

人工智能



机器学习 Machine Learning



主要内容

□ 什么是机器学习

□ 有监督学习

- 分类问题 是什么 图像识别，语音识别
- 回归问题 是多少 预测收入，预测房价

□ 无监督学习

- 降维 抽象和压缩
- 聚类 怎么划分

□ 强化学习

- 怎么做决策，如博弈与游戏，自动驾驶，机器人控制

□ 神经网络和深度学习

机器学习（Machine Learning）

- (机器学习系统)利用**经验**来改善**计算机系统**自身的**性能**。

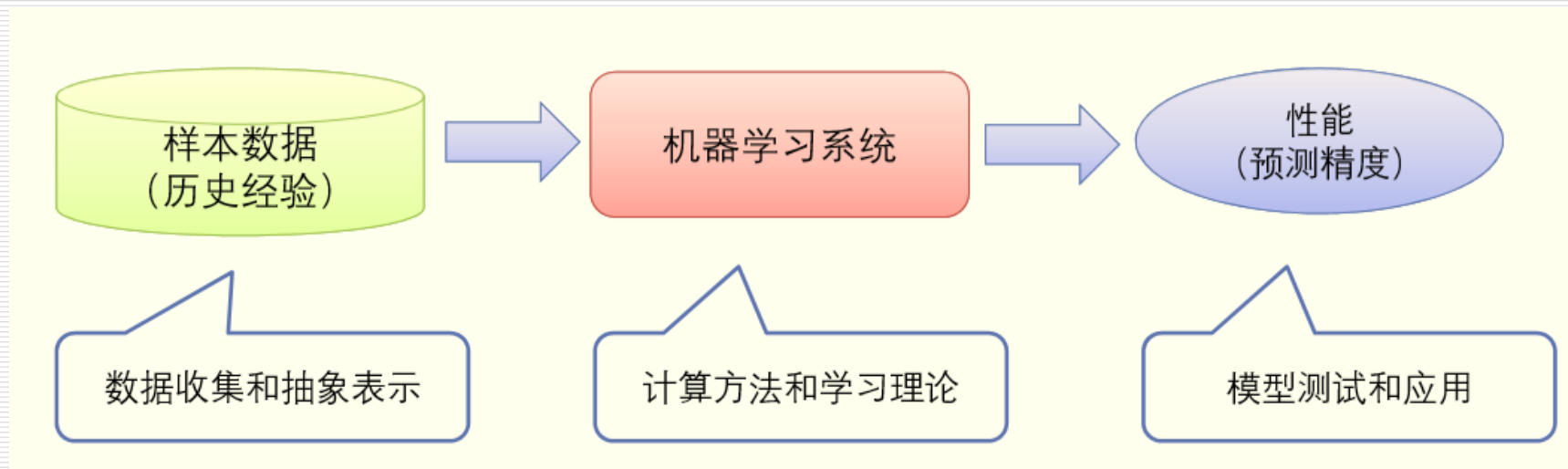
—— Mitchell, Machine Learning

- 学习系统使用样本数据来建立并更新模型，并以可理解的符号形式表达，使经过更新后的模型处理同源数据的能力得以提升。

—— Donald Michie (1991)

机器学习框架

- ❑ 经验。。。计算机系统。。。性能
- ❑ 样本数据。。。模型。。。能力



图像分类的机器学习

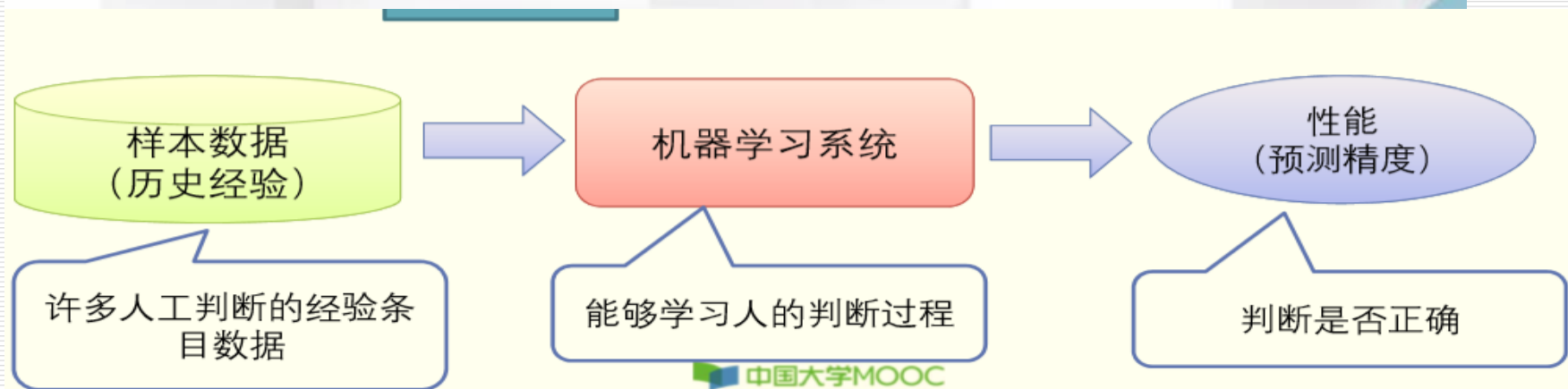
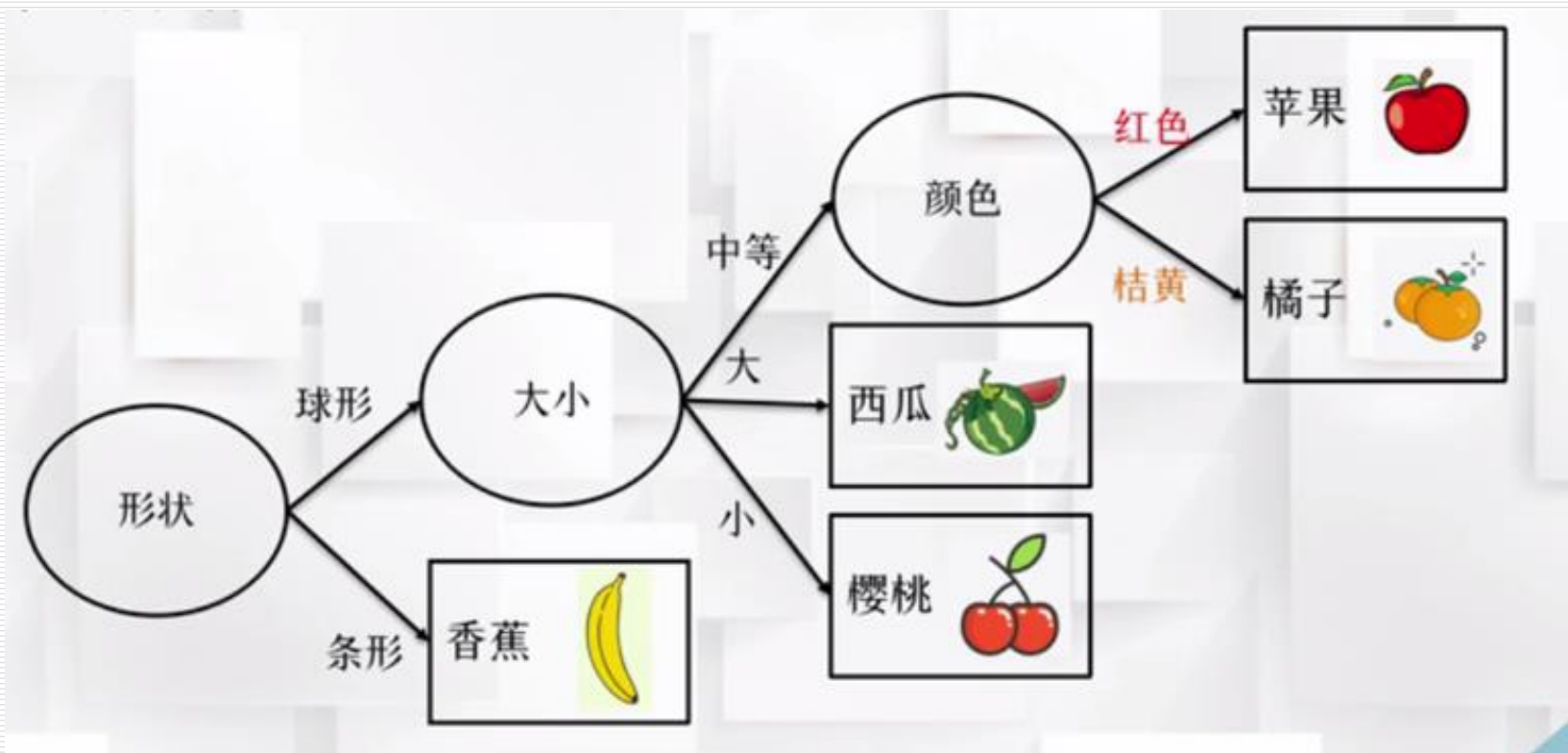


身高预测的机器学习

如何根据年龄、性别、
体重等等特征，来预测
小孩的身高？



归纳规律



机器学习在人工智能中的定位

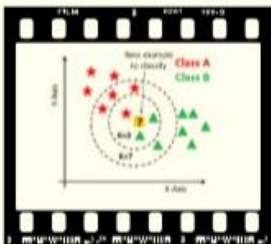
- 从方法与技术研究发展的角度，人工智能经历了两个阶段
 - 1980之前-逻辑推理，人工建立规则、知识库
 - 1980之后-机器学习，模仿人的学习能力，从实例中进行学习获得经验和知识
- ✓ 从以“推理”为重点到以“知识”为重点，再到以“学习”为重点
- ✓ 机器可以自动“学习”的算法，即从数据中自动分析获得规律，并利用规律对未知数据进行预测的算法。目前，机器学习=“分类”
- ✓ 人工智能 > 机器学习 > 深度学习



