



随着“智能制造”概念的普及，人工智能技术的研究与应用变的越来越被人们关注。人工智能在制造中的运用已经成为实现制造的知识化、自动化、柔性化以实现对市场的快速响应的关键。机器学习无疑是最有希望实现这个“智能”的研究方向之一。斯坦福大学的“Stanford Engineering Everywhere”免费提供学校里最受欢迎的工科课程给全世界的学生和教育工作者。得益于这个项目，我们有机会和全世界站在同一个数量级的知识起跑线上。

课程共 20 集，网易公开课已经全部翻译完成。讲师：Andrew Ng。

【第 1 集】机器学习的动机与应用

简介：机器学习的动机与应用、Logistic 类、机器学习的定义、监督学习概观、学习理论概述、非监督学习概述和强化学习概述。



【第 2 集】监督学习应用

简介：监督学习应用——自主推导，ALVINN 系统，线性回归，梯度下降，组梯度下降，随机梯度下降，标准方程推导。



【第 3 集】欠拟合与过拟合的概念

简介：欠拟合与过拟合的概念，参数化及非参数化算法概念，局部加权回归，对于线性模型的概率解释，Logistic 回归，感知器。



【第 4 集】牛顿方法

简介：介绍了牛顿方法，可以代替梯度上升算法用来计算函数的最大值；之后以高斯分布和伯努利分布为例介绍了指数分布函数族；最后以指数分布函数族为基础，引出了广义线性模型，可以通过指定概率分布直接推导出模型。



【第 5 集】生成学习算法

简介：一类新的学习算法——生成学习算法，并详细地讲解了该算法的一个例子：高斯判别分析；之后对生成学习算法与之前的判别学习算法进行了对比；最后介绍了一个适合对文本进行分类的算法——朴素贝叶斯算法，并结合该算法讲述了一种常用的平滑技术——Laplace 平滑。



【第 6 集】朴素贝叶斯算法

简介：先介绍了两种朴素贝叶斯算法的事件模型；之后介绍了神经网络算法；在最后介绍了两个重要的概念：函数间隔和几何间隔。基于这两个概念提出了一个线性分类算法：最大间隔分类器算法。该算法用于引出一个非常重要的非线性分类算法：支持向量机。



【第 7 集】最优间隔分类器问题

简介：首先提出了原始的优化问题：最优间隔分类器问题；之后介绍了对偶问题的



概念和 KKT 条件；基于原始优化问题的对偶问题的分析，介绍了 SVM 算法；最后对 SVM 算法进行了评价，以引出下节课对核方法的介绍。

【第 8 集】顺序最小优化算法

简介：核的概念——它在 SVM 以及许多学习算法中都有重要的应用；之后介绍了 l_1 norm 软间隔 SVM——它是一种 SVM 的变化形式，可以处理非线性可分隔的数据；最后介绍了 SMO 算法——一种高效的可以解决 SVM 优化问题的算法。



【第 9 集】经验风险最小化

简介：主要介绍了模型选择中的一种常见现象——偏差方差权衡。为了解释该概念，首先介绍了两个重要的引理——联合界引理和 Hoeffding 不等式；之后定义了两个重要的概念——训练误差和一般误差，并提出了一种简化的机器学习算法模型——经验风险最小化（ERM）；最后基于这些概念对 ERM 结果的理论上界进行了证明，并基于上界分析对偏差方差权衡进行了解释。



【第 10 集】特征选择

简介：VC 维的概念——该概念能够将关于 ERM 一般误差的界的结论推广到无限假设类的情形；模型选择问题——具体介绍了交叉验证方法以及几种变形；特征选择问题——具体介绍了两类方法：封装特征选择和过滤特征选择。



【第 11 集】贝叶斯统计正则化

简介：贝叶斯统计和规范化；简单介绍了在线学习的概念；机器学习算法设计中的问题诊断技巧；两种分析技巧：误差分析与销蚀分析；两种应用机器学习算法的方式与适用场景。



【第 12 集】K-means 算法

简介：无监督学习的内容。首先介绍了 k -means 聚类算法；混合高斯模型，它是最大期望算法（EM）的一种特例；引入了 Jensen 不等式，利用 Jensen 不等式引出了 EM



算法的一般形式。

【第 13 集】高斯混合模型

简介：对混合高斯模型在 EM 算法下的结论进行了推导，并且介绍了 EM 算法在混合贝叶斯模型中的应用。最后介绍了因子分析算法。该算法可以进行高维数据下样本数目较少的情况下的模型拟合。



【第 14 集】主成分分析

简介：本讲继续上一讲的内容，详细地讲解了因子分析问题对应的 EM 算法的步骤推导过程，并重点提出了其中应该注意的问题。之后介绍了主成分分析（PCA）的算法原理和主要应用。该算法是一种常用的降低数据维度的算法。



