绪论

------2020.2 BY YANGZHONGXIU

内容

- 课程介绍
- 基本术语
- 假设空间
- 归纳偏好
- 机器学习三要素
- 典型应用案例
- 发展现状

课程介绍

机器学习：致力于研究如何通过计算机的手段，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”形式存在，因此，**机器学习所研究的主要内容是关于在计算机上从数据中产生“模型”（model）的算法，即“学习算法”（learning algorithm）。**

有了学习算法，我们把经验数据提供给它，它就能基于这些数据产生模型；在面对新的情况时，模型会给我们提供相应的判断。

如果说计算机科学是研究关于“算法”的学问，那么机器学习就是研究关于“学习算法”的学问。

课程介绍

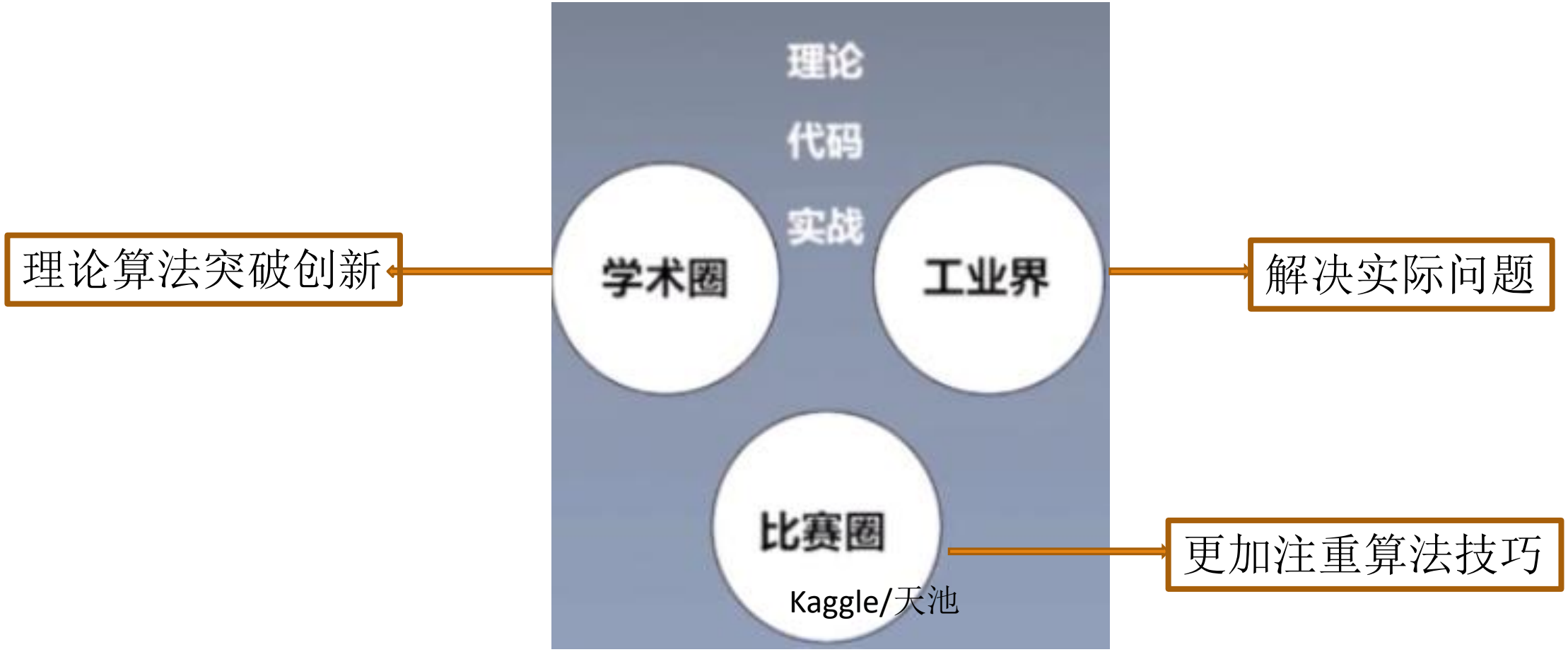
课程内容

- 模型评估和选择
- 线性模型
- 决策树
- 支持向量机
- 贝叶斯分类器
- 聚类
- 神经网络
- 集成学习
- 强化学习

机器学习知识结构图


通用知识
数学基础：微积分、线性代数、概率论、最优化、信息论
随机过程、矩阵论、控制论、泛函分析、微分几何、动力系统
计算机基础：数据结构、数据库、操作系统、计算机网络
分布式存储与计算
流式计算
中间件
AI算法基础：机器学习、深度学习
细分领域基础：计算机视觉、自然语言、语音处理、推荐系统、计算广告、结构化数据挖掘

机器学习路径



达到目标


实践高度
企业项目
AI大赛
项目实战
小练习
AI科学家
算法工程师
我们带你走到这
调包侠
顶级实践
顶级理论
AI入门
调参方法
算法推理
《西瓜书》
AI论文

基本术语

数据集（data set）： 一组记录的集合。其中每条记录是关于一个事件或对象的描述，称为一个“**示例（instance）**”或者“**样本（sample）**”，反映事件或者对象在某方面的表现或者性质的事项。

数据集是机器学习的基础。机器学习就是在已知数据集上进行学习，没有数据集就没有学习。数据集的选择对于学习效果也有很大决定性。

	A	B	C	D
1	国家	2019年国际	2018世界杯	2015亚洲杯
2	中国	73	40	7
3	日本	60	15	5
4	韩国	61	19	2
5	伊朗	34	18	6
6	沙特	67	26	10
7	伊拉克	91	40	4
8	卡塔尔	101	40	13
9	阿联酋	81	40	6
10	乌兹别克斯	88	40	8
11	泰国	122	40	17
12	越南	102	50	17
13	阿曼	87	50	12
14	巴林	116	50	11
15	朝鲜	110	50	14
16	印尼	164	50	17
17	澳洲	40	30	1
18	叙利亚	76	40	17
19	约旦	118	50	9
20	科威特	160	50	15
21	巴勒斯坦	96	50	16

