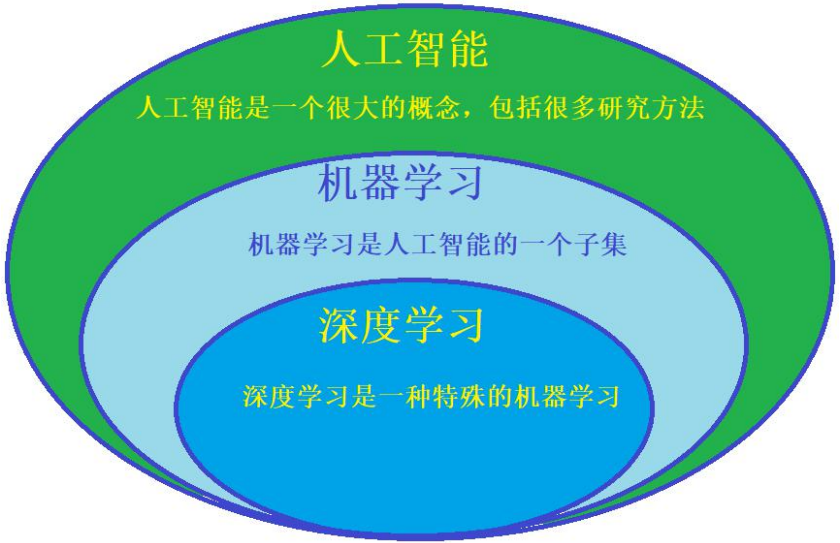
PART B

机器学习概论

B 人工智能、机器学习、深度学习的关系

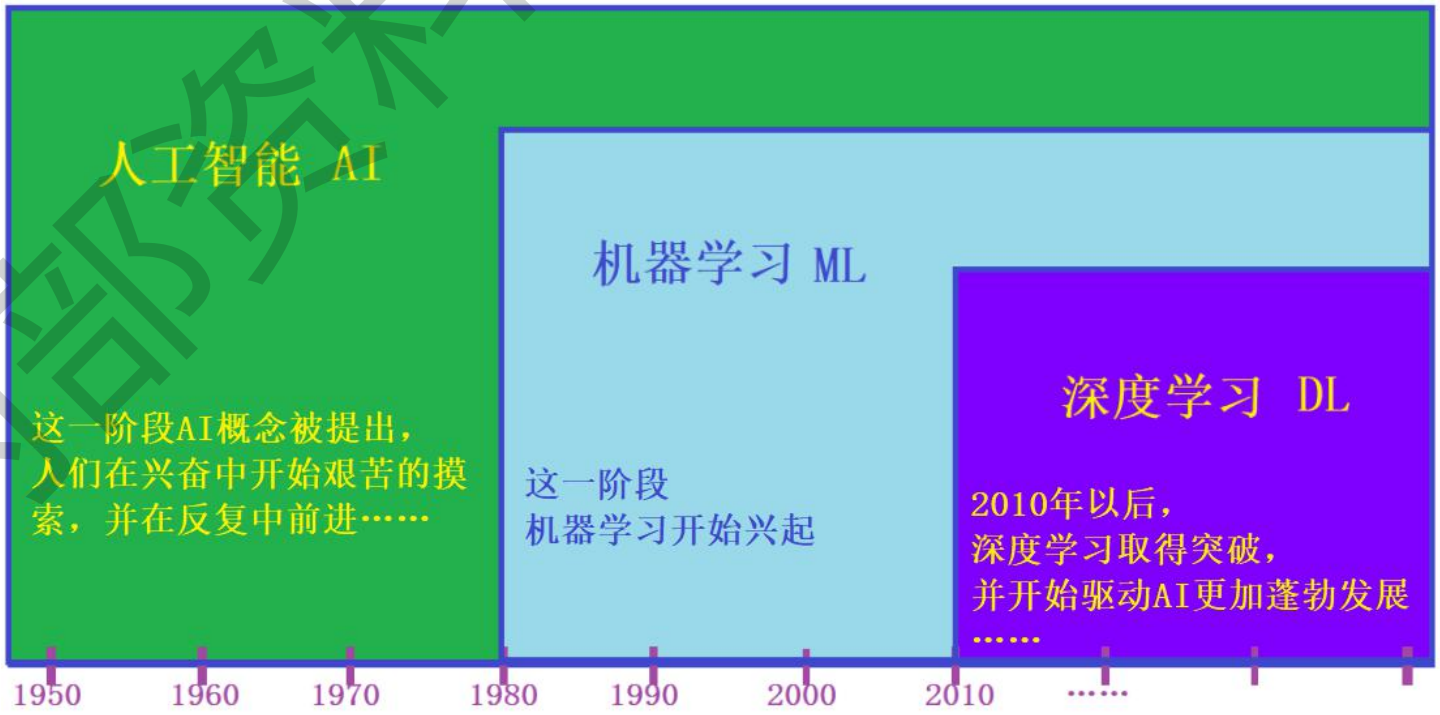
人工智能、机器学习、深度学习三者关系：

- 人工智能是一个非常广泛的概念，
- 里面其中一个子集就是机器学习，
- 机器学习的一个子集是深度学习。



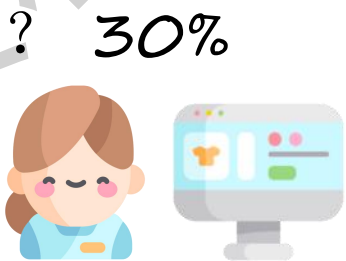
B 人工智能、机器学习、深度学习的关系

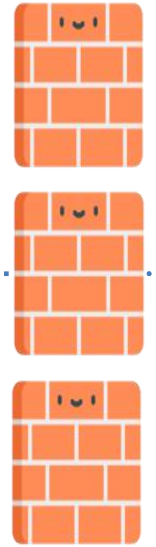
从人工智能的发展历程也可以展现三者之间的关系



B 图灵测试

艾伦.图灵 (Alan Turing, 1912-1954),
英国数学家, 计算机科学之父





