

北京联盟手机版

深度 | 机器学习初学者最常见的5个错误：你该怎么避开它们？

发布日期：2016-10-03 13:46



选自MLMastery[北京联盟](http://www.010lm.com/) <http://www.010lm.com/>

作者：Jason Brownlee[北京联盟](http://www.010lm.com/) <http://www.010lm.com/>

机器之心编译[北京联盟](http://www.010lm.com/) <http://www.010lm.com/>

参与：吴攀、李亚洲[北京联盟](http://www.010lm.com/) <http://www.010lm.com/>

本文作者 Jason Brownlee 是一位学术研究者、作者、专业开发者和机器学习实践者，他非常热衷于传播机器学习的学习方法和经验。机器之心还曾编译过他的文章《给默片配音、自动生成手写体，盘点 8 个最具启发意义的深度学习应用》和《提升深度学习模型的表现，你需要这 20 个技巧》。本文介绍了机器学习初学者最常出现的 5 种错误以及避免它们的方法。[北京联盟](http://www.010lm.com/) <http://www.010lm.com/>

我一直在帮助初学者入门机器学习，但我也一直看到初学者在思想和行动上屡屡犯错。[北京联盟](http://www.010lm.com/) <http://www.010lm.com/>

在这篇文章中，你将看到我总结的初学者上手机器学习时最常出现的 5 种错误。我坚定地认为任何人都可以入门机器学习并且很好地应用它们。希望这篇文章能够帮助你避开或解决你学习和实践中的一些错误。北京联盟 <http://www.010lm.com/>

下面就让我们开始吧！北京联盟 <http://www.010lm.com/>

一、不要从理论开始入手北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

传统的机器学习教学是自下而上的：北京联盟 <http://www.010lm.com/>

努力掌握数学背景知识北京联盟 <http://www.010lm.com/>

努力学习机器学习理论北京联盟 <http://www.010lm.com/>

努力从头开始实现算法北京联盟 <http://www.010lm.com/>

未知？？？（一些神奇事情发生）北京联盟 <http://www.010lm.com/>

最后开始使用机器学习（这是你的目标）北京联盟 <http://www.010lm.com/>

这种方法非常慢，而且很困难。这是为想要拓展该领域前沿的学者设计的，并不适合只是想要得到结果的实践者。北京联盟 <http://www.010lm.com/>

陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

如果你在想以下问题，你就掉入了这个陷阱：北京联盟 <http://www.010lm.com/>

我需要先完成线性代数的课程；北京联盟 <http://www.010lm.com/>

我需要回去先拿个博士学位；北京联盟 <http://www.010lm.com/>

我必须先读完教科书。北京联盟 <http://www.010lm.com/>

避开陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

4 年的数学或深奥的算法理论学习能够怎样帮助你实现目标？北京联盟 <http://www.010lm.com/>

你多半会停下来。会失败。不会离你的目标更进一步。北京联盟 <http://www.010lm.com/>

解决的方法是鼓捣模型。北京联盟 <http://www.010lm.com/>

如果机器学习对市场的有价值的贡献是准确的预测，那么你就学习对问题进行建模并得出准确的预测吧。从现在开始！北京联盟 <http://www.010lm.com/>

然后努力掌握它，做到非常擅长。北京联盟 <http://www.010lm.com/>

如果你需要，就阅读、窃取、收割别人的理论，但你只需要你的目标所需要的——只要它能帮助你实现更好的价值。北京联盟 <http://www.010lm.com/>

二、不要学习所有的机器学习北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

机器学习是一个非常大的研究领域。北京联盟 <http://www.010lm.com/>

它是指计算机学习过程的自动化，与人工智能有很大的交叠。北京联盟 <http://www.010lm.com/>

从深奥的学习理论到机器人技术——这个领域的范围非常广。你不可能将它们整个拿下。北京联盟 <http://www.010lm.com/>

陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

如果你在想下面这些，你就落套了：北京联盟 <http://www.010lm.com/>

我需要学习一个新网站上提到的每一种新技术；北京联盟 <http://www.010lm.com/>

我首先需要学习计算机视觉、自然语言处理、语音等等；北京联盟 <http://www.010lm.com/>

我需要了解每项技术的每个方面。北京联盟 <http://www.010lm.com/>

避开陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

选择一个小方向并且保持专注。北京联盟 <http://www.010lm.com/>

然后进一步收窄。北京联盟 <http://www.010lm.com/>

机器学习中最有价值的领域是预测建模——创建数据的模型以做出预测。北京联盟 <http://www.010lm.com/>

就从这里开始。北京联盟 <http://www.010lm.com/>

接下来，专注于一种与你最相关或你最感兴趣的预测建模。北京联盟 <http://www.010lm.com/>

然后保持专注。北京联盟 <http://www.010lm.com/>

也许你是根据技术选择的（比如深度学习），也许你是根据问题类型选择的（比如推荐系统）；也许你不确定，只是随便选了一个。不管怎样，努力掌握它或至少做到精通。北京联盟 <http://www.010lm.com/>