亚洲足球成绩数据集

Rank	No.	President	Height (in)	Height (cm)
1	16	Abraham Lincoln	6 ft 4 in	193 cm
2	36	Lyndon B. Johnson	6 ft 3 ½ in	192 cm
3	45	Donald Trump	6 ft 3 in	191 cm
4	3	Thomas Jefferson	6 ft 2 ½ in	189 cm
5	1	George Washington	6 ft 2 in	188 cm
	21	Chester A. Arthur	6 ft 2 in	188 cm
	32	Franklin D. Roosevelt	6 ft 2 in	188 cm
	41	George H. W. Bush	6 ft 2 in	188 cm
	42	Bill Clinton	6 ft 2 in	188 cm
10	7	Andrew Jackson	6 ft 1 in	185 cm
	35	John F. Kennedy	6 ft 1 in	185 cm
	40	Ronald Reagan	6 ft 1 in	185 cm
	44	Barack Obama	6 ft 1 in	185 cm
14	5	James Monroe	6 ft 0 in	183 cm
	10	John Tyler	6 ft 0 in	183 cm
	15	James Buchanan	6 ft 0 in	183 cm
	20	James A. Garfield	6 ft 0 in	183 cm
	29	Warren G. Harding	6 ft 0 in	183 cm
	38	Gerald Ford	6 ft 0 in	183 cm
20	27	William Howard Taft	5 ft 11 ½ in	182 cm
	31	Herbert Hoover	5 ft 11 ½ in	182 cm
	37	Richard Nixon	5 ft 11 ½ in	182 cm