?

人工

机器人

B 机器学习的工作过程

B 机器学习的工作过程

f( )=狗

f( )=狗

f( )=狗

f( )=狗

f( )= ?

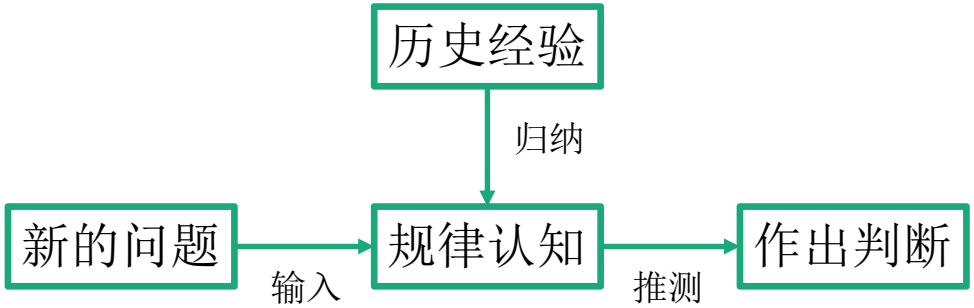
f( )= ?

难点：

泛化问题

B 机器学习的工作过程

类比人类的学习过程



B 机器学习的工作过程

输入数据：历年高考真题、模拟题

构建模型：解题方法

新的数据：今年高考（新题）

表现评估：高考成绩

难点：需要在 未见过 的任务上表现良好

机器学习目标：找到对应场景的函数

- 语音识别
- 图像识别
- 下棋

f() = “How are you”

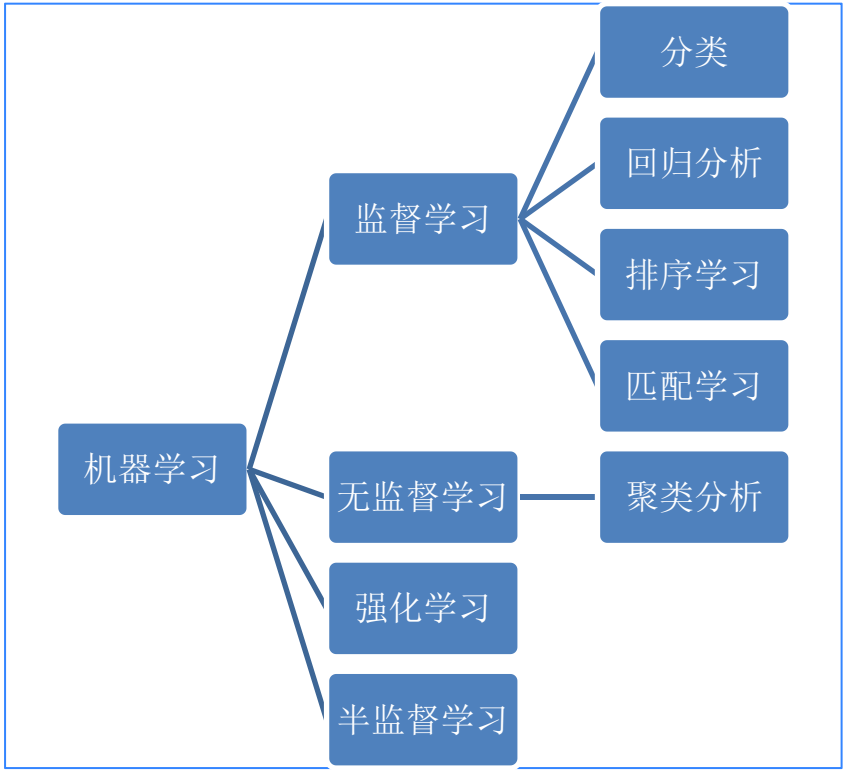
f() = “cat”

