



# 人工智能：学习的框架

饶洋辉

计算机学院,

中山大学

[raoyangh@mail.sysu.edu.cn](mailto:raoyangh@mail.sysu.edu.cn)

<http://cse.sysu.edu.cn/node/2471>

课件来源：北京大学王文敏教授

- 学习框架（Learning framework）指的是可以用来实现机器学习范式及任务的算法设计途径和概念框架。
- 它属于算法设计的方法与概念层面上的问题，而非数据层面或物理层面的问题，亦不同于软件框架。
- 在一种机器学习的框架下，涵盖若干种子框架，每种种子框架又对应于若干种算法。
- 本章中有些子框架也被称为模型（Model），这是为了沿用学术界常用的术语所致。



## ☐ 学习框架概述

☐ 概率框架

☐ 统计框架

☐ 几何框架

☐ 联结框架

☐ 逻辑框架

# 学习框架概述

## ● 学习框架

设计机器学习算法时，依赖于所遵循的学习框架（Learning frameworks）。

换言之，框架的选用、改进、乃至融合等，是机器学习的一个重要环节。

基于此，本章将机器学习算法加以抽象，归纳出若干种学习框架。每种框架涵盖若干种子框架，每种子框架又对应于若干种算法。

可以将学习框架分为六大类：

- ❖ 概率框架（Probabilistic frameworks）
- ❖ 统计框架（Statistical frameworks）
- ❖ 几何框架（Geometric frameworks）
- ❖ 联结框架（Connected frameworks）
- ❖ 逻辑框架（Logical frameworks）
- ❖ 决策框架（Decision-making frameworks）

# 学习框架概述

## • 代表性的学习框架

框架	简短描述	子框架	典型算法
概率	采用概率来表示随机变量之间的条件相关性	生成	贝叶斯网络
		判别	条件随机场
统计	将统计学作为构建机器学习算法的框架	参数	线性回归
		非参数	k-近邻
几何	以几何概念为主要特征的机器学习算法	超平面	支持向量机
		距离	单链聚类
		流形	等距映射
联结	基于联结思想的人工神经网络方式	浅层与深层	感知机与CNN
		前馈与非前馈	CNN与RNN
逻辑	以逻辑或规则为基础的学习算法	基于逻辑	归纳逻辑编程
		基于规则	关联规则
决策	用于序贯决策的马尔科夫决策过程	基于模型	动态规划
		模型无关	强化学习



- 学习框架概述

- 概率框架

- 统计框架

- 几何框架

- 联结框架

- 逻辑框架

# 概率框架

- 关于概率框架

概率框架（Probabilistic Framework）以概率论为基础，将学习任务归结为随机变量的概率分布。

概率论的两个学派：

- ❖ 频率学派（Frequentist）：

将概率看作是重复许多次的频次，例如反复投掷硬币确定其正反面出现的几率；

- ❖ 贝叶斯学派（Bayesian）：

给定一组特定数据，概率是对事件中的不确定性进行量化，旨在推断出最接近事实的真值。

# 概率框架

## • 贝叶斯规则

基于贝叶斯规则（Bayes rule），可以从数据得到假设，再利用假设进行预测。

给定（已知）输入变量集合 $X$ ，（未知）输出变量集合为 $Y$ ，要训练学习算法得到一个假设函数  $h: X \rightarrow Y$ ，如用条件概率表示，则为  $P(Y|X)$ ，依据贝叶斯规则，则有：

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

若将其中的 $Y$ 和 $X$ 替换成 hypothesis（假设）和 data（数据），得到：

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{data}|\text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

还可以写成：

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

已知先验概率和证据，若再知道似然性，则可计算后验概率。这就是贝叶斯规则的精髓。



# 概率框架

## • 生成与判别模型

被广泛用于监督学习。以机器学习的分类任务为例：

### ❖ 生成模型（generative models）

- 对每个类的实际分布（actual distribution）进行建模
- 学到的是一种联合概率分布（joint probability distribution），即 $P(Y, X)$

### ❖ 判别模型（discriminative models）

- 对类之间的决策边界（decision boundary）进行建模
- 学到的是一种条件概率（conditional probability），即 $P(Y|X)$

