

1、机器学习定义

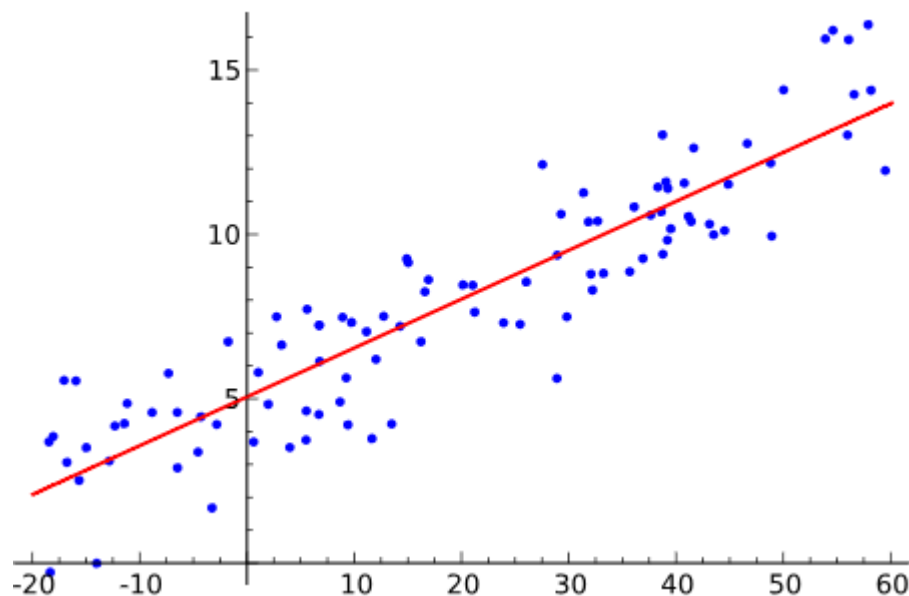
使用经验 E ，在完成任务 T 时，性能指标 P 相比以前得到增长，则认为从经验 E 有所学习。

2、机器学习分类

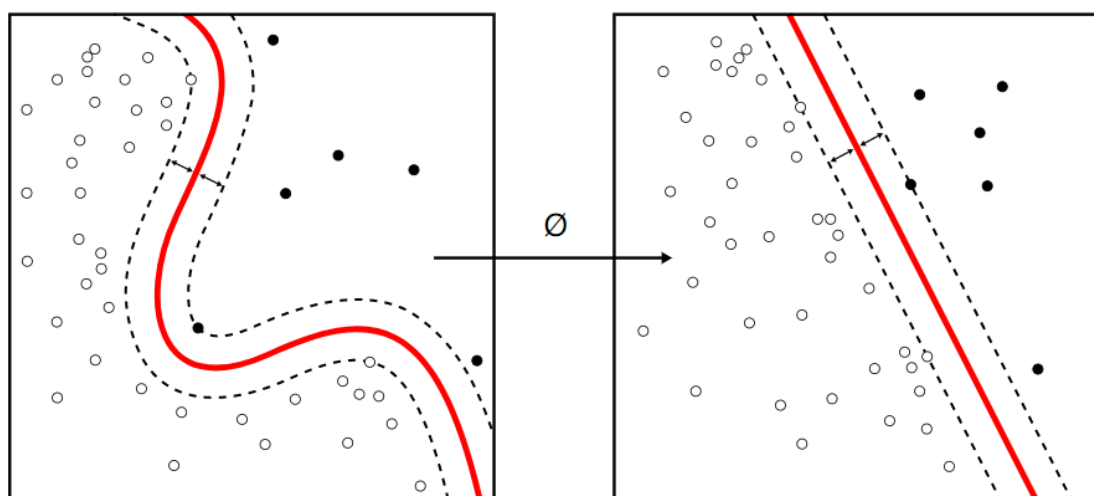
A、监督学习

教会计算机该去做什么，要完成这样似乎不可能的事情，首先你得有评判标准，即已知输入得到的正确的输出，此后，计算机根据输入与输出自动产生某种关系映射，在下次需要根据输入得到输出时，这个映射就有用武之地了。

监督学习分回归和分类（区别于无监督学习的聚类）。回归问题中，我们根据已有的输入可能得到无限多种输出。分类问题，根据输入将结果映射到离散的类别中。



回归问题

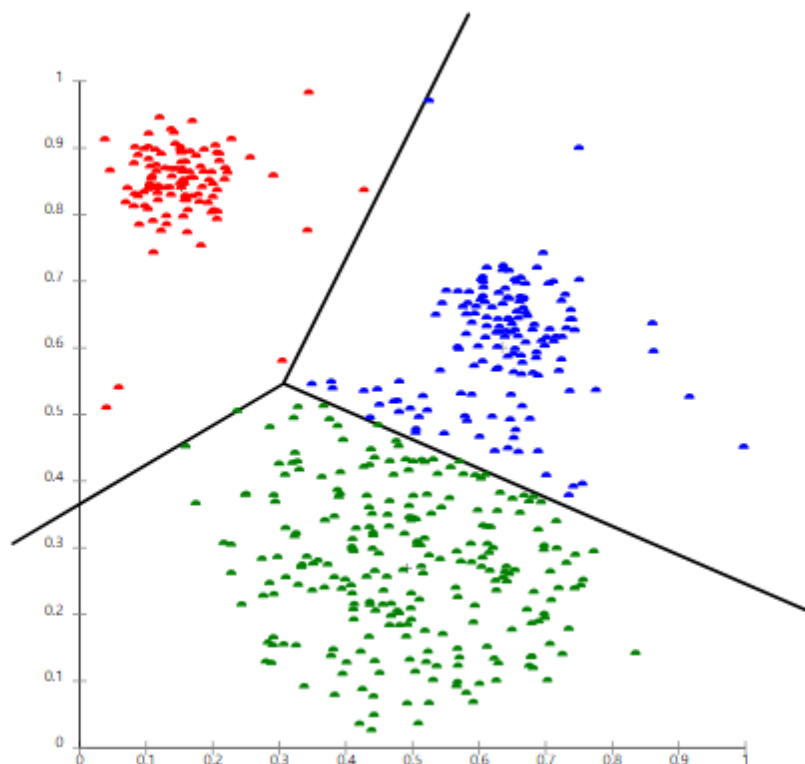


分类问题

B、无监督学习

我们并不知道将这一种输入得到的结果划分为哪一类，也不知道一共应划分多少种类，只能由计算机根据每个测试对象所包含的变量，自己判断它们之间的联系。

无监督学习包含聚类问题和非聚类问题。聚类问题，将数据集中相似度较大的数据划归为一组。非聚类问题，没有准确的定义，给一个简单的例子，从混杂的声音中区分出几种不同发声源的声音。



聚类问题

大体来讲，不论监督还是无监督，计算机都应根据某种映射方式或者区分方式，自动的由输入得到输出。

监督学习是由编程人员给定了训练集中隐含的某种关系，计算机再由此推导其他的输出。通常情况，回归问题中，我们选定多项式拟合训练集的方式，得到这种关系；分类问题，常见的有决策树算法。

而对于无监督学习，计算机想要根据测试对象的变量自动划分组，就必须得有一个评判标准，这个评判标准由编程人员给定，比如相似度度量方式中的欧氏距离、杰卡德相似系数和余弦相似度等，都可以判断样本之间相似程度。