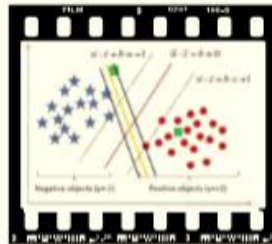
ML的发展历程



塞缪尔写出第一个跳棋程序，能够从人机对弈中学习规则。



最近邻方法被提出，计算机具有了简单识别类型的能力



支持向量机取得巨大成功，统计学习方法成为主流。



邓力提出的深度学习语音识别方法取得突破，语音技术走向成熟。

1952

1957

1967

1997

1999

2006

2010

2016

Frank 设计出了第一个神经网络——感知机



深蓝击败卡斯帕罗夫，计算机学习能力达到新高



Geoffrey Hinton 正式提出了深度学习概念，开启机器学习新纪元



谷歌AlphaGo打败了围棋专业选手李世石、柯洁

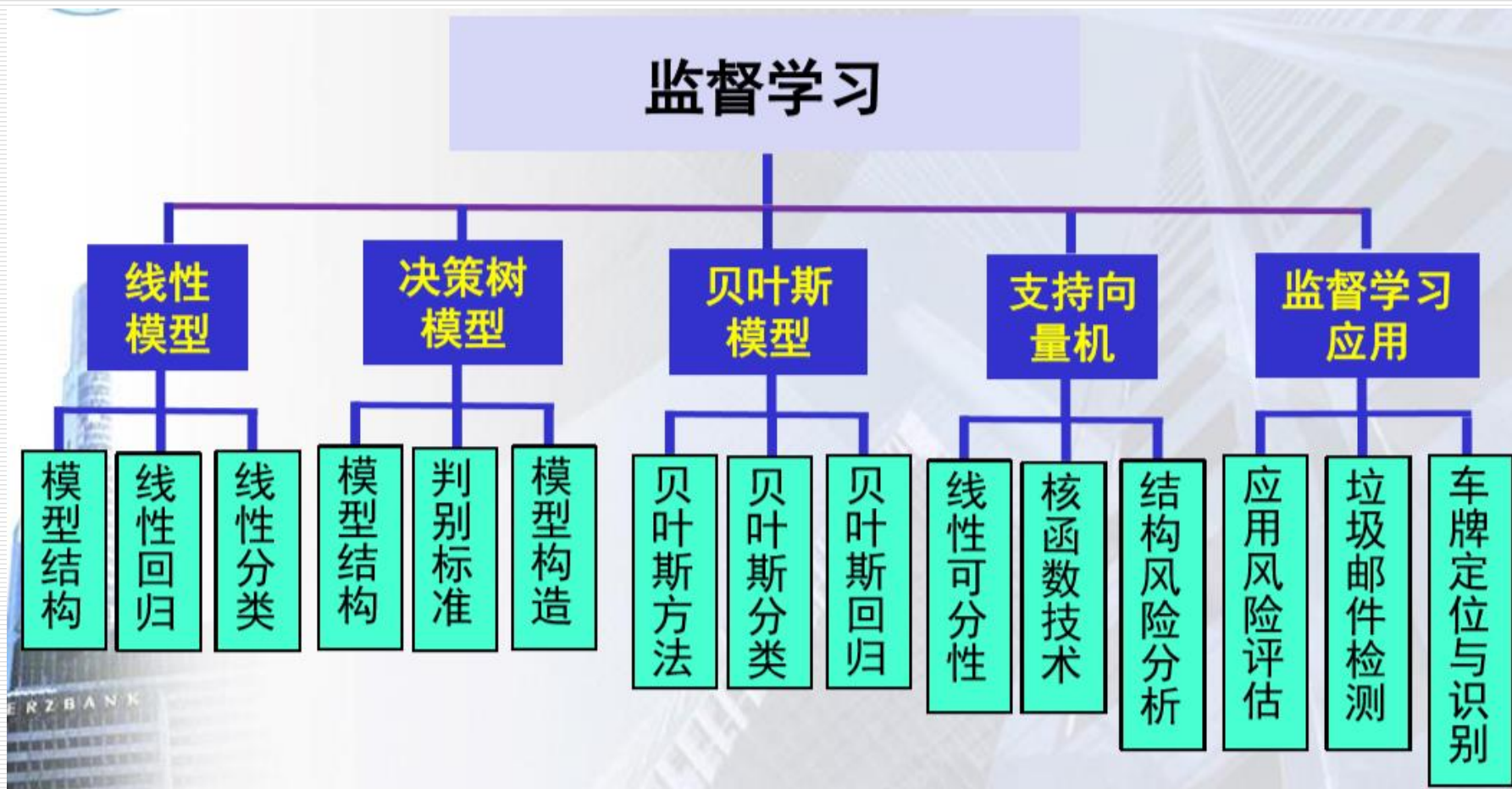


ML的编程语言

Platform	2019 % share	2018 % share	% change
Python	65.8%	65.6%	0.2%
R Language	46.6%	48.5%	-4.0%
SQL Language	32.8%	39.6%	-17.2%
Java	12.4%	15.1%	-17.7%
Unix shell/awk	7.9%	9.2%	-13.4%
C/C++	7.1%	6.8%	3.7%
Javascript	6.8%	na	na
Other programming and data languages	5.7%	6.9%	-17.1%
Scala	3.5%	5.9%	-41.0%
Julia	1.7%	0.7%	150.4%
Perl	1.3%	1.0%	25.2%
Lisp	0.4%	0.3%	46.1%

基本术语

- 有监督学习、无监督学习、强化学习



有监督学习

- 有监督学习的基本框架
- 有监督学习的典型方法
 - K邻近算法
 - 朴素贝叶斯算法
 - 决策树算法
 - 支持向量机

有监督学习

有监督机器学习的数学描述

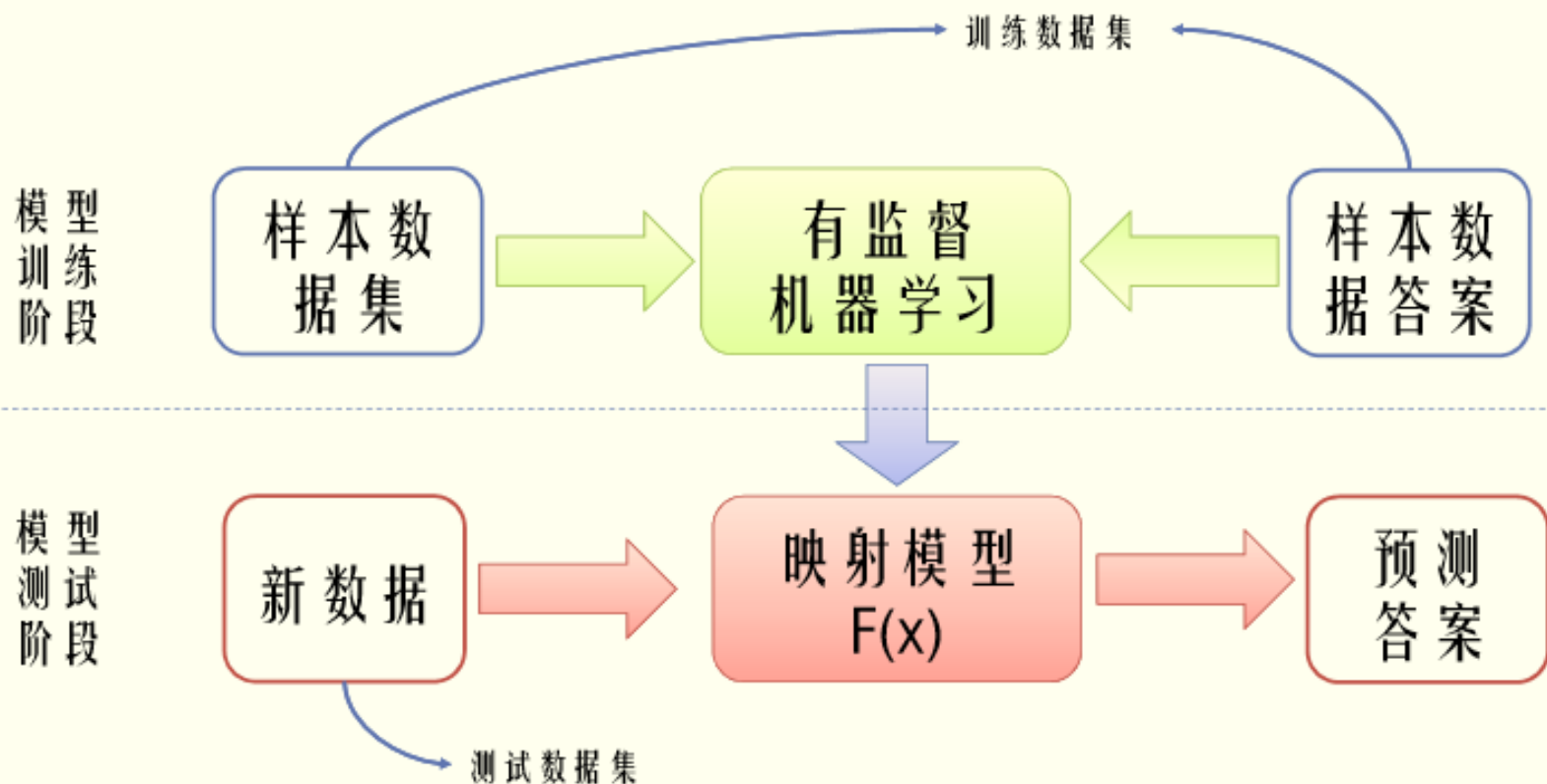
- 给定一些训练样本: $\{(x_i, y_i), 1 \leq i \leq N\}$;
- 其中 x_i 是输入, y_i 是需要预测的目标,
- 计算机自动寻找一个决策函数 $f(\cdot)$ 来建立 x 和 y 之间的关系:

$$\hat{y} = f(\phi(x), \theta),$$

其中: ϕ 是一种函数变换, θ 是模型参数

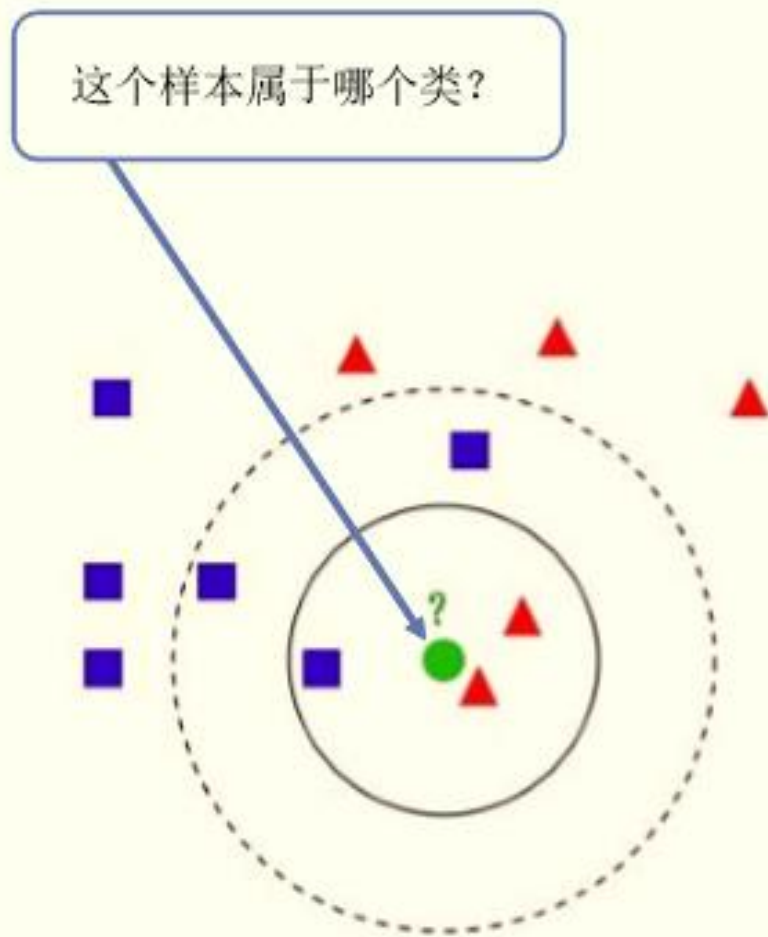
- 选取参数 θ 使得 \hat{y} 与预测目标 y 之间的差别尽可能小

有监督学习系统的运行过程



K邻近算法

- 给定一个训练数据集，无需训练
- 对新的输入实例，在训练数据集中找到与该实例最邻近的K个实例
- 根据这K个实例的类别做投票，哪个类的实例最多，就把该输入实例分类到这个类中。

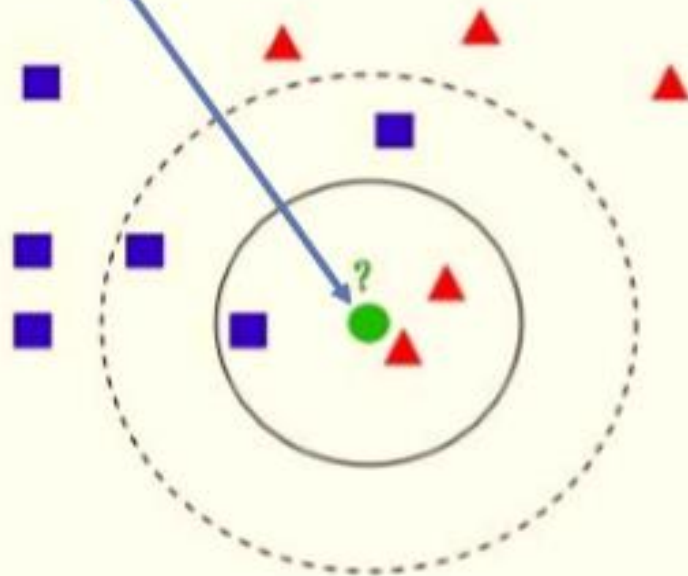


K邻近算法

- 对于右图的例子：
 - 如果 $K=3$ ，绿点最邻近的3个点分别是2红、1蓝，投票后判定绿色点属于红色类。
 - 如果 $K=5$ ，绿点最邻近的5个点分别是2红色3蓝，投票后判定绿色点属于蓝色类。

K值的选择，对分类结果有很大影响！

这个样本属于哪个类？



K邻近算法

- 我们用身高与脚码作为特征，判断性别。现有数据如下：

- A [(179, 42), 男]
- B [(178, 43), 男]
- C [(165, 36) 女]
- D [(177, 42), 男]
- E [(160, 35), 女]

$$AF = \sqrt{(167-179)^2 + (43-42)^2} = \sqrt{145}$$

$$BF = \sqrt{(167-178)^2 + (43-43)^2} = \sqrt{121}$$

$$CF = \sqrt{(167-165)^2 + (43-36)^2} = \sqrt{53}$$

$$DF = \sqrt{(167-177)^2 + (43-42)^2} = \sqrt{101}$$

$$EF = \sqrt{(167-160)^2 + (43-35)^2} = \sqrt{103}$$

- 现在测试样本 F (167, 43)。

- 令K=3，分别算出F离训练样本距离，如右图

- 选取最近的3个为C, D, E, 2女1男，因此判断为女性。

- 很不合理！这是因为，特征“身高”的绝对值太大，很容易主导数据波动，从而影响结果。通常需要做数据的“归一化”