生成模型	判别模型
朴素贝叶斯（Naïve Bayes）	逻辑回归（Logistic regression）
贝叶斯网络（Bayesian networks）	标量向量机（Scalar Vector Machine）
马尔科夫随机场（Markov random fields）	$k$ -近邻（ $k$ -Nearest neighbour）
隐藏马尔科夫模型（Hidden Markov Models）	条件随机场（Conditional Random Fields）

# 概率框架

## • 概率图模型

概率图模型（probabilistic graphical model）是用图来表示随机变量之间条件相关结构的概率模型。

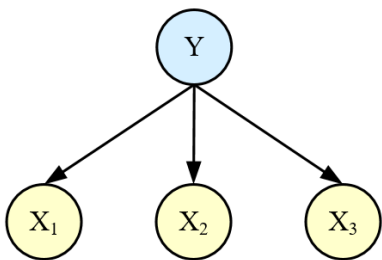
根据边的性质不同，可以将概率图模型分为两类：

### ❖ 有向概率图模型

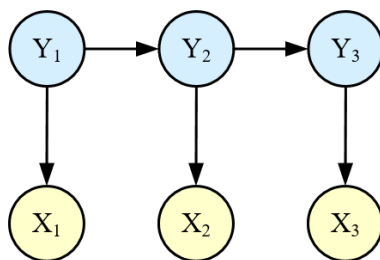
代表性的有向概率图模型是贝叶斯网络（Bayesian network），

### ❖ 无向概率图模型

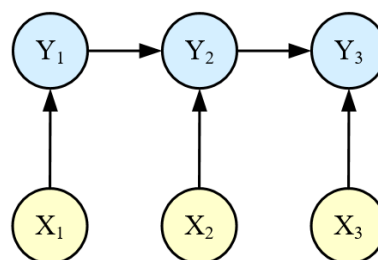
代表性的无向图模型则是马尔科夫网络（Markov network），又称马尔科夫随机场（Markov Random Field）。



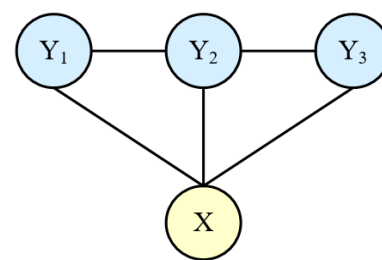
(a) 朴素贝叶斯



(b) 隐马尔科夫模型



(c) 最大熵马尔科夫模型



(d) 条件随机场



- 学习框架概述

- 概率框架

- 统计框架

- 几何框架

- 联结框架

- 逻辑框架

# 统计框架

- 关于统计框架

统计（Statistical）源于统计学（Statistics）和泛函分析（Functional analysis）。  
概率论是统计学的理论基础，两者又相辅相成。

有人打比方：

给定一桶球，两种颜色各占一定的比例，从中摸出若干球：

- ❖ 概率论是已知桶里的球，去推断手中球的颜色；
- ❖ 统计学则是已知手中球的颜色，再去计算桶里的球。



从参数的角度来看，统计框架又包含两个子框架：

- ❖ 参数模型（Parametric model）
- ❖ 非参数模型（Non-Parametric model）



# 统计框架

- 参数模型

**定义：** 设存在一个  $K$  维的有限参数向量  $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ 。若在给定该参数向量的条件下，特征预测  $x$  独立于观测数据  $\mathcal{D}$ ，即：  $P(x|\theta, \mathcal{D}) = P(x|\theta)$ ，则称为参数模型。

参数模型有固定个数的参数。由于  $x$  与观测数据  $\mathcal{D}$  无关，因此即使  $\mathcal{D}$  的量是无限的，该模型的复杂性也是有界的。

典型的参数模型之一是高斯混合模型（Gaussian mixture model）。它由  $K$  个部分混合而成，其中，第  $i$  个部分具有均值  $\mu_i$  和协方差  $\sigma_i$ ：

$$P(x|\theta) = \sum_{i=1}^K \phi_i \mathcal{N}(x|\mu_i, \sigma_i)$$
$$\mathcal{N}(x|\mu_i, \sigma_i) = \frac{1}{\sqrt{(2\pi)^K |\sigma_i|}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \sigma_i^{-1} (x - \mu_i)\right)$$

