

同濟大學  
TONGJI UNIVERSITY

王睿智

[ruizhiwang@tongji.edu.cn](mailto:ruizhiwang@tongji.edu.cn)

## 人工智能技术与应用

1.1 人工智能定义

1.2 人工智能、机器学习、深度学习

1.3 机器学习定义

1.4 机器学习系统的类型

1.5 机器学习工作流程

1.6 获取开放数据



## 大语言模型：

**GPT 系列** (OpenAI)

**BERT系列** (Google)

**LLaMA系列** (Meta)

**文心一言** (百度)

**讯飞星火** (科大讯飞)

**盘古** (华为)

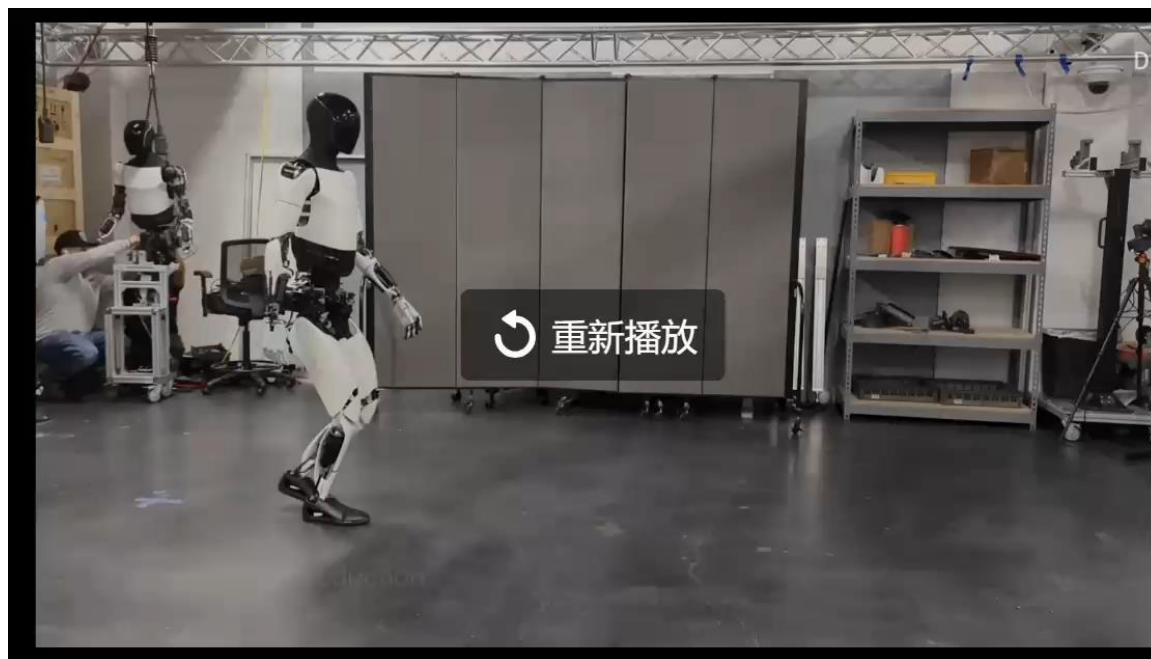
**通义千问** (阿里)



## 文生视频大模型

Sora(OpenAI, 2024)





特斯拉人形机器人-Optimus Gen-2



星辰智能AI机器人S1





ChatGPT, 如同其他基于GPT (Generative Pre-trained Transformer) 架构的模型, 基于Transformer结构。以下是一些关于它工作原理的基本点:

1. **Transformer 架构**: 这是近年来在自然语言处理领域中非常流行的一种深度学习模型架构。它主要基于自注意力机制, 能够捕捉句子中的长距离依赖关系。
2. **预训练和微调**:
  - **预训练**: 模型首先在大量的文本数据上进行预训练。在这个阶段, 模型学会了语言的基础知识、语法、事实和一些常识。通常, 这是通过让模型预测文本中的下一个词来完成的。
  - **微调**: 在预训练之后, 模型会在特定的任务数据上进行微调。例如, 为了创建ChatGPT, 模型可能会在与人类对话的数据集上进行微调, 以更好地回应用户的输入。
3. **生成文本**: 给定一个输入, GPT模型会生成一个输出序列。这是通过在每个时间步骤选择一个词, 并将这个词加入到输入序列中, 然后再次询问模型下一个词是什么, 直到达到特定的长度或生成一个结束标记为止。
4. **自注意力机制**: Transformer 的核心是自注意力机制, 它允许模型在生成文本时考虑到输入序列中的其他词。这意味着, 当模型生成一个回应时, 它不仅仅考虑到最后一个词, 而是考虑到整个句子或段落。
5. **参数**: 大型的GPT模型, 如GPT-3或ChatGPT, 有数十亿到数百亿的参数。这使它们能够存储大量的信息, 并在多种任务上表现出色。

