



其实，这句话用在机器学习领域，上面的论述也是大致适用的。机器学习的学习对象是数据，数据是否有标签，就是机器学习所处的“环境”，“环境”不一样，其表现出来的“性情”也有所不同，大致可分为三类：

(1) 监督学习 (Supervised Learning) :用数据挖掘大家韩家炜 (Jiawei Han) 老师的观点来说，**监督学习基本上就是“分类 (classification)”的代名词**[1]。它从有标签的训练数据中学习，然后给定某个新数据，预测它的标签 (given data, predict labels)。这里的标签 (label)，其实就是某个事物的分类。

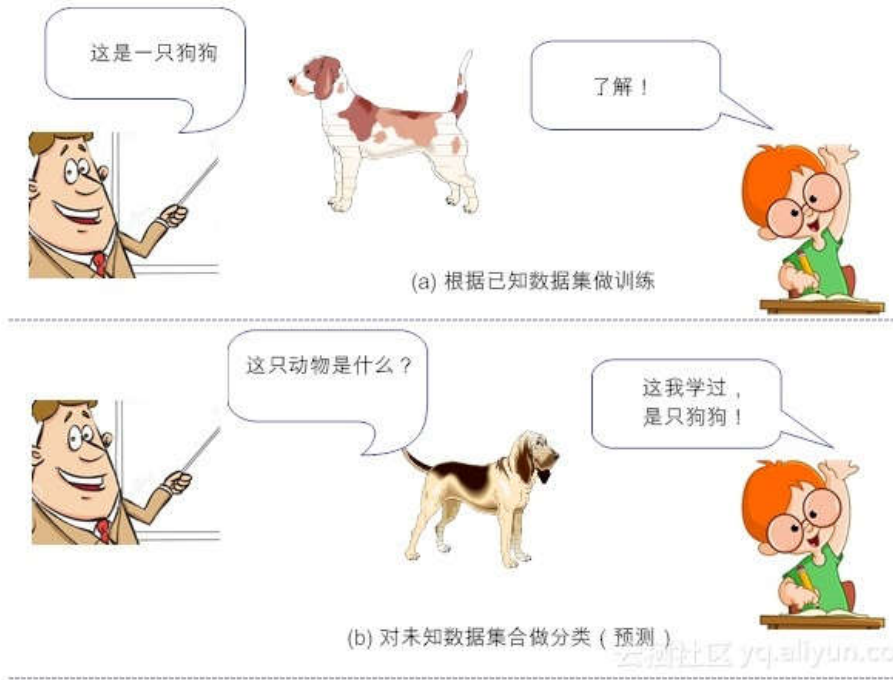


图4-1 监督学习

比如说，小时候父母告诉我们某个动物是猫、是狗或是猪，然后我们的大脑里就会形成或猫或狗或猪的印象，然后面前来了一条“新”小狗，如果你能叫出来“这是一条小狗”，那么恭喜你，你的标签分类成功！但如果你说“这是一头小猪”。这时你的监护人就会纠正你的偏差，“乖，不对呦，这是一头小狗”，这样一来二去的训练，就不断更新你的大脑认知体系，聪明如你，下次再遇到这类新的“猫、狗、猪”等，你就会天才般的给出正确“预测”分类（如图1所示）。简单来说，监督学习的工作，就是通过有标签的数据训练，获得一个模型，然后通过构建的模型，给新数据添加上特定的标签。

事实上，整个机器学习的目标，都是使学习得到的模型，能很好地适用于“新样本”，而不是仅仅在训练样本上工作得很好。通过训练得到的模型，适用于新样本的能力，称之为“泛化 (generalization) 能力”。

(2) 非监督学习 (Unsupervised Learning) ：与监督学习相反的是，非监督学习所处的学习环境，都是非标签的数据。韩老师接着说，**非监督学习，本质上，就是“聚类 (cluster)”的近义词。**

人工智能究竟是不是“虚假式”繁荣？精选问答  
手环~ (/roundtable/489810)  
【赠送参会码】PostgreSQL 2018 中国技术  
告 (/roundtable/473635)