- 参考下文：介绍为什么需要缩放特征、各种缩放方法以及什么时候应该对特征进行尺度缩放，帮助我们对特征缩放有更深入的了解
- <https://mp.weixin.qq.com/s/yIkP4yFv5c19gBKWL-vFng>

朴素贝叶斯分类 (Naïve Bayesian)

设D是训练元组和它们相关联的类标号的集合。通常，每个元组用一个n维属性向量 $X = (x_1, x_2, \dots, x_n)$ 表示，描述由n个属性对元组的n个测量

假设有m个类 C_1, C_2, \dots, C_m

给定元组X，分类法将预测X属于具有最高后验概率的类（在条件X下）

根据贝叶斯定理 $P(C_i | \mathbf{X}) = \frac{P(\mathbf{X} | C_i)P(C_i)}{P(\mathbf{X})}$

由于 $P(\mathbf{X})$ 对所有类为常数，所以需将下式最大化 $P(C_i | \mathbf{X}) = P(\mathbf{X} | C_i)P(C_i)$

类条件独立 $P(\mathbf{X} | C_i) = \prod_{k=1}^n P(x_k | C_i) = P(x_1 | C_i)P(x_2 | C_i) \cdots P(x_n | C_i)$

朴素贝叶斯分类 (Naïve Bayesian)

类

- C1:buys_computer = 'yes'
- C2:buys_computer = 'no'

age	income	student	credit_rating	<u>buys_computer</u>
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

希望分类的元组

- X = (age <=30, Income = medium, Student = yes, Credit_rating = Fair)

使用朴素贝叶斯分类预测类标号

age	income	buys_computer
student	credit_rating	buys_computer
no	fair	no
no	excellent	no
no	fair	yes
no	fair	yes
yes	fair	yes
yes	excellent	no
yes	excellent	yes
no	fair	no
yes	fair	yes
yes	fair	yes
yes	excellent	yes
no	excellent	yes
yes	fair	yes
no	excellent	no

$P(C_i)$

- $P(\text{buys_computer} = \text{"yes"}) = 9/14 = 0.643$
- $P(\text{buys_computer} = \text{"no"}) = 5/14 = 0.357$

为计算 $P(X|C_i)$ ，计算下面的条件概率

- $P(\text{age} = \text{"<=30"} | \text{buys_computer} = \text{"yes"}) = 2/9 = 0.222$
- $P(\text{age} = \text{"<=30"} | \text{buys_computer} = \text{"no"}) = 3/5 = 0.600$
- $P(\text{income} = \text{"medium"} | \text{buys_computer} = \text{"yes"}) = 4/9 = 0.444$
- $P(\text{income} = \text{"medium"} | \text{buys_computer} = \text{"no"}) = 2/5 = 0.400$
- $P(\text{student} = \text{"yes"} | \text{buys_computer} = \text{"yes"}) = 6/9 = 0.667$
- $P(\text{student} = \text{"yes"} | \text{buys_computer} = \text{"no"}) = 1/5 = 0.200$
- $P(\text{credit_rating} = \text{"fair"} | \text{buys_computer} = \text{"yes"}) = 6/9 = 0.667$
- $P(\text{credit_rating} = \text{"fair"} | \text{buys_computer} = \text{"no"}) = 2/5 = 0.400$

朴素贝叶斯分类

$X = (\text{age} \leq 30, \text{income} = \text{medium}, \text{student} = \text{yes}, \text{credit_rating} = \text{fair})$

$$P(X | C_i) = \prod_{k=1}^n P(x_k | C_i) = P(x_1 | C_i) P(x_2 | C_i) \cdots P(x_n | C_i)$$

$P(X|C_i)$

- $P(X|\text{buys_computer} = \text{"yes"}) = 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$
- $P(X|\text{buys_computer} = \text{"no"}) = 0.6 \times 0.4 \times 0.2 \times 0.4 = 0.019$

$P(X|C_i) * P(C_i)$

- $P(X|\text{buys_computer} = \text{"yes"}) * P(\text{buys_computer} = \text{"yes"}) = 0.028$
- $P(X|\text{buys_computer} = \text{"no"}) * P(\text{buys_computer} = \text{"no"}) = 0.007$

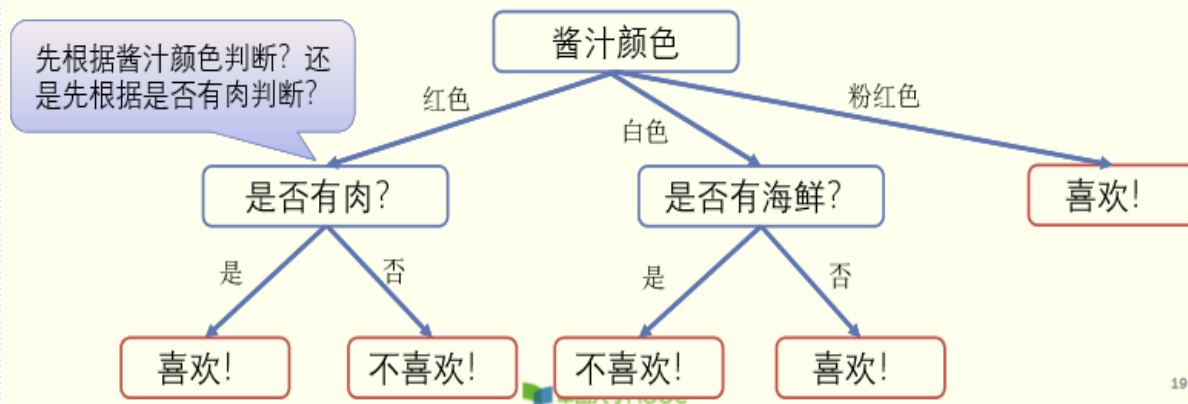
因此，对于元组X，朴素贝叶斯分类预测元组X的类为 $\text{"buys_computer} = \text{yes"}$

决策树算法

- 很多情况下我们的样本是下面这样的表格统计数据。
- 如何从中归纳出规律？如何根据新的数据快速判断结论？

序号	酱汁颜色	是否有肉	是否有海鲜	结论
1	红色	是	否	喜欢
2	红色	否	否	不喜欢
3	红色	否	是	不喜欢
4	白色	否	是	不喜欢
5	粉红色	否	否	喜欢
6	红色	是	否	喜欢
7	白色	否	否	喜欢
8	白色	否	否	喜欢
9	红色	否	否	不喜欢

- 规则归纳问题，适合用决策树来表示，如意大利面决策树。
- 关键问题是：如何从表格构建出这样的决策树？
- 换言之：如何选择各个属性决策的顺序？



决策树算法

□ ID3算法，昆兰（Quinlan, 1986）

□ 算法思想



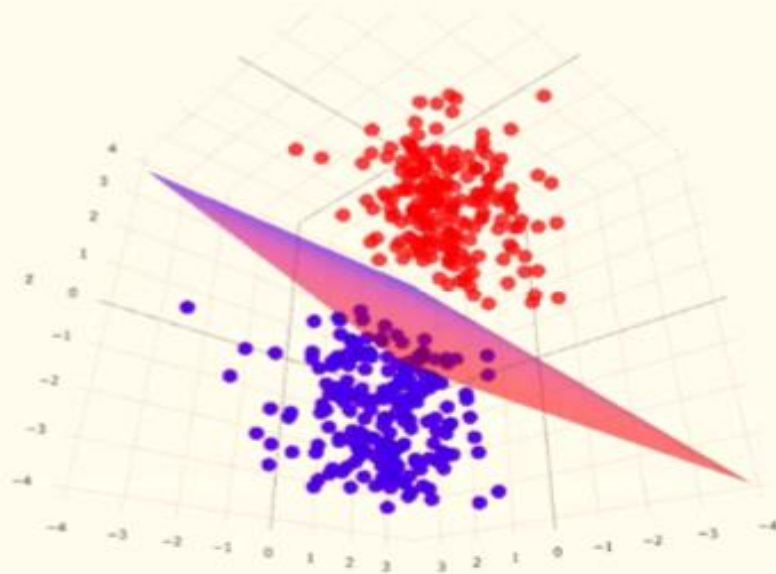
- 属性在决策树中的位置不同，决策树的效率是不同的。
 - 如果一个属性对于所有样本都没有区分能力（比如有是否能吃这个属性），那么对于决策毫无用处。
 - 如果一个属性恰好能将样本数据一分为二，则这是一个好的属性，应该尽量在决策早期就使用。
 - 如果根据一个属性做判断，样本仍然有若干种情况，则该属性不应该出现在决策早期。
-
- 昆兰提出利用“信息增益”来对属性进行排序。
 - 如果一个属性执行后，使得数据集上的信息增益最大（最有序），则该属性应该优先被执行（更接近根节点）。

支持向量机模型

- ❑ SUPPORT VECTOR MACHINES (SVM)
- ❑ 1995年，Corinna Cortes和Vapnik 提出使用“软间隔”思想;
- ❑ 机器学习领域大名鼎鼎，直接导致2000年代统计机器学习的繁荣。

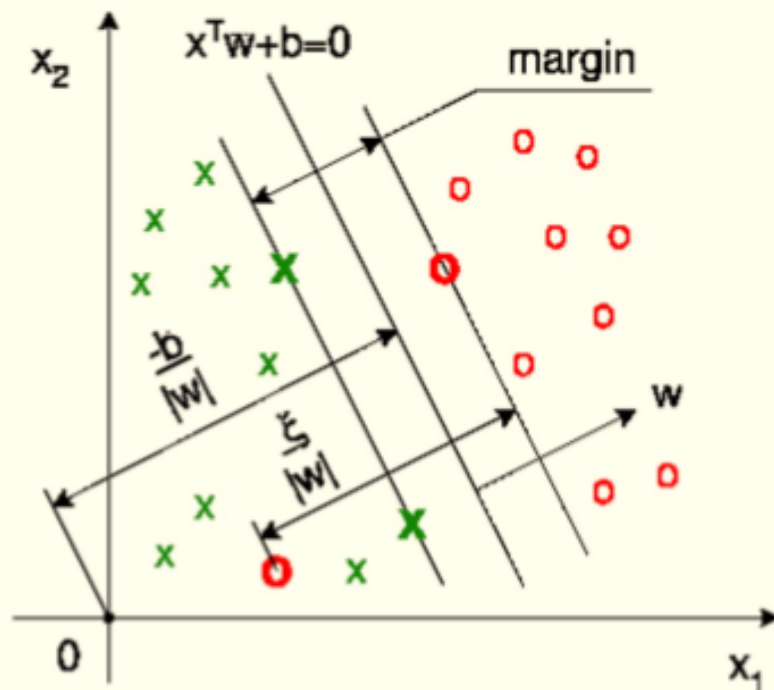
- 基本思想

- 对于二分类问题，将样本数据表示为空间中的点。
- 使用“平面”来切割空间，实现分类。
- 如何选择“平面”，成为关键问题。



支持向量机模型

- 我们认为，一个样本点与分割平面的距离越远，则分类越明确。
- 如果我们能找到一个平面，使得在其两侧，两类样本之间的间隔尽可能大，则该平面就是优秀的分割平面。
- 为达到间隔最大的目标，只需要看距离平面最近的几个样本，它们代表着间隔。称为“支持向量”。



