



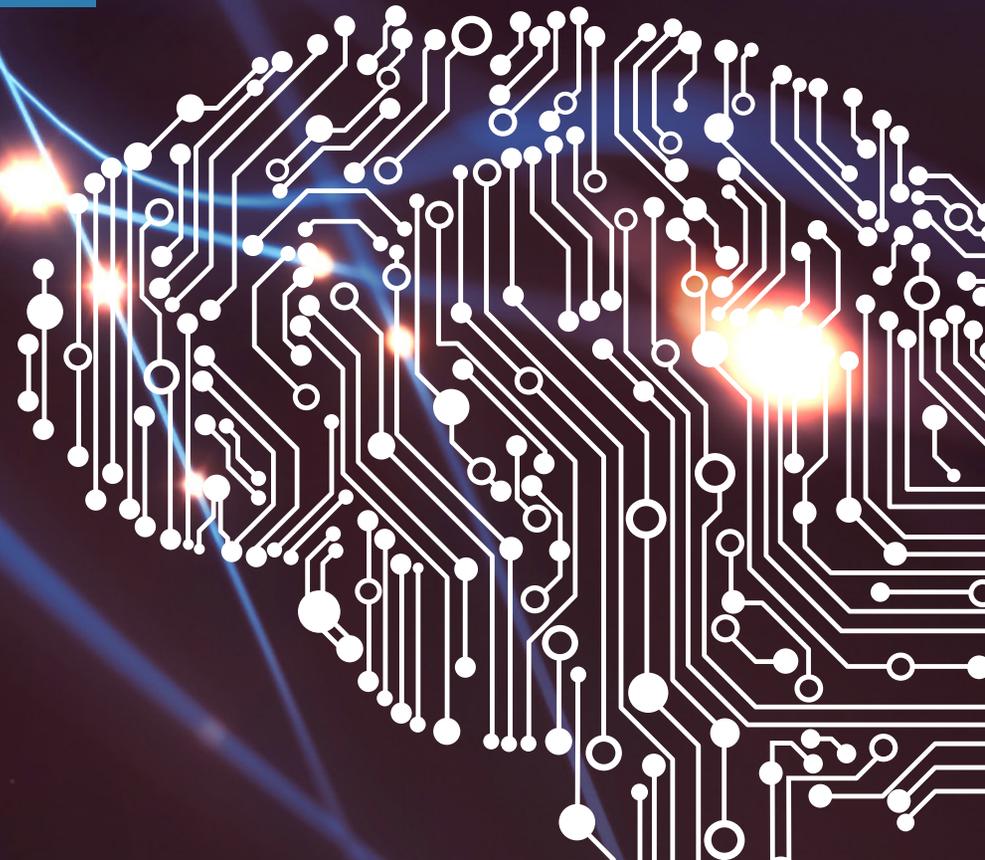
机器学习简介

```
%% Generalized Linear Model - Logistic Regression  
glm = GeneralizedLinearModel.fit(Xtrain,double(Ytrain),  
    'linear','Distribution','binomial','link','logit');
```

```
%% Discriminant Analysis  
da = ClassificationDiscriminant.fit(Xtrain,Ytrain,  
    'discrimType','quadratic');
```

```
%% Classification Using Nearest Neighbors  
knn = ClassificationKNN.fit(Xtrain,Ytrain,...  
    'Distance','seuclidean');
```

```
%% Ensemble Learning: TreeBagger  
opts = statset('UseParallel',true);  
  
tb = TreeBagger(150,Xtrain,Ytrain,'method','classification',...  
    'Options',opts,'OOBVarImp','on','cost',[0 1; 5 0]);
```



机器学习是什么？

机器学习教计算机执行人和动物与生俱来的活动：从经验中学习。机器学习算法使用计算方法直接从数据中“学习”信息，而不依赖于预定方程模型。当可用于学习的样本数量增加时，这些算法可自适应提高性能。

