

开篇词 | 入门机器学习，已迫在眉睫

大家好，我是李烨。现就职于微软（Microsoft），曾在易安信（EMC）和太阳微系统（Sun Microsystems）任软件工程师。先后参与过聊天机器人、大数据分析平台等项目的开发。在未来的 21 天里，我将通过这个训练营课程与同学们分享机器学习的相关知识。

我们的第一个打卡知识点是“热身”，这部分需要大家学习这篇开篇词和下一篇热身课文章，阅读完文章后，大家记得通过社群里的打卡小程序链接来提交学习笔记完成打卡。对提交学习笔记有疑问的同学可以看一下我们的指导手册或咨询助教。

### 课程背景

首先，我们来看下当前机器学习领域招聘市场行情。



上面表格中所有带有“算法”、“人工智能”、“数据挖掘”、“视觉”字样的职位，都需要掌握机器学习相关知识。

在产品和服务中应用机器学习模型，已经逐步成为互联网行业的通行方法。甚至很多传统软件企业，也开始尝试应用机器学习。说得更直接些，人工智能正处在炙手可热的风口浪尖上，作为程序员不会机器学习都不好意思去找工作。

很多技术开发者迫切希望****快速进入人工智能领域****，****从事工程或者算法****等相关工作，这也是我推出这门课程的初衷。

### 课程大纲

《21 天入门机器学习》以我的达人课《机器学习极简入门课》为学习素材，通过 21 天的社群互动带领大家入门机器学习。本训练营课程共 6 大主题、 42 讲，包含 17 个核心知识点，另配有 3 个可实际操作的训练项目（及数据）供同学们上手实践。

《21 天入门机器学习》面向机器学习初学者，从机器学习、深度学习最基本的原理及学习意义入手，以模型为驱动，详解几个经典机器学习模型的原理、数学推导、训练过程和优化方法。

训练营主题与学习计划：

****开营仪式****

* 学时 1 天，1 次打卡任务。

****第一部分：绪论****

授人以鱼不如授人以渔。本部分从意义和作用出发，给出相应的学习方法和与理论配套的编程练习。

学时 2 天，1 次打卡任务。

****第二部分：基本原理****

深谙其理，才能灵活应变。本部分带大家了解什么是机器学习、机器如何自己学习，以及机器学习三要素：数据、模型、算法之间的关系。

模型是机器学习的核心，那么模型是怎么得到的呢？本部分也将讲解模型的获取（训练）和评价（验证/测试）过程，相应数据集合的划分以及具体的评价指标。

这部分知识和后面讲述的具体模型结合起来，就可以实践了！

学时 2 天，1 次打卡任务。

****第三部分：有监督学习 I****

* 抓住关键，个个击破。本部分重在详细讲解有监督学习中经典的线性回归、朴素贝叶斯、逻辑回归、决策树模型。这几个模型不仅基础、经典、常用，而且数学工具特别简单。
* 学时 4 天，4 次打卡任务。

****第四部分：有监督学习 II****

* 百尺竿头，更上一层楼。本部分主要讲述支持向量机、支持向量回归、隐马尔科夫和条件随机场模型，从支持向量机开始，数学工具的需求较之前上了一个台阶，难度明显加大。
* 学时 5 天，5 次打卡任务。

****第五部分：无监督学习****

* 无须标注，方便运行。本部分重在讲解无监督学习中的聚类、高斯混合及主成分分析等模型。训练数据无须标注，方便在各种数据上随时进行尝试，是这些模型的特征。在现实中，经常用来作为有监督的辅助手段。
* 学时 5 天，4 次打卡任务。

****第六部分：深度学习****

* 超越自我，实现蜕变。本部分重在讲解深度学习基本原理、深度学习与机器学习的关联与衔接、以及深度学习目前的应用领域，为读者下一步学习“深度学习”奠定基础。
* 学时 1 天，1 次打卡任务。

****结营仪式****

* 学时 1 天。

整个课程以经典模型为驱动，讲述每一个模型所解决的问题域，模型原理和数学推导过程。作为课程的主体，第三部分和第四部分讲解得每个模型，都附有实例和相应的 Python 代码。每个例子的数据量都非常小——这样设计就是为了让读者可以用人脑模拟计算机，根据刚刚学到的模型算法对这些极小量数据进行“模拟训练/预测”，以此来加深对模型的理解。

### 你将收获什么

****AI 技术岗位求职知识储备****

如果大家真的有意投身到人工智能领域，从事相关技术性工作，通过技术笔试、面试则是必要条件。在面试中被要求从头解释某一个机器学习模型的运行原理、推导过程和优化方法，是目前非常常见的一种测试方法。机器学习模型虽然很多，但是经典、常用的很有限。如果能学会本课程所讲解的经典模型，你将足以挑战这些面试题。

****触类旁通各大模型与算法****

各种机器学习模型的具体形式和推导过程虽然有很大差别，但在更基础的层面上有许多共性。掌握共性之后，再去学新的模型、算法，就会高效得多。虽然本课的第二部分集中描述了部分一般性共同点，但真要理解个中含义，还需要以若干具体模型为载体，从问题发源，到解决方案，再到解决方案的数学抽象，以及后续数学模型求解的全过程，来了解体味。这也就是本课以模型为驱动的出发点。

****极简版实例体验实际应用****

运用到实践中去，是我们学习一切知识的目的。机器学习本身更是一种实操性很强的技术，学习它，原本就是为了应用。反之，应用也能够促进知识的深化理解和吸收。本课虽然以原理为核心，但也同样介绍了：划分数据集、从源数据中提取特征、模型训练过程、模型的测试和评估等方法和工具。

****配套数据+代码快速实操上手****

本课程中各个实例的 Python 代码及相应数据，大家可以下载、运行、改写、参考。

### 课程寄语

我希望本课的读者在知识和技巧的掌握之外，能够将学习到的基本规律运用到日常生活中，更加理性地看待世界。

再遇到“人工智能产品”，能够根据自己的知识，去推导：How it works——

* 它背后有没有用到机器学习模型？
* 如果有的话，是有监督模型还是无监督模型？
* 是分类模型还是回归模型？
* 选取的特征是哪些？
* 如果由你来解决这个问题，有没有更好的方法？

我们自己用来判断万事万物的“观点”、“看法”、“洞察”，实际上都是我们头脑中一个个“模型”对所闻所见（输入数据）进行“预测”的结果。这些模型自身的质量，直接导致了预测结果的合理性。

从机器学习认识客观规律的过程中，我们可以知道，模型是由数据和算法决定的。对应到人脑，****数据是我们经历和见过的万事万物，而算法则是我们的思辨能力。****

作为人类，我们不必被动等待一个外来的主宰者，完全可以主动训练自己的思维模型，通过改进算法和增大数据量及数据多样性来提升模型质量——如果能在这方面给读者朋友们带来些许启发，我实在不胜荣幸。

最后，预祝每一位订阅课程的朋友，能够通过学习找到心仪的工作，如果大家有任何疑问和建议，也欢迎通过读者圈与我交流，我们共同学习，共同进步。

[邀请好友一起学，获得 25% 返现奖励](https://gitbook.cn/m/mazi/columns/5bc6ac7442d7d32f50f19a98/topics/5bc6bc0c42d7d32f50f19bc9)

[IMG_258](https://gitbook.cn/m/mazi/comp/column?columnId=5bc6ac7442d7d32f50f19a98%26tag=2#catalog)