有监督学习

□ 分类 VS. 预测



分类

- 预测类标号（离散值）
- 根据训练数据集和类标号属性，构建模型来分类现有数据，并用来分类新数据



预测

- 建立连续函数值模型，比如预测空缺值



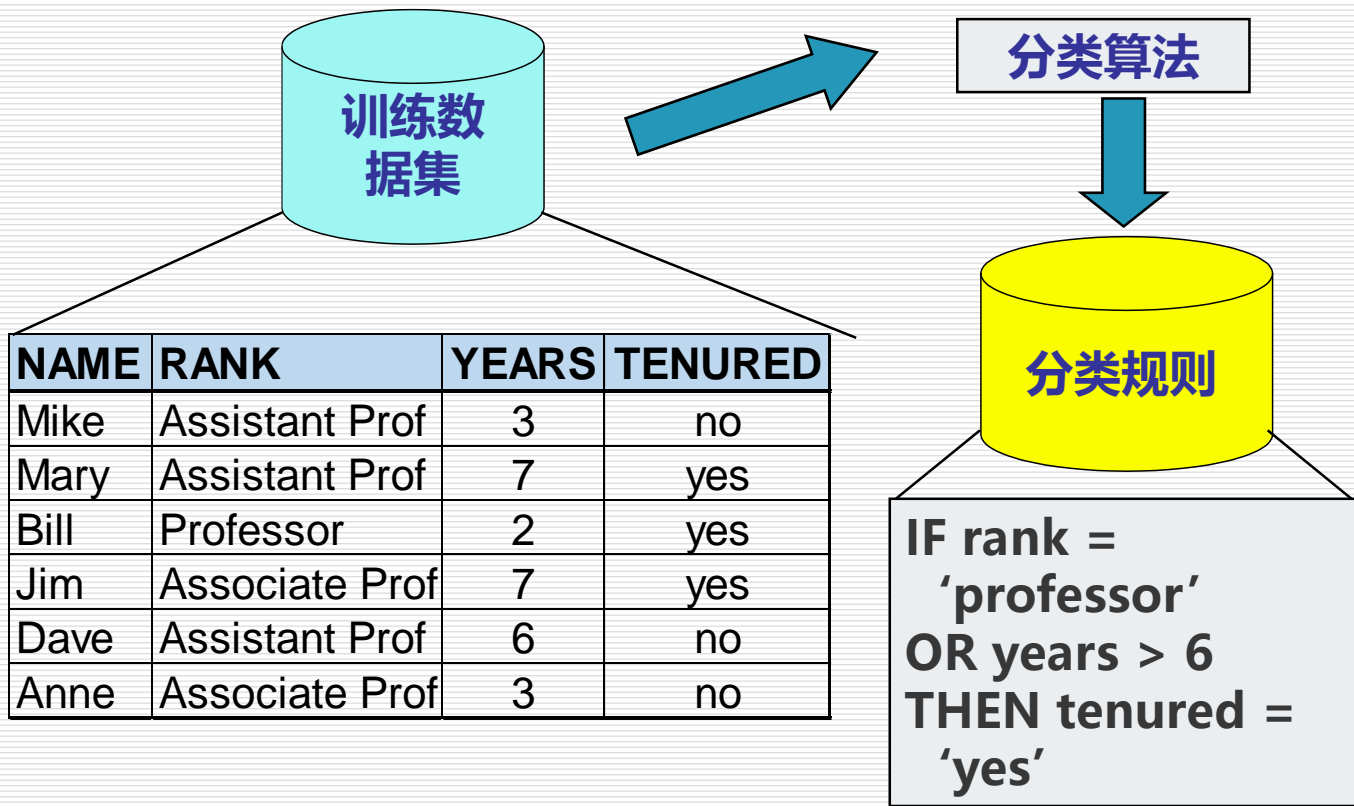
典型应用

- 信誉证实（分类为低，中，高风险）
- 医疗诊断（肿瘤是良性还是恶性）
- 性能预测
- 目标市场

有监督学习

第一阶段-训练，建立一个分类模型，描述预定数据类或概念集

- **假定每个元组属于一个预定义的类，由一个类标号属性确定**
 - ✓ 训练数据集：由为建立模型而被分析的数据元组形成
 - ✓ 训练样本：训练数据集中的单个样本（元组）
- **学习模型可以由分类规则、判定树或数学公式的形式提供**



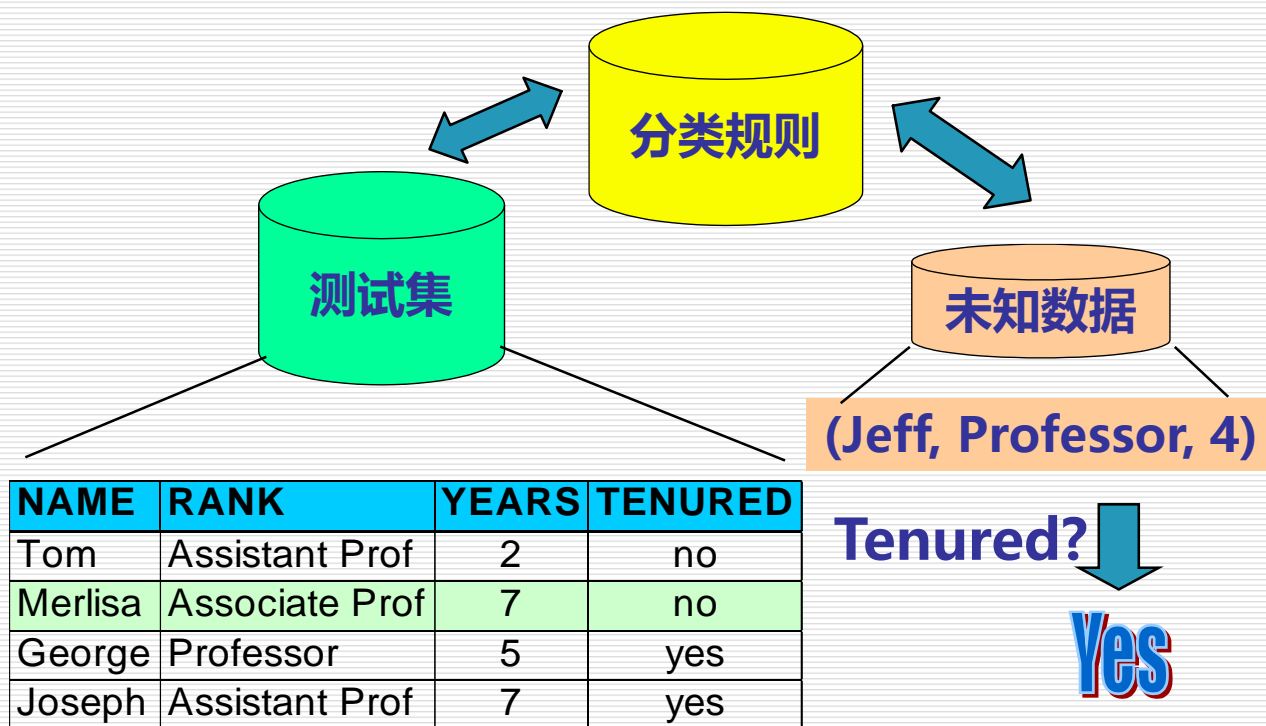
有监督学习

第二阶段-测试，使用模型，对将来的或未知的对象进行分类

➤ 评估模型的预测准确率

- ✓ 测试集：要独立于训练样本集，避免“过分拟合”的情况
- ✓ 对每个测试样本，将已知的类标号和该样本的学习模型类预测比较
- ✓ 准确率：被模型正确分类的测试样本的百分比

➤ 如果准确率可以接受，那么使用该模型来分类标签为未知的样本



有监督学习

模型评估与选择

评价指标: 我们如何测量精度?其他指标要考虑吗?

分类器的准确率最好在检验集上估计

评估一个分类准确率的方法

- 保持方法, 随机二次抽样
- 交叉验证
- 自助法

模型选择

- 统计显著性检验
- 基于成本效益和ROC 曲线

有监督学习

模型评估与选择-评估分类器性能的度量

正组元 (P) : 感兴趣的主要类的元组。

负组元 (N) : 其他元组。

真正例 (True Positive, TP) : 是指被分类器正确分类的正元组。

真负例 (True Negative, TN) : 是指被分类器正确分类的负元组。

假正例 (False Positive, FP) : 是被错误地标记为正元组的负元组。

假负例 (False Negative, FN) : 是被错误地标记为负元组的正元组。

。

混淆矩阵

实际的类\预测的类	yes	no
yes	TP	FN
no	FP	TN

有监督学习

模型评估与选择-精度、召回率、F 度量

精度: 精确性的度量, 即标记为正类元组实际为正类所占的百分比

$$precision = \frac{TP}{TP + FP}$$

F 度量 (F_1 或 F分数): 把精度和召回率集中到一个度量中

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

F_β : 精度和召回率的加权度量

➤ 它赋予召回率权重是赋予精度的 β 倍

$$F_\beta = \frac{(1 + \beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$$

有监督学习

模型评估与选择-保持方法, 随机二次抽样



保持方法

- 给定的数据随机的划分为**两个独立的集合**
 - ✓ 训练集, 通常2/3的数据被分配到训练集
 - ✓ 检验集, 通常1/3的数据被分配到检验集
- 随机二次抽样: 保持方法的变形将保持方法重复k次, 总准确率估计取每次迭代准确率的平均值



交叉验证(k-折交叉验证)

- 初始数据随机地划分成k个互不相关的子集, 每个子集的大小大致相等
- 在第i次迭代, 分区 D_i 用作检验集, 其他做为训练集
- 留一: 每次只给检验集“留出”一个样本
- 分层交叉验证: 折被分层, 使的每个折中样本的类分布与在初始数据中的大致相同

有监督学习

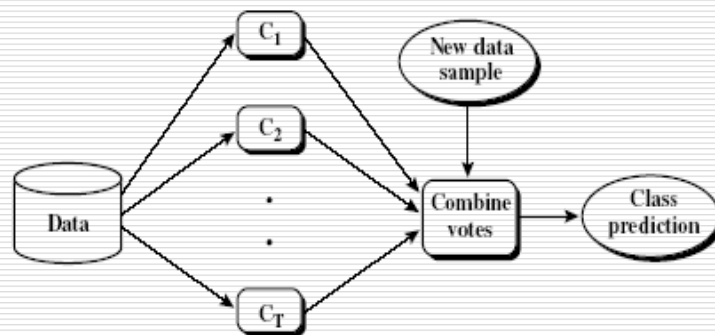
提高分类准确率的技术

组合分类方法

- 利用模型的组合来提高准确率
- 把 k 个学习得到的模型, M_1, M_2, \dots, M_k , 组合在一起 旨在创建一个改进的复合分类模型 M^*

常见的组合分类方法

- 装袋 (Bagging)
- 提升 (Boosting)

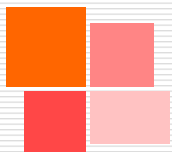


有监督学习

提高类不平衡数据的分类准确率



类不平衡问题:正类多, 负类少, 例如“欺诈”



传统的分类算法旨在最小化分类误差, 不适合类不平衡数据



提高类不平衡数据分类准确率的方法

- 过抽样
- 欠抽样
- 阈值移动
- 组合技术

有监督学习

□ 有监督数据的标注数据稀缺 昂贵

□ 标注 标注 标注！



有监督学习（用于分类）

- 模型的学习在被告知每个训练样本属于哪个类的“指导”下进行
- 新数据使用训练数据集中得到的规则进行分类



无监督学习（用于聚类）

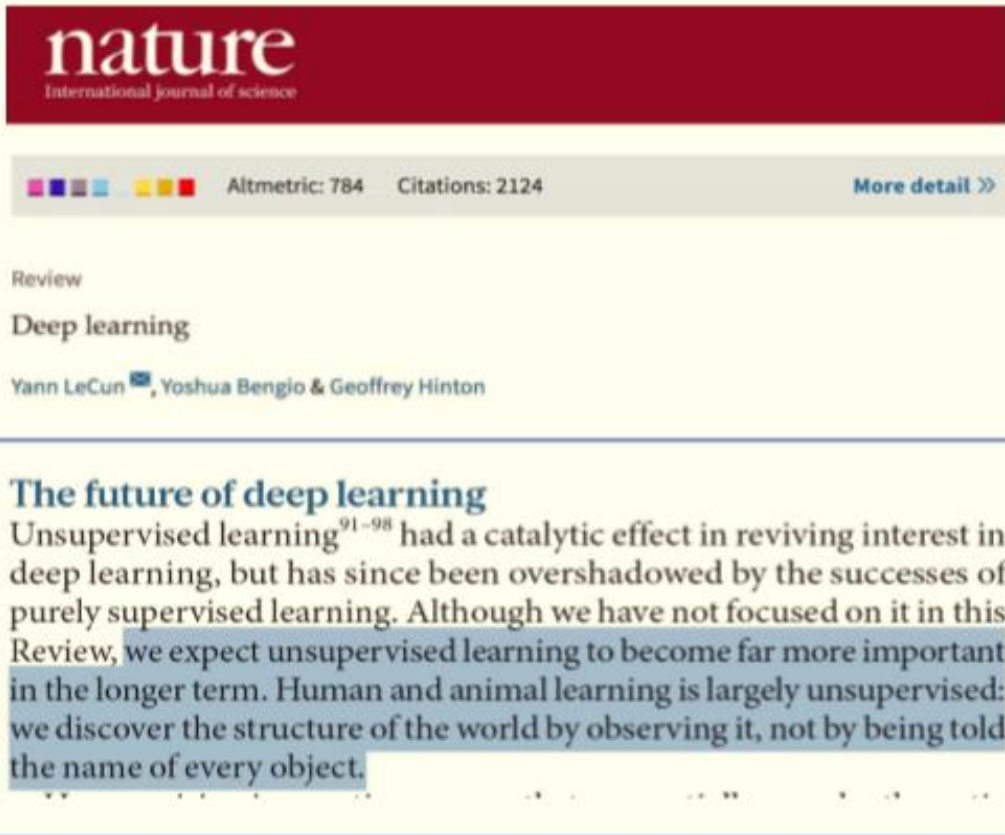
- 每个训练样本的类编号是未知的，要学习的类集合或数量也可能是事先未知的
- 通过一系列的度量、观察来建立数据中的类编号或进行聚类