开年Hi购季

开发者分会

海量学习资源，一站式开发者福利

{/}

<https://www.aliyun.com/acts/product-section-2019/developer>

### 【深度学习之美】“机器学习”三重门，“中庸之道”趋若人（入门系列之四）



(a) 在非标签数据集中做归纳



(b) 对未知数据集做归类（预测）

图4-2 非监督学习

简单来说，给定数据，从数据中学，能学到什么，就看数据本身具备什么特性了（given data, learn about that data）。我们常说的“物以类聚，人以群分”说得就是“非监督学习”。这里的“类”也好，“群”也罢，事先我们是不知道的。一旦我们归纳出“类”或“群”的特征，如果再要来一个新数据，我们就根据它距离哪个“类”或“群”较近，就“预测”它属于哪个“类”或“群”，从而完成新数据的“分类”或“分群”功能。

(3) 半监督学习（Semi-supervised Learning）：这类学习方式，既用到了标签数据，又用到了非标签数据。有句骂人的话，说某个人“有妈生，没妈教”，抛开这句话骂人的含义，其实它说的是“无监督学习”。但我们绝大多数人，不仅“有妈生，有妈教”，还“有小学教，有中学教，有大学教”，“有人教”，这就是说，有人告诉我们事物的对与错（即对事物打了标签），然后我们可据此改善自己的性情，慢慢把自己调教得更“有教养”，这自然就属于“监督学习”。但总有那么一天我们要长大。而长大的标志之一，就是自立。何谓“自立”？就是远离父母、走出校园后，没有人告诉你对与错，一切都要基于自己早期已获取的知识为基础，从社会中学习，扩大并更新自己的认知体系，然后遇到新事物时，我们能“泰然自若”处理，而非茫然“六神无主”。

从这个角度来看，现代人类成长学习的最佳方式，当属“半监督学习”！它既不是纯粹的“监督学习”（因为如果完全是这样，就会扼杀我们的创造力，我们的认知体系也就永远不可能超越我们的父辈和师辈）。但我们也不属于完全的“非监督学习”（因为如果完全这样，我们会如“无根之浮萍”，会花很多时间“重造轮子”。前人的思考，我们的阶梯，这话没毛病！）。

那么到底什么是“半监督学习”呢？下面我们给出它的形式化定义：

给定一个来自某未知分布的有标记示例集  $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ ，其中  $x_i$  是数据， $y_i$  是标签。对于一个未标记示例集  $U = \{x_{l+1}, x_{l+1}, \dots, x_{l+u}\}$ ， $l < u$ ，于是，我们期望学得函数  $f: X \rightarrow Y$  可以准确地对未标识的数据  $x_i$  预测其标记  $y_i$ 。这里均为  $d$  维向量， $y_i \in Y$  为示例  $x_i$  的标记。



图4-3 半监督学习

形式化的定义比较抽象，下面我们列举一个现实生活中的例子，来辅助说明这个概念。假设我们已经学习到：

- (1) 马晓云同学（数据1）是个牛逼的人（标签：牛逼的人）
- (2) 马晓腾同学（数据2）是个牛逼的人（标签：牛逼的人）
- (3) 假设我们并不知道李晓宏同学（数据3）是谁，也不知道他牛逼不牛逼，但考虑他经常和二马同学共同出没于高规格大会，都经常会被达官贵人接见（也就是说他们虽独立，但同分布），我们很容易根据“物以类聚，人以群分”的思想，把李晓宏同学打上标签：他也是一个很牛逼的人！

这样一来，我们的已知领域（标签数据）就扩大了（由两个扩大到三个！），这也就完成了半监督学习。事实上，半监督学习就是以“已知之认知（标签化的分类信息）”，扩大“未知之领域（通过聚类思想将未知事物归类为已知事物）”。但这里隐含了一个基本假设——“聚类假设（cluster assumption）”，其核心要义就是：“相似的样本，拥有相似的输出”。

事实上，我们对半监督学习的现实需求，是非常强烈的。其原因很简单，就是因为人们能收集到的标签数据非常有限，而手工标记数据需要耗费大量的人力物力成本，但非标签数据却大量存在且触手可及，这个现象在互联网数据中更为凸显，因此，“半监督学习”就显得尤为重要性[2]。