美国历届总统身高数据集

Ｗｉｎｅ数据集

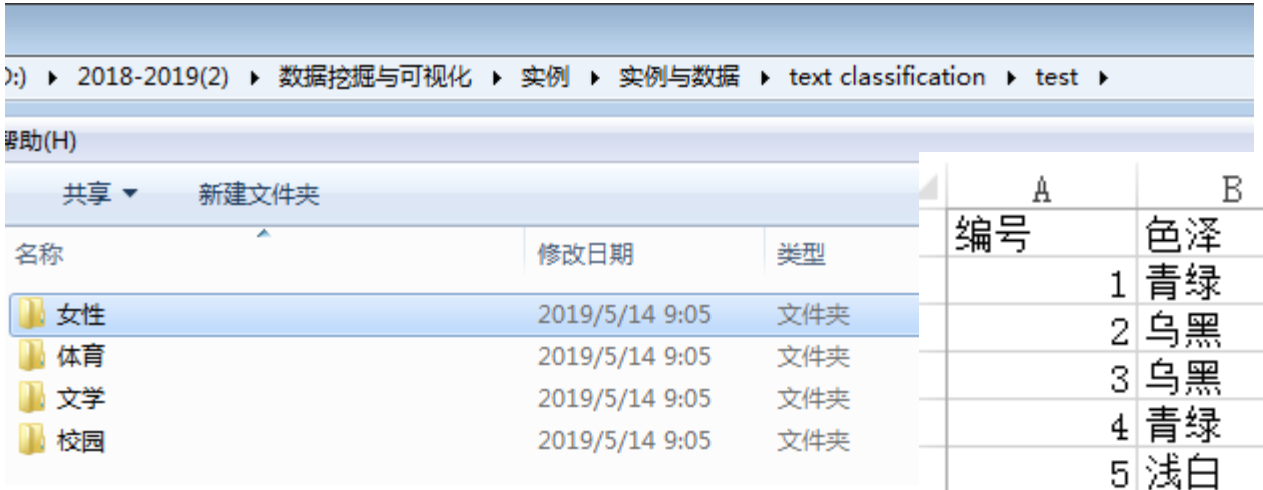
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
1,14.23,1.71,2.43,15.6,127,2.8,3.06,.28,2.29,5.64,1.04,3.92,106
1,13.2,1.78,2.14,11.2,100,2.65,2.76,.26,1.28,4.38,1.05,3.4,1050
1,13.16,2.36,2.67,18.6,101,2.8,3.24,.3,2.81,5.68,1.03,3.17,1185
1,14.37,1.95,2.5,16.8,113,3.85,3.49,.24,2.18,7.8,.86,3.45,1480
1,13.24,2.59,2.87,21,118,2.8,2.69,.39,1.82,4.32,1.04,2.93,735
1,14.2,1.76,2.45,15.2,112,3.27,3.39,.34,1.97,6.75,1.05,2.85,145
1,14.39,1.87,2.45,14.6,96,2.5,2.52,.3,1.98,5.25,1.02,3.58,1290
1,14.06,2.15,2.61,17.6,
1,14.83,1.64,2.17,14,97
1,13.86,1.35,2.27,16,98
1,14.1,2.16,2.3,18,105,
1,14.12,1.48,2.32,16.8,
1,13.75,1.73,2.41,16,89
1,14.75,1.73,2.39,11.4,
1,14.38,1.87,2.38,12,10
1,13.63,1.81,2.7,17.2,1
1,14.3,1.92,2.72,20,120
1,13.83,1.57,2.62,20,11
1,14.19,1.59,2.48,16.5,
1,13.64,3.1,2.56,15.2,1
1,14.06,1.63,2.28,16,12
1,12.93,3.8,2.65,18.6,1
1,13.71,1.86,2.36,16.6,
```

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

1) Alcohol 酒精度
2) Malic acid 苹果酸
3) Ash 灰分
4) Alcalinity of ash 灰的碱性
5) Magnesium Mg含量(mg/L)
6) Total phenols 苯酚总量
7) Flavanoids 黄烷类(mg/L)
8) Nonflavanoid phenols 非黄烷类(mg/L)
9) Proanthocyanins 原花色素类(mg/L)
10) Color intensity酒的色密度
11) Hue 色调
12) OD280/OD315 of diluted wines 经稀释后的酒的蛋白质的光谱度量
13) Proline 脯氨酸(mg/L)

Alcohol,Malic acid,Ash,Alcalinity of ash,Magnesium, Total



文本分类数据集

	A	B	C	D	E	F	G	H
	编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
	1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
	2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
	3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
	4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
	5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
	6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
	7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
	8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
	9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
	10	青绿	硬挺	清脆	清晰	平坦	软粘	否
	11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
	12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
	13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
	14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
	15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
