

概念

机器学习致力于研究如何通过**计算**的手段，利用**经验**来改善系统自身的性能。在计算机系统中，**经验**通常以**数据**的形式存在

假设用P来评估计算机程序在某任务类T上的性能，若一个程序通过利用经验E在T中任务上获得了性能改善，则我们就说关于T和P，该程序对E进行了学习

基本术语

数据集 / 属性或特征 / 特征向量 / 维数 / 标签 / 分类 / 回归 / 聚类 / 监督学习 / 无监督学习 / 泛化

归纳偏好

任何一个有效的机器学习算法必有其归纳偏好，否则它将被假设空间中看似在训练集上“等效”的假设所迷惑，而无法产生确定的学习结果。归纳偏好可以看作学习算法自身在一个可能很庞大的假设空间中对假设进行选择的启发式。

我们需要清楚地意识到：脱离**具体**问题，空泛地谈论“什么学习算法更好”毫无意义，因为若考虑所有潜在的问题，则所有学习算法都一样好。要谈论算法的优劣，必须要针对具体的学习问题；在某些问题上表现好的学习算法，在另一些问题上却可能不尽人意，学习算法自身的**归纳偏好**于问题是否相配，往往会起到决定作用。