f() = “5-5” （落子的位置）

B 机器学习分类

机器学习分类：

- 监督学习
- 无监督学习
- 半监督学习
- 强化学习



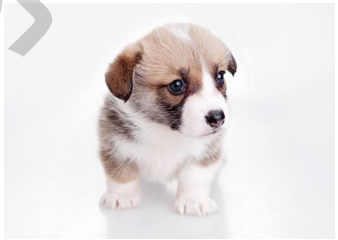
B 监督学习

有监督学习（supervised learning）：从给定的**有标注的训练数据集**中学习出一个函数（模型），用这个学习出的函数（模型）来对新数据进行预测。

如何理解“有监督学习”？比如考试试题，有标准答案。

猫

狗



狗

乌龟

B 监督学习

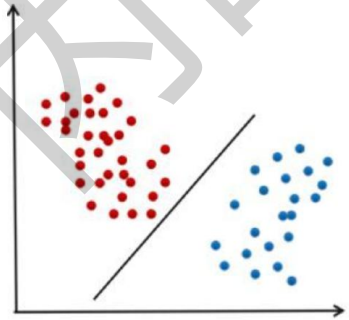
有监督学习（supervised learning）： 从给定的**有标注的训练数据集**中学习出一个函数（模型），用这个学习出的函数（模型）来对新数据进行预测。