无监督学习

无监督学习

- 有明确的学习目标
- 有样本数据
- 不需人工标注知识
- 提升难度大

从长远来看，我们希望无监督学习变得更加重要。人类和动物学习在很大程度上是无人监督的：我们通过观察它来发现世界的结构，而不是通过被告知每个物体的名称。——作者



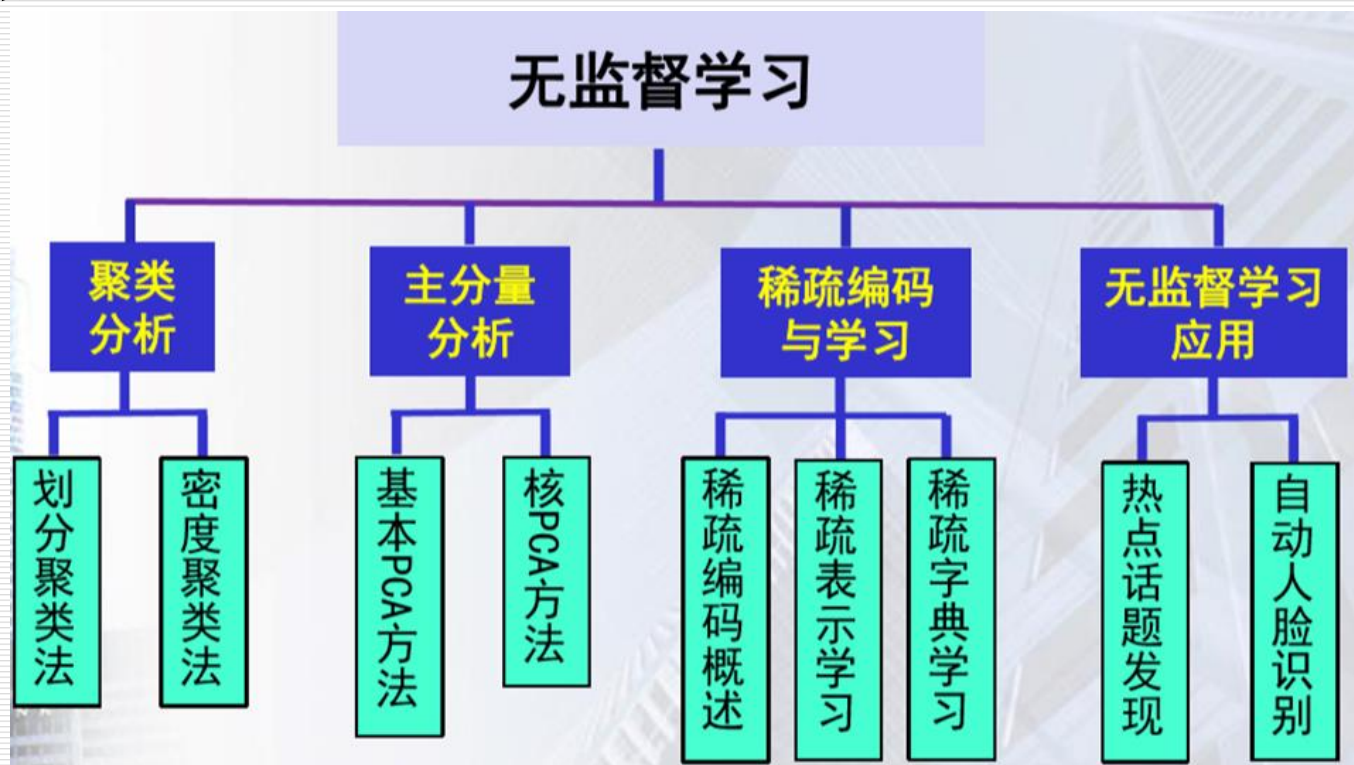
- ✓ 不需人工标注知识—计算机自主去总结、归纳
- ✓ 无监督学习更接近人类学习的过程

无监督学习

□ 聚类

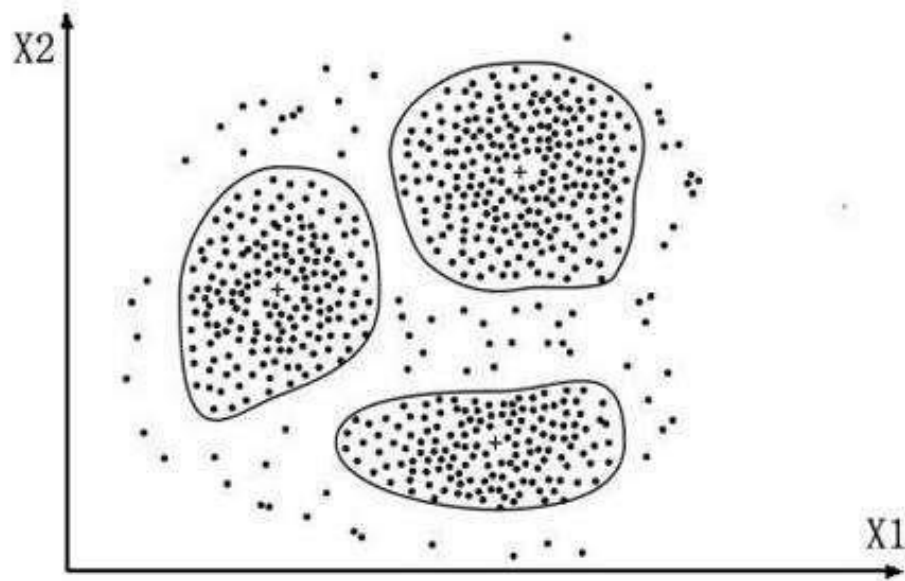
□ K-means聚类

□ 自动编码器



聚类

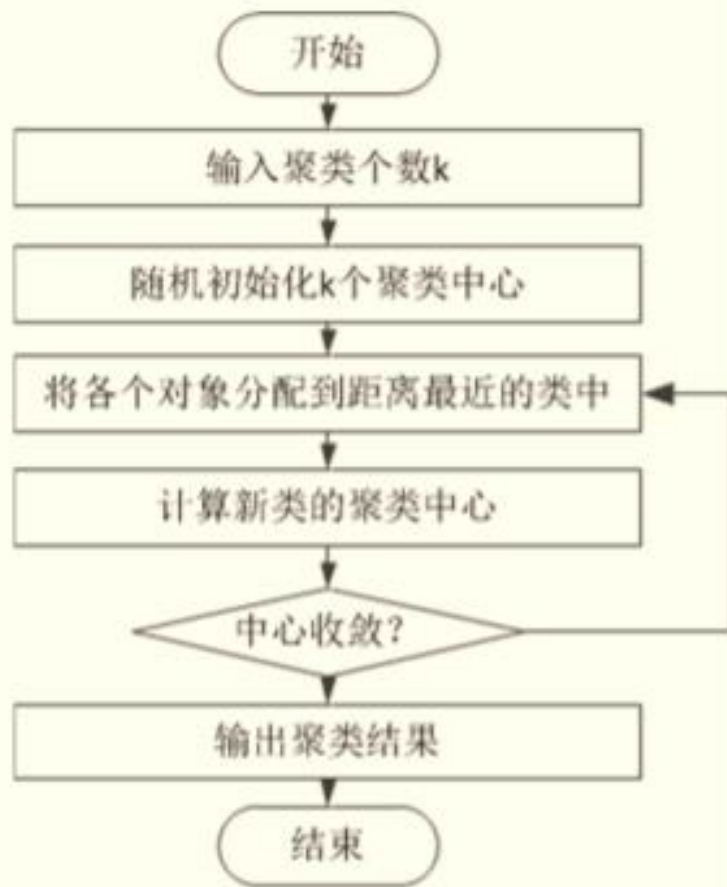
- 人类和许多动物都有“归堆”的能力。这就是聚类。
- 聚类算法根据某种“相似程度”或者“距离”，把距离近或者相似度高的实例归到一类。
- 与分类类似，大部分聚类也需要事先指定聚类的数目。



K-means聚类算法

- 思路:

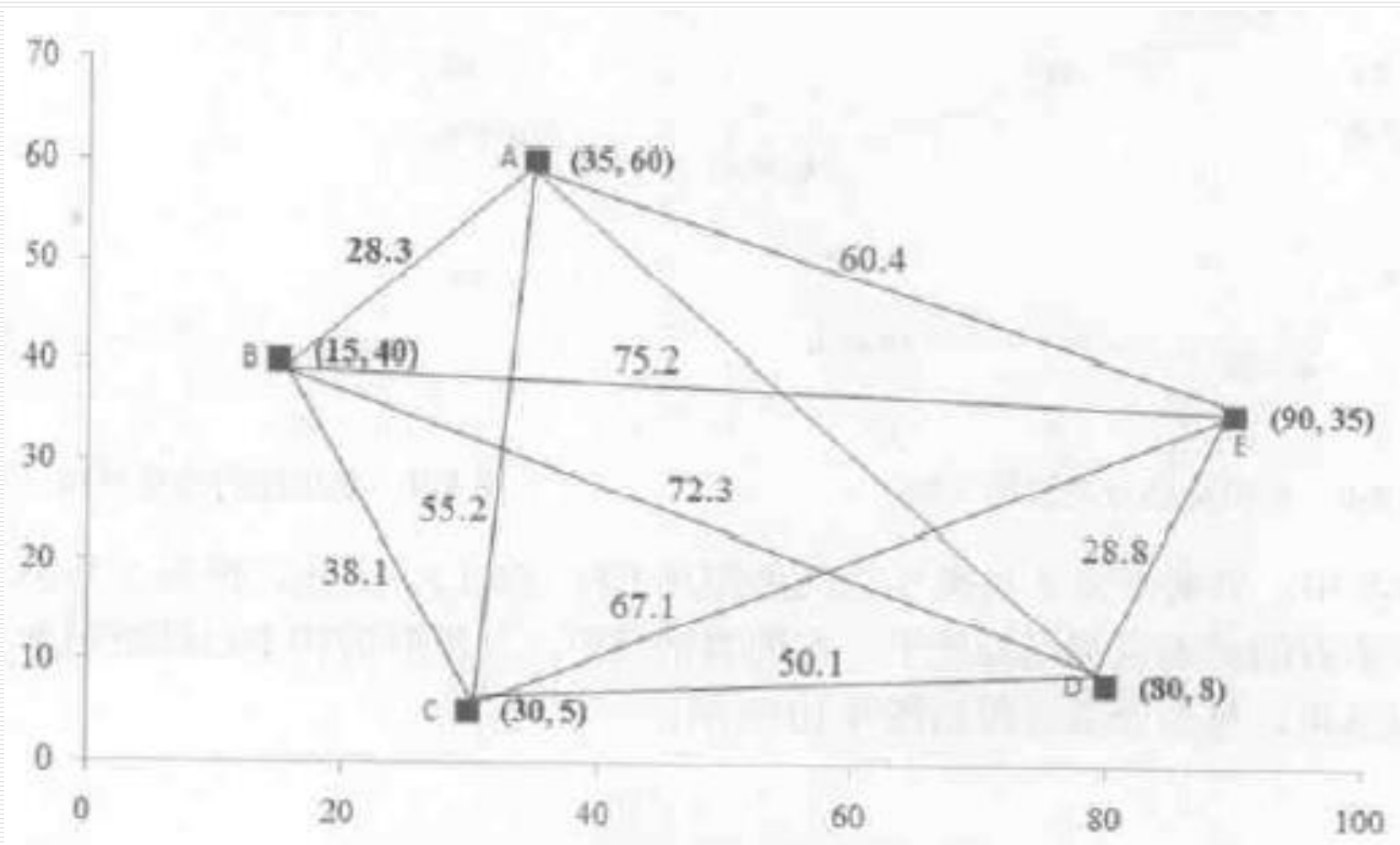
- 首先假设每个类别都聚集在某个“中心点”附近。
- 最初时, 任选3个样本作为“中心点”。
- 逐一计算剩下的样本与“中心点”的距离, 距离谁近, 就归到哪类。
- 根据最新的类别划分情况, 重新计算“中心点”。
- 循环该过程, 直至中心点不再改变。



