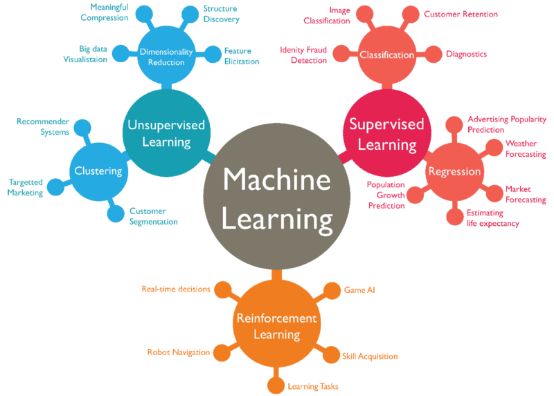
**机器学习的核心：**

使用算法解析数据，从中学习，然后对世界上的某件事情做出决定或预测”。这意味着，与其显式地编写程序来执行某些任务，不如教计算机如何开发一个算法来完成任务。

**分类：**

监督学习、非监督学习、强化学习



**监督学习：**监督学习涉及一组标记数据。计算机可以使用特定的模式来识别每种标记类型的新样本。

分类：

1. 回归。在回归中，机器使用先前的(标记的)数据来预测未来。天气应用是回归的好例子。使用气象事件的历史数据(即平均气温、湿度和降水量)，你的手机天气应用程序可以查看当前天气，并在未来的时间内对天气进行预测。
2. **分类。**在分类中，机器被训练成将一个组划分为特定的类。分类的一个简单例子是电子邮件帐户上的垃圾邮件过滤器。过滤器分析你以前标记为垃圾邮件的电子邮件，并将它们与新邮件进行比较。如果它们匹配一定的百分比，这些新邮件将被标记为垃圾邮件并发送到适当的文件夹。那些比较不相似的电子邮件被归类为正常邮件并发送到你的邮箱。

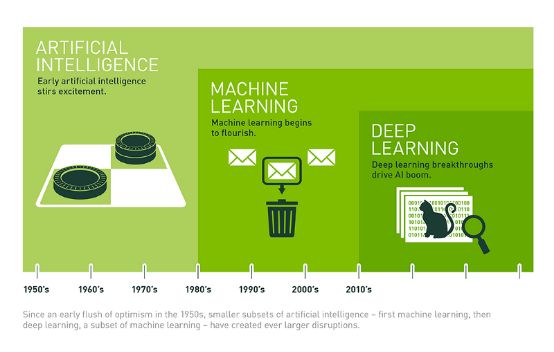
**无监督学习：**在无监督学习中，数据是无标签的。由于大多数真实世界的数据都没有标签，这些算法特别有用。

分类：

1. 聚类。聚类用于根据属性和行为对象进行分组。这与分类不同，因为这些组不是你提供的。聚类的一个例子是将一个组划分成不同的子组(例如，基于年龄和婚姻状况)，然后应用到有针对性的营销方案中。
2. **降维。**降维通过找到共同点来减少数据集的变量。大多数大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个人历史和经验来做出决定。强化学习的经典应用是玩游戏。与监督和非监督学习不同，强化学习不涉及提供“正确的”答案或输出。相反，它只关注性能。这反映了人类是如何根据积极和消极的结果学习的。很快就学会了不要重复这一动作。同样的道理，一台下棋的电脑可以学会不把它的国王移到对手的棋子可以进入的空间。然后，国际象棋的这一基本教训就可以被扩展和推断出来，直到机器能够打(并最终击败)人类顶级玩家为止。

**深度学习是机器学习的一个子集，专注于模仿人类大脑的生物学和过程。**



**机器学习历史**

机器学习最早的发展是Thomas Bayes 在1783年发表的同名理论，贝斯定理发现了给定有关类似事件的历史数据的事件的可能性。这是机器学习的贝叶斯分支的基础，它寻求根据以前的信息寻找最可能发生的事件。换句话说，Bayes定理只是一个从经验中学习的数学方法，是机器学习的基本思想。

