

第03课：如何学习“机器学习”

学习“机器学习”这件事，我走过很长的一段弯路。有一些心得、体会和方法，在此和大家分享。

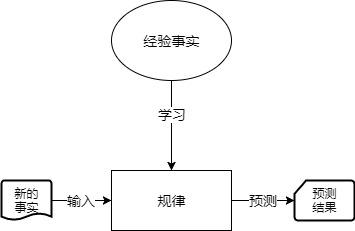
### 以模型为驱动，了解“机器学习”的本质

本课力图引导大家以模型为驱动，进行学习。

我选择了几种经典模型（例如：线性回归、逻辑回归、决策树、支持向量机、条件随机场、K 均值等），作为入门的进阶。

初次学习模型，不必贪多。但就这些选定的模型，一定要搞清楚其问题域、模型函数、目标函数、训练算法……潜入到数学公式推导的层面，对每一步的公式变换和对应的物理意义有所理解，然后再去实践。

这一个个的模型，就是机器学习的“肌肉”，我们要通过观察学习这一块块肌肉，以其为载体来了解机器学习的核心——将事物转化为数值，将关系、变换转化为运算，****以事实（数据）为依据，以规章（算法）为准绳，通过计算来掌握事物的客观规律（模型）。****



要知道，这些经典模型固然经典，但是到了实际应用中，它们并非神圣不可改变。也不是说机器学习就仅限于这些经典模型。

只有掌握了机器学习精髓的人，才可能灵活运用现有模型去解决实际问题，甚至进一步针对具体问题得出自己的模型和算法。

### 反复学习，从记忆到掌握

当然，达到这种程度并非一蹴而就，总要从最简单的模型开始。

根据我的经验，即使要真正掌握本课列出的这几个最常用且相对简单的模型，也需要反复学习。

有可能第一遍看完有点似懂非懂，或者感觉自己明白了，但要自己从头推导又卡壳了。

就像刚学完又忘掉的生字，或是背了一半后半部分怎么也想不起来的九九乘法口诀。这都是非常正常的现象。究其原因，就是还没有真正掌握。

****多学几遍****，从头到尾掌握整件事情的逻辑，知道一个模型是怎么从最初设定目标，到一步步实现目标的——当真正掌握了这个过程之后，再回头看一个个具体的模型，每一步都是顺理成章的。

掌握了事物内在的逻辑，再去记忆，就容易多了。

学习标准设置得不必太高，比如可以这样：

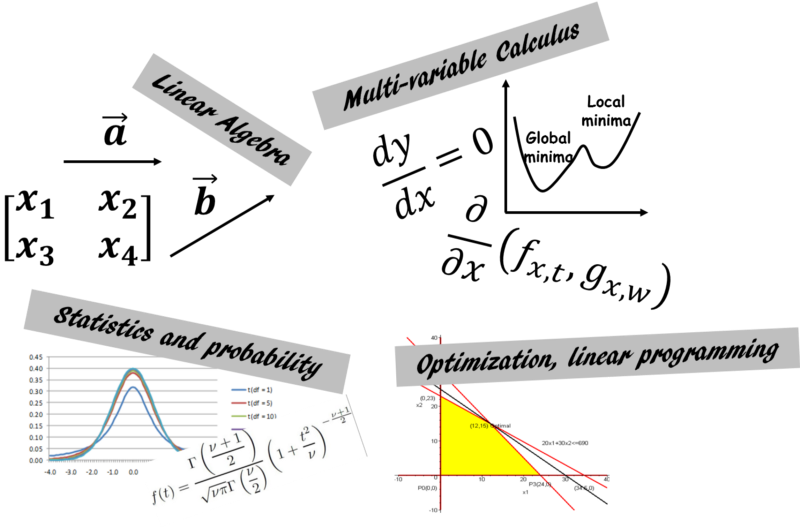
第一遍学，只要求自己能完全掌握机器学习数据、模型、算法相互作用的基本关系，训练过程和评价标准。具体的模型，掌握最简单的线性回归就可以。

只要从头到尾掌握了第一个模型，做到：能够从模型函数开始推导出目标函数，再用梯度下降算法求解，用（伪）代码实现梯度下降求解过程。

第一遍学习掌握一个模型；第二遍学习就不难掌握2-3个模型；第三遍就有可能掌握本课列出的大部分模型………如此****由易到难，螺旋式推进****。

对于一些本身就比较复杂的数学模型，比如条件随机场、隐马尔科夫这类涉及场论和势函数的模型，如果直接入手，往往会卡在模型函数本身上。但是当有了前面几个模型的基础，了解了更抽象层面的原理，掌握起来就容易多了。