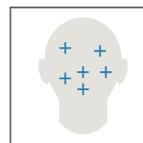
更多数据、更多问题、更多解答

机器学习算法可从能够带来洞察力的数据中发现自然模式，帮助您更好地制定决策和做出预测。医疗诊断、股票交易、能量负荷预测及更多行业每天都在使用这些算法制定关键决策。媒体网站依靠机器学习算法从数百万种选择中筛选出为您推荐的歌曲或影片。零售商利用这些算法深入了解客户的购买行为。

实际环境中的应用程序

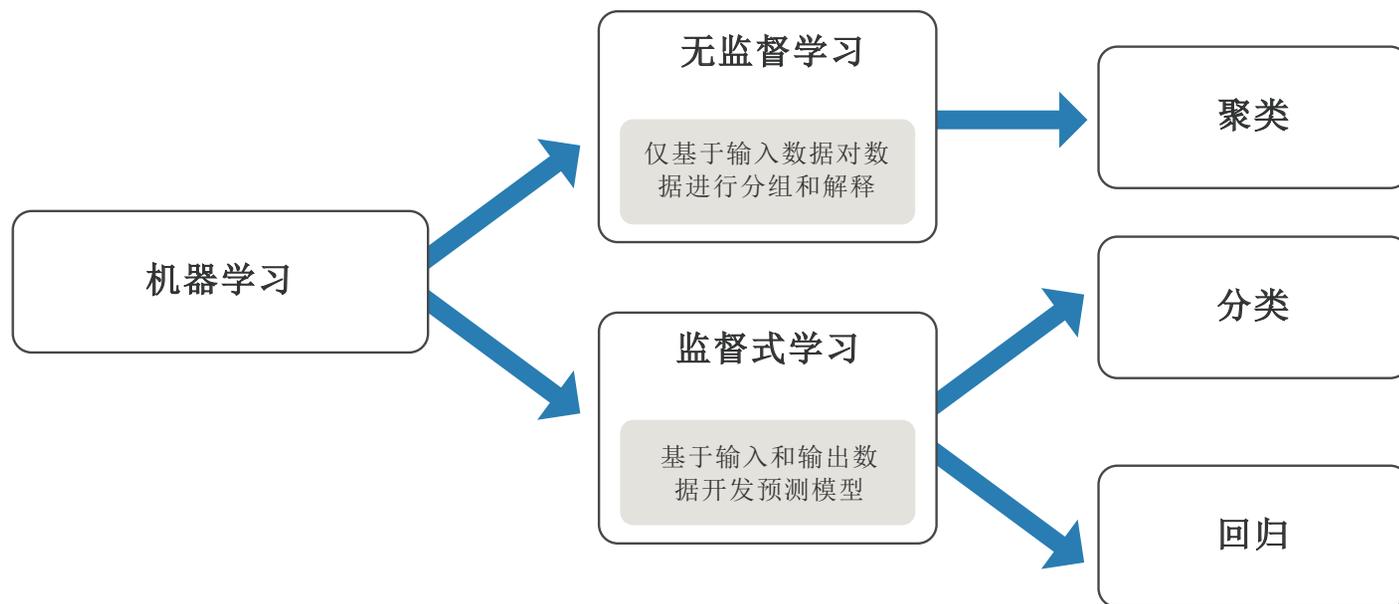
随着大数据的增加，机器学习对于解决以下领域的问题变得尤为重要：

- 计算金融学，用于信用评估和算法交易
- 图像处理和计算机视觉，用于人脸识别、运动检测和对象检测
- 计算生物学，用于肿瘤检测、药物发现和 DNA 顺序分析
- 能源生产，用于预测价格和负荷
- 汽车、航空航天和制造业，用于预见性维护
- 自然语言处理



机器学习的工作原理

机器学习采用两种类型的技术：监督式学习和无监督学习。监督式学习根据已知的输入和输出训练模型，让模型能够预测未来输出；无监督学习从输入数据中找出隐藏模式或内在结构。