【第 15 集】奇异值分解

简介：主成分分析 PCA，及举出利用 PCA 找出相似文档的例子；SVD（奇异值分解）；无监督算法和因子分析；ICA（独立成分分析算法）和 CDF（累积分布函数），并复习了高斯分布的知识；最后举了几个应用 ICA 的例子。



【第 16 集】马尔可夫决策过程

简介：主要介绍了监督学习；然后引出强化学习的知识，用“使直升机飞翔”的例子阐述强化学习；介绍了马氏决策过程（MDP），由此引出来的两个解决最优策略和最优回报的算法；最后重点介绍了“值迭代”和“策略迭代算法”的实施，以及比较了它们的优缺点。



【第 17 集】离散与维数灾难

简介：继续马氏决策过程（MDP），以及解决状态 MDP 的算法，主要详细介绍了拟合值迭代算法（fitted value iteration）和近似政策迭代（approximate policy iteration）这两种算法，并通过具体的例子和求解的方式来原因说明这两种算法。



[第18集] 线性二次型调节控制

简介：控制 NVP 算法，非线性动力学系统；在动力系统的模型和线性二次型调节控制 (linear quadratic regulation control)，导出一些处理情况的函数；还包含线性模型的建立，非线性模型的线性化的知识。



[第19集] 微分动态规划

简介：强化学习算法，引入调试强化学习算法，介绍 Kalman 滤波器微分动态规划，卡尔曼滤波与 LQR 控制结合的一种算法 (LQG 控制算法，线性二次高斯)，并比



较了高斯分布和卡尔曼滤波的效率问题。

[第20集] 策略搜索

简介：学习和复习了强化学习算法，讲述了一些 POMDPs(部分可观察马氏决策过程)的知识，完全可观察 MDP 的知识；接下来介绍了策略搜索算法 (其中包括两种算法：Reinforced 和 Pegasus)；最后，介绍了与这门课程相关的一些课程，并给学生提出一些希望。 



名词解释

机器学习 (Machine Learning, ML) 是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析和算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。

机器学习是人工智能的核心，是使计算机具有智能的根本途径，其应用遍及人工智能的各个领域，主要使用归纳、综合而不是演绎。

一、研究意义

顾名思义，机器学习是研究如何使用机器来模拟人类学习活动的一门学科。更为严格的提法是：机器学习是一门研究机器获取新知识和新技能，并识别现有知识的学问。这里所说的“机器”，指的就是计算机，电子计算机，中子计算机、光子计算机或神经计算机等。

机器学习有下面几种定义：“机器学习是一门人工智能的科学，该领域的主要研究对象是人工智能，特别是如何在经验学习中改善具体算法的性能”。“机器学习是对能通过经验自动改进的计算机算法的研究”。“机器学习是用数据或以往的经验，以此优化计算机程序的性能标准。”一种经常引用的英文定义是：A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

学习是人类具有的一种重要智能行为，但究竟什么是学习，长期以来却众说纷纭。社会学家、逻辑学家和心理学家都各有其不同的看法。

机器能否像人类一样能具有学习能力呢？1959 年美国的塞缪尔 (Samuel) 设计了一个下棋程序，这个程序具有学习能力，它可以在不断的对弈中改善自己的棋艺。4 年后，这个程序战胜了设计者本人。又过了 3 年，这个程序战胜了美国一个保持 8 年之久的常胜不败的冠军。这个程序向人们展示了机器学习的能力，提出了许多令人深思的社会问题与哲学问题。

机器的能力是否能超过人的，很多持否定意见的人的一个主要论据是：机器是人造的，其性能和动作完全是由设计者规定的，因此无论如何其能力也不会超过设计者本人。这种意见对不具备学习能力的机器来说的确是对的，可是对具备学习能力的机器就值得考虑了，因为这种机器的能力在应用中不断地提高，过一段时间之后，设计者本人也不知它的能力到了何种水平。

二、主要策略

学习是一项复杂的智能活动，学习过程与推理过程是紧密相连的。按照学习中使用推理的多少，机器学习所采用的策略大体上可分为 4 种——机械学习、通过传授学习、类比学习和通过事例学习。学习中所用的推理越多，系统的能力越强。

三、基本结构

环境向系统的学习部分提供某些信息；学习部分利用这些信息修改知识库，以增进系统执行部分完成任务的效能；执行部分根据知识库完成任务，同时把获得的信息反馈给学习部分。在具体的应用中，环境、知识库和执行部分决定了具体的工作内容，学习部分所需要解决的问题完全由上述三部分确定。 