	16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
	17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

西瓜数据集

令 $D = \{x_1, x_2, \ldots, x_m\}$ 表示包含m 个示例的数据集，每个示例由d 个属性描述(例如上面的西瓜数据使用了3 个属性)，则每个示例 $x_i = (x_{i1}; x_{i2}; \ldots; x_{id})$ 是d 维样本空间X 中的一个向量 $x_i \in \mathcal{X}$， 其中 x_{ij} 是 x_i 在第 j 个属性上的取值, d 称为样本 x_i 的"维数" (dimensionality).

基本术语

训练 / 学习，训练数据 / 训练样本，训练集

从数据中学得模型的过程称为“**学习**”或者“**训练**”。这个过程通过执行某个学习算法来完成。

训练过程中使用的数据成为“训练数据”，其中每个样本称为一个“**训练样本**”。

训练样本的组成的集合称为“**训练集**”

基本术语

分类，回归

我们打算预测的结果是离散值，比如“好瓜 / 坏瓜”，“一流 / 二流 / 三流球队”，此类学习任务称为“分类”（classification）。

如果预测的是连续值，例如：西瓜的熟度0.85、0.37等，此类学习任务称为“回归”（regression）。比如：房价的预测等，也属于回归。

基本术语

测试样本，测试集

通过对样本数据学习得到“模型”后，需要对该模型进行测试其与“真实”的逼近程度，需要进行预测。预测的过程称为“测试”。被用于预测的样本数据成为“测试样本”，其数据的集合称为“测试集”。

基本术语

模型 / 假设，真相 / 真实

计算机通过训练数据学习得到的**模型**是关于数据的某种潜在的规律**又称为“假设”**。这种**潜在的规律自身则称为“真相”或者“真实”**。

学习过程就是为了找出或者逼近真相。

基本术语

聚类

聚类是一种机器学习方法，将训练集中数据分成若干组，每组成为一个“簇”；这些自动形成的簇可能对应一些潜在的概念划分，这样的学习过程有助于我们了解数据内在的规律。

基本术语

有监督学习，无监督学习

根据训练数据是否拥有标记信息，学习任务可大致分为两大类：监督学习和无监督学习。

基本术语

泛化（generalization）能力

学习得到的模型是用于新样本的能力，称为“泛化”能力。具有强泛化能力的模型能很好的适用于整个样本空间。

我们对于模型的追求之一就是极好的泛化能力。

假设空间

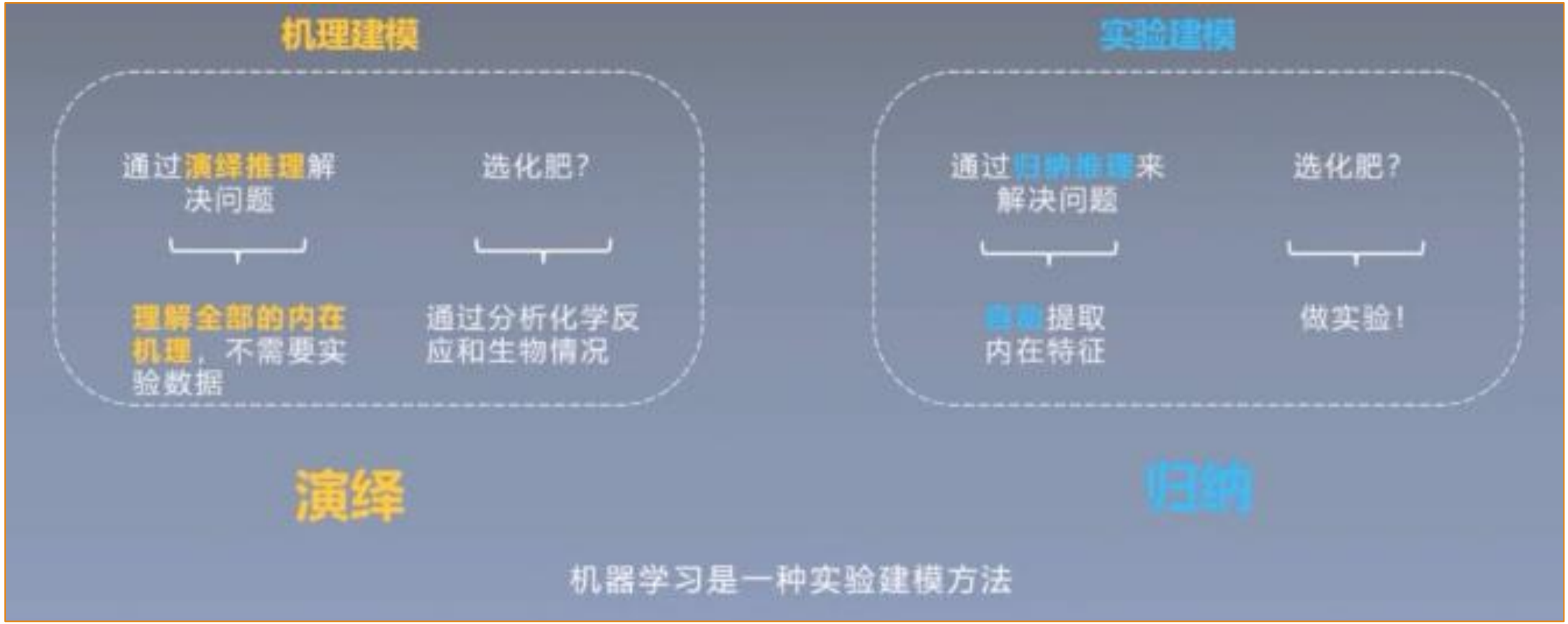
科学推理

归纳(induction) → 实验建模

从特殊到一般的泛化过程，即从具体的事实归结出一般性质规律。

演绎（deduction） → 机理建模

从一般到特殊的“特化”(Specialization)过程，即从基础原理推演出具体状况。



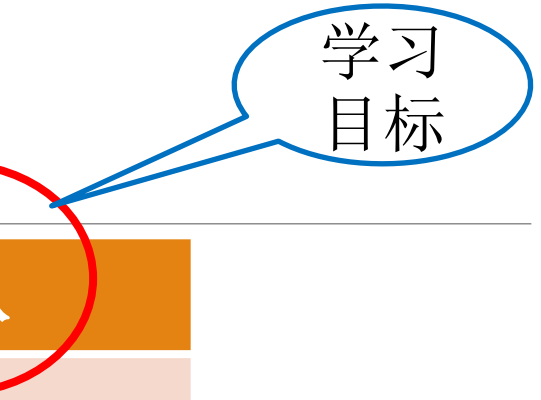
归纳建模方法的局限性:
优势:只要有大量数据就能归纳推理出一定结论。
劣势:不能做因果推断。
不能像演绎推理A->B 归纳推理只能推理出A与B相关

假设空间

归纳学习有狭义和广义之分，广义的归纳学习大体相当于从样例中学习，狭义的归纳学习则要求从训练数据中学得概念，因此又称为“**概念学习**”或“概念形成”。概念学习技术目前研究和应用均比较少（原因是太难）。现在常用的技术大多是产生“黑箱”模型。

概念学习中最基本的是布尔概念学习，即对“是”，“不是”这样的可表示为0/1布尔值的目标概念学习。

假设空间



编号	色泽	根蒂	敲声	好瓜
1	青绿	蜷缩	浊响	是
2	乌黑	蜷缩	浊响	是
3	青绿	硬挺	清脆	否
4	乌黑	稍蜷	沉闷	否

我们根据色泽、根蒂、敲声来判断是不是好瓜。我们学得概念即为：“好瓜是某种色泽、某种根蒂、某种敲声的瓜”这样的概念。用布尔表达式即表示为：

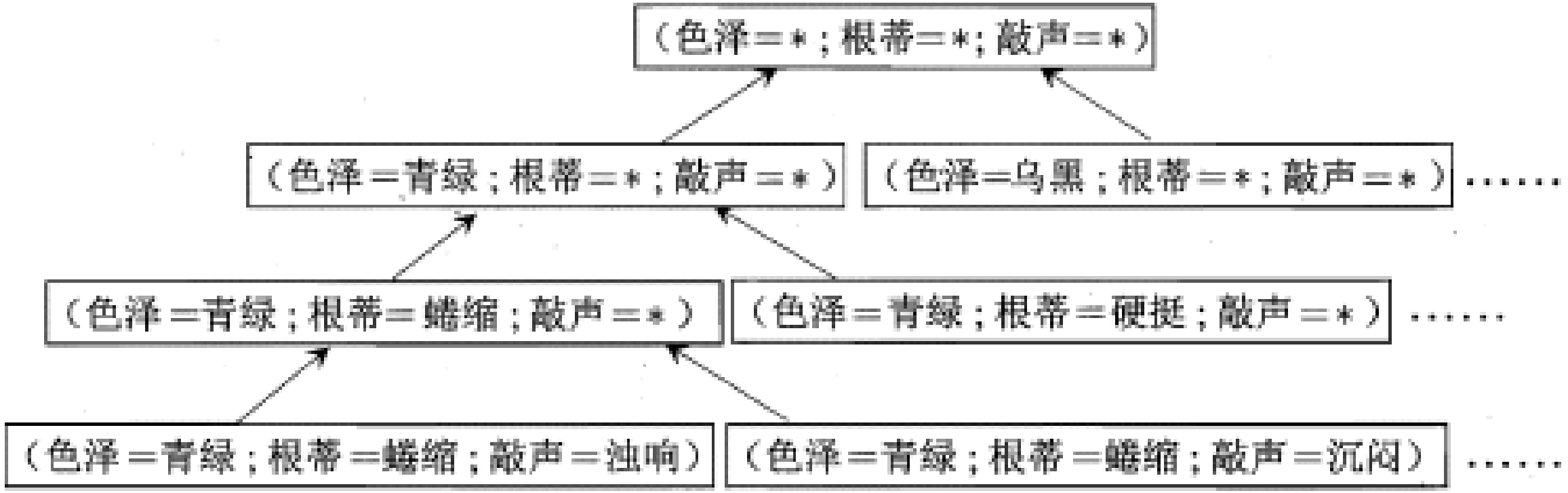
好瓜 ↔ （色泽＝？）∧（根蒂＝？）∧（敲声＝？）

假设空间

我们可以把学习过程看作一个在所有假设（hypothesis）组成的空间中进行搜索的过程，搜索目标是找到与训练集“匹配”的假设，即能将训练集中的样本判断正确的假设。

若西瓜的色泽、根蒂、敲声分别有３，２，２种选择，再加上这个特征可以没有，因此，西瓜的假设空间有

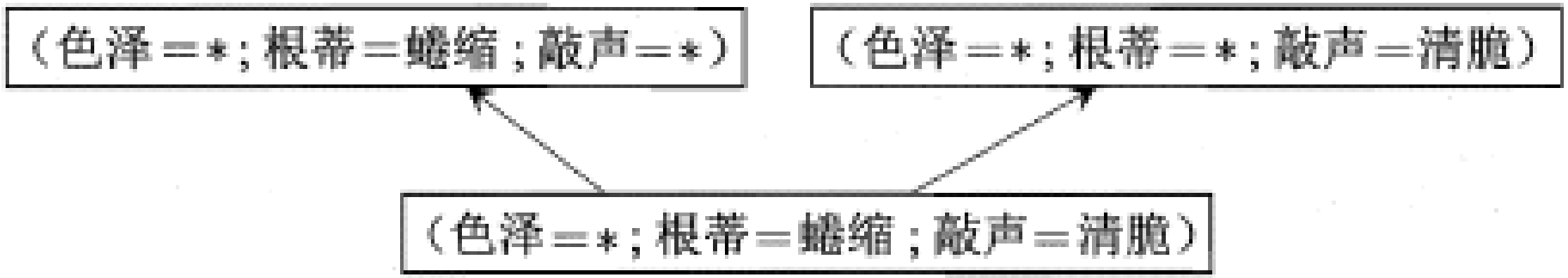
４＊３＊３＋１＝３７：


（色泽＝*；根蒂＝*；敲声＝*）
（色泽＝青绿；根蒂＝*；敲声＝*）
（色泽＝乌黑；根蒂＝*；敲声＝*）……
（色泽＝青绿；根蒂＝蜷缩；敲声＝*）
（色泽＝青绿；根蒂＝硬挺；敲声＝*）……
（色泽＝青绿；根蒂＝蜷缩；敲声＝浊响）
（色泽＝青绿；根蒂＝蜷缩；敲声＝沉闷）……

西瓜问题的假设空间

假设空间

注意：现实问题中我们面临很大的假设空间，但学习过程是基于有限样本训练集进行的，因此，有可能多个假设与训练集已知，即存在一个与训练集已知的“假设集合”，我们称之为“版本空间”



西瓜问题的版本空间

归纳偏好

奥卡姆剃刀（Occam’s razor）原则：

