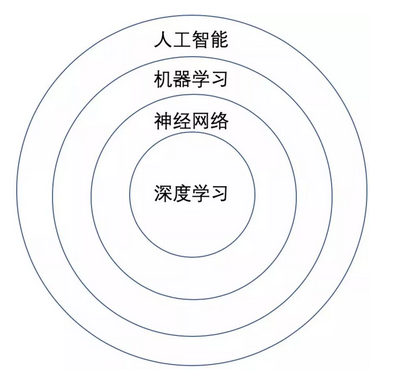
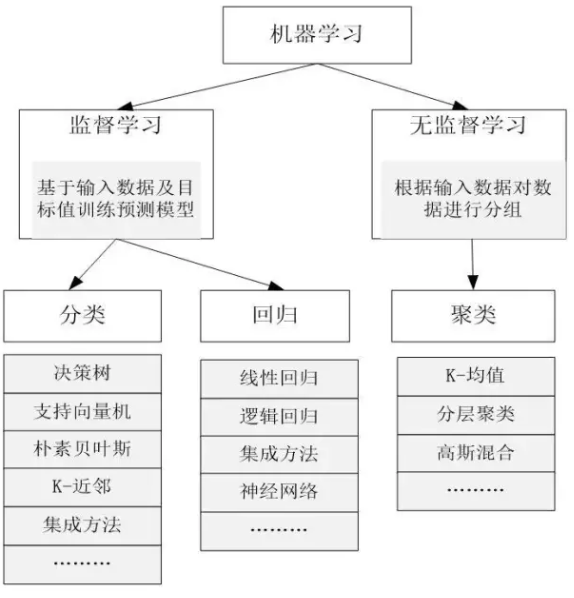
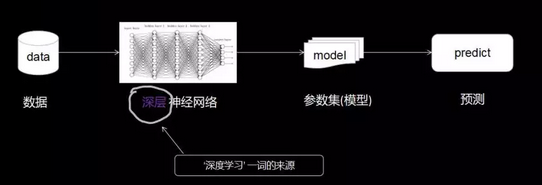
# 概述

 人工智能，机器学习和深度学习的关系：

人工智能：能够独立推理思考、有自我意识、有自己的世界观的机器（程序）。

机器学习：对于某类任务T和性能度量P，如果计算机程序在T上以P衡量的性能随着经验E而自我完善，那么就成这个计算机程序从经验E学习（机器学习）。

 机器学习按照方法主要可以分为两大类：监督学习和无监督学习（还有强化学习）。其中监督学习主要由分类和回归等问题组成，无监督学习主要由聚类和关联分析等问题组成。**深度学习则属于监督学习当中的一种**。

 深度学习：对于某类任务T和性能度量P，如果计算机程序在T上以P度量的性能随着经验E而自我完善，那么就称这个计算机程序从经验E学习。

# 分类

## 监督学习

监督学习的两种主要类型是分类和回归。

在分类中，训练的机器将把一组数据分成特定的类。比如邮箱的垃圾邮件过滤器，过滤器分析之前标记为垃圾邮件的邮件，并将其与新邮件进行比较。如果达到某个百分比，则这些新邮件会被标记为垃圾邮件，并发送到相应的文件夹；不像垃圾邮件的将被归类为正常并发送到收件箱。

第二种是回归。在回归中，机器使用先前标注的数据来预测未来。比如天气应用。利用天气的相关历史数据（即平均温度，湿度和降水量），手机的天气应用可以查看当前天气，并对一定时间范围内的天气进行预测。