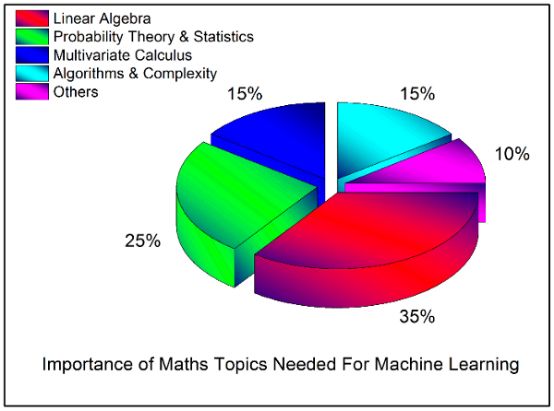
1950年，计算机科学家 Alan Turing发明了所谓的图灵测试，计算机必须通过文字对话一个人，让人以为她在和另一个人说话。

Donald Michie 在1963年推出的强化学习的tic-tac-toe程序。

1997年达到巅峰，当时IBM国际象棋电脑深蓝（Deep Blue）在一场国际象棋比赛中击败了世界冠军加里·卡斯帕罗夫(Garry Kasparov)。

机器学习最大的突破是2006年的深度学习。深度学习是一类机器学习，目的是模仿人脑的思维过程，经常用于图像和语音识别。

**机器学习中的数学知识**



Top 3线性代数概念：

1.矩阵运算；

2.特征值/特征向量；

3.向量空间和范数

Top 3微积分概念：

1.偏导数；

2.向量-值函数；

3.方向梯度

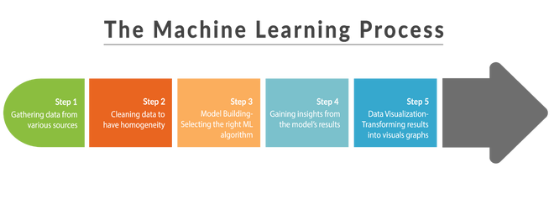
Top 3统计概念：

1.Bayes定理；

2.组合学；

3.抽样方法

**机器学习过程：**

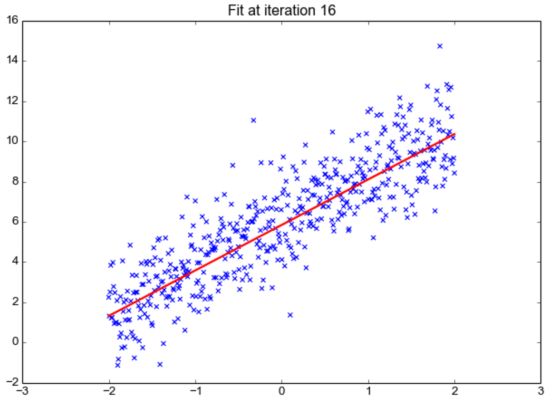


We don’t have better algorithms, we just have more data.—Peter Norvig

**机器学习经典算法**

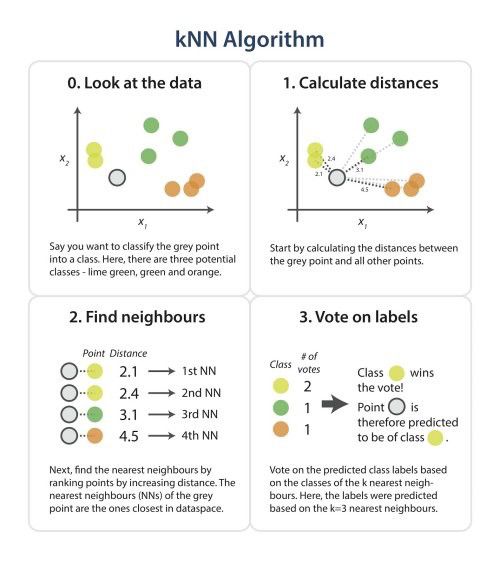
**回归算法**

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，Logistic回归专门用来预测离散值。这两种（以及所有其他回归算法）都以它们的速度而闻名，它们一直是最快速的机器学习算法之一。



