

机器学习算法分类



- 机器学习的算法很多。很多时候困惑人们的是，很多算法是一类算法，而有些算法又是从其他算法中延伸出来的。
- 这里，我们从两个方面来给大家介绍，
 - 第一个方面是学习的方式
 - 第二个方面是算法的类似性

学习方式



- 根据数据类型的不同，对一个问题的建模有不同的方式。在机器学习或者人工智能领域，人们首先会考虑算法的学习方式。
- 在机器学习领域，有几种主要的学习方式。将算法按照学习方式分类是一个不错的想法，这样可以让人们在建模和算法选择的时候考虑能根据输入数据来选择最合适的算法来获得最好的结果。

监督式学习

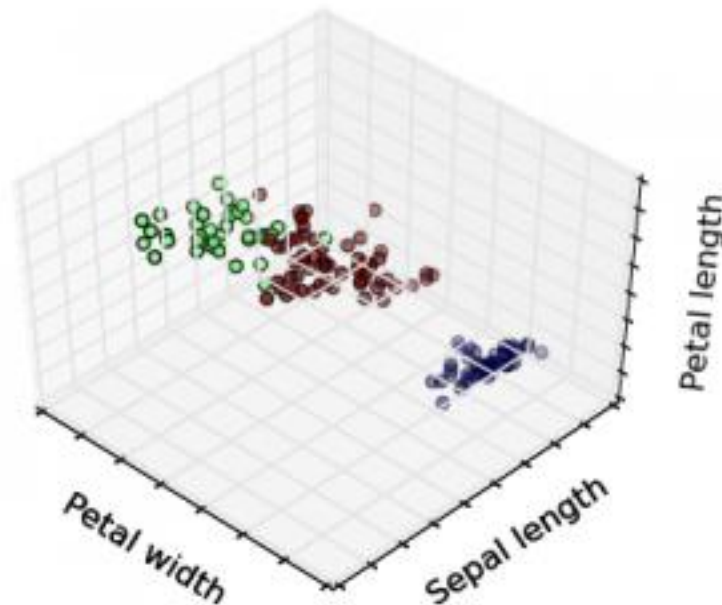


- 在监督式学习下，输入数据被称为“训练数据”，每组训练数据有一个明确的标识或结果，如防垃圾邮件系统中“垃圾邮件”“非垃圾邮件”，对手写数字识别中的“1”，“2”，“3”，“4”等。
- 在建立预测模型的时候，监督式学习建立一个学习过程，将预测结果与“训练数据”的实际结果进行比较，不断的调整预测模型，直到模型的预测结果达到一个预期的准确率。
- 监督式学习的常见应用场景如分类问题和回归问题。常见算法有逻辑回归（Logistic Regression）和反向传递神经网络（Back Propagation Neural Network）。

非监督式学习



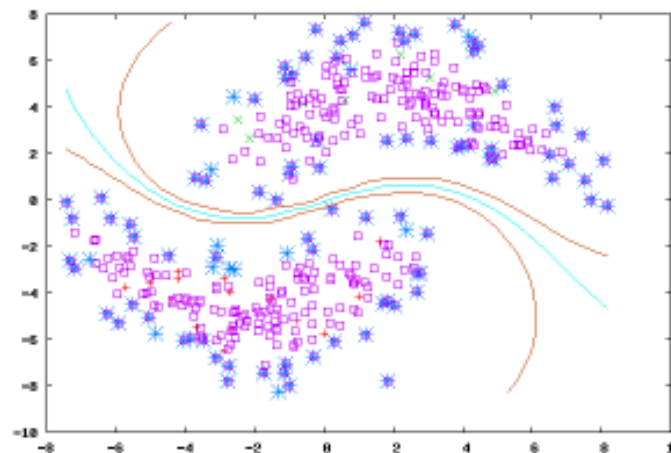
- 在非监督式学习中，数据并不被特别标识，学习模型是为了推断出数据的一些内在结构。
- 常见的应用场景包括关联规则的学习以及聚类等。常见算法包括Apriori算法以及k-Means算法。



半监督式学习



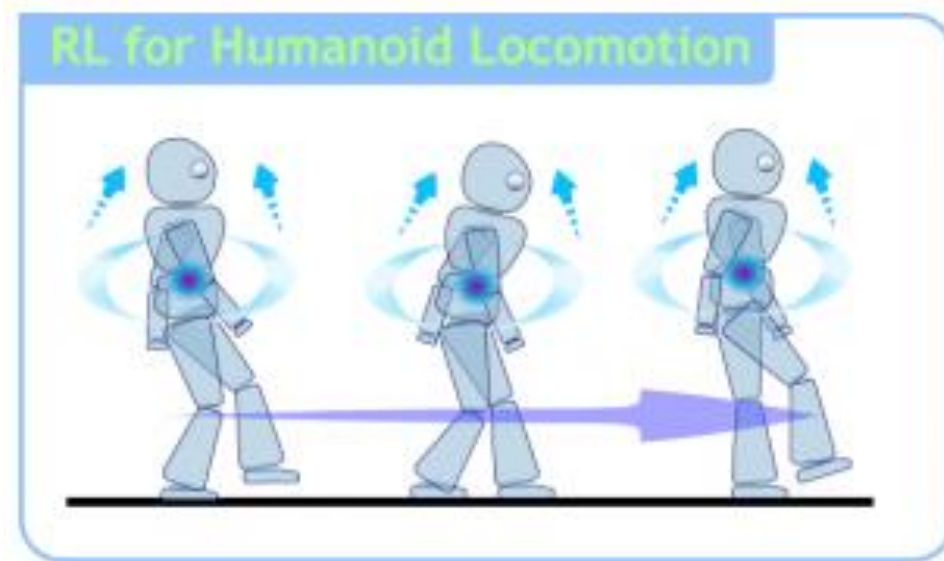
- 在此学习方式下，输入数据部分被标识，部分没有被标识，这种学习模型可以用来进行预测，但是模型首先需要学习数据的内在结构以便合理的组织数据来进行预测。
- 应用场景包括分类和回归，算法包括一些对常用监督式学习算法的延伸，这些算法首先试图对未标识数据进行建模，在此基础上再对标识的数据进行预测。如图论推理算法（Graph Inference）或者拉普拉斯支持向量机（Laplacian SVM.）等。



强化学习



- 在这种学习模式下，输入数据作为对模型的反馈，不像监督模型那样，输入数据仅仅是作为一个检查模型对错的方式，在强化学习下，输入数据直接反馈到模型，模型必须对此立刻作出调整。
- 常见的应用场景包括动态系统以及机器人控制等。常见算法包括Q-Learning以及时间差学习 (Temporal difference learning) 。



学习方式总结



- 在企业数据应用的场景下，人们最常用的可能就是监督式学习和非监督式学习的模型。
- 在图像识别等领域，由于存在大量的非标识的数据和少量的可标识数据，目前半监督式学习是一个很热的话题。
- 而强化学习更多的应用在机器人控制及其他需要进行系统控制的领域。

算法类似性



- 根据算法的功能和形式的类似性，我们可以把算法分类，比如说基于树的算法，基于神经网络的算法等等。
- 当然，机器学习的范围非常庞大，有些算法很难明确归类到某一类。而对于有些分类来说，同一分类的算法可以针对不同类型的问题。这里，我们尽量把常用的算法按照最容易理解的方式进行分类。