人类的知识，其实都是这样，以“半监督”的滚雪球的模式，越扩越大。“半监督学习”既用到了“监督学习”，也吸纳了“非监督学习”的优点，二者兼顾。

如此一来，“半监督学习”就有点类似于我们中华文化的“中庸之道”了。

的确如此吗？下面我们就聊聊机器学习的“中庸之道”。

4.2从“中庸之道”看机器学习

说到“中庸之道”，很多人立马想到的就是“平庸之道”，把它的含义理解为“不偏不倚、不上不下、不左不右、不前不后”。其实，这是一个很大的误解！

据吴伯凡先生介绍[3]，“中”最早其实是一个器具，它看上去像一个槌子，为了拿起方便，就用手柄穿越其中，即为“中”。

这个“中”可不得了，它非常重要，且只有少数人才能使用，那都是谁来用呢？答案就是古代的军事指挥官。在“铁马金戈风沙腾”的战场上，军旗飘飘，唯有一人高高站在战车上，手握其“中”，其他将士都视其“中”而进退有方（见图4-4第二行第一字），而手握其“中”的人，称之为“史”（见图4-4第一行第一字）。所以现在你知道了吧，其实“史”最早的本意，就是手握指挥大权的“大官”。

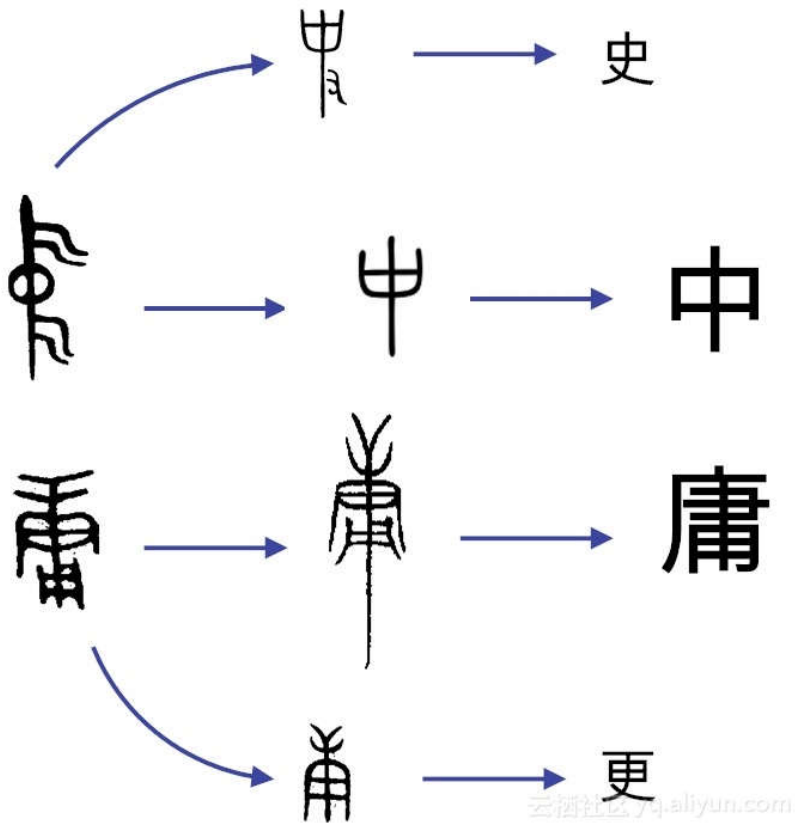


图4-4 中庸之道，蕴意为何？

再后来，“中”就有各种各样的引申含义。在中原地带的人，在他们的语言里头到现在还保留一些古代遗风，比如说河南人说“对”或者“是”的时候，他说的是“中（zhóng）”，当他们说“中（zhóng）”的时候，就表示事情是正确的，是可行的。

其实，“中”还有一个读音叫“中（zhòng）”，比如说成语里就有“正中下怀”、“百发百中”等，这时“中（zhòng）”的含义就是恰到好处，不偏离原则，坚守关键点。

下面再来说说“庸”。“庸”的上半部是“庚”，“庚”同音于“更”，即“变化”之意。而“庸”的下半部是“用”，“用”之本意为“变化中的不变”，即为“常”。在编程语言中，我们常说“常量”，说的就是不变化的量。所以，“庸”的最佳解释应该是“富有弹性的坚定”。

