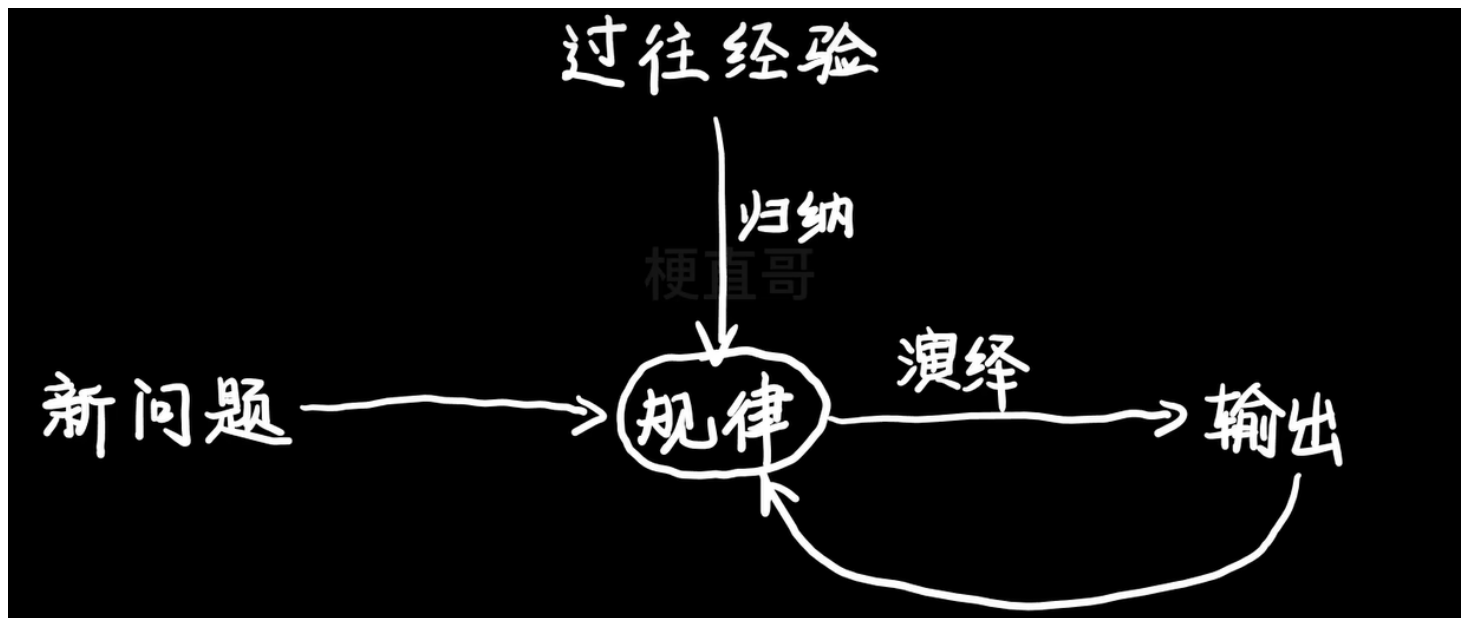


机器学习

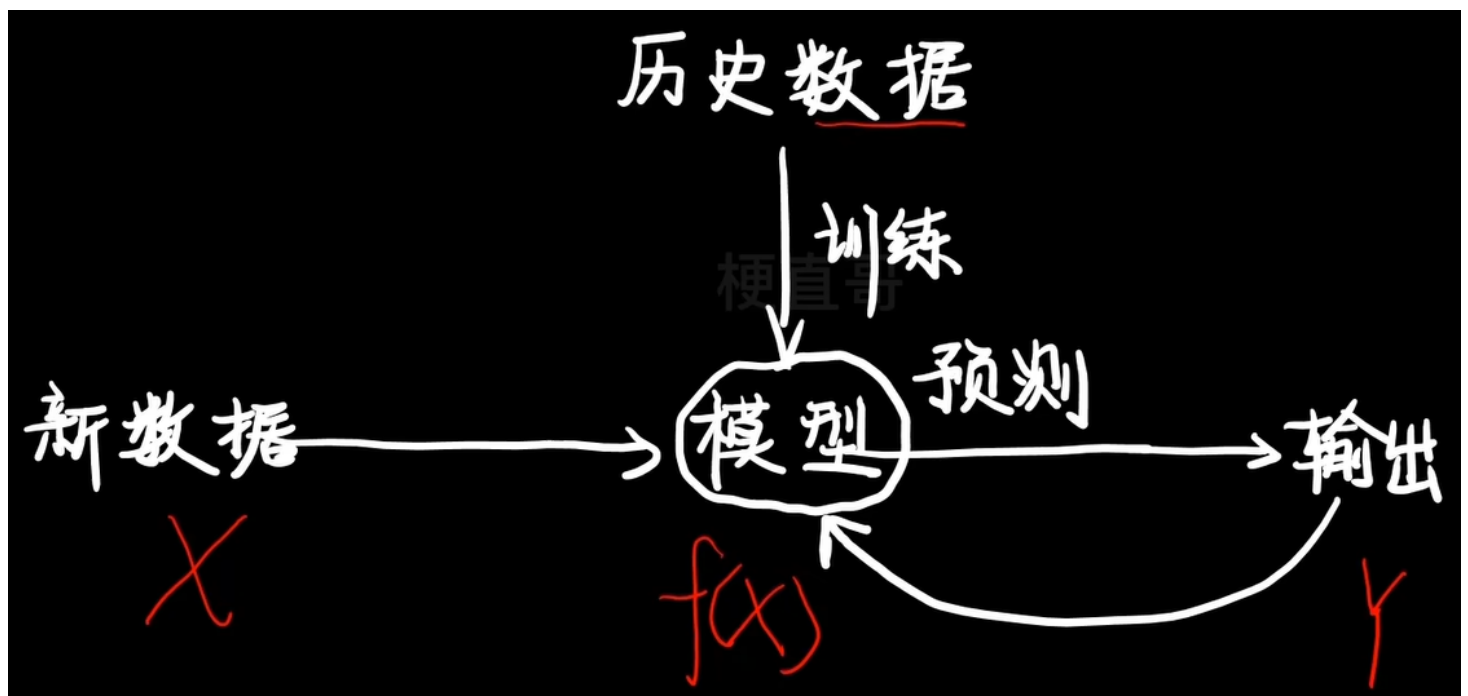
def: 计算机从数据中学习规律并改善自身进行预测的过程

why to learn: 人与计算机理解方式不同; 规则系统不够用; 逻辑判断专家参与耗时耗力

人类学习:



机器学习:



数据到模型的需要

1. 数据准备（一般占项目耗时的80%以上，像个excel表格，有输入特征与输出标签）
 - i. 对原始数据进行清洗和预处理（纠正缺失值和进行数据格式的转化）；
 - ii. 然后根据二八原则把整个数据集拆分成20%叫做测试集，80%叫做训练集。先80%的数据训练成模型，再用20%的数据检测模型好还是坏，然后再去选择模型
2. 模型学习
 - i. 根据目标问题进行**算法**的选择
 - ii. 对算法进行参数的调优，每种算法都涉及到很多超参数，不可能存在一个普遍适用于所有问题的解决方案
 - iii. 进行特征的选择，寻找最优的特征子集，减少特征的个数，提高模型的精确度，减少运行的时间
3. 模型评价（本质上机器学习就是给定一组x的值，预测y这样的过程）
 - i. 分类
 - a. 准确率
 - b. 灵敏度（召回率）
 - c. 特异性
 - d. 相关系数
 - ii. 回归
 - a. 确定系数
 - b. 均方误差
 - c. 均方根误差

实际应用

打开网购页面，迎面而来的一大批商品都与你有关，背后就有某种**推荐算法**

- 商品画像
- 用户画像
- 商品推荐
- 搜索算法
- 以图搜图
- 商品排序
- 人脸检测
- 活体检测
- 人脸识别

机器学习适用的条件

- 问题存在某种模式
- 难以用人工总结定义
- 有足够样本数据

(强人工智能：机器能解决人都解决不了的问题)