如何理解“有监督学习”？比如考试试题，有标准答案。

常见的有监督学习任务：**分类**和**回归**

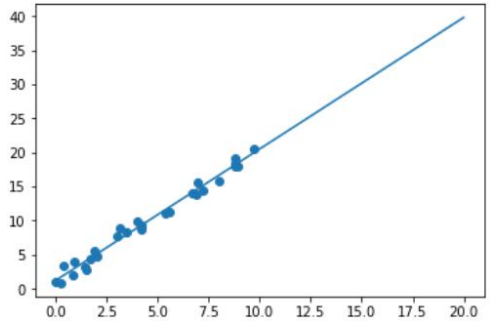
分类： 输出的是类别标签

输入：猫的图片；狗的图片
输出：猫？狗？



回归： 输出的是实数

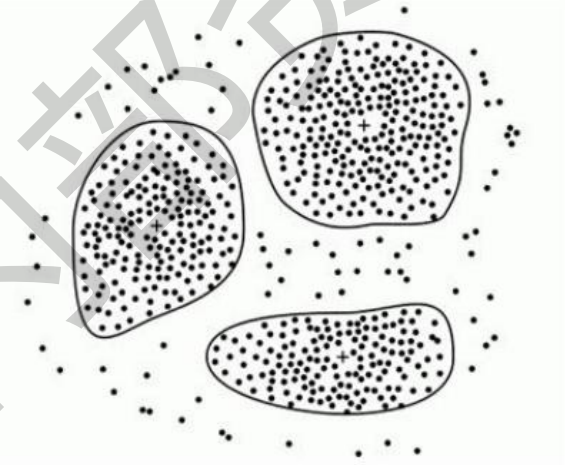
输入：房屋面积
输出：房屋的价钱

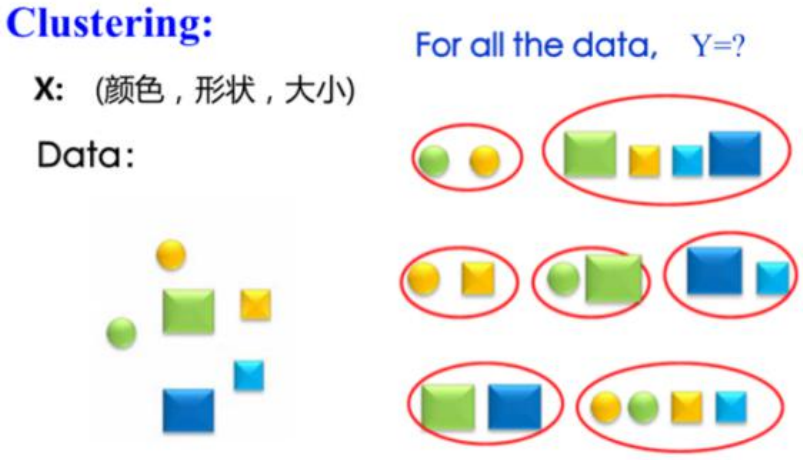


B 无监督学习

无监督学习（unsupervised learning）：没有标注的训练数据集，所有数据只有特征向量没有标签，需要根据样本统计规律来进行分析。比如聚类问题。

如何理解“无监督学习”？比如考试试题，没有标准答案，那么就意味着可能有多种结果。





B 无监督学习

无监督学习（unsupervised learning）：没有标注的训练数据集，所有数据只有特征向量没有标签，需要根据样本统计规律来进行分析。比如聚类问题。

如何理解“无监督学习”？比如考试试题，没有标准答案，那么就意味着可能有多种结果。

系统需要将如下新闻稿件进行归纳聚类进而推送给对应的编辑

火箭队大胜凯尔特人队

北京国安主场大胜

热火队的主教练换帅

鲁能泰山3:1力克对手

杨幂最新作品获得观众好评

本山大叔暗示今年可能上春晚

沈腾与马丽不为人知的故事

金州勇士获得新秀球员未来可期

B 半监督学习

半监督学习（semi-supervised）：在训练阶段结合了**大量未标记的数据**和**少量标签数据，**进行数据的分类学习。

如何理解“半监督学习”？比如考试试题，给了很少部分题的标准答案，但大部分题是没有标准答案的。

一句话解释：

- 有监督是所有的训练文本为人工标记的；
- 半监督是一部分是有标记的，剩下的为无标记的（一般无标记>>有标记）；
- 无监督就是全部都是无标记的。

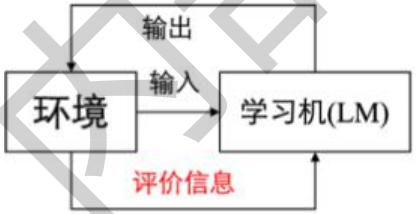
B 强化学习

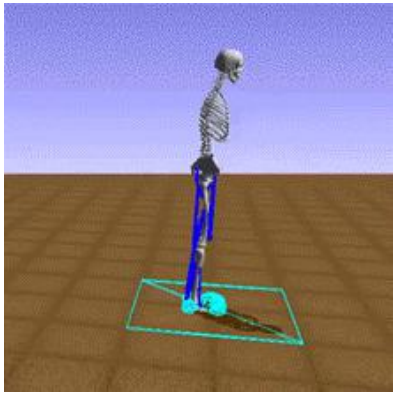
强化学习（Reinforcement Learning）：在学习的过程中，**外部环境对输出只给出评价信息而非正确答案**，学习机通过受奖励的动作来改善自身的性能。

强化学习就是让计算机实现从一开始什么都不懂，通过不断地尝试，从错误中学习，最后找到规律。

举个通俗的例子：

你要训练一只小老鼠，让他学会在迷宫中找到出口。那么在训练时，如果他走出了正确的路线，就会给它奖励（糖），走错了，就给他适当惩罚。久而久之，他就本能的学会了如何找到出口的路。

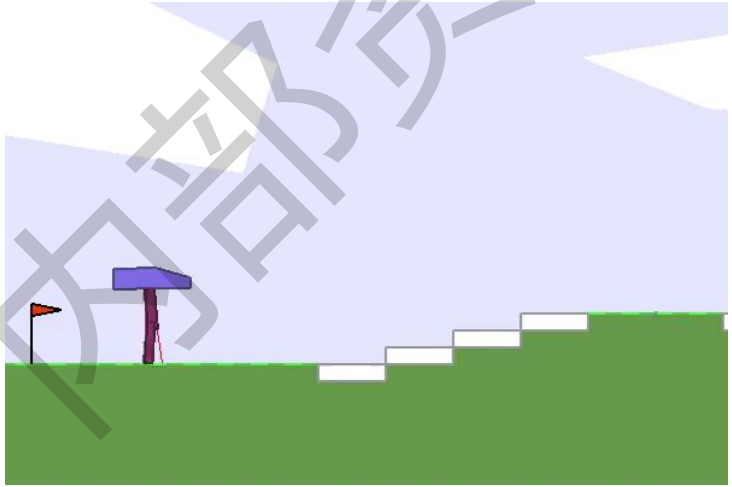


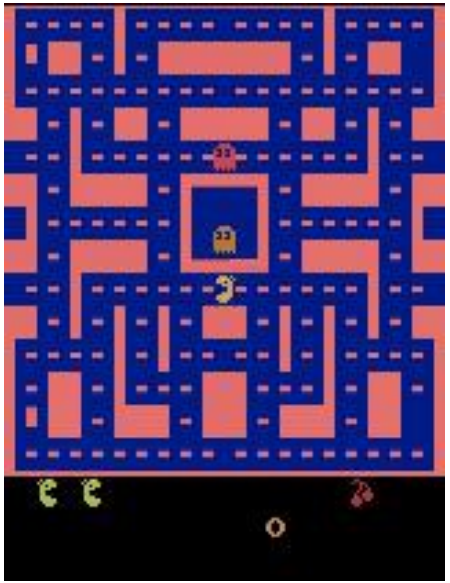


B 强化学习

强化学习（Reinforcement Learning）：在学习的过程中，**外部环境对输出只给出评价信息而非正确答案**，学习机通过受奖励的动作来改善自身的性能。