## 无监督学习

在无监督学习中，数据是未标注的。由于现实中，大多数的数据都是未标注的，因此这些算法特别有用。

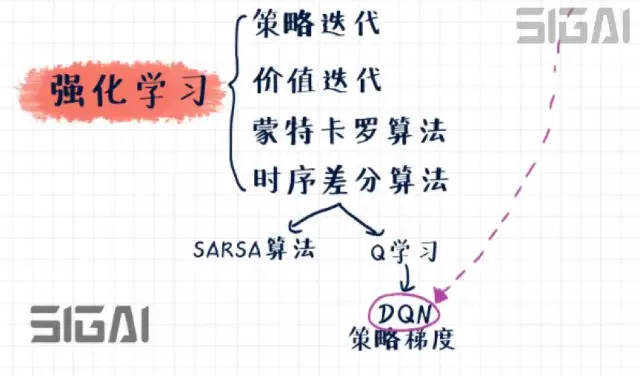
无监督学习分为聚类和降维。

聚类用于根据属性和行为对象进行分组。这与分类不同，因为这些组不会提供给你。聚类将一个组划分为不同的子组（例如，根据年龄和婚姻状况），然后进行有针对性的营销。

另一方面，降维涉及通过查找共性来减少数据集的变量。大多数数据可视化使用降维来识别趋势和规则。

## 强化学习

强化学习使用机器的历史和经验来做出决策。强化学习的经典应用是游戏。

与监督和无监督学习相反，强化学习不注重提供“正确”的答案或输出。相反，它专注于性能，这类似人类根据积极和消极后果进行学习。如果孩子碰到了热炉，他很快就会学会不再重复这个动作。同样在国际象棋中，计算机可以学习不将王移动到对手的棋子可以到达的地方。根据这个原理，在游戏中机器能够最终击败顶级的人类玩家。