公元 14 世纪，来自奥卡姆的威廉（William of Ockham）对当时无休无止的关于“共相”、“本质”之类的争吵感到厌倦，于是著书立说，宣传只承认确实存在的东西，认为那些空洞无物的普遍性要领都是无用的累赘，应当被无情地“剃除”。他所主张的“思维经济原则”，概括起来就是**“如无必要，勿增实体。”**因为他叫威廉，来自奥卡姆，人们为了纪念他就把这句话称为“奥卡姆剃刀”。

归纳偏好

归纳偏好可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或“价值观”。

一般常用的原则来引导算法“正确性”确立偏好如：

奥卡姆剃刀原则：如有多个假设与观察一致，则选择最简单的那个。

奥卡姆剃刀并非唯一可行原则，并且奥卡姆剃刀的解释本身也可能有多个解释。

算法的归纳偏好是否与问题本身匹配，大多说情况直接决定了算法是否取得好性能。

归纳偏好

NFL定理（No Free Lunch Theorem没有免费午餐定理）

事实证明：当所有状况出现的机会相同，或所有问题同等重要，无论学习算法多聪明或者多少笨，它们的期望性能都相同。

脱离具体问题研究“什么算法更好”毫无意义。在某些问题上表现好的学习算法，在另外一些问题上不能尽如人意，学习算法自身的归纳偏好与问题是否匹配，往往起决定性作用。

机器学习三要素

通过对机器学习探索，发现其实无论用什么方法想要达到什么目的，其最终都是要求的一个能对新数据进行预测的公式，该公式可能是以概率的形式出现，即P（Y|X）；也可能是以函数的形式出现，即y=f（x）。

首先我们得明确我们求解思路，而思路可以归咎为以下公式：

公式（方法）= 模型 + 策略 + 算法

机器学习三要素

- **模型**

模型就是我们要求的，可以由输入产生正确输出的函数或者概率模型。求出这个模型是我们最终的目标。因此我们第一步要确定模型的范围，也就是确定假设空间。

条件概率分布的公式表达如下：

$$F = P|P(Y|X)$$

决策函数的公式表达为：

$$F = \{f|Y = f\theta(X), \theta \in R^n\}$$

上面所说的确定模型的范围就是指确定假设空间。

机器学习三要素

- **策略**

由于假设空间是模型的集合，而我们要从集合中选择具体的模型，我们就应该考虑选择的指标与依据。 我们所用到的策略方法有损失函数与风险函数两种。

- **损失函数**

损失函数用来度量预测错误的程度。常用的损失函数有0-1损失函数（等于设定值损失为零，不等于损失为1），平方损失函数（设定值与预测值的差的平方），绝对损失函数（设定值与预测值的差的绝对值）。 但是损失函数一般是用来度量模型对于一个样本的预测与分类的准确度。一般我们进行训练时，需要很多样本。

- 0-1损失函数 0-1 loss function

$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$$

- 平方损失函数 quadratic loss function

$$L(Y, f(X)) = (Y - f(X))^2$$

- 绝对损失函数 absolute loss function

$$L(Y, f(X)) = |Y - f(X)|$$

- 对数损失函数 logarithmic loss function 或对数似然损失函数 loglikelihood loss function

$$L(Y, P(Y \mid X)) = -\log P(Y \mid X)$$

机器学习三要素

- **风险函数**

若有多个样本，则可以通过求出每个样本的损失，然后求这些样本的平均损失，这个平均损失，就是模型的经验风险。风险函数可以度量模型对于多个样本的预测的准确度，除了经验风险，还有结构风险。

结构化风险是为了防止模型的过拟合，加入了一个正则化项

损失函数：模型一次预测好坏的度量

风险函数：模型平均意义下的预测好坏度量。

我们的学习目标就是：**选择期望风险最小的模型**。

期望风险：模型关于联合分布的期望损失。

由于联合分布【P(X,Y)】式未知的，我们也没法求，因此我们常用【经验风险最小】替代【期望风险最小】，因为根据大数定理，当样本容量趋于无穷时，经验风险也就趋于期望风险。

经验风险：模型关于训练样本的平均损失

$$R_{emp}(f)=\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))$$

因此我们的目标就是求上述公式值取最小时对应的参数值。即：

$$\min_{f\in\mathcal{F}}\ \frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))$$

但通过该方法求得的参数往往会造成过拟合，因此达不到对新数据的预测效果，因此常常用“结构风险最小化”来求参数，这样可以达到避免过拟合的效果。结构风险最小化其实也就是在经验风险最小公式的后面加入表示模型复杂程度的正则化项，可以表示为：

$$R_{\text{srm}}(f)=\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))+\lambda J(f)$$

$$\min_{f\in\mathcal{F}}\ \frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))+\lambda J(f)$$

