

一、The learning problem 机器学习问题

1.1 Course Introduction 课程简介

1.2 What is Machine Learning 什么是机器学习

三个ML（机器学习的英文缩写）的关键要素：

1. 存在一个模式或者说表现可以让我们对它进行改进提高；

2. 规则并不容易那么定义；

3. 需要有数据。

1.3 Applications of Machine Learning 机器学习的应用

简述下机器学习可以应用在衣食住行育乐，包含了人类生活的方方面面，所以机器学习的应用场景很广泛很有市场。

1.4 Components of Machine Learning 机器学习的组成部分

从一个银行是否会发信用卡给用户的例子引出了机器学习可以分为哪几个部分（组件）。

1. 输入(input): $x \in X$ （代表银行所掌握的用户信息）

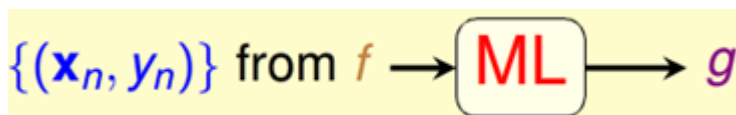
2. 输出(output): $y \in Y$ （是否会发信用卡给用户）

3. 未知的函数，即目标函数(target function): $f: X \rightarrow Y$ （理想的信用卡发放公式）

4. 数据或者叫做资料(data)，即训练样本(training examples): $D = \{(), (), \dots, ()\}$ （银行的历史记录）

5. 假设(hypothesis)，即前面提到的技能，能够具有更好地表现: $g: X \rightarrow Y$ （能够学习到的公式）

可以通过一个简单的流程图表示，如图1-1所示。

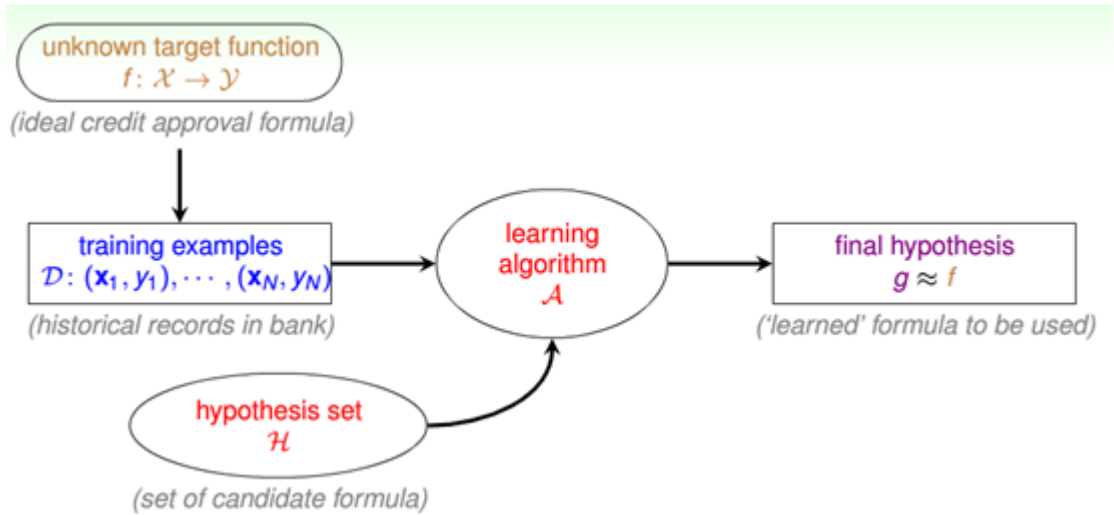


1-1 机器学习的简单流程图

从图中可以清楚机器学习就是从我们未知但是却存在的一个规则或者公式f中得到大量的数据或者说资料（训练样本），在这些资料的基础上得到一个近似于未知规则g的过程。

这么说还是有点抽象，特别是目标函数f又是未知的，那为什么还能找到一个假设g能够接近f呢？

还是以的一个更加详细的流程图来说明这一问题，如图1-2。



1-2 详细的机器学习流程图

这个流程图和图1-1有些不同，其中ML被更详细的定义为机器学习算法（learning algorithm）一般用A表示。还多出来一个新的项目，就是假设空间或者叫做 假设集合（hypothesis set）一般用H表示，它是包含各种各样的假设，其中包括好的假设和坏的假设，而这时A的作用就体现了，它可以从H这个集合中挑选出它认为最好的假设作为g。

注：

- 1、这里还要说明的是机器学习的输入在这个流程图中就变成了两个部分，一个是训练样本集，而另一个就是假设空间H。
- 2、还有一点需要注意的是，我们所说的机器学习模型在这个流程图中也不仅仅是算法A，而且还包含了假设空间H。
- 3、要求得g来近似于未知目标函数f。
- 4、给出了机器学习的一个更准确点的定义，就是通过数据来计算得到一个假设g使它接近未知目标函数。

1.5 Machine Learning and Other Fields 机器学习与其他各个领域

1.5.1 ML VS DM （Data Mining）机器学习与数据挖掘者叫知识发现（KDD Knowledge Discovery in Dataset）

上一节中已经给出了机器学习的概念，因此只介绍下数据挖掘的概念，就是从大量的数据中找出有用的信息。

从定义出发，我们可以将两者之间的关系分为3种。

1. 两者是一致的：能够找出的有用信息就是我们要求得的近似目标函数的假设。
 2. 两者是互助的：能够找出的有用信息就能帮助我们找出近似的假设，反之也可行。
 3. 传统的数据挖掘更关注与从大量的数据中的计算问题。
- 总的来时，两者密不可分。

1.5.2 M L VS AI （artificial intelligence）机器学习与人工智能。

人工智能的大概概念就是电脑能够表现出一些智慧行为。

从定义可以得到，机器学习是实现人工智能的一种方式。

1.5.3 ML VS statistic 机器学习与统计。

统计也需要通过数据，来做一个未知的推论。

因此统计是一种实现机器学习的方法。

传统的统计学习更关注与数学公式，而非计算本身。