监督式学习

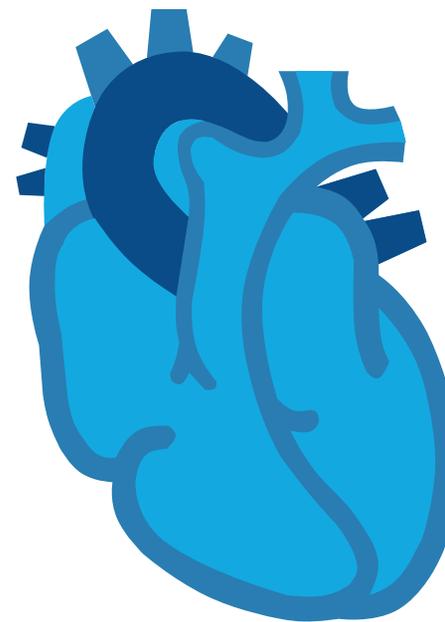
监督式机器学习旨在构建能够根据存在不确定性的证据做出预测的模型。监督式学习算法接受已知的输入数据集和对数据的已知响应（输出），然后训练模型，让模型能够为新输入数据的响应生成合理的预测。

监督式学习采用分类和回归技术开发预测模型。

- **分类技术可预测离散的反应** — 例如，电子邮件是真正邮件还是垃圾邮件，肿瘤是恶性还是良性的。分类模型可将输入数据划分成不同类别。典型的应用包括医学成像、语音识别和信用评估。
- **回归技术可预测连续的反应** — 例如，电力需求中温度或波动的变化。典型的应用包括电力系统负荷预测和算法交易。

使用监督式学习预测心脏病发作

假设临床医生希望预测某位患者在一年内是否会心脏病发作。他们有以前就医的患者的相关数据，包括年龄、体重、身高和血压。他们知道以前的患者在一年内是否出现过心脏病发作。因此，问题在于如何将现有数据合并到模型中，让该模型能够预测新患者在一年内是否会出现心脏病发作。

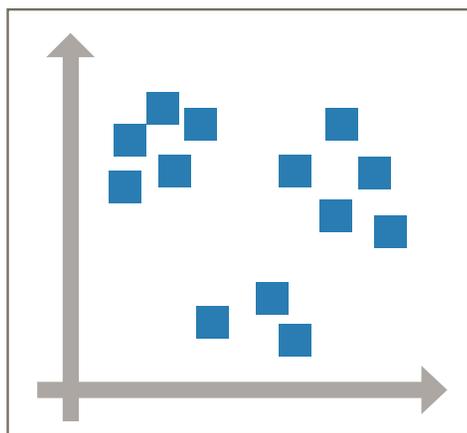


无监督学习

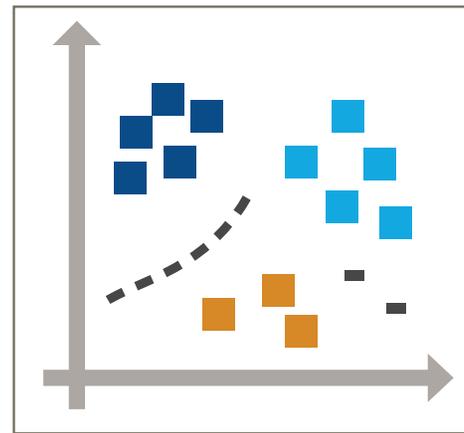
无监督学习可发现数据中隐藏的模式或内在结构。这种技术可根据包含未标记响应的输入数据的数据集执行推理。