### 数学需要多精深？



很多同学，想学机器学习。但是一上来就看模型，看到一大堆炫酷的公式，往往感觉很吓人。

有些人也就因此萌生退意，要么放弃，要么只用现成工具把模型当作黑盒使用。

其实，学习经典模型，并不需要多么精深的数学功底，只要掌握大学理工科本科所教授的数学知识就基本够用了。

#### 基础数学概念

在学习的最初阶段，只要：

* 学过坐标系；
* 了解函数、向量和矩阵的概念；
* 能从直观角度对求导、求微与函数的关系有所理解；
* 并掌握最简单的矩阵运算。

就可以对本课中的经典机器学习模型有一定深度的感性认识了。

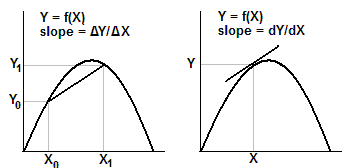
本课中的公式都是经过精简的，务必要掌握。如果数学工具实在掌握得太少，最起码也要读懂一组公式推导中的第一个和最后一个式子：

读懂它们都针对哪些变量进行了怎么样的运算；

这些变量和运算的物理意义是什么；

第一个式子是对哪种真实场景的数学描述；

最后推导的结果又具备怎样的特征。



初次学习者，可以暂且掌握到这个深度。

#### 两条路径反刍数学知识

当然，仅仅掌握到这种程度，勉强可以套用现成的模型，却不可能达到灵活运用的程度。

所以，同学们应力求理解每一步推导过程。

如果有可能，可以事先复习一下大学本科的高数（数学分析）、概率统计和线性代数。或者，在学习模型的过程中，一旦遇到了数学上的阻碍，就回头去查找相应知识。

#### 制定数学知识速查手册

很多数学问题，之所以让人头大，其实并不是真的有多难，而是符号系统比较复杂，运算繁复，或者运算所表达的物理意义多样。

很多时候造成困扰是因为想不起来这里用到什么定理，哪个公式，或者这样操作表达的含义是什么。

如果把常用的细小知识点都记录下来，按主题整理成速查手册（小字典），需要用的时候快速查找对应的知识点，这样我们的“机器学习”学习之路就顺畅不少。

下面两个是我自己制作的例子，供大家参考：

* [机器学习常用微积分知识速查手册](http://gitbook.cn/books/59ee907516fc0231837614e3/index.html)
* [机器学习常用线性代数知识速查手册](http://gitbook.cn/books/59ed598e991df70ecd5a0049/index.html)

### 日常学习 Tips

#### 关联

把新学到的东西和日常的工作、生活联系起来，进行思考。比如：将理论代入现实，不同领域间事物的类比，相似内容的对比等。

以身边的实例来检测理论。不仅能够加深理论知识的理解，而且有助于改进日常事物的处理方法。

#### 记录

准备一个笔记本，纸质版或电子版均可。有什么发现、感想、疑问、经验等等，全都记下来。

如果是对某个话题、题目有比较完整的想法，最好能够及时整理成文，至少记录下要点。

隔一段时间把笔记整理一下，把分散的点滴整理成块，一点点填充自己的“思维地图”。

#### 分享

知识技能这种东西，学了，就得“炫耀”——把学习到的新知识、理论、方法，分享给更多的人。如此一来，倒逼自己整理体系、记忆要点。这可以说是与人方便自己方便的最佳实例。

把自己的感想、体会、经验分享出来的同时，也锻炼了自己的逻辑思维能力和归纳总结能力。一举多得，何乐而不为？

以上谈到的关联、记录、分享的内容，大家都可以通过社群内的学习笔记来提交。我会挑选优秀的内容进行点评和回复。

[邀请好友一起学，获得 25% 返现奖励](https://gitbook.cn/m/mazi/columns/5bc6ac7442d7d32f50f19a98/topics/5be138a6665e8a7d734a30b1)

[IMG_260](https://gitbook.cn/m/mazi/comp/column?columnId=5bc6ac7442d7d32f50f19a98%26tag=2#catalog)