可标注的，有可能因为各种原因无法实现人工标注或标注成本太高，此时便可采用无监督学习。无监督学习最典型的例子是聚类。

3）**半监督学习**是一种介于监督学习和无监督学习之间的学习方式。半监督学习是训练数据中有少部分样本是被标记的，其他大部分样本并未被标记。半监督学习可以用来进行预测，模型需要先学习数据的内在结构，以便得到更好的预测效果。

4）**强化学习**是智能体（agent）采取不同的动作（action），通过与环境的交互不断获得奖励指导，从而最终获得最大的奖励。监督学习中数据标记的标签用于检验模型的对错，并不足以在交互的环境中学习。而在强化学习下，交互数据可以直接反馈到模型，模型可以根据需要立即做出调整。强化学习不同于无监督学习，因为无监督学习旨在学习未标记数据间的内在结构，而强化学习的目标是最大化奖励。

1.1.3 机器学习应用开发步骤

开发机器学习应用时，读者可以尝试不同的模型算法，采用不同的方法对数据进行处理，这个过程十分灵活，但也并非无章可循。本节会对机器学习应用开发中的经典步骤进行逐一介绍。

（1）定义问题

在开发机器学习应用之前，先要明确需要解决的是什么问题。在实际应用中，很多时候我们得到的并非是一个明确的机器学习任务，而只是一个需要解决的问题。首先要将实际问题转化为机器学习问题，例如解决公司员工不断收到垃圾邮件的问题，可以先对邮件进行分类，通过机器学习算法将垃圾邮件识别出来，然后对其进行过滤。由此，我们将一个过滤垃圾邮件的现实问题转化为了机器学习的二分类问题（判断是否是垃圾邮件）。

（2）数据采集

数据对于机器学习是至关重要的，数据采集是机器学习应用开发的基础。数据采集有很多种方法，最简单的就是人工收集数据，例如预测房屋价格，可以从和房屋相关的网站上获取数据、提取特征并进行标记（如果需要）。人工收集数据耗时较长且非常容易出错，所以通常是其他方法都无法实现时才会采用。除人工收集数据外，还可以通过网络爬虫从相关网站收集数据，从传感器收集实测数据（如压力传感器的压力数据），从某些 API 获取数据（如交易所的交易数据），从 App 或 Web 端收集数据等。对于某些领域，也可直接采用业界的公开数据集，从而节省时间和精力。

（3）数据清洗

通过数据采集得到的原始数据可能并不规范，需对数据进行清洗才能满足使用需求。例如，去掉数据集中的重复数据、噪声数据，修正错误数据等，最后将数据转换为需要的格