强化学习就是让计算机实现从一开始什么都不懂，通过不断地尝试，从错误中学习，最后找到规律。

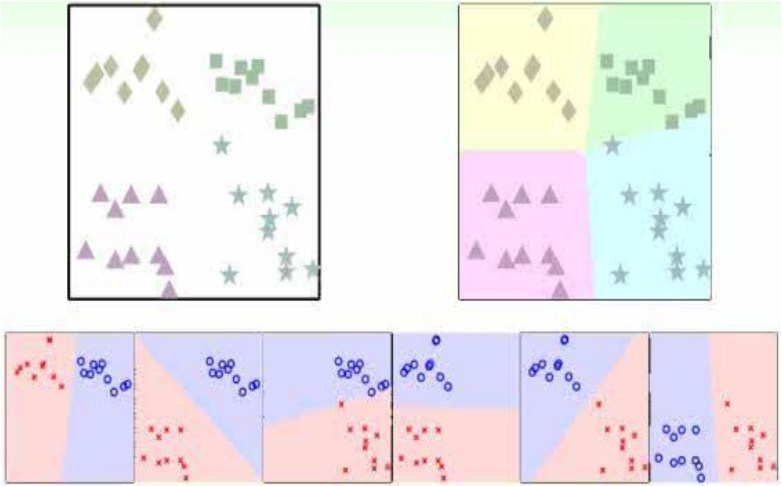




B 分类问题

分类问题

- 通常情况下，数据集有N个训练对象x_1，…，x_n对于每个对象，我们还提供了一个标签 t_n 描述对象 n 属于哪个类别。t_n通常取整数值。
- 每个对象都是一个 D 维向量。
- **我们的目标**：对于给定的对象 x_{new} ,预测他的类别 t_{new}。



贝叶斯分类器

贝叶斯分类器的分类原理是通过某对象的先验概率，利用贝叶斯公式计算出其后验概率，即该对象属于某一类的概率，选择具有最大后验概率的类作为该对象所属的类。

逻辑回归

利用已知的自变量来预测一个离散型因变量的值（像二进制值0/1，是/否，真/假）。简单来说，它就是通过拟合一个逻辑函数（logit-fuction）来预测一个事件发生的概率。

K-近邻算法

该方法的思路是：如果一个样本在特征空间中的k个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别，则该样本也属于这个类别。

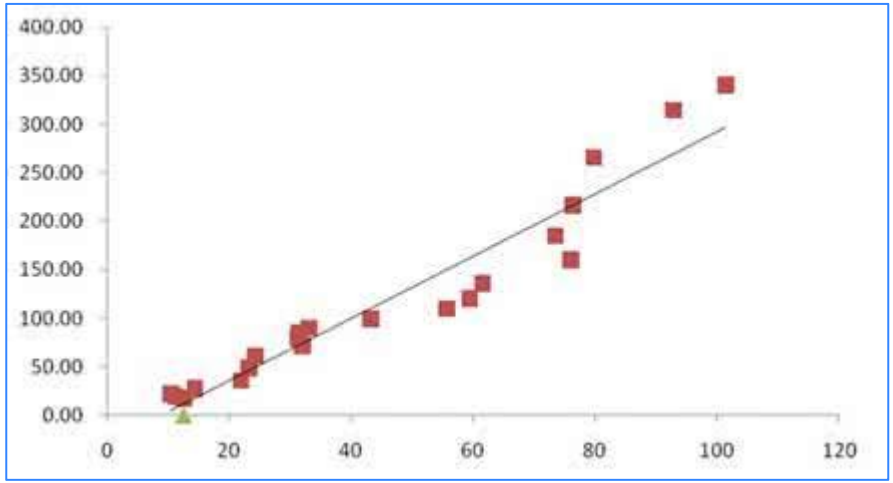
支持向量机

支持向量机方法是建立在统计学习理论的VC维理论和结构风险最小原理基础上的，根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折中，以求获得最好的推广能力 。