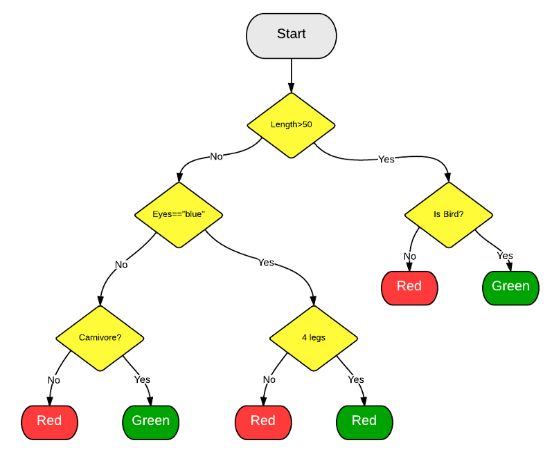
**基于实例的算法**

基于实例的分析使用提供数据的特定实例来预测结果。最著名的基于实例的算法是k-最近邻算法，也称为KNN。KNN用于分类，比较数据点的距离，并将每个点分配给它最接近的组。



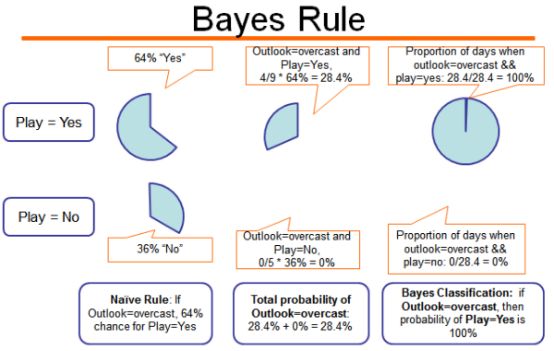
**决策树算法**

决策树算法将一组“弱”学习器集合在一起，形成一种强算法，这些学习器组织在树状结构中，相互分支。一种流行的决策树算法是随机森林算法。在该算法中，弱学习器是随机选择的，这往往可以获得一个强预测器。在下面的例子中，我们可以发现许多共同的特征(就像眼睛是蓝的或者不是蓝色的)，它们都不足以单独识别动物。然而，当我们把所有这些观察结合在一起时，我们就能形成一个更完整的画面，并做出更准确的预测。



**贝叶斯算法**

丝毫不奇怪，这些算法都是基于Bayes理论的，最流行的算法是朴素Bayes，它经常用于文本分析。例如，大多数垃圾邮件过滤器使用贝叶斯算法，它们使用用户输入的类标记数据来比较新数据并对其进行适当分类。

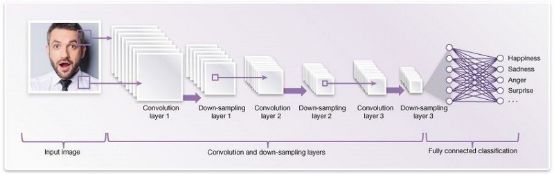


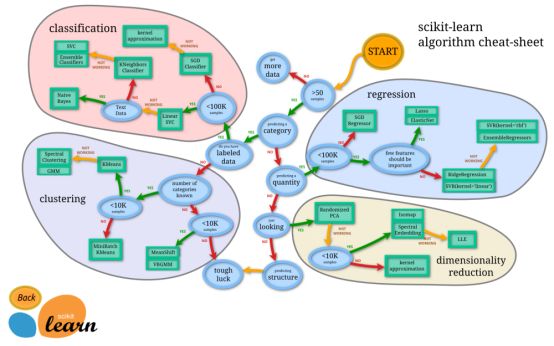
**聚类算法**

聚类算法的重点是发现元素之间的共性并对它们进行相应的分组，常用的聚类算法是k-means聚类算法。在k-means中，分析人员选择簇数(以变量k表示)，并根据物理距离将元素分组为适当的聚类。