其实求解最优模型，也就是求解最优化模型里的参数，而最优参数就是在结构风险取最小值时，对应的一些列参数的取值。如何利用算法解决上述公式，就是后面“算法”

机器学习三要素

- **算法**

算法是指学习模型的具体的计算方法，也就是求模型中的具体的参数的方法。一般会用到最优化得算法，比如最小二乘法、梯度下降等。

机器学习应用举例

人工智能
机器学习
深度学习 关系

机器学习比较活跃的四大领域

数据挖掘---->发现数据之间的关系

计算机视觉---->像人一样看懂世界

自然语言处理----->像人一样看懂文字

机器人决策---->像人一样具有决策能力

数据挖掘案例---案例1

id	性别	年龄	*总蛋白	白蛋白	*球蛋白	白球比例	甘油三酯	总胆固醇	血小板计数		血小板体积	比积	中性粒细胞%	淋巴细胞%	血糖
1	男	41	76.88	49.6	27.28	1.82	1.31	4.43	166		17.4	0.164	54.1	34.2	6.06
2	男	41	79.43	47.76	31.67	1.51	2.81	4.06	277		10.3	0.26	52	36.7	5.39
3	男	46	86.23	48	38.23	1.26	0.99	4.13	241		16.6	0.199	48.1	40.3	5.59
id	性别	年龄	70.98	44.02	26.96	1.63	1.06	6.89	252		10.8	0.26	41.7	46.5	4.3
5	女	48	78.05	41.83	36.22	1.15	0.97	5.37	316		14	0.35	56.6	33.1	5.42
6	女	74	76.46	45.85	30.61	1.5	2.45	6.65	249		17	0.211	42.9	47	5.97
7	男	31	80.82	46.4	34.42	1.35	1.51	5.6	246		13.3	0.27	52.9	32	5.11
8	男	55	74.17	41.63	32.54	1.28	1.32	5.92	282	...	13	0.3	52.8	36.9	5.94
9	男	39	76.12	49.31	26.81	1.84	2.07	5.5	275		11.4	0.27	53.1	35.9	5.66
10	女	35	75.94	46.64	29.3	1.59	1.25	3.98	247		12.6	0.27	65.6	27.3	5.48
11	男	47	75.84	43.71	32.13	1.36	1.96	4.32	178		13.5	0.19	54	35.7	4.44
12	男	50	73.8	48.32	25.48	1.9	3.32	4.47	259		12	0.27	57.8	33.1	5.48
.															.
.															.
.															.
5641	男	51	75.7	48.68	27.02	1.8	1.87	4.48	166		17.2	0.164	58.2	33.9	5.16
5642	女	38	76.12	44.87	31.25	1.44	1.42	4.78	571		10.1	0.58	55.9	27.6	5.13
5643	女	33	83.11	47.37	35.74	1.33	0.94	3.49	196		14.2	0.23	45.8	47.4	4.37

已知： 如下表格,共5643个样本,每个样本是一个人的体检信息。

目的： 根据这些体检值，求一个数学函数f，使其f(特征1，特征2，...，特征n) = 一个人的血糖值

数据挖掘案例---案例2

id	SNP1	身高	BMI分类	VAR00007	wbc		BUN	CHO	TG	SNP55	有/无糖尿病
1	2	158	0	1.574846	7.56		2.21	4.97	2.4	3	1
2	3	163	0	1.564441	9.38		2.25	6.61	3.93	2	0
3	1	170	0	1.888584	8.26		3.86	6.35	2.63	1	1
4	1	160	0	1.551809	12.88		2.54	5.65	1.18	2	1
5	1	158	0	1.652497	9.31		2.79	6.01	2.12	2	0
6	2	156	0	1.490654	9.63	. . .	2.07	7.62	2.59	1	1
7	2	158	0	1.564441	10.13		2.89	5.54	2.41	2	0
8	3	165	0	1.597365	7.21		2.89	6.89	1.69	3	1
9	2	160	0	1.490654	8.62		3.54	6.53	1.83	3	1
10	1	165	0	1.715598	6.54		3	5.57	1.53	3	1
11	2	164	0	1.621366	11.17		3.71	6.14	2.93	3	0
12	2	157	0	1.621366	6.8		3.82	6.92	2.12	3	0
13	2	160	0	1.528228	8.39		4.07	6.11	2.56	2	0
14	3	167	2	1.621366	11.72		2.93	6.06	2.88	3	1
.										.	.