聚类是一种最常用的无监督学习技术。这种技术可通过探索性数据分析发现数据中隐藏的模式或分组。

聚类的应用包括基因序列分析、市场调查和对象识别。



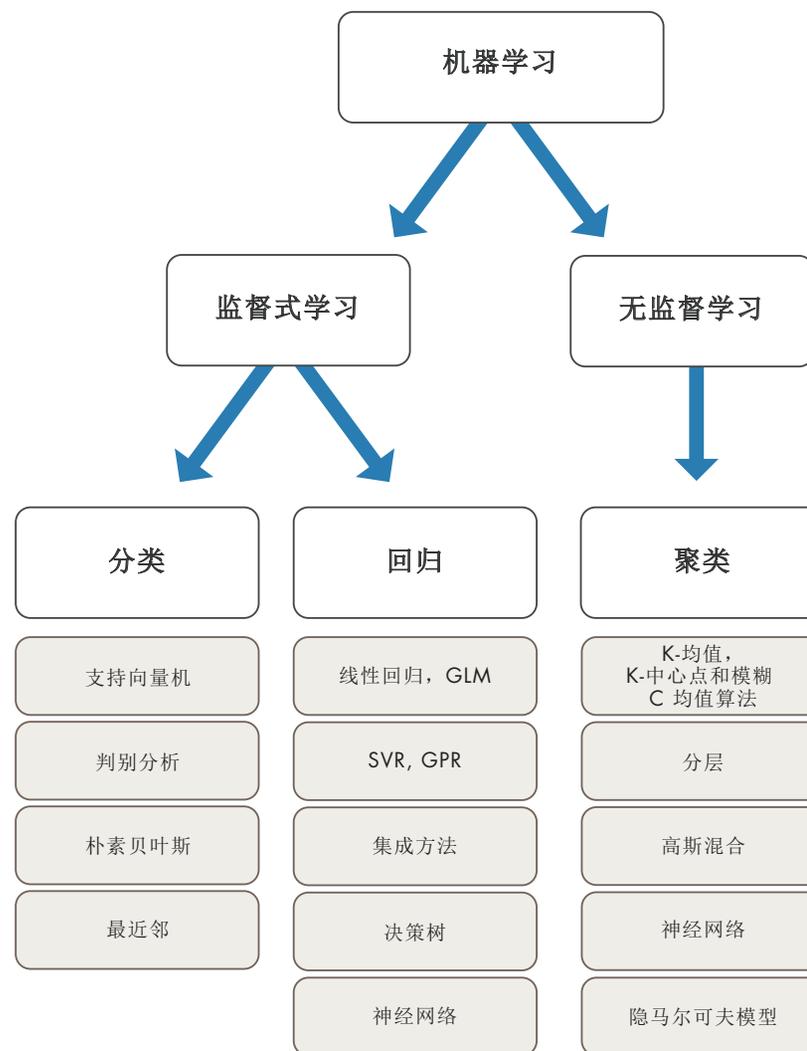
为数据中的模式执行聚类



如何确定使用哪种算法？

选择正确的算法看似难以驾驭——需要从几十种监督式和无监督机器学习算法中选择，每种算法又包含不同的学习方法。

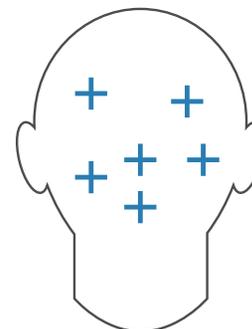
没有最佳方法或万全之策。找到正确的算法只是试错过程的一部分——即使是经验丰富的数据科学家，也无法说出某种算法是否无需试错即可使用。但算法的选择还取决于您要处理的数据的大小和类型、您要从数据中获得的洞察力以及如何运用这些洞察力。



何时应该使用机器学习？

当您遇到涉及大量数据和许多变量的复杂任务或问题，但没有现成的处理公式或方程式时，可以考虑使用机器学习。例如，如果您需要处理以下情况，使用机器学习是一个很好的选择：

- 手写规则和方程式太过复杂——例如人脸识别和语音识别。
- 任务的规则始终在变化——例如事务处理记录的欺诈检测。
- 数据本身在不断变化，程序也必须适应这种变化——例如自动交易、能量需求预测和购物趋势预测等。



实际环境中的示例

创建可分析艺术作品的算法

美国罗格斯大学艺术与人工智能实验室的研究人员曾经想知道计算机算法能否像人类一样根据风格、流派和艺术家将绘画作品轻松归类。

开始时,他们通过识别视觉特征来对绘画作品的风格分类。他们开发的绘画风格分类算法在数据库中的准确度达到 60%,远超过普通非专业人士。

研究人员假定可用于对风格分类(监督式学习问题)的视觉特征也能用于确定艺术影响力(无监督学习问题)。

他们将经过训练的分类算法应用到 Google 图像,用于确定具体对象。他们对跨度长达 550 年的 66 位不同艺术家的 1,700 幅绘画作品测试了此算法。此算法可以可靠地识别出相关的作品,包括迭戈·委拉斯开兹的《教皇英诺森十世肖像》对弗朗西斯·培根的《教皇英诺森十世肖像的习作》产生的影响。



