如何研究学习一个机器学习算法

机器学习算法都是一个个复杂的体系，需要通过研究来理解。学习算法的静态描述是一个好的开始，但是这并不足以使我们理解算法的行为，我们需要在动态中来理解算法。

机器学习算法的运行实验，会使你对于不同类型问题得出的实验结论，并对实验结论与算法参数两者的因果关系有一个直观认识。

在这篇文章中，你将会知道怎么研究学习一个机器学习算法。你将会学到5个简单步骤，你可以用来设计和完成你的第一个机器学习算法实验

你会发现机器学习实验不光是学者们的专利，你也可以；你也会知道实验是通往精通的必经之路，因为你可以从经验中学到因果关系的知识， 这是其它地方学不到的。

什么是研究机器学习算法

当研究一个机器学习算法的时候，你的目标是找到可得到好结果的机器算法行为，这些结果是可以推广到多个问题或者多个类型的问题上。

你通过对算法状态做系统研究来研究学习机器学习算法。这项工作通过设计和运行可控实验来完成

一旦你完成了一项实验，你可以对结论作出解释和提交。这些结论会让你得以管窥在算法变化中因果关系。这就是算法行为和你获得的结论间的关系。

怎样研究学习机器学习算法

在这一部分，我们将学到5个简单的步骤，你可以通过它来研究学习一个机器算法

1.选择一个算法

选择一个你有疑问的算法

这个算法可能是你正在某个问题上应用的，或者你发现在其他环境中表现很好，将来你想使用

就实验的意图来说，使用现成的算法是有帮助的。这会给你一个底线：存在bug几率最低

自己实现一个算法可能是了解算法过程的一个好的方式，但是，实验期间，会引入额外的变量，比如bug，和大量必须为算法所做的微观决策

2.确定一个问题

你必须有一个你试图寻找答案的研究问题。问题越明确，问题越有用

给出的示例问题包括以下几个方面：

KNN算法中，作为样本空间中的一部分的K值在增大时有什么影响？

在SVM算法中，选择不同的核函数在二分类问题上有什么影响 ？

在二分类问题中，逻辑回归上的不同参数的缩放有什么影响 ？

在随机森林模型中，在训练集上增加任意属性对在分类准确性上有什么影响？

针对算法，设计你想回答的问题。仔细考虑，然后列出5个逐渐演变的问题，并且深入推敲那个最精确的

3.设计实验

从你的问题中挑选出关键元素然后组成你的实验内容。 例如，拿上面的示例问题为例：“二元分类问题中逻辑回归上的不同的参数缩放有什么影响？”

你从这个问题中挑出来用来设计实验的元素是：

属性缩放法：你可以采用像正态化、标准化，将某一属性提升至乘方、取对数等方法

逻辑回归：你想使用哪种已经实现的逻辑回归。

二元分类问题：存在数值属性不同的二分类问题标准。需要准备多种问题，其中一些问题的规模是相同的（像电离层），然而其他一些问题的属性有不同的缩放值（像糖尿病问题）。

性能： 类似分类准确性的模型性能分数是需要的

花时间仔细挑选你问题中的组成元素以便为你的问题给出最佳解答。

4. 进行试验并且报告你的结论

完成你的实验

如果算法是随机的，你需要多次重复实验操作并且记录一个平均数和标准偏差

如果你试图寻找在不同实验（比如带有不同的参数）之间结果的差异，你可能想要使用一种统计工具来标明差异是否统计上显著的（就像学生的t检验）

一些工具像R和scikit-learn/SciPy完成这些类型的实验，但是你需要把它们组合在一起，并且为实验写脚本。其他工具像Weka带有图形用户界面，你所使用的工具不要影响问题和你实验设计的严密

总结你的实验结论。你可能想使用图表。单独呈现结果是不够的，他们只是数字。你必须将数字和问题联系起来，并且通过你的实验设计提取出它们的意义

对实验问题来说，实验结果又暗示着什么呢？

保持怀疑的态度。你的结论上有留什么样的漏洞和局限呢。不要逃避这一部分。知道局限性和知道实验结果一样重要

5. 重复

重复操作

继续研究你选择的算法。你甚至想要重复带有不同参数或者不同的测试数据集的同一个实验

。你可能想要处理你试验中的局限性

不要只停留在一个算法上，开始建立知识体系和对算法的直觉

通过使用一些简单工具，提出好的问题，保持严谨和怀疑的态度，你对机器算法行为的理解很快就会到达世界级的水平

研究学习算法不仅仅是学者才能做的

你也可以学习研究机器学习算法。

你不需要一个很高的学位，你不需要用研究的方式训练，你也不需要成为一名学者

对每个拥有计算机和浓厚兴趣的人来说，机器学习算法的系统研究学习是开放的。事实上，如果你主修机器学习，你一定会适应机器学习算法的系统研究。知识根本不会自己出来，你需要靠自己的经验去得到

当谈论你的发现的适用性时，你需要保持怀疑和谨慎

你不一定提出独一无二的问题。通过研究一般的问题，你也将会收获很多，例如根据一些一般的标准数据集总结出一个参数的普遍影响。你保不住会发现某些具有最优方法的常例的局限性甚至反例。

行动步骤

在本篇文章中，通过可控实验你知道了研究学习机器学习算法行为的重要性。你掌握了简单的5个步骤，你可以在一个机器学习算法上设计和运行你的第一项实验

采取行动。使用你在这篇博文中学到的步骤，来完成你的第一个机器学习实验。一旦你完成了一个，甚至是很小的一个，你将会获得自信，工具、能力来完成第二个以及更多

我很乐意听到你第一个实验的消息。留下评论，分享你的结论、你的收获