那么“中庸”放在一起是什么意思呢？那就是告诉我们“在变化中保持不变”。其中，所谓“变化”，就是我们所处的环境变化多端，所以我們也需要“随机应变，伺机而动”。而所谓“不变”就是要我们“守住底线，中心原则不变”。二者在一起，“中庸之道”就是要告诉我们要在灵活性（变）和原则性（不变）之间，保持一个最佳的平衡。

那说了半天，这“中庸之道”和机器学习有啥关系呢？其实这就是一个方法论问题。“监督学习”，就是告诉你“正误之道”，即有“不变”之原则。而“非监督学习”，就有点“随心所欲，变化多端”，不易收敛，很易“无根”，“不用临池更相笑，最无根蒂是浮萍。”

那“中庸之道”的机器学习应该是怎样的呢？自然就是“半监督学习”，做有弹性的坚定学习。这里的“坚定”自然就是“监督学习”，而“有弹性”自然就是“非监督学习”。

“有弹性”的变化，不是简单的加加减减，而是要求导数（变化），而且还可能是导数的导数（变化中的变化）。只有这样，我们才能达到学习最本质的需求——性能的提升。在机器学习中，我们不正是以提高性能为原则，用梯度（导数）递减的方式来完成的吗？

所以，你看看，我们老祖先的方法论，其实是很牛逼的。只不过是历时太久远了，其宝贵的内涵，被时间的尘埃蒙蔽了而已。

现在，我们经常提“文化自信”，哈哈，你看我这个例子算不算一个？



#### 4.3 小结

【深度学习之美】“机器学习”三重门，“中庸之道”趋若人（入门系列之四）  
我们刚刚学习了深度学习中的监督学习、非监督学习和半监督学习。它们之间核心区别在于是否（部分）使用了标签数据。

然后我们又从老祖先的“中庸之道”，谈了谈机器学习的发展方向，不管是从人类自己的学习方式，还是“中庸之道”核心本质，“半监督学习”一定是未来机器学习的大趋势。

我们这样说是依据的，因为人工智能的最高标准，不正是要模拟学习人类的智能吗？而人类就是通过“半监督学习”获取最妙、最高的智能啊，所以你有什么理由不相信“机器学习（包括深度学习）”不是朝着这个方向发展的呢？

不管你信不信，反正我是信了！

好了，就此打住吧！在下一个小节中，我们真的该聊聊具体的神经网络学习算法了。我们知道，“人之初，性本善”，那么“神经”之初，又是什么呢，自然就是“感知机”了。在下一小节，我们就非常务实地聊聊“感知机”的学习算法（并附上源代码），它可是一切神经网络学习（包括深度学习）的基础，请你关注！

#### 4.4 请你思考

通过上面的学习，请你思考如下问题：

(1) 深度学习算法既有监督学习模式的，也有非监督学习模式的？它有没有半监督学习模式的？如果有，请你分别列举一二？

(2) 阿尔法狗再次“咬伤”了人类的最佳棋手柯洁，不出意外地再次“震惊世人”，有人说阿尔法狗是深度学习的典范之作？仅仅如此吗？除了深度学习之外,它还结合使用了什么技术？

(3) 中国古代的铜钱，也体现有“中庸之道”，你知道是什么吗？

写下你的感悟，祝你每天都有收获！

#### 【参考文献】

[1] Han J. Data Mining: Concepts and Techniques[M]. Morgan Kaufmann Publishers Inc. 2005.

[2] 周志华.机器学习.北京：清华大学出版社.2016.1

[3]吴伯凡.中庸之美.得到.2017.2

文章作者：张玉宏（著有《品味大数据 (<http://product.dangdang.com/24048575.html>)》、本文节选自《深度学习之美 (<https://item.jd.com/12382640.html>)》（最通俗易懂的深度学习入门）2018年7月出版）

审校：我是主题曲哥哥。

掌握第一手干货，欢迎关注阿里云云栖社区官方微博：<http://weibo.com/u/1939498534> (<http://weibo.com/u/1939498534>)

**(未完待续)**

## 系列文章：

Hello World感知机，懂你我心才安息(深度学习入门系列之五) (<https://yq.aliyun.com/articles/93540>)

损失函数减肥用，神经网络调权重（深度学习入门系列之六） (<https://yq.aliyun.com/articles/96427>)

山重水复疑无路，最快下降问梯度（深度学习入门系列之七） (<https://yq.aliyun.com/articles/105339>)