接下来才开始进入下一个领域。北京联盟 <http://www.010lm.com/>

三、不要在算法上花费太多时间北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

机器学习实际上就是算法。北京联盟 <http://www.010lm.com/>

算法实在太多了。每一个算法都是一个复杂的系统而且拥有自己的小研究领域和生态系统。北京联盟 <http://www.010lm.com/>

你可以将自己迷失在算法之中，这么做的人被称为学者。北京联盟 <http://www.010lm.com/>

陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

如果你在说这些，你就已经在陷阱里了：北京联盟 <http://www.010lm.com/>

在我使用它之前我需要知道它的工作原理；北京联盟 <http://www.010lm.com/>

我首先需要深入理解超参数；北京联盟 <http://www.010lm.com/>

在进行微调时，我需要解释其因果关系。北京联盟 <http://www.010lm.com/>

避开陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

算法并非结果。它们只是实现结果的方法。北京联盟 <http://www.010lm.com/>

事实上，机器学习算法是一种商品。北京联盟 <http://www.010lm.com/>

切换它们。在你的问题上实验各种各样的算法。你可以对其中一些进行调节，但不要把自己困在这一步。北京联盟 <http://www.010lm.com/>

使用一个系统过程（systematic process）。设计调试实验并将它们的执行和分析自动化。北京联盟 <http://www.010lm.com/>

机器学习的关键是好好利用算法，而应用算法不只是鼓捣算法本身。北京联盟 <http://www.010lm.com/>

你的目标应该是为每一个项目得出一个结果——一组预测或一个能给出预测的模型。北京联盟 <http://www.010lm.com/>

四、不要从头开始实现所有东西北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

从头开始实现算法能够教会你很多东西。北京联盟 <http://www.010lm.com/>

有时候你需要实现一项技术，因为没有合适的或可用的实现。北京联盟 <http://www.010lm.com/>

但是，通常来说，你不一定要也不应该这么做。北京联盟 <http://www.010lm.com/>

很抱歉地说：你的实现结果可能很糟糕。北京联盟 <http://www.010lm.com/>

有各种漏洞北京联盟 <http://www.010lm.com/>

速度很慢北京联盟 <http://www.010lm.com/>

非常吃内存北京联盟 <http://www.010lm.com/>

无法很好地解决边缘情况北京联盟 <http://www.010lm.com/>

甚至可能是错的北京联盟 <http://www.010lm.com/>

陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

如果你在做下面这些事，你已经掉入陷阱了：北京联盟 <http://www.010lm.com/>

你在编写载入 CSV 文件的代码（搞什么鬼！？）；北京联盟 <http://www.010lm.com/>

你在编写线性回归这种标准算法的代码；北京联盟 <http://www.010lm.com/>

你在编写用于交叉验证或超参数调整的代码。北京联盟 <http://www.010lm.com/>

避开陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

别这么做就好了！北京联盟 <http://www.010lm.com/>

使用一个已被成千上万的开发者所使用过的能够处理所有边缘情况的库，众所周知它是正确的；

北京联盟 <http://www.010lm.com/>

使用一个经过了高度优化的库，它能充分利用你的硬件的每一个计算周期和你的内存的每一个字节；

北京联盟 <http://www.010lm.com/>

为你的项目使用一个图形用户界面，从而完全避开代码。北京联盟 <http://www.010lm.com/>

每次都自己实现你想要使用的算法是非常缓慢的入门机器学习的方式。北京联盟

<http://www.010lm.com/>

如果你是为了学习而实现它们，那么就要对自己诚实：你的实现还不够好，你不能通过这样应用

机器学习来带来价值。北京联盟 <http://www.010lm.com/>

五、不要总改变使用的工具北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

目前有很多的很棒的机器学习工具。事实上，很棒的工具加上数据的可用性和更快的硬件使得机器学习的复兴成为了可能。但你可能陷入这样一个坑中：跳入每个你偶然发现的新工具中。北京

