. 对于重点内容，只需要认真研究理解《Deep Learning》中的4.3节和4.4节，这两节信息量很大，请务必认真阅读每一句话。如果这两节都感到寸步难行的话，请补习最优化的前置课程哦。
3. 对于高级内容，《Numerical Optimazation》是极其合适的，这本书很偏工程实践，讲了很多practical的问题。也是我们学校最优化课的教材。这本书貌似没有中文版，不过相信您的数学基础都那么好啦，看英文资料也没有问题哒~

- 主要意义

小夕在指导(一)中提到的机器学习瓶颈就是指的这门课！

小夕第一次学机器学习时，以为各个机器学习模型是孤立的，有的用梯度下降，有的二次规划的，当时也不

知道，结果学完之后机器学习体系特别散。

直到上了这门课，才恍然大悟，竟然有这么一个数学体系将机器学习中的“寻最优参数”（即最优化）问题全都聚拢到一起了！

所以，这门课之于机器学习的重要性，小夕只能说重中之重呀。

机器学习-下

- 前置课程
 - 线性代数
 - 概率与统计
 - 最优化算法
- 主要参考资料
 - 《统计学习方法》李航
 - 《数据挖掘导论》Pang-Ning Tan等
- 重点内容：
 - 感知机模型
 - K近邻模型
 - 朴素贝叶斯模型
 - 决策树模型
 - 支持向量机模型
 - 集成分类器
 - Bagging
 - Boosting
 - Random Forest
- 学习方法
 1. 在确信自己的前置课程已经完成的前提下，针对重点内容的前5项，抱起《统计学习方法》怒啃吧！不要有心理障碍，当你像接受python一样接受数学语言时，你会发现这本书读起来很轻松~
 - a. 对于Logistic模型，也就是书上的第六章可以不用看。完成阶段一后应对Logistic模型有足够的了解了（只要你别偷懒！），更深入的了解会在下一阶段展开。小夕总觉得这一章讲的半浅不深的。建议跳过。
 - b. 对于决策树模型，也就是书上的第五章，需要预先学习信息论中的一点知识，起码弄懂熵、条件熵、自信息、互信息的概念。如果感觉书上的讲解不易理解，可以求助《数据挖掘导论》或者Google。
 - c. 对于SVM模型，记得初学时在98、99页的位置遇到过一个问题，感觉书上的解释不够详细，导致没理解SVM。但是小夕实在想不起来了QAQ。懂了之后再找这个点好难。所以如果你们读到这里后遇到跟小夕一样的问题，请及时求助Google或者纸质资料哦（小夕当时是看《信息检索导论》理解的这个点，这本书的第15章讲的SVM挺棒，不过你们没有文本方面的前置知识小夕不清楚会不会影响理解QAQ）
 - d. 如果依然感到某个模型的数学描述太难理解，则拿起《数据挖掘导论》找到对应的模型学习一遍，再返回来看。（你想说《数据挖掘导论》上也看不懂怎么办？咳咳，请返回计划1。）
 - e. 如果你感觉某个模型的数学描述看懂了，但是总感觉意犹未尽，好像还缺点什么，甚至会有一种小茫然，那么恭喜你，你的大脑试图在模型之间建立联系，它已经迫不及待的要开启下一轮计划了~暂时没有这种感觉也没关系，毕竟不同人的学习模式不一定一样嘛。
 2. 重点内容的前五项完成后，第六项在《数据挖掘导论》中解决（第五章第六节）。
- 主要意义

在阶段一我们已经浅浅的尝了一下整个机器学习大框架。除了了解了机器学习的基本概念、常用模型外，也对正则化技术、交叉验证思想等有了初步的认识，完成了科普阶段的积累。

但是小夕讲了，我们不做大忽悠哦~因此经过本课程勤恳的钻研，您就可以揭开机器学习的大部分面纱啦！你会发现看似神奇的学习算法，人工智能中看似神奇的学习能力，原来不过是几个数学公式+工业界的小trick嘛

~~

至此,您可以称得上机器学习的从业者了。动手欲强的人估计此时已经将书上的数学语言翻译成您喜欢的编程语言了吧(¯▽¯)

下阶段预告

1. python
2. 机器学习任务实践
3. 模式识别与深度学习-上

文末,还是想再说一声“谢谢你们(*´▽`*)”。小夕会将你们带给小夕的感动延续下去,再转交到你们手里。



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!