实际环境中的示例

优化大型建筑中的 HVAC 能耗

在办公大楼、医院及其他大型商业楼宇内使用的暖通空调系统 (HVAC) 通常效率低下, 原因在于这些系统未考虑不断变化的气候模式、多变的能耗或建筑物的热性能。

Building IQ 的基于云的软件平台可解决这个问题。该平台采用先进的算法和机器学习方法连续处理来自功率计、温度计和 HVAC 压力传感器的数千兆字节信息以及天气和能耗。更为特殊的是, 机器学习可用于对数据分段和确定天然气、电力、蒸汽和太阳能对加热和冷却流程的相对贡献量。Building IQ 平台将大型商业楼宇内使用的 HVAC 在正常运行期间的能耗降低了 10% - 25%。

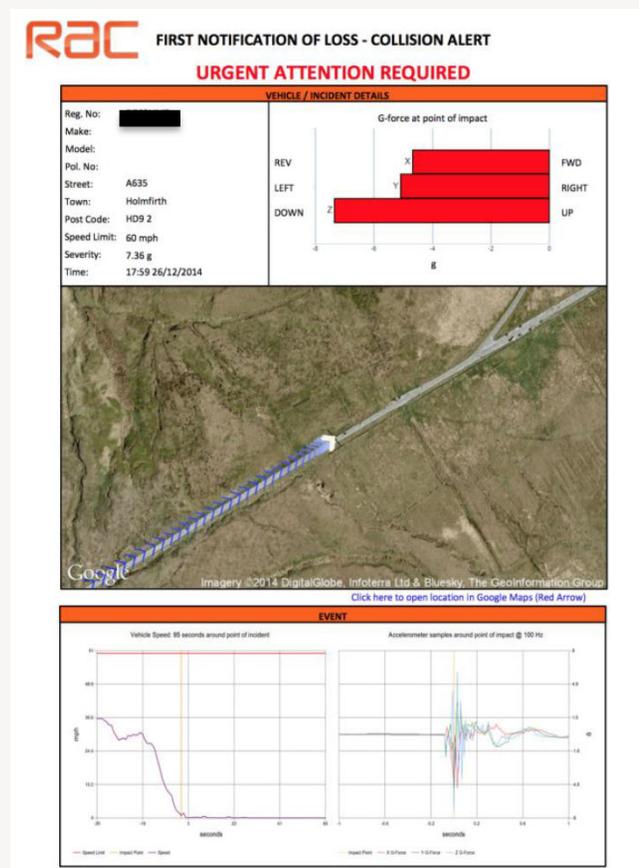


实际环境中的示例

检测低速撞车

RAC 是英国最大的汽车公司之一，拥有超过 8 百万名成员，为私人 and 商业汽车司机提供道路救援、保险和其他服务。

为了能够快速响应道路事故、减少碰撞事故和保险费用，RAC 开发了车载撞击感应系统，该系统采用先进的机器学习算法检测低速碰撞，并且可将这些行为与更常见的驾驶行为（例如驶过路面减速带或路面凹坑）区分开。独立测试数据显示 RAC 系统在碰撞检测测试中获得的准确度达到 92%。



了解更多

准备更深入地钻研？深入了解这些资源以了解有关机器学习方法、示例和工具的更多信息。

▶ 观看

[机器学习一点通](#) 34:34

[使用信号处理和机器学习功能进行传感器数据分析](#) 42:45

📖 阅读

[机器学习博客：社会网络分析、文本挖掘、贝叶斯推理及更多](#)

[Netflix Prize 与生产机器学习系统：行家看点](#)

[机器学习的挑战：选择最佳模型并避免过度拟合](#)

🔍 深入了解

[应用 MATLAB 的机器学习示例](#)

[机器学习解决方案](#)

[使用分类学习器应用程序进行数据分类](#)