K-means聚类算法

- 关键问题：初始选择的3个点，很重要！
 - 选择两两距离最远的3个点。
 - 随机划分样本成3堆，以3堆的平均重心作为初始点。
- 关键问题：距离计算方法。
 - 聚类大量依赖于距离或者相似度计算。因此确定一种距离非常重要。

聚类



表格表示初始状态

表 8-3

样本点之间的距离（一）

	A	B	C	D	E
A					
B	28.3				
C	38.1	55.2			
D	72.3	68.7	50.1		
E	75.2	60.4	67.1	28.8	

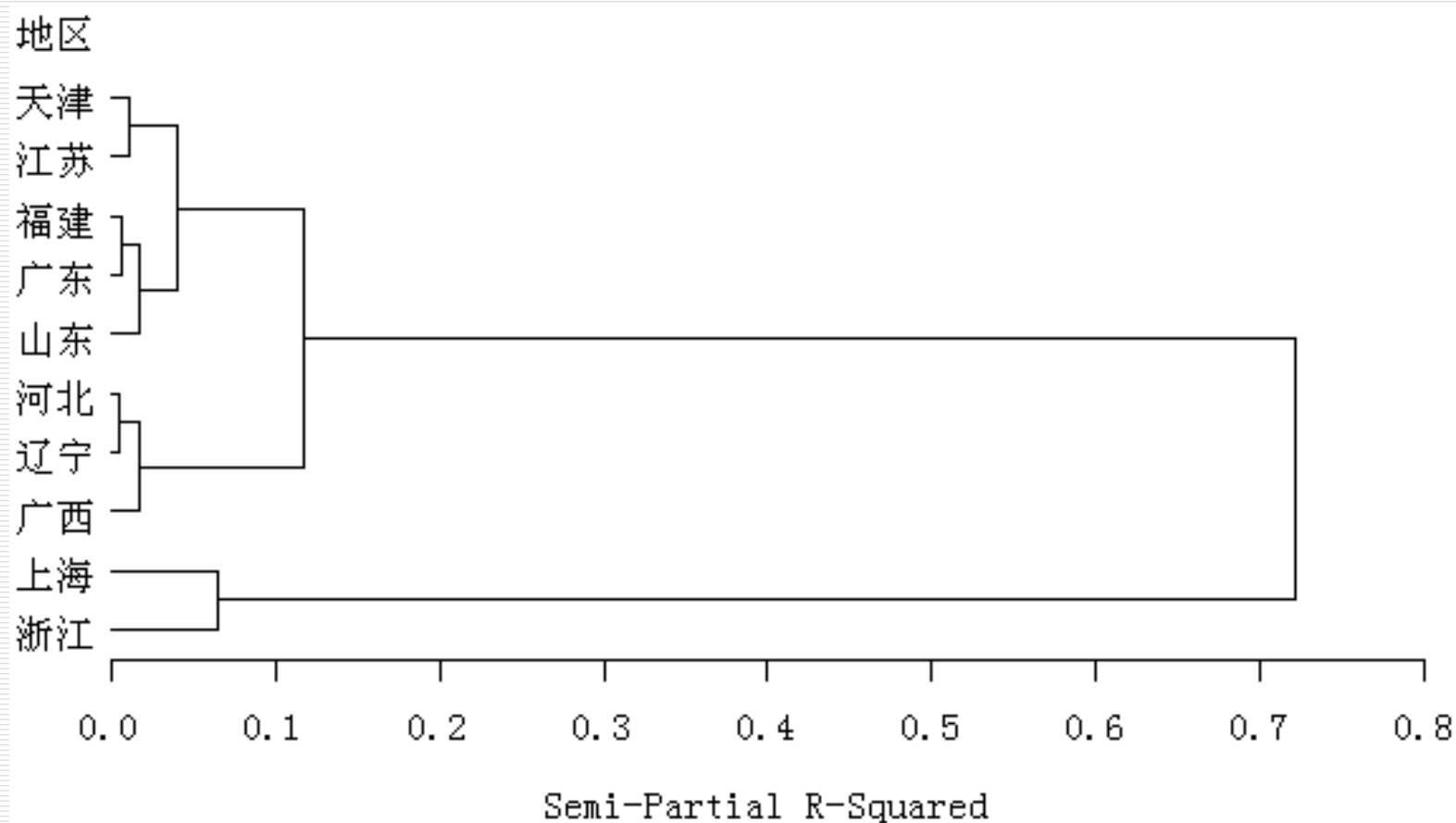


表 8-4

样本点之间的距离（二）

	AB (4)		C	D	E
AB (4)					
C	38.1	55.2			
D	72.3	68.7	50.1		
E	75.2	60.4	67.1	28.8	

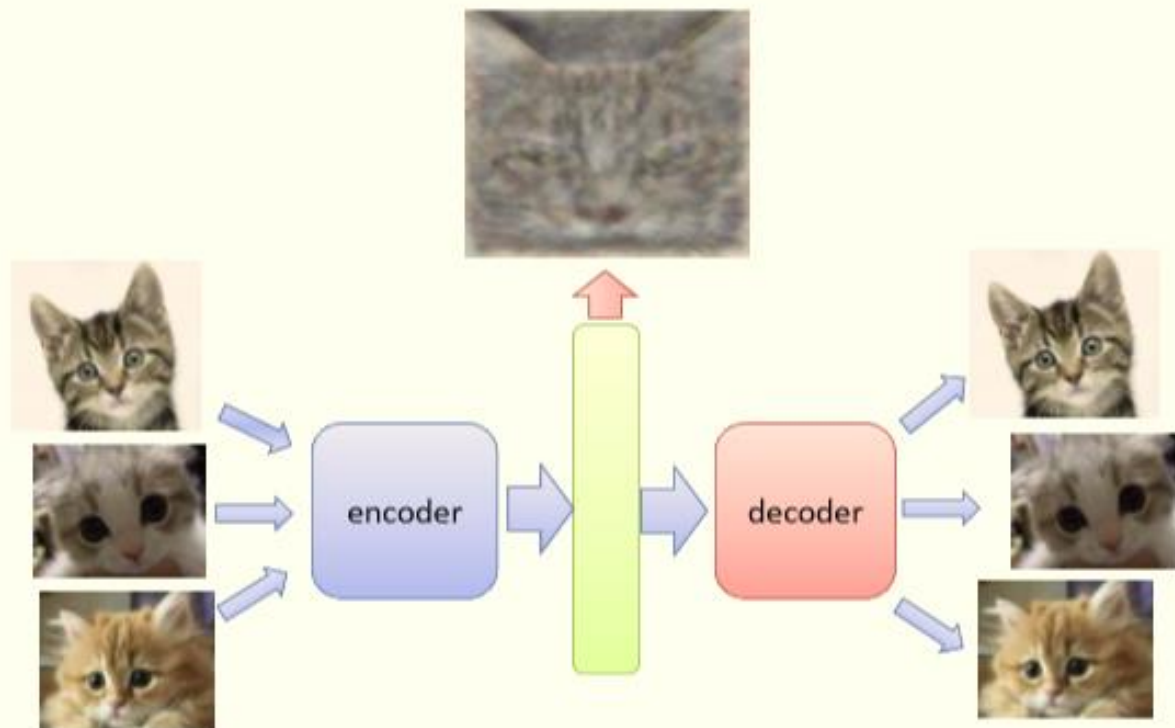
聚类



自动编码器

无监督学习方法：自动编码器

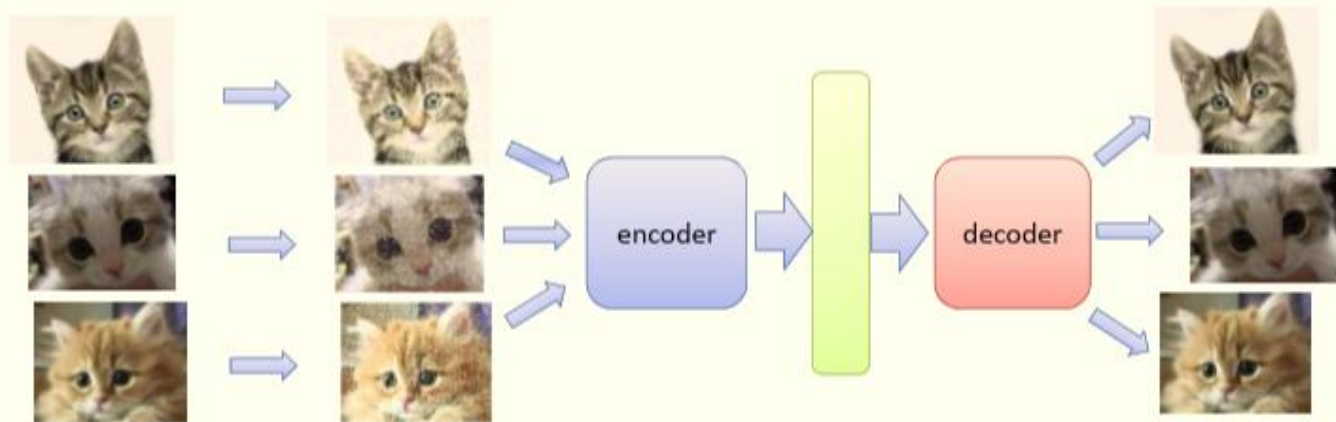
- 自动编码器
 - 原始数据经过神经网络“编码”，再根据编码信息“解码”，还原原有信息。
 - 适应大量样本后，编码器本身能够表示关于问题的本质特征。



自动编码器的应用

(1) 图像抗噪

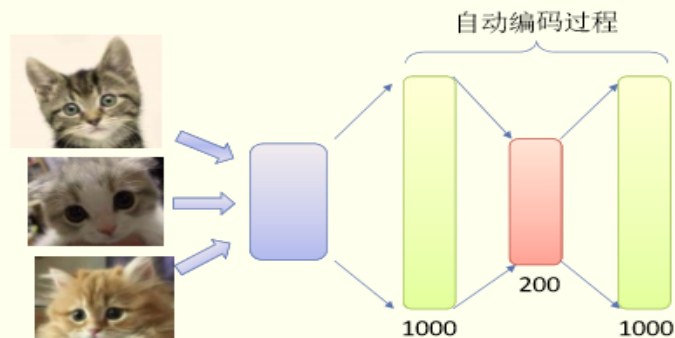
– 原始图像添加噪声之后再进入编码器，试图生成没有噪声的图像



(2) 数据降维

– 原始数据生成的模型往往仍然规模比较大

– 此时可以使用自动编码器做一次降维，不损失信息的情况下，降低空间使用。



半监督学习

- 全监督学习：方法丰富、研究充分、性能好、成本高
- 无监督学习：方法简单、数据成本低、性能难以提升

□ 半监督学习

- 一部分数据带有人工标记知识
- 另一部分数据无标记
- 通过学习有标记数据，逐渐扩展无标记数据

迁移学习

- 人类具有举一反三的能力。
 - 打乒乓球—打网球
 - 下国际象棋—下中国象棋
- 计算机是否可以实现类似的功能？
 - 任务A 与 任务B 具有某种相似性
 - 利用任务A的学习经验，解决任务B
 - 即迁移学习。
- 样本迁移：
 - 寻找任务A标注数据中，直接能够用于任务B的数据
 - 训练任务B的模型
- 特征迁移：
 - 估计任务A和B之间的“差距”，设计一个变换方程，将任务A的数据转换为任务B的数据，再进行训练
- 模型迁移：
 - 使用任务A训练得到的模型，经过某种变换，直接应用于任务B