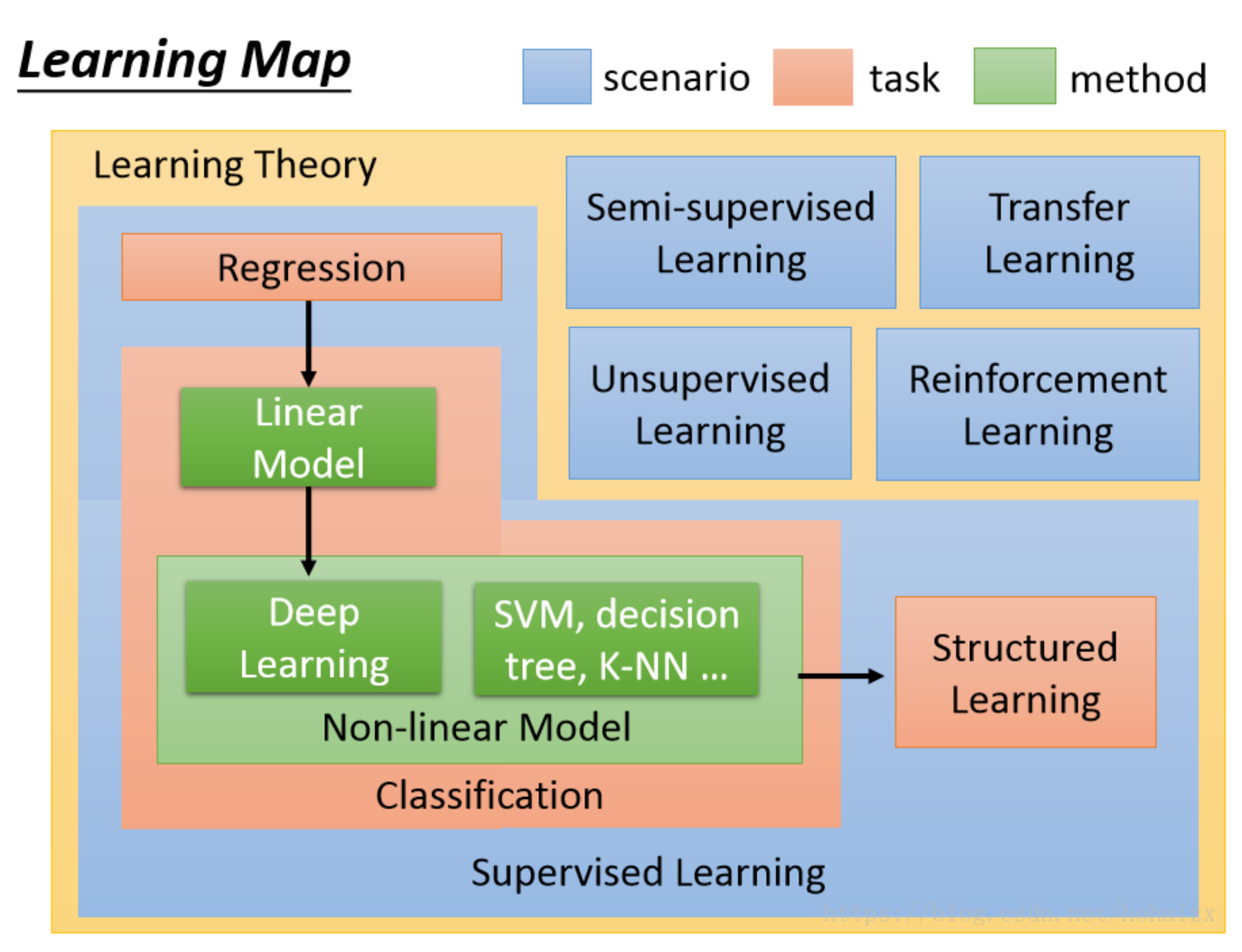
**深度学习和神经网络算法**

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、且极其复杂的神经网络，使用少量的标记数据和更多的未标记数据。神经网络和深度学习有许多输入，它们经过几个隐藏层后才产生一个或多个输出。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连接的方式。此外，随着算法的运行，隐藏层往往变得更小、更细微。





****

****