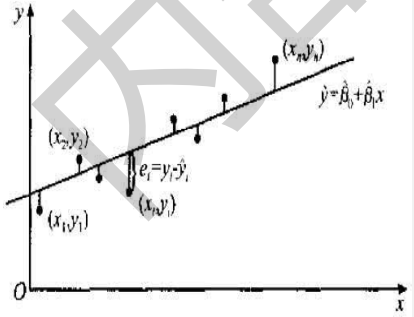
B 回归分析

回归分析

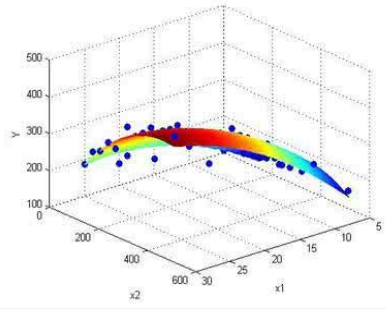
回归分析是一种预测性的建模技术，它研究的是**因变量（目标）和自变量（预测器）之间的关系。**这种技术通常用于预测分析，时间序列模型以及发现变量之间的因果关系。



单元回归



多元回归



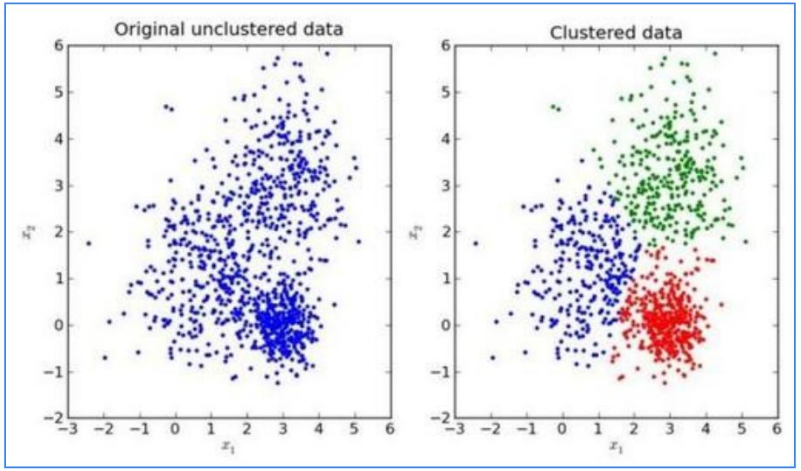
最小二乘法

它通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。

最大似然法

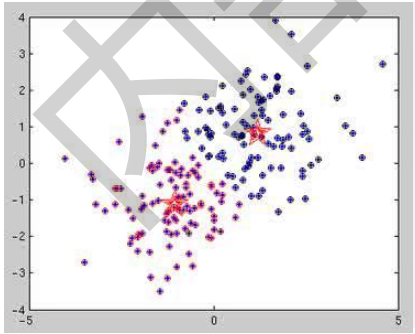
算法思想：当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的概率最大。

B 聚类分析



聚类分析：

聚类分析的目标是，创建满足处于同一组内的对象相似，不同组内对象相异的对象分组。

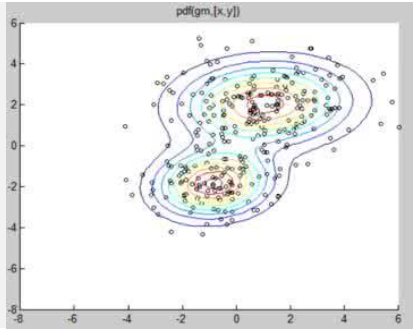


K-均值算法

算法是硬聚类算法，是典型的基于原型的目标函数聚类方法的代表，它是数据点到原型的某种距离作为优化的目标函数，利用函数求极值的方法得到迭代运算的调整规则。

混合模型

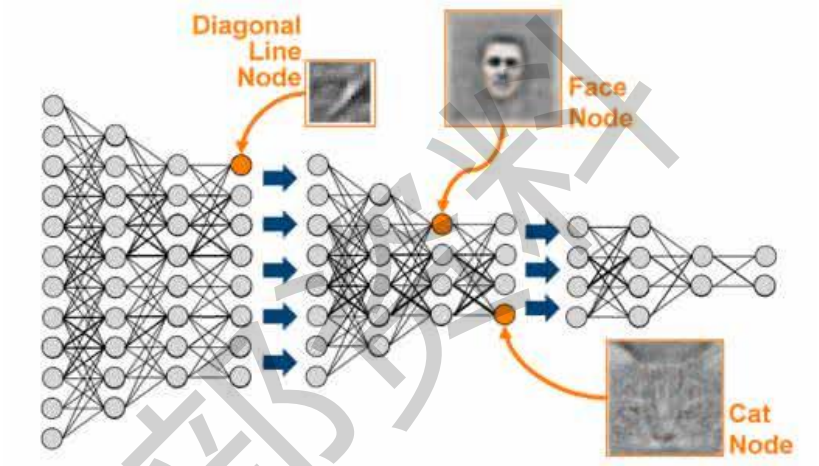
统计混合学将每个类表示为一个概率密度，这种归纳引出了一个强大的方法，我们可以在几乎任何类型的数据集中医各种图形来建模聚类。