# 原理

线性代数概念Top 3：

1. 矩阵运算

2. 特征值/特征向量

3. 向量空间和范数

微积分概念Top 3：

1. 偏导数

2. 向量值函数

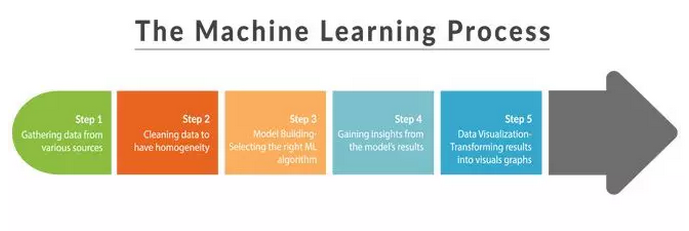
3. 方向梯度

统计概念Top 3：

1. 贝叶斯定理

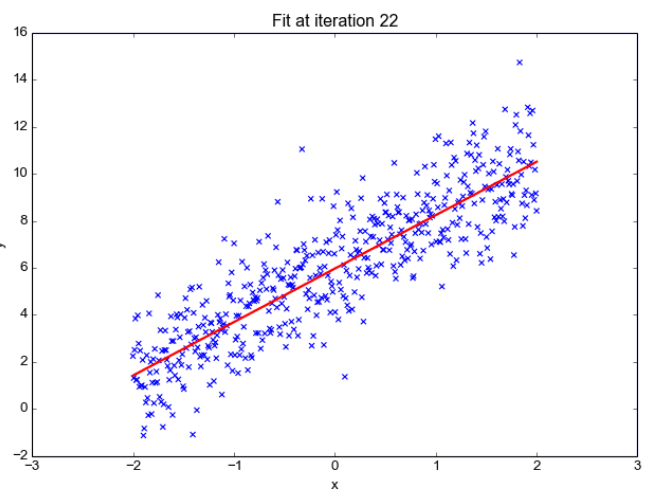
2. 组合学

3. 抽样方法

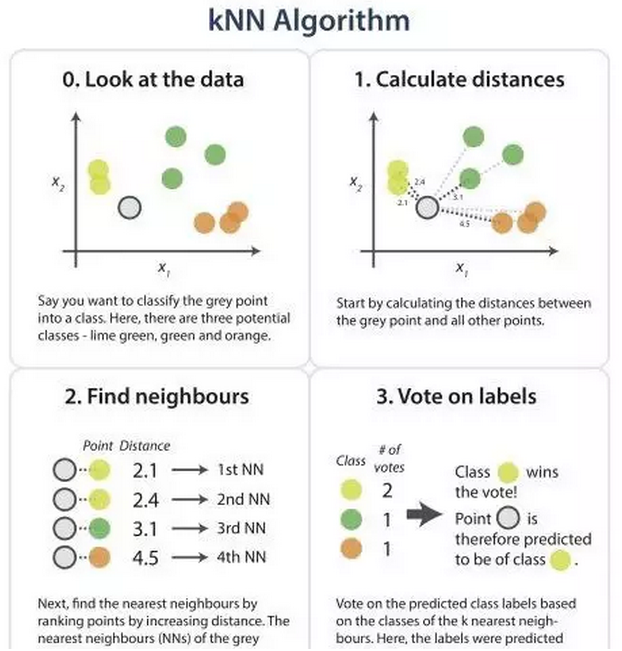
一旦掌握了基本的数学概念，就可以入门机器学习了，有5个主要步骤：

# 算法

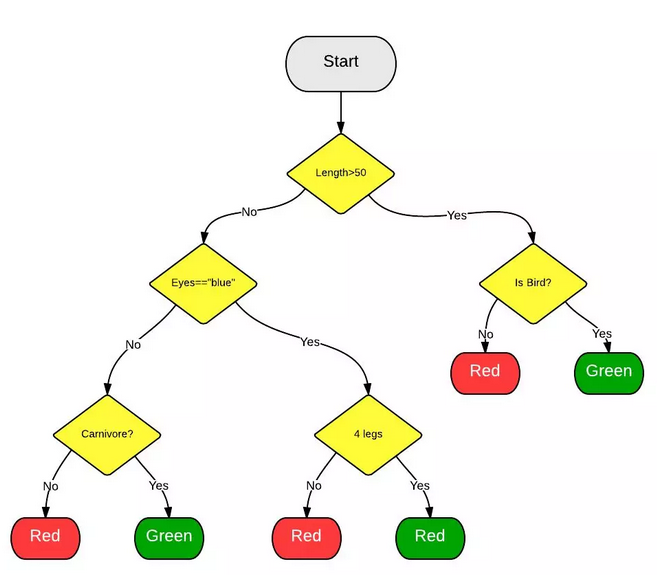
## 回归算法

这可能是最流行的机器学习算法，线性回归算法是基于连续变量预测特定结果的监督学习算法。另一方面，逻辑回归专门用于预测离散值。这些算法都以其速度而闻名，它们被认为是最快的机器学习算法之一。

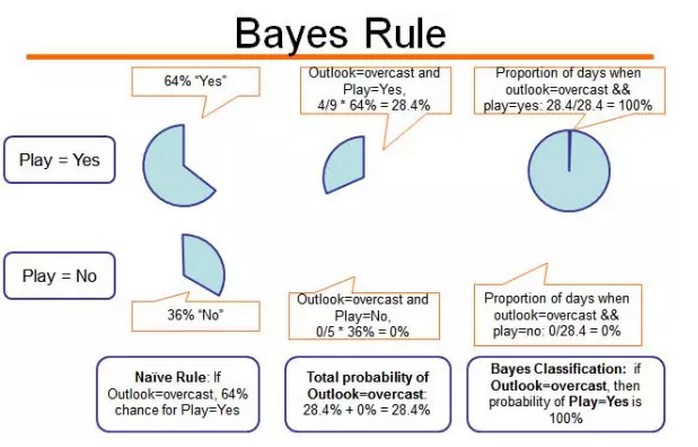
## 基于实例的算法

基于实例的分析根据提供数据的特定实例来预测结果。最著名的基于实例算法是k-Nearest Neighbor，也称为kNN。用于分类中，kNN比较数据点的距离并将每个点分配给它最接近的组。

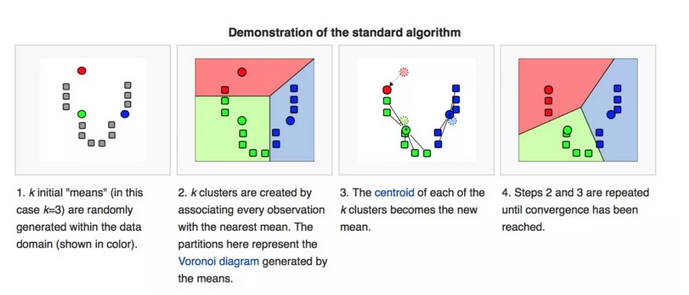
## 决策树算法

决策树算法聚集“弱”学习元素，让它们一起形成强大的算法，这些元素以树状结构组成。其中比较流行的决策树算法是随机森林算法。在该算法中，弱学习元素是随机选择的。在下面的例子中，我们可以发现许多共同的特征（比如眼睛为蓝色或非蓝色），这都无法对动物种类进行辨别。然而，当我们将所有这些观察结果结合在一起时，我们能够形成更完整的理解并进行更准确的预测。