图像风格的迁移学习

- 给定某种风格的图画样本，如梵高油画集合
- 训练模型，能够表达梵高图片中的某些风格
- 给定不具有该风格的图像
- 不改变内容，仅将梵高风格添加到图片中

- 使用星空作为源数据



强化学习

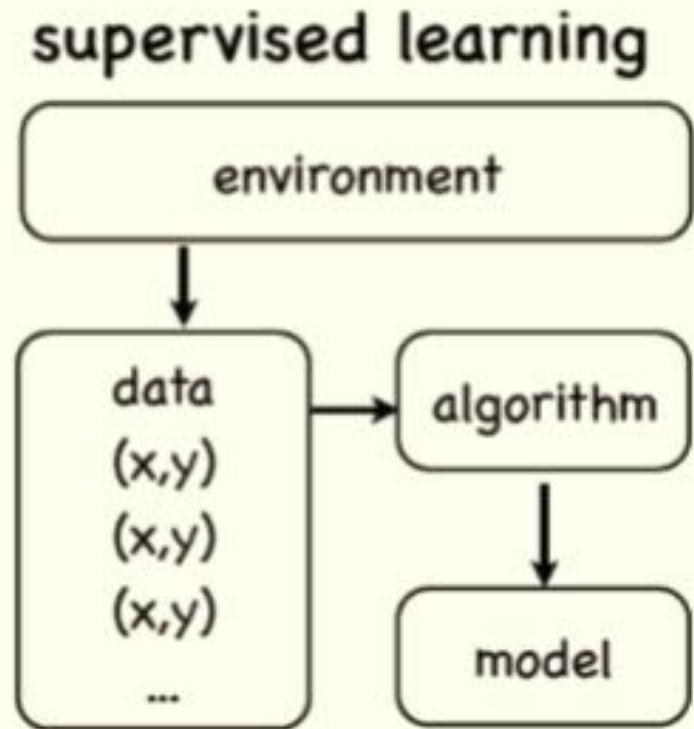
强化学习的基本思想

- 首先，强化学习(Reinforcement Learning, RL)并不是一个独立的方法，而是一种机器学习的模式，或者说是一种思路。
- 强化学习认为，计算机单纯通过感知环境，与环境交互，并且从交互中获得评价反馈，就可以适应所处的环境。从我们看来，就相当于计算机学习得到了环境下的某些知识，实现了机器学习。
- 强化学习的思想，来源于自然界生物的学习能力。
- 比如我们常常觉得小猫天然会爬树、小狗天然会游泳。
- 但实际上小猫、小狗的这种能力，并不是天然的，而是在树上、水里通过适应环境学习得到的。这种学习是不需要正确答案来“教”的，单纯靠适应环境、不断尝试就可以实现。

强化学习

强化学习的基本思想

- 强化学习与之前介绍的有监督学习，有类似的地方，也有区别。
- 典型的有监督学习，可以看作一个“开环”的流程。我们从“环境”中提取一些样例数据，数据中带有监督信息。
- 然后我们根据输入值和监督信息来设计算法，训练模型去学习监督信息。

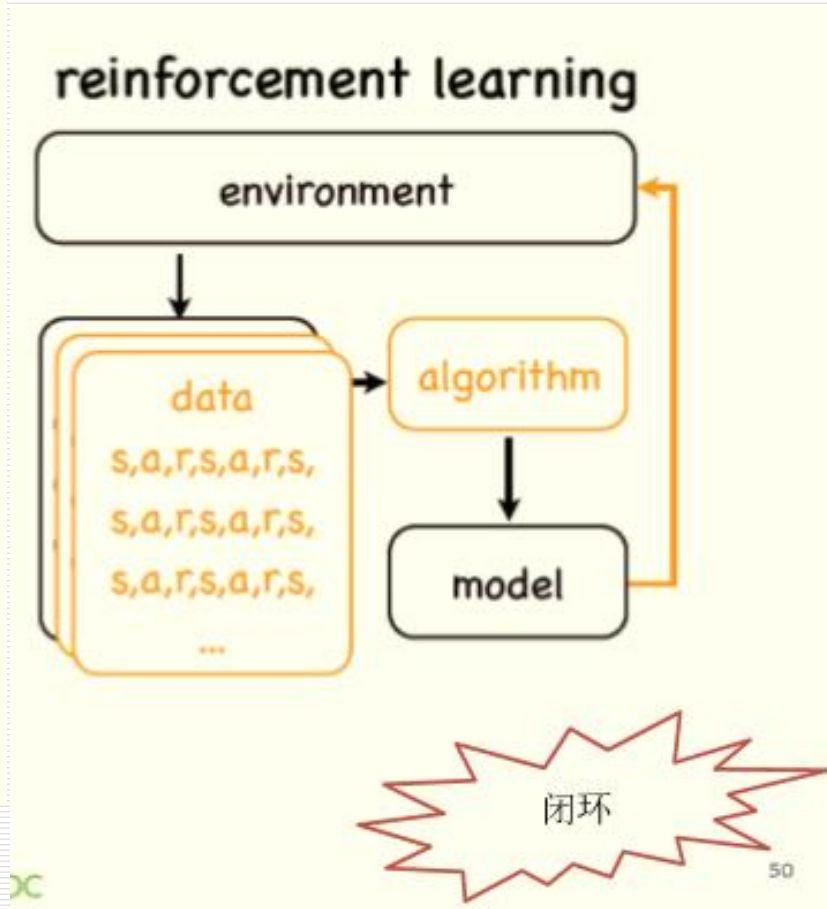


开环

强化学习

强化学习的基本思想

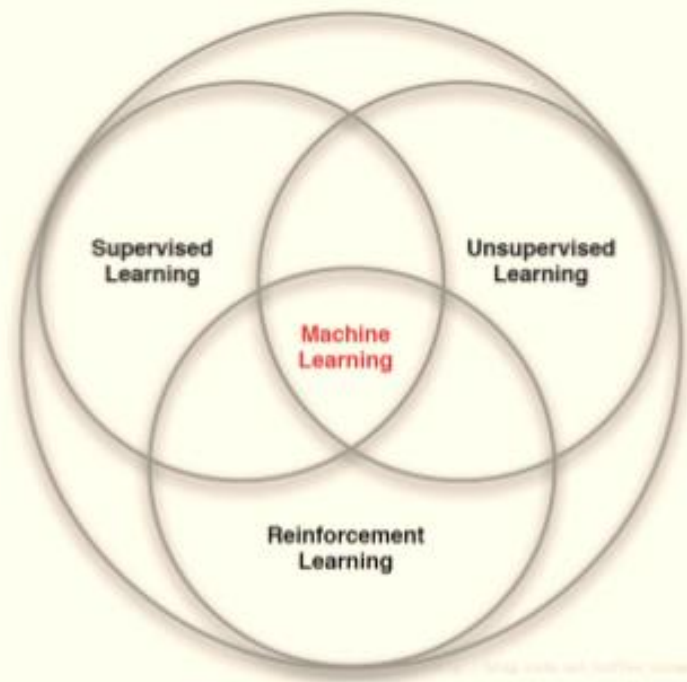
- 强化学习中，我们从环境中得到的是合理的行动方式，行为。
- 这些行为没有监督信息，即我们只知道在这个环境中可以干什么，但不知道后果。
- 经过算法计算后，模型把计算结果返回给“环境”，这时环境才给我们“奖赏”，即动作产生的后果。
- 我们用这个奖赏来训练、更新模型。可以看出，强化学习的整个过程形成一个“闭环”。



强化学习

强化学习的基本思想

- 同时，强化学习与无监督学习也不完全一样。
- 环境所提供的“奖励”，实际上就是监督信号。
- 因此，强化学习近年来成为机器学习和人工智能的一个新的研究领域。



强化学习

□ 例：小孩学习走路

- 我们设想一下，一个一岁的小孩学习走路的过程：
- （1）小孩作为一个独立的个体。
- （2）小孩在确定的环境中，有明确的目标：从当前位置走到妈妈怀里。
- （3）小孩能够掌握基本的走路动作，比如抬腿迈步、挥手等。
- （4）如果成功走到妈妈怀里，则得到奖励：笑脸、怀抱；反之则受到惩罚：摔跤、找不到妈妈，等。
- 通过反复多次练习，小孩能够学会走路。



强化学习

□ 学习走路的关键要素：

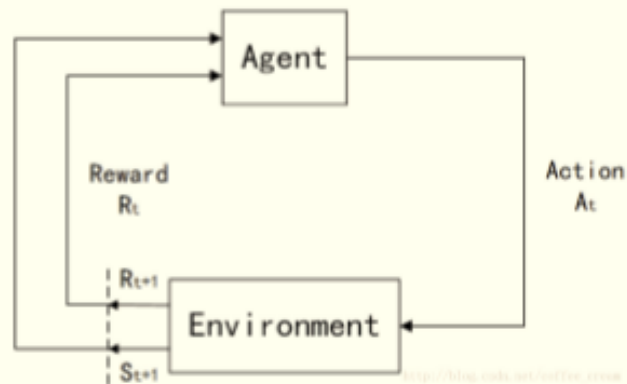
- 主体（小孩）
- 环境（地板、墙、起点、目的地）
- 动作（迈步、挥手）
- 奖惩（笑脸、怀抱、摔跤）

强化学习

强化学习的基本思想

我们把图中的各个要素做进一步抽象，就可以强化学习的结构图。其中：

- agent为智能体，对应于例子中的小孩。agent具有明确的目标，并且能感知自己所处的环境和状态。
- agent根据目标来指导自己的行为，在每一步，agent选择一个可行的动作，并通知环境自己的动作。
- 环境根据agent的动作，推断两件事情：
 - (1) agent执行完该动作之后会产生什么后果，是奖励，还是惩罚？
 - (2) 执行完动作之后，agent会处于什么样的状态？
- Agent接收到环境的反馈，更新自己的状态，同时用得到的奖励来更新自己的模型参数。
- 重复以上过程。



深度学习

特征的自学习

- 传统的机器学习方法：



- 通过传感器获取数据，然后经过预处理、特征提取、特征选择、再到推理、预测或识别
- 特征提取与选择的好坏对最终算法的确定性起了非常关键的作用。而特征的样式目前一般都是靠人工提取特征。而手工选取特征费时费力，需要专业知识，很大程度上靠经验和运气，那么机器能不能自动的学习特征呢？深度学习的出现就这个问题提出了一种解决方案

深度学习



深度学习

- 深度学习的历史可以追溯到20世纪40年代，自2006年，深度学习（Deep Learning）已经成为机器学习研究中的一个新兴领域，通常也被叫做深层结构学习或分层学习。其动机在于建立、模拟人脑进行分析学习的神经网络，它模拟人脑的机制来解释数据
- 深度学习的概念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，来发现数据的分布式特征表示

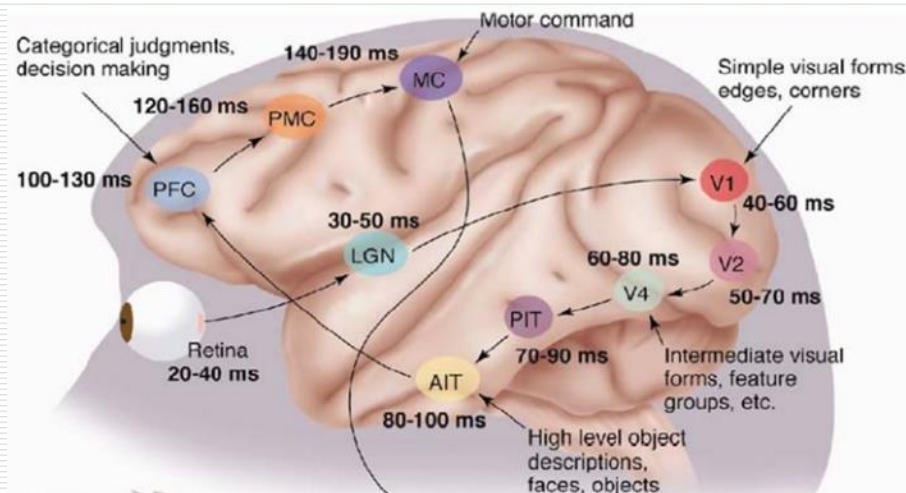
深度学习

- 是一个由线性阈值元件组成的单层（或多层）神经元的神经网络
- 当输入的加权和大于或等于阈值时，输出为**1**，否则为**0**
- 模型假定神经元中间的耦合程度（即加权系数 W ）可变，这样，该模型可以学习
- 当感知器用于两类模式的分类时，相当于在高维样本空间中，用一个超平面将两类样本分开
- 神经网络的学习过程就是神经网络参数的设定过程
- 一个神经元网络结构确定之后，需要对一系列参数（权重、阈值等）进行有效的设定。这个过程叫做学习或训练过程，此时的方法叫学习算法