# 统计框架

- 非参数模型

**定义：**设数据分布不能用固定的参数来表示，而是用一个函数 $\theta$ 来定义，且 $\theta$ 能够随训练数据 $D$ 的量而增减，这就是非参数模型。

非参数模型不是无参数，而是不能用固定参数进行参数化，因而也被称为可变参数模型。

非参数模型	举例
功能分布（Distributions on functions）	高斯过程（Gaussian process）
聚类（Clustering）	中餐馆过程（Chinese restaurant process）
稀疏二元矩阵（Sparse binary matrices）	印度自助餐过程（Indian buffet process）
生存分析（Survival analysis）	贝塔过程（Beta processes）
测度分布（Distributions on measures）	完全随机测度（Completely random measures）

- 学习框架概述
- 概率框架
- 统计框架
- 几何框架
- 联结框架
- 逻辑框架

# 几何框架

- 关于几何框架

几何可分为欧式几何、微分几何、拓扑学、代数几何等。

- ❖ 欧式几何：传统的线、面、距离等

- ❖ 非欧几何：测地线距离、流形（manifold）等

在机器学习中，几何框架意味着从几何角度出发设计学习算法，即采用欧式几何的线、面、距离，或非欧几何的测地线距离、流形等来构建学习算法。

对于几何模型来说，若数据位于二维或三维空间，其主要优点是易于可视化。



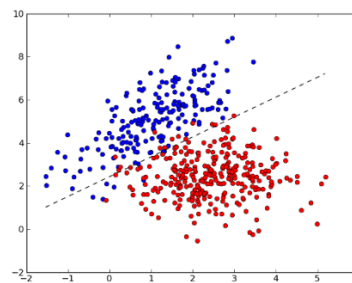
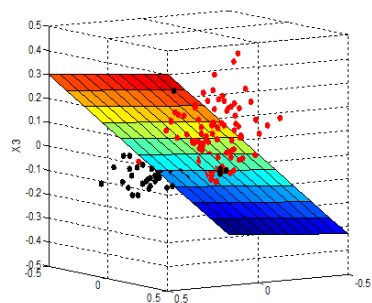
# 几何框架

## • 超平面

超平面（hyper-plane）是一个统称，与维度有关。

在几何学中，超平面是其维数小于其周围空间的一个子空间：

- ❖ 如果空间是三维的，则其中的超平面是二维的平面；
- ❖ 倘若空间是二维的，其中的超平面则是一维的线。



三维与二维空间中的超平面

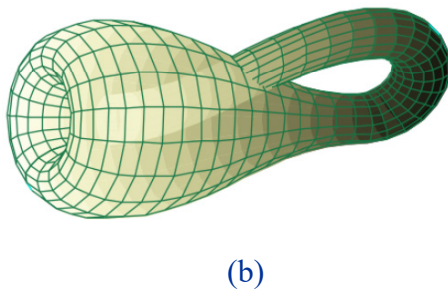
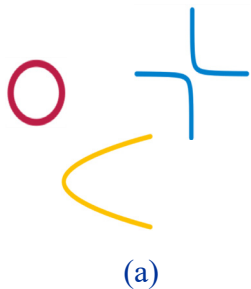
# 几何框架

- 流形

流形（manifold）源自德国数学家波恩哈德·黎曼（Bernhard Riemann）的德语术语 Mannigfaltigkeit。数学家江泽涵先生将其翻译成流形，据说出自于文天祥的《正气歌》。