联盟 <http://www.010lm.com/>

陷阱北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

如果发现自己符合以下三点那你就陷入了该陷阱中：北京联盟 <http://www.010lm.com/>

使用每个你听到过的新工具北京联盟 <http://www.010lm.com/>

发现自己每周或每个月都学习一项新工具或语言北京联盟 <http://www.010lm.com/>

学习一个库时半途而废并转向新的库北京联盟 <http://www.010lm.com/>

避开陷阱北京联盟 <http://www.010lm.com/>

有策略的学习和使用新工具，在解决机器学习问题的系统处理过程中融合你学到的新工具。如果你选择一个主要的大型平台并坚持使用它，那在解决机器学习问题时会更加高效，至少也要足够精通该平台。北京联盟 <http://www.010lm.com/>

这里我推荐 3 个顶尖的平台：北京联盟 <http://www.010lm.com/>

Weka Workbench北京联盟 <http://www.010lm.com/>

Python Ecosystem北京联盟 <http://www.010lm.com/>

R Platform北京联盟 <http://www.010lm.com/>

该领域也有其他的更专门化的工具，根据个人兴趣和专业的不同可跟进。北京联盟 <http://www.010lm.com/>

总结北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

在该博文中，你可以看到初学者进入机器学习常见的 5 个错误。总结如下：北京联盟 <http://www.010lm.com/>

不要从理论开始入手北京联盟 <http://www.010lm.com/>

不要学习所有的机器学习北京联盟 <http://www.010lm.com/>

不要在算法上花费太多时间北京联盟 <http://www.010lm.com/>

不要从头开始实现所有东西北京联盟 <http://www.010lm.com/>

不要总改变使用的工具北京联盟 <http://www.010lm.com/>

你陷入了这些坑中了吗？北京联盟 <http://www.010lm.com/>

?本文由机器之心编译，转载请联系本公众号获得授权。北京联盟 <http://www.010lm.com/>

?-----北京联盟 <http://www.010lm.com/>

加入机器之心（全职记者/实习生）：hr@almosthuman.cn北京联盟 <http://www.010lm.com/>北京联盟 <http://www.010lm.com/>

投稿或寻求报道：editor@almosthuman.cn 北京联盟 <http://www.010lm.com/> 北京联盟 <http://www.010lm.com/>

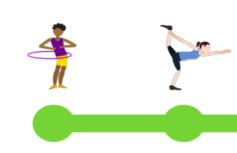
广告商务合作：bd@almosthuman.cn 北京联盟 <http://www.010lm.com/> 北京联盟 <http://www.010lm.com/>



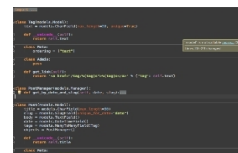
住家保姆



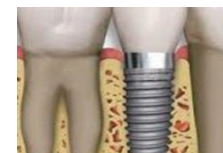
Python机器学习



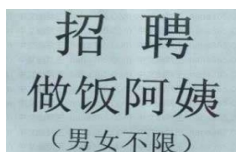
activity



python.



地包天



做饭阿姨



病痛



严重失眠




住家保姆价格




人工智能课程




精彩推荐

 男子街头摆摊 只要顾客 "粗心" 就能赚钱


男子街头摆摊 只要顾客 "粗心" 就能赚

 南方妹子遇北方大葱疯了 真是没见过还有比人高的大葱


南方妹子遇北方大葱疯了 真是没见过还

 央视曝光！你海淘的商品可能全是假货，它们都来自...


央视曝光！你海淘的商品可能全是假货，

 高校避孕套免费领 大学生:质量太差不要

高校避孕套免费领 大学生:质量太差不要

 女子发现男友与其他女性深夜共处 气急之下将其杀死

女子发现男友与其他女性深夜共处 气急

 是亲生的吗？6岁女孩遭母亲虐待 头发被扯掉一半

是亲生的吗？6岁女孩遭母亲虐待 头发被

微软赢得克里奥大奖2016“年度最佳广告主奖”

深度 | 机器学习初学者最常见的5个错误：你该怎么避开

最近3个月Apple Pay已开始支持这些银行了

硅谷一周融资速报：24 家初创公司获得总计超 5.14 亿

这些美景你没有忘了吧！

国庆文艺冲击波来袭！花香静谧、诗情画意，太棒啦！

如果世界形成中美两极格局，对中国而言将意味着什么？

全国散文名家德阳采风笔会顺利闭幕

安哥拉一航班因行李工人受困货舱迫降 无人伤亡

恢宏不羁的豪放派赏宋词①

PC版

Copyright © 2012-2016 北京联盟 版权所有