.						. . .				.	.
.										.	.
1000	3	160	0	1.6245	7.07		3.03	4.94	1.67	3	1

已知： 如下表格,共1000个样本,每个样本是一个人的体检信息。

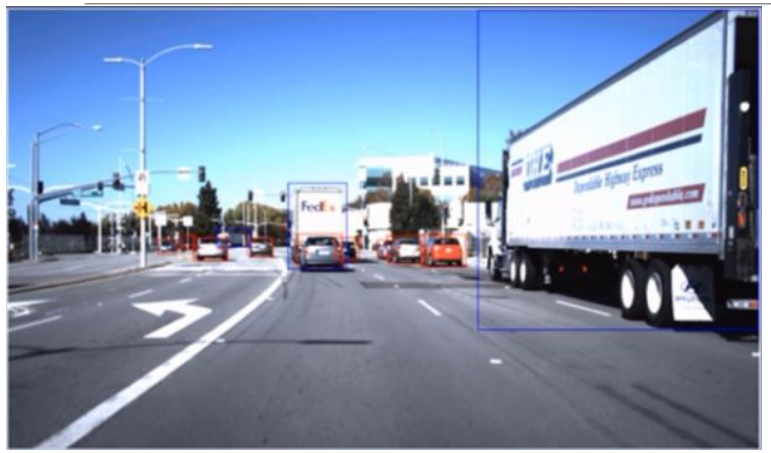
目的： 根据这些体检值，求一个数学函数f，使其f(特征1，特征2，...，特征n) = 一个人是否患糖尿病

计算机视觉案例---案例3 图像分类

已知：图片示例样本。

目的：求一个机器学习模型，使其

计算机视觉案例---案例4 目标检测



已知： 图片示例样本。

目的： 求一个机器学习模型，使其

已知：如右图所示，为2个示例样本

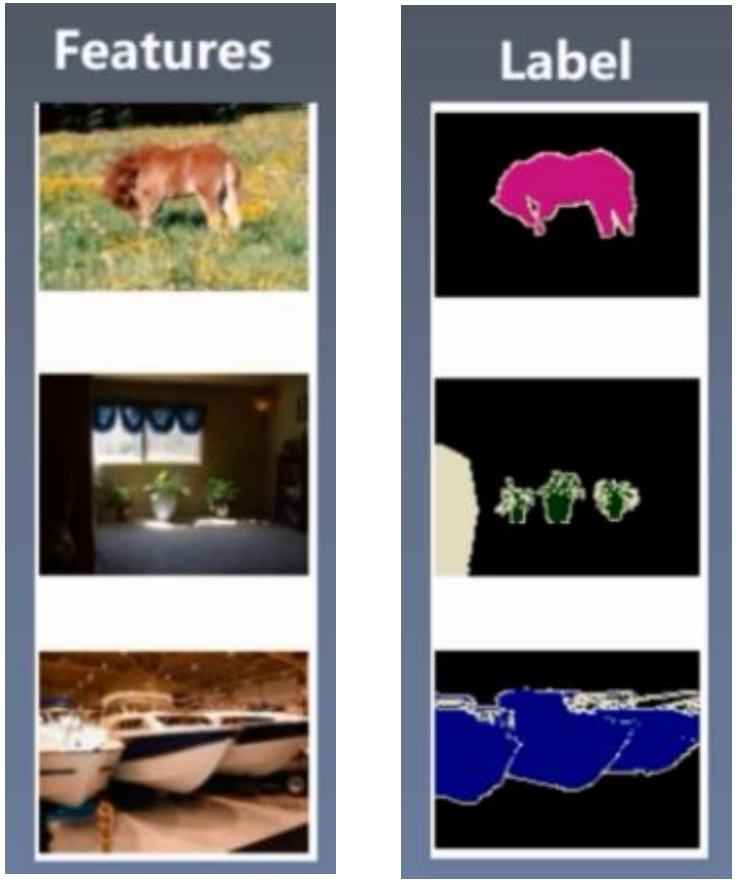
Features	Label (xmin, xmax, ymin, ymax, class)
原始图片0001	(20, 60, 50, 80, 4) ... (5, 20, 30, 60, 5)
⋮	⋮
原始图片5w	(100, 150, 100, 60, 7) ... (321, 299, 10, 50, 10)

计算机视觉案例---案例5 语义分割

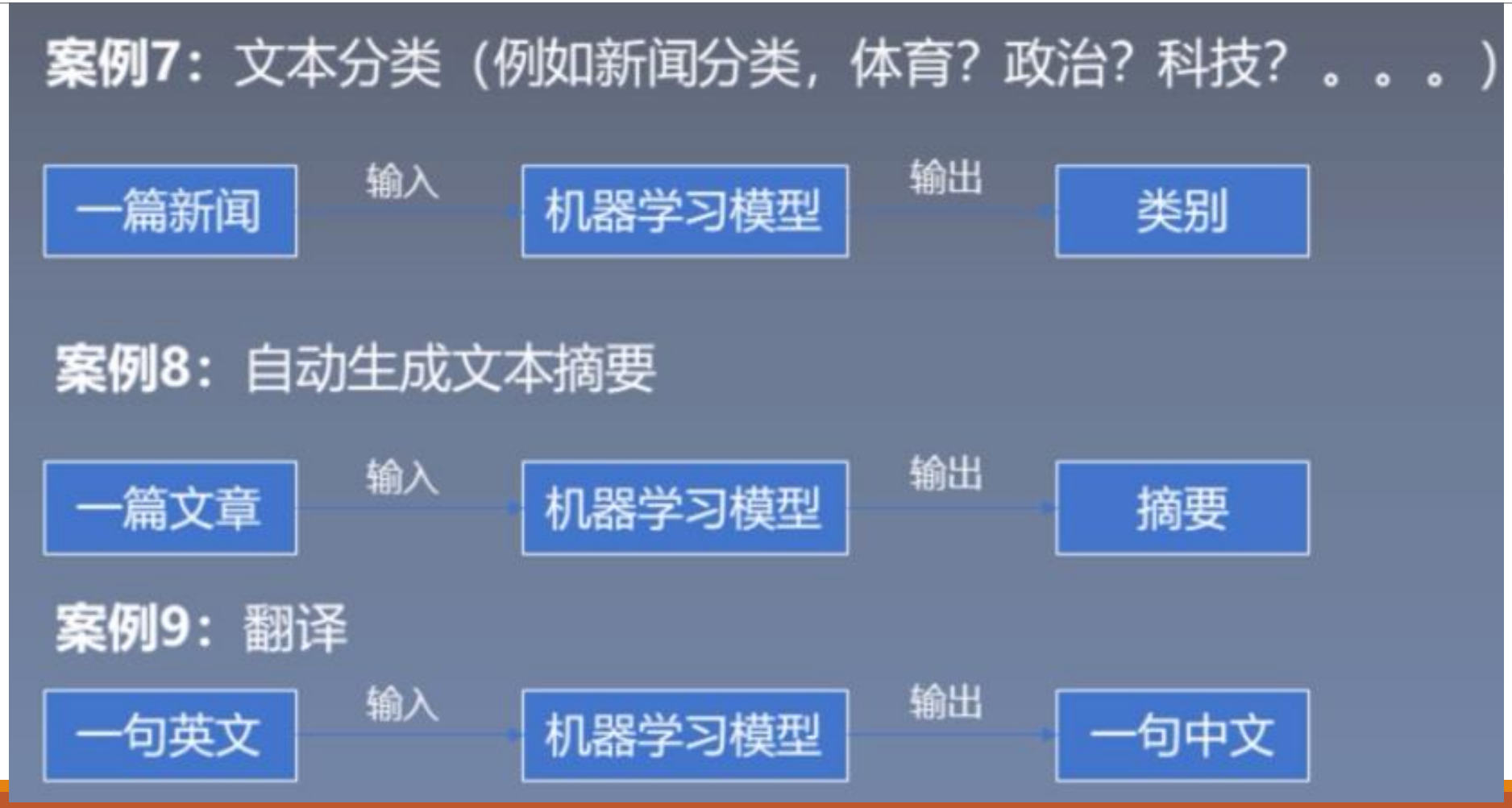
已知：右图示例三个样本。

对每个象素点进行分类

目的：求一个机器学习模型F，使其



自然语言处理案例


案例7：文本分类（例如新闻分类，体育？政治？科技？。。。）
一篇新闻
输入
机器学习模型
输出
类别
案例8：自动生成文本摘要
一篇文章
输入
机器学习模型
输出
摘要
案例9：翻译
一句英文
输入
机器学习模型
输出
一句中文

自然语言处理案例---问答系统

query	passage_text	人工答案
中国最大的内陆盆地是哪个	中国新疆的塔里木盆地，是世界上最大的内陆盆地，东西长约1500公里，南北最宽处约600公里。盆地底部海拔1000米左右，面积53万平方公里。	塔里木盆地
	中国最大的固定、半固定沙漠天山与昆仑山之间又有塔里木盆地，面积53万平方公里，是世界最大的内陆盆地。盆地中部是塔克拉玛干大沙漠，面积33.7万平方公里，为世界第二大流动性沙漠。	

已知：如右图所示，为一个实例样本

求出：一个机器学习模型f，使其

query+text

输入

机器学习模型

输出

答案

人机对话


例如：微软小冰
一句话
输入
机器学习模型
输出
一句话
你好
你坏
你他妈坏
你就是想搞事吧？
咋的啊
可不咋地.
哈哈
有件事我想了很久了，总感觉你的性格特别敏感，你是天蝎座还是双鱼座的？
调戏你

图像语义

已知：如下图所示，共4个实例样本

求出：一个机器学习模型f，使其

image

输入

机器学习模型

输出

text

自动驾驶

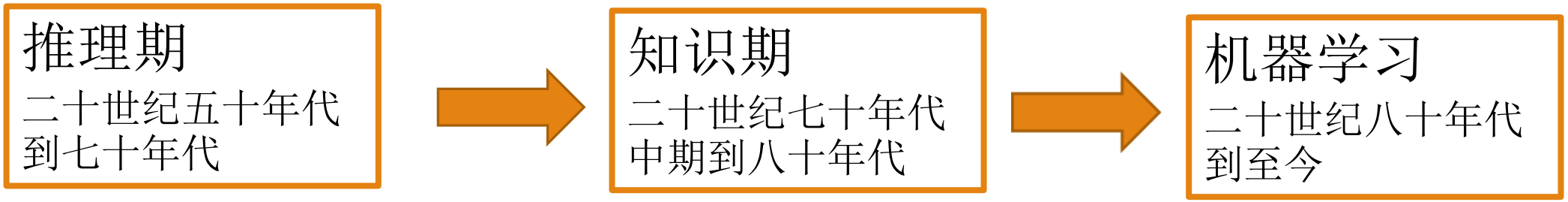
已知：如右图所示，为2个示例样本

原始图片0001	Label (车辆的控制信号)
⋮	⋮
原始图片5w	Label (车辆的控制信号)

求：一个机器学习模型f，使其 相机 一张图片 机器学习模型 控制信号 车辆

机器学习发展现状

机器学习是人工智能(artificial intelligence)研究发展到一定阶段的必然产物



机器学习发展现状

在二十世纪八十年代，“从样例中学习”的一大主流是**符号主义学习**，其代表包括决策树(decision tree)和基于逻辑的学习。

典型的**决策树学习**以信息论为基础，以信息熵的最小化为目标，直接模拟了人类对概念进行判定的树形流程。

基于逻辑的学习的著名代表是**归纳逻辑程序设计**(Inductive Logic Programming，简称ILP)，可看作机器学习与逻辑程序设计的交叉，它使用一阶逻辑(即谓词逻辑)来进行知识表示，通过修改和扩充逻辑表达式(例如Prolog表达式)来完成对数据的归纳.符号主义学习占据主流地位与整个人工智能领域的发展历程是分不开的

机器学习发展现状

ILP 具有很强的知识表示能力，可以较容易地表达出复杂数据关系，而且领域知识通常可方便地通过逻辑表达式进行描述，因此，ILP 不仅可利用领域知识辅助学习，还可通过学习对领域知识进行精化和增强;

由于表示能力太强，直接导致学习过程面临的假设空间太大、复杂度极高。因此，**问题规模稍大就难以有效进行学习，九十年代中期后这方面的研究相对陷入低潮**。

机器学习发展现状

目前人工智能的主要学派有下列三家：

(1) 符号主义(symbolicism)，又称为逻辑主义、心理学派或计算机学派，其原理主要为物理符号系统(即符号操作系统)假设和有限合理性原理。

(2) 连接主义(connectionism)，又称为仿生学派或生理学派，其主要原理为神经网络及神经网络间的连接机制与学习算法。

(3) 行为主义(actionism)，又称为进化主义或控制论学派，其原理为控制论及感知-动作型控制系统。

机器学习发展现状

1. 符号主义