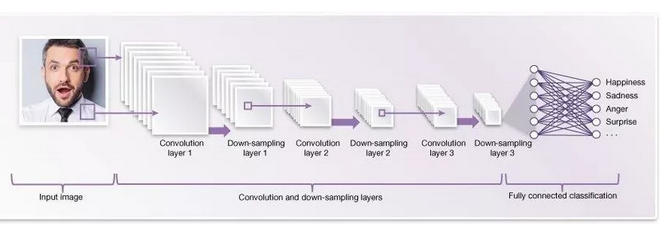
## 贝叶斯算法

这些算法基于贝叶斯定理的，最受欢迎的是朴素贝叶斯算法，它经常用于文本分析。例如，大多数垃圾邮件过滤器都使用贝叶斯算法。它们使用按类别标记的用户输入数据来比较新数据，并对其进行适当分类。

## 聚类算法

聚类算法专注找到元素间的共性，并相应地对它们进行分组。常见的聚类算法是K-Means聚类。根据K-Means，分析人员选择聚类的数量（由变量K表示），算法将元素按物理距离分组到适当的聚类中。

## 深度学习和神经网络算法

人工神经网络算法基于生物神经网络的结构。深度学习采用神经网络模型并对其进行更新。它们是大型且极其复杂的神经网络，使用少量标注数据和大量未标注数据。神经网络和深度学习具有许多输入，这些输入在产生一个或多个输出之前要通过若干隐藏层。这些连接形成一个特定的循环，模仿人脑处理信息和建立逻辑联系的方式。此外，随着算法的运行，隐藏层通常会变得更小、更细微。

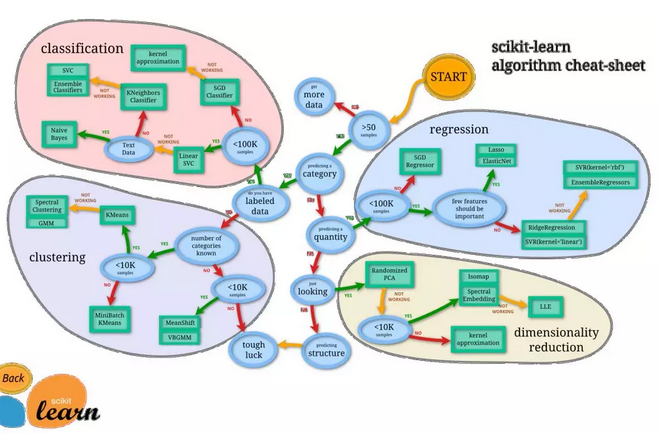
## 降维算法

## 集成算法

包括提升树boosting tree和XGBoost

## 其他算法

下面的图表注明了主要的机器学习算法，它们的类别以及之间的关系。

一旦选择并运行算法，你还需要一个非常重要的步骤：对结果进行可视化。虽然与算法编程相比，这看似很简单而没有技术含量。但出色的可视化能力对于数据科学家来说是至关重要的。即使你得出的分析见解再好，一旦没有人能理解也是毫无价值的。

# 应用

我们需要明确的是，机器学习有潜力能够改变世界。通过Google Brain和斯坦福机器等研究团队的努力，我们正朝着真正的人工智能迈进。但是，机器学习即将影响的领域有哪些呢？

1. 物联网

物联网或IOT，指家庭和办公室中与网络连接的物理设备。其中一个流行的物联网设备是智能灯泡，其销售额在过去几年中飙升。随着机器学习的进步，物联网设备比以往更智能，更复杂。

物联网相关的机器学习应用主要有两方面，提高设备性能和收集数据。

提高设备性能非常简单，我们可以使用机器学习来定制环境，比如用面部识别软件识别谁是房间里，并相应地调整温度和空调。

收集数据更加简单，通过连接网络的设备（如Amazon echo），亚马逊等公司将收集的用户数据提供给广告商，这些数据包括你会看哪些节目，起床和睡觉的时间，你家里有几口人等等。

2. 聊天机器人

在过去几年，聊天机器人的数量激增，复杂的语言处理算法在不断改进。公司在自己的移动应用和第三方应用上使用聊天机器人，从而提供比更快、更高效的客户服务。例如，如果要从H＆M订购衬衫，你可以告诉他们的聊天机器人你想要的款式和尺寸，轻松订购产品。

3. 自动驾驶汽车

如今，雪佛兰、优步和特斯拉等几家大公司正在开发自动驾驶汽车。这些汽车使用通过机器学习进行导航、维护和安全程序。比如交通标志传感器，它使用监督学习算法来识别交通标志，并与标注数据集进行比较。因此，汽车看到停车标志时，系统将进行确认并停车。