**定义：**一个流形是一个拓扑空间，其中的每个点局部近似于欧氏空间。精确地说，一个  $n$  维流形上的每个点都可以找到一个与  $n$  维欧氏空间同胚（homeomorphic）的相邻点。

满足上述定义的流形很多，例如图(a)所示的线和圆、图(b)所示的曲面、等等。



流形在机器学习中，既可以用于降维，还可以用于聚类。



- 学习框架概述

- 概率框架

- 统计框架

- 几何框架

- 联结框架

- 逻辑框架

# 联结框架

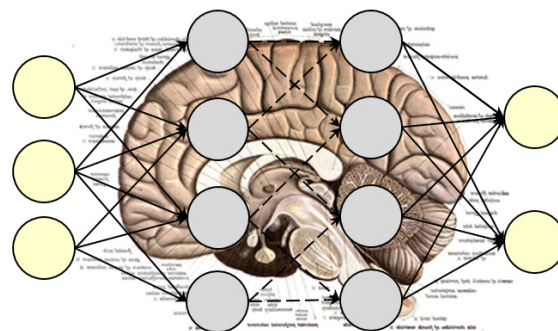
- 关于联结框架


联结框架（Connected frameworks）指的是基于联结的理念、即被称为联结主义（connectionism）的人工神经网络。

第1章介绍过联结主义。

它以神经网络理论为核心，主张通过模拟生物神经网络的结构和功能来建立其计算模型。

人工神经网络可看作是大脑的一种人工表征，用于模拟人类的学习过程。



- 
- 学习框架概述
  - 概率框架
  - 统计框架
  - 几何框架
  - 联结框架
  - 逻辑框架

# 逻辑框架

- 关于逻辑框架

逻辑框架（Logical frameworks）是机器学习中对问题的因果链假设的描述。

- ❖ 基于逻辑（logic-based）：

- 归纳逻辑编程（Inductive logic programming, ILP）
- 逻辑学习机（Logic learning machine）

- ❖ 基于规则（rule-based）：

- 关联规则学习（Association rule learning）
- 决策树（Decision tree）
- 随机森林（Random forest）

# 逻辑框架

## • 基于逻辑

归纳逻辑编程（Inductive Logic Programming, ILP）的形式化定义

设 $B$ 为背景知识（background knowledge）并采用霍恩子句（Horn clauses）的形式。又设 $E^+$ 和 $E^-$ 为正例和反例。一个假设（hypothesis） $h$ 是满足如下条件的逻辑命题：

必要性（Necessity）	$B$	$\not\models$	$E^+$
充分性（Sufficiency）	$B \wedge h$	$\models$	$E^+$
弱一致性（Weak consistency）	$B \wedge h$	$\not\models$	false
强一致性（Strong consistency）	$B \wedge h \wedge E^-$	$\not\models$	false

其模式（schema）如下：

$$positive\ example + negative\ example + background\ knowledge \Rightarrow hypothesis$$

例如：给定如下背景知识（ $B$ ），可通过推论得到一个简单的正例（ $E^+$ ）：

$B$  :       $parent(Helen, Mary) \wedge parent(Tom, Eve) \wedge female(Helen) \wedge female(Mary) \wedge female(Eve)$

$E^+$  :       $daughter(Mary, Helen) \wedge daughter(Eve, Tom)$

# 逻辑框架

- 基于逻辑

归纳逻辑编程是一种基于逻辑的机器学习方法。使用类Prolog语言的逻辑编程（logic programming, LP）形式，将背景知识、实例和假设统一表示。

## 逻辑编程 vs. 归纳逻辑编程

### 逻辑编程（LP）

逻辑与控制的结合。其中的逻辑，以户编写的程序方式出现。

LP = Logic + Control

### 归纳逻辑编程（ILP）

逻辑、统计与控制的结合。增加的统计部分使其表现力与计算能力都大幅度提高。

ILP = Logic + Statistics + Computational Control

归纳逻辑编程是归纳机器学习（inductive machine learning）的理论基础。

是一种从正例、反例、以及背景知识进行推论、得到假设的方法。



# 逻辑框架

- 基于规则

## 关联规则学习（Association rule learning）的形式化定义

设  $I = \{i_1, i_2, \dots, i_m\}$  是称为项（items）的二值属性的集合， $T = \{t_1, t_2, \dots, t_n\}$  是交易的集合，其中每个交易  $t_i \subseteq I$  有一个唯一的交易ID。一个关联规则是由两个不同的项集（itemsets） $X$  和  $Y$  所组成，表现为  $X \Rightarrow Y$  的形式， $X$  和  $Y$  分别称为关联规则的前提和结论。其中  $X, Y \subseteq I$ ，且  $X \cap Y = \emptyset$ 。

### 支持度（support）

$\text{supp}(X)$  是在交易中出现的频率， $\text{supp}(X \Rightarrow Y)$  是该规则中所有项集的支持度：

$$\text{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|} = P(X) \qquad \text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y) = P(X \cap Y)$$

### 置信度（confidence）

$\text{conf}(X \Rightarrow Y)$  是以该规则的前提为条件、求其结论的概率：

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \Rightarrow Y)}{\text{supp}(X)} = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)} = \frac{P(X \cap Y)}{P(X)} = P(Y|X)$$