认为人工智能源于数理逻辑。数理逻辑从19世纪末起得以迅速发展，到20世纪30年代开始用于描述智能行为。计算机出现后，又在计算机上实现了逻辑演绎系统。其有代表性的成果为启发式程序LT逻辑理论家，它证明了38条数学定理，表明了可以应用计算机研究人的思维过程，模拟人类智能活动。正是这些符号主义者，早在1956年首先采用“人工智能”这个术语。后来又发展了启发式算法>专家系统>知识工程理论与技术，并在20世纪80年代取得很大发展。符号主义曾长期一枝独秀，为人工智能的发展作出重要贡献，尤其是专家系统的成功开发与应用，为人工智能走向工程应用和实现理论联系实际具有特别重要的意义。在人工智能的其他学派出现之后，符号主义仍然是人工智能的主流派别。这个学派的代表人物有纽厄尔(Newell)、西蒙(Simon)和尼尔逊(Nilsson)等。

机器学习发展现状

2. 连接主义

认为人工智能源于仿生学，特别是对人脑模型的研究。它的代表性成果是1943年由生理学家麦卡洛克(McCulloch)和数理逻辑学家皮茨(Pitts)创立的脑模型，即MP模型，开创了用电子装置模仿人脑结构和功能的新途径。它从神经元开始进而研究神经网络模型和脑模型，开辟了人工智能的又一发展道路。20世纪60~70年代，连接主义，尤其是对以感知机(perceptron)为代表的脑模型的研究出现过热潮，由于受到当时的理论模型、生物原型和技术条件的限制，脑模型研究在20世纪70年代后期至80年代初期落入低潮。直到Hopfield教授在1982年和1984年发表两篇重要论文，提出用硬件模拟神经网络以后，连接主义才又重新抬头。1986年，鲁梅尔哈特(Rumelhart)等人提出多层网络中的反向传播(BP)算法。此后，连接主义势头大振，从模型到算法，从理论分析到工程实现，伟神经网络计算机走向市场打下基础。现在，对人工神经网络(ANN)的研究热情仍然较高，但研究成果没有像预想的那样好。

机器学习发展现状

3. 行为主义

认为人工智能源于控制论。控制论思想早在20世纪40~50年代就成为时代思潮的重要部分，影响了早期的人工智能工作者。维纳(Wiener)和麦克洛克(McCulloch)等人提出的控制论和自组织系统以及钱学森等人提出的工程控制论和生物控制论，影响了许多领域。控制论把神经系统的工作原理与信息理论、控制理论、逻辑以及计算机联系起来。早期的研究工作重点是模拟人在控制过程中的智能行为和作用，如对自寻优、自适应、自镇定、自组织和自学习等控制论系统的研究，并进行“控制论动物”的研制。到20世纪60~70年代，上述这些控制论系统的研究取得一定进展，播下智能控制和智能机器人的种子，并在20世纪80年代诞生了智能控制和智能机器人系统。行为主义是20世纪末才以人工智能新学派的面孔出现的，引起许多人的兴趣。这一学派的代表作者首推布鲁克斯(Brooks)的六足行走机器人，它被看作是新一代的“控制论动物”，是一个基于感知-动作模式模拟昆虫行为的控制系统。

机器学习发展现状

二十世纪九十年代中期之前，"从样例中学习"的另一主流技术是基于神经网络的**连接主义学习**.连接主义学习在二十世纪五十年代取得了大发展.

但因为早期的很多人工智能研究者对符号表示有特别偏爱，例如图灵奖得主ESimon 曾断言人工智能是研究“对智能行为的符号化建模”，所以当时连接主义的研究未被纳入主流人工智能研究范畴。尤其是连接主义自身也遇到了很大的障碍，正如图灵奖得主M. Minsky 丰11 S. Papert 在1969 年指出，(当时的)神经网络只能处理线性分类, 甚至对“异或”这么简单的问题都处理小了。1983 年，J. J. Hopfield 利用神经网络求解“流动推销员问题”这个著名的NP 难题取得重大进展，使得连接主义重新受到人们关注. 1986 年，D. E. Rumelhart 等人重新发明了著名的BP 算法，产生了深远影响。

机器学习发展现在

连接主义与符号主义

符号主义学习能产生明确的概念表示

连接主义学习产生的是“黑箱”模型

因此从知识获取的角度来看，连接主义学习技术有明显弱点；然而，由于有BP 这样有效的算法，使得它可以在很多现实问题上发挥作用。事实上， BP 一直是被应用得最广泛的机器学习算法之一。连接主义学习的最大局限是其“试错性” 简单地说，其学习过程涉及大量参数，而参数的设置缺乏理论指导，主要靠于工“调参”。夸张一点说，参数调节上失之毫厘，学习结果可能谬以千里

机器学习发展现状

二十世纪九十年代中期"**统计学习**" (statistical learning) 闪亮登场并迅速占据主流舞台，代表性技术是支持向量机(Support Vector Machine，简称SVM) 以及更一般的"核方法" (kernel methods).

机器学习发展现状

二十一世纪初，连接主义学习又卷土重来，掀起了以“**深度学习**”为名的热潮。所谓深度学习，狭义地说就是“很多层”的神经网络.在若干测试和竞赛上，尤其是涉及语音、图像等复杂对象的应用中，深度学习技术取得了优越性能。以往机器学习技术在应用中要取得好性能，对使用者的要求较高；而深度学习技术涉及的模型复杂度非常高，以至于只要下工夫“调参”把参数调节好，性能往往就好。因此，深度学习虽缺乏严格的理论基础，但它显著降低了机器学习应用者的门槛，为**机器学习技术走向工程实践带来了便利**。

本章小结

- 课程介绍

  （开课目的，课程要求，课程内容）

- 基本概念

（数据集，训练集，训练样本，测试集，测试样本，有监督学习，无监督学习，泛化能力）

- 假设空间
- 归纳偏好
- 发展历程

延伸阅读:符号主义、连接主义、行为主义