机器学习常用算法

PART C

人工智能与深度学习



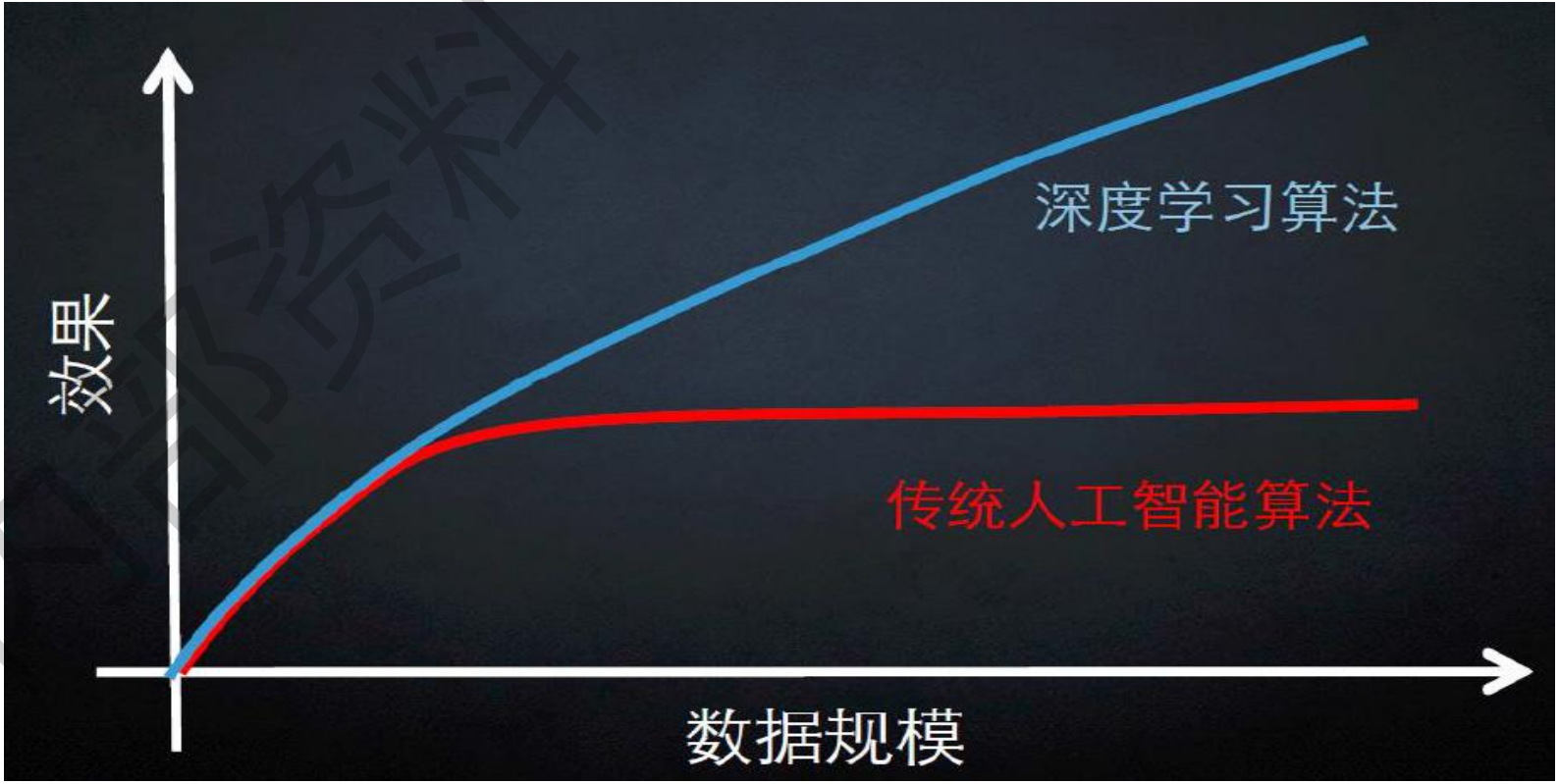
深度学习

- 深度学习是机器学习研究中的一个新的领域；
- 其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本。
- 其源于人工神经网络的研究。
- 含多隐层的多层感知器就是一种深度学习结构。
- 深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。