深度学习

人脑的视觉机理

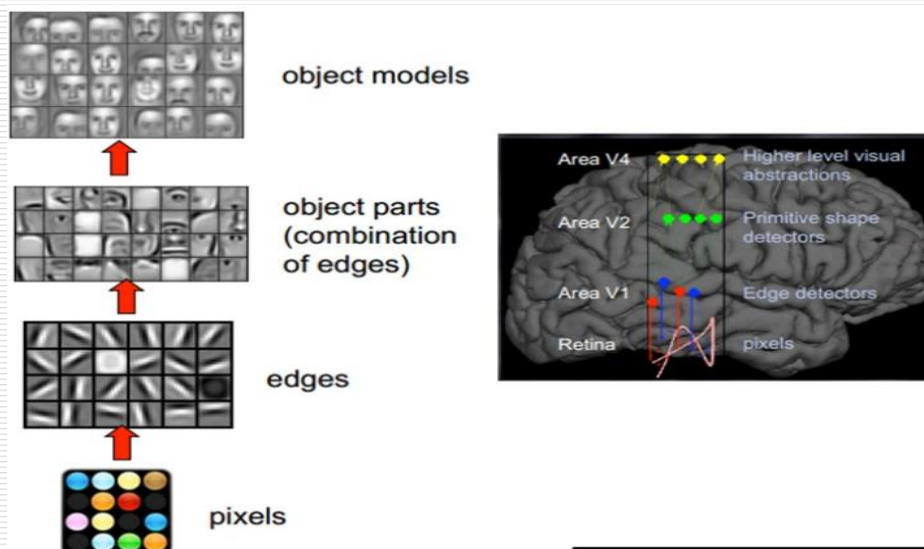
- 1981年的诺贝尔医学奖获得者 David Hubel和Torsten Wiesel发现了视觉系统的信息处理机制，他们发现了一种被称为“方向选择性细胞的神经元细胞，当瞳孔发现了眼前的物体的边缘，而且这个边缘指向某个方向时，这种神经元细胞就会活跃



深度学习

人脑的视觉机理

- 由此可知人的视觉系统的信息处理是分级的，高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图，抽象层面越高，存在的可能猜测就越少，就越利于分类

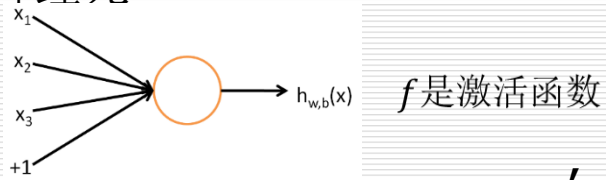


深度学习

人工神经网络

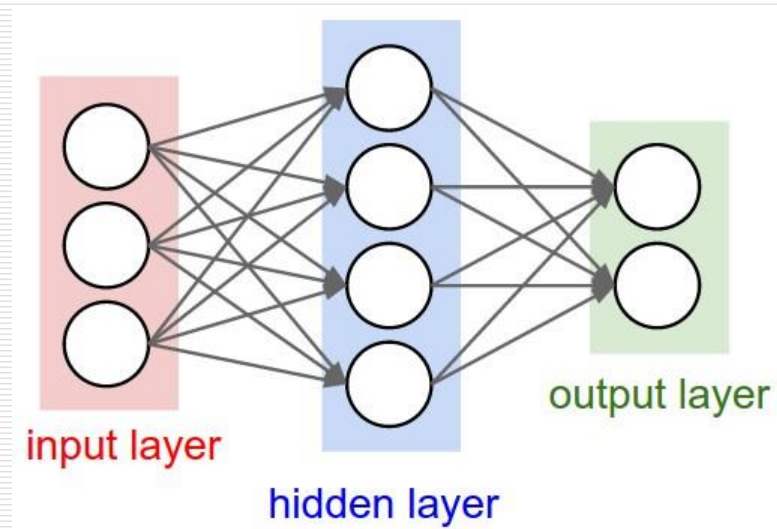
- 人工神经网络（Artificial neural network, ANN），简称神经网络（NN）

- 神经网络其实就是按照一定规则连接起来的多个神经元



- 神经元：

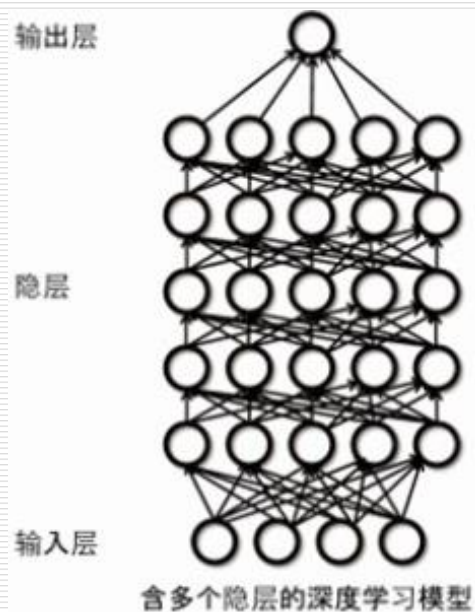
- 每个连接都有一个权值



一个全连接的神经网络

深度学习

深度学习

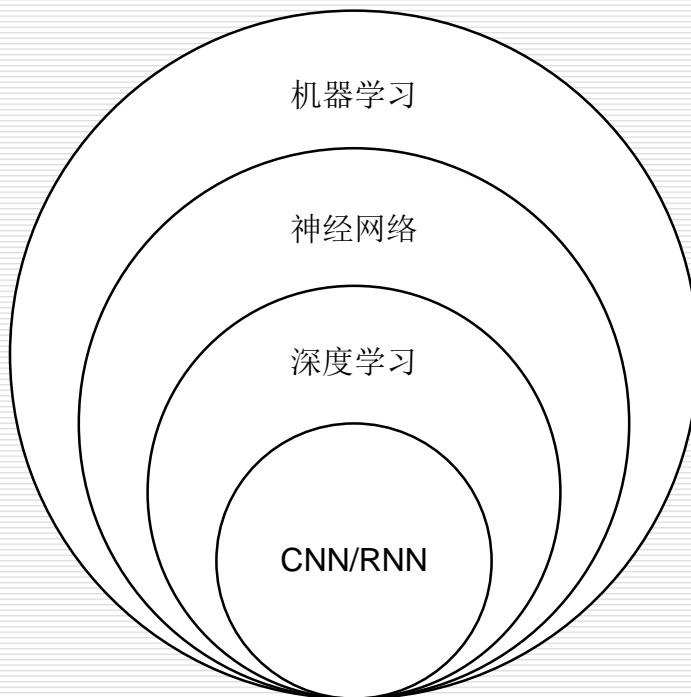


- 本质：通过构建多隐层的模型和海量训练数据，来学习更有用的特征，从而最终提升分类或预测的准确性。“深度模型”是手段，“特征学习”是目的
- 与浅层神经网络区别：
 - 1) 强调了模型结构的深度，通常有5-10多层的隐层节点；
 - 2) 明确突出了特征学习的重要性，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，从而使分类或预测更加容易。与人工规则构造特征的方法相比，利用大数据来学习特征，更能够刻画数据的丰富内在信息

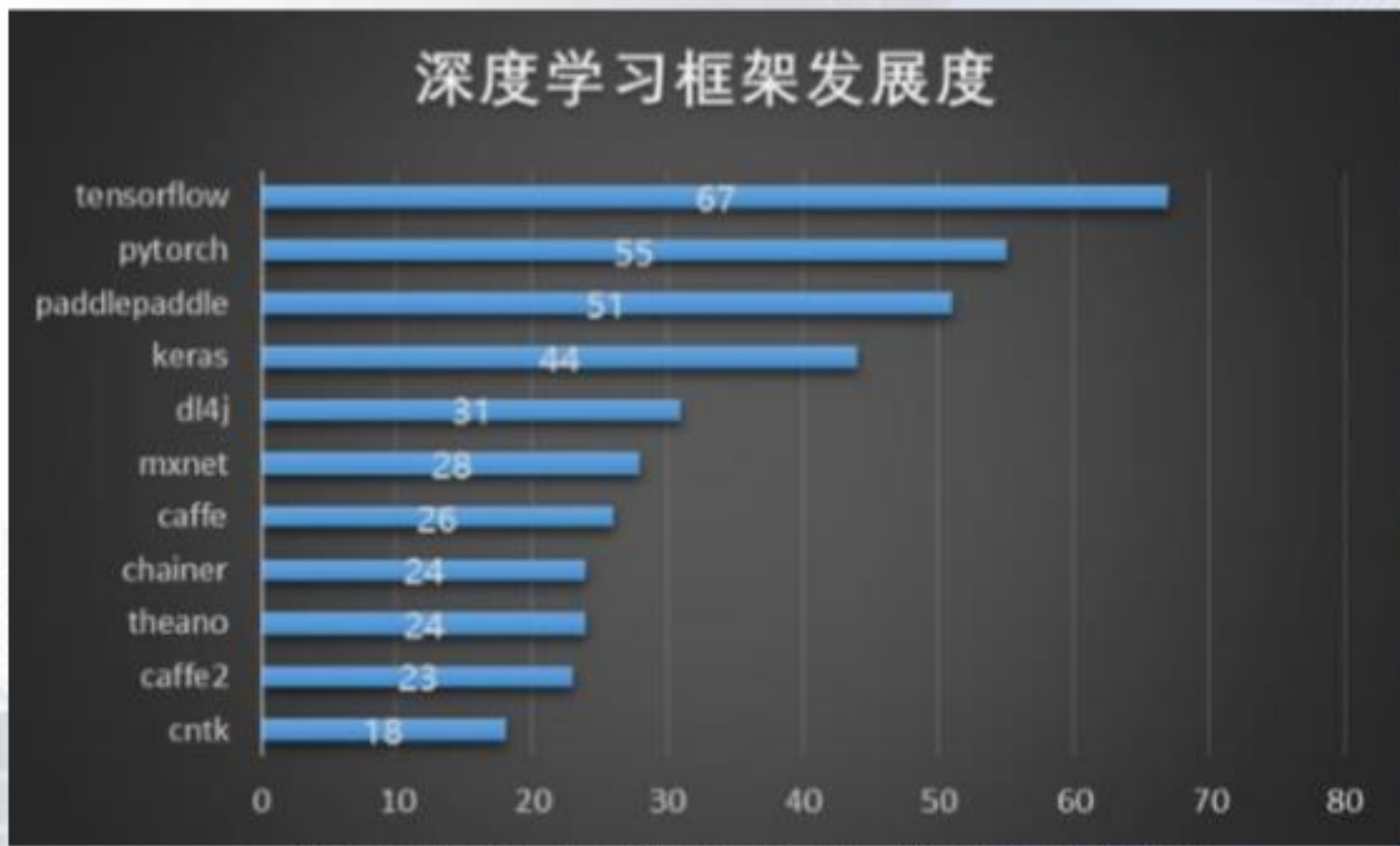
深度学习

深度学习的几种常用模型

- Convolutional Neural Networks （卷积神经网络）
- Recurrent Neural Networks （循环神经网络）



十大深度学习框架



2019年十大深度学习框架发展程度

深度学习给NLP带来什么

- ❑ **性能从追赶到超越。**几乎所有的自然语言处理任务中深度学习的性能都可以达到或者是超越原有办法。
- ❑ **语义变得更为重要。**有了分布式的离散表示，深度学习让语义变得可计算，我们可以得到字词句子篇章的向量，就使得语义真正可以用起来。
- ❑ **不断出现的新“玩法”、新应用场景。**比如说看图说话，甚至是给一个意图，让计算机画画，做诗、作词、写歌、写程序。将来的程序员可能也是有一定危机的岗位，计算机有可能代替程序员写代码。对于写作方面，现在可以写新闻稿，可以写天气预报，不久的将来可能可以写小说等等，这些都是计算机用深度学习可以做的事情。
- ❑ **门槛降低。**深度学习使得自动问答和机器翻译的门槛降低。也许一个一年级的研究生，看了几本深度学习的书本，甚至就能够发表很好的文章，提出很好的算法。所以说这是一个竞争很激烈的行业。

本章内容

- 概述机器学习
 - 什么是机器学习
 - 机器学习的框架
 - 定位、发展历程、编程语言、基本术语
- 有监督学习
 - K邻近算法、朴素贝叶斯算法、决策树算法、支持向量机
 - 分类问题 是什么 图像识别，语音识别
 - 回归问题 是多少 预测收入，预测房价
- 无监督学习
 - 降维 抽象和压缩
 - 聚类 怎么划分
- 迁移学习
- 强化学习
 - 怎么做决策，如博弈与游戏，自动驾驶，机器人控制
- 人工神经网络和深度学习