全连接神经网络（DNN）

卷积神经网络（CNN）

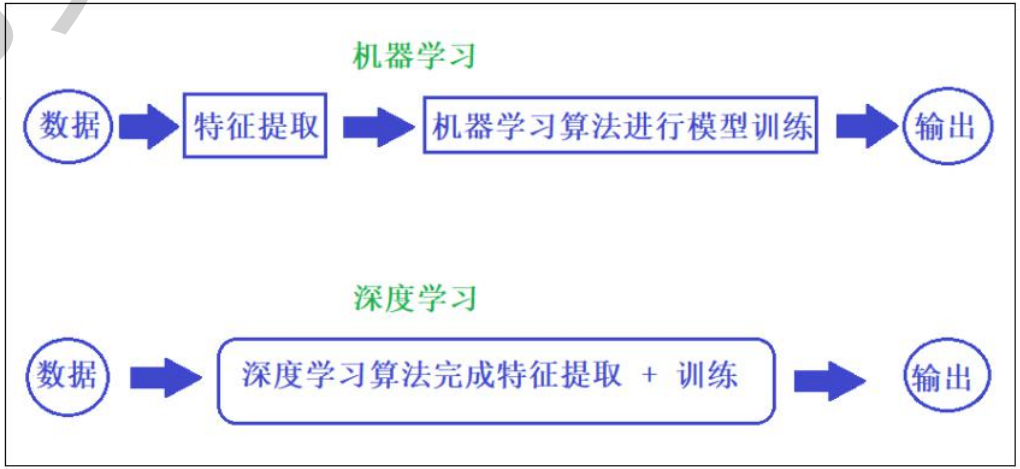
循环神经网络(RNN)


效果
深度学习算法
传统人工智能算法
数据规模

机器学习深度学习的区别

机器学习和深度学习的主要区别在于**特征提取**。

- 在传统的机器学习算法中，我们首先需要用一些算法（比如PCA、LDA等）来进行特征的提取，然后再用机器学习算法（如svm等）进行模型训练。
- 在深度学习中，特征由算法本身自动完成提取，通常不需要我们另外写一个算法来进行特征提取。比如CNN网络中，CNN的作用就实现了特征的提取。



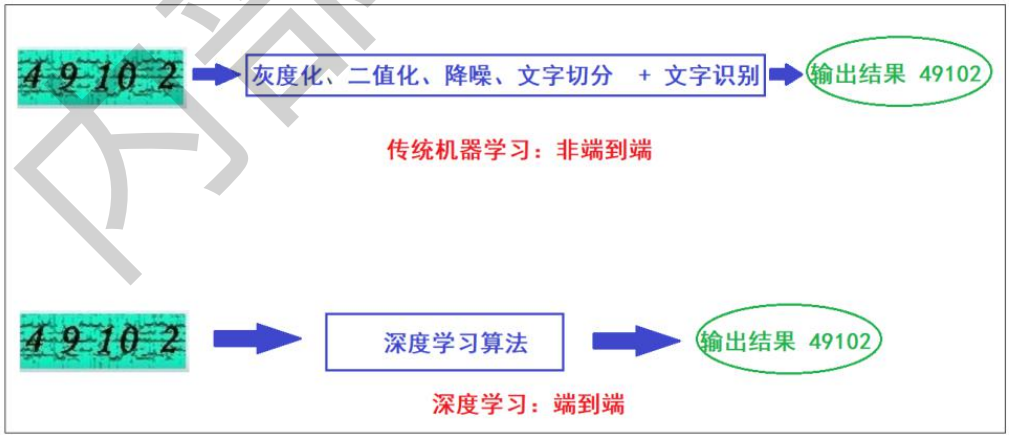
机器学习深度学习的区别

机器学习与深度学习的区别还表现在**解决问题的方式**。

- 传统机器学习通常会将问题分解为多个子问题，并把逐个子问题解决后，最后结合所有子问题的结果获得最终结果。

- 而深度学习提倡直接的端到端的解决问题。

比如在做OCR（文字识别）任务时：



图：深度学习实现了端到端的学习

机器学习深度学习的区别

机器学习与深度学习的区别还表现在**可解释性**。

- 机器学习的可解释性很强，许多传统的机器学习算法有明确的数学规则，解释起来相对容易。比如说线性回归，逻辑回归、决策树、svm等这些算法解释起来就很容易。

- 但是深度学习的可解释性就没有那么强了。深度神经网络更像是一种“黑箱子”，网络里面具体每一层是怎么操作的，神经元做了什么，很多时候是不明确的。深度学习的可解释性是一个热门研究话题。

THANK YOU