简而言之, ChatGPT利用了预训练的语言模型的能力, 经过在对话数据上的微调, 使其能够与人类进行流畅的对话。

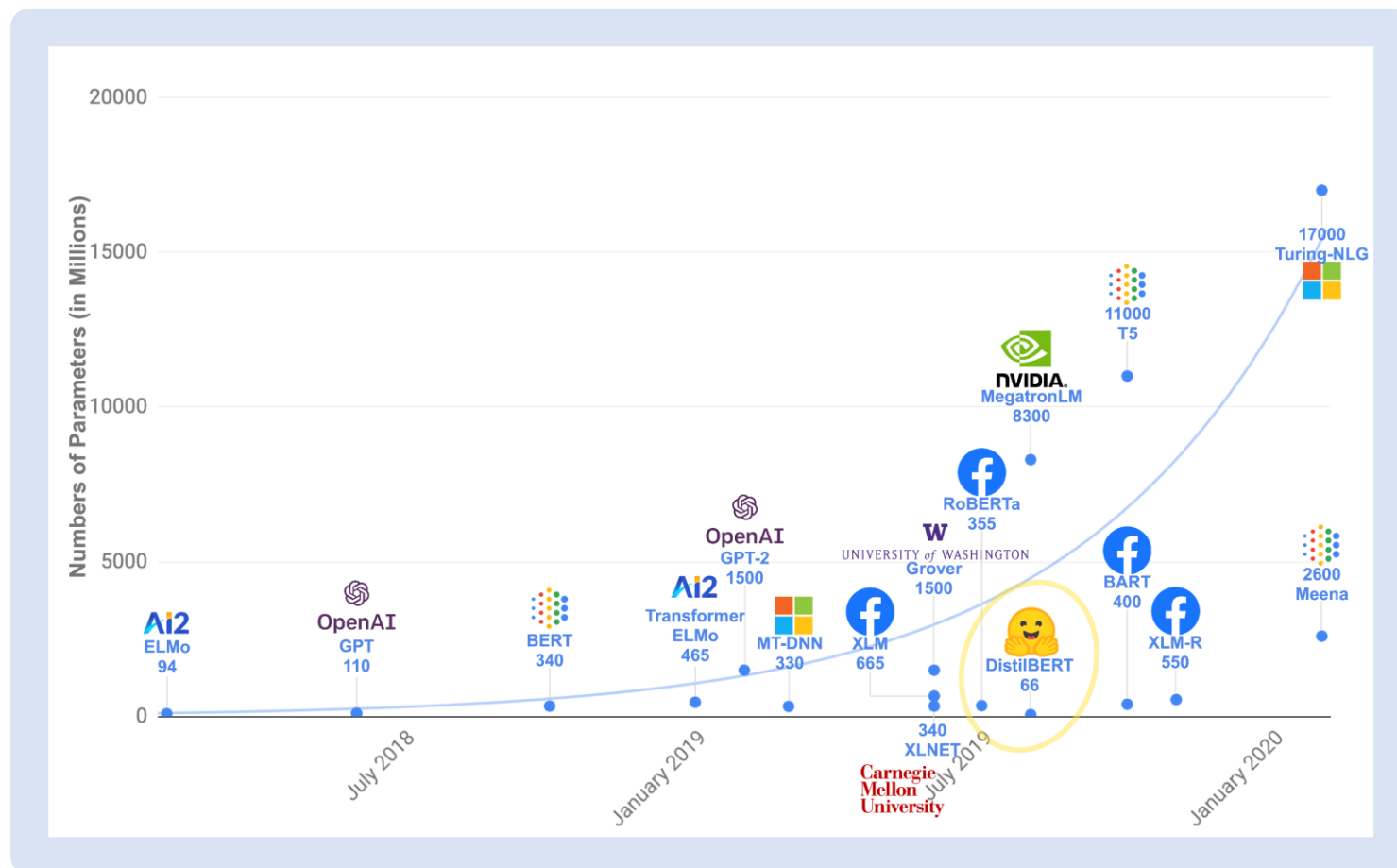
Sora是一种基于**扩散模型**和**Transformer**结构的视频生成系统。

以下是关于Sora原理的更详细解释：

1. **扩散模型 (Diffusion Model)**：扩散模型的核心思想是通过逐步去除视频中的噪声来生成视频。具体来说，Sora从一个看似静态噪声的视频片段开始，然后通过多个步骤逐步移除这些噪声，最终将视频从最初的随机像素转化为清晰的图像场景。
2. **Transformer结构**：Transformer结构在处理序列数据方面非常有效，它能够捕捉数据中的长距离依赖关系。在Sora中，视频被处理成有时序的向量，Transformer擅长预测一个向量接下来的向量，这使得它能够有效地处理视频数据。对于Sora而言，视频被分解为一系列称为patch的基本单位，这些patch类似于自然语言处理中的token，是构成视频的基本元素。
3. **VQ-VAE-2架构**：Sora可能还采用了VQ-VAE-2架构，这是一种改进的变分自编码器（VAE），它通过向量量化的方法来处理潜在空间的表示。在VQ-VAE-2中，潜在空间被量化为一组离散的向量，每个输入数据点都被映射到最接近的这些向量之一。这种改进使得模型能够更好地处理复杂数据结构，如高分辨率图像或视频。
4. **视频压缩网络**：Sora利用视频压缩网络对输入的图片或视频进行压缩，再通过空间时间补丁将其分解为基本元素，从而在压缩的潜在空间上实现训练和视频生成。
5. **灵感来源**：Sora的训练过程受到了大语言模型的启发，它采用了扩散型变换器模型，并结合了大规模训练数据，以提高生成视频的质量和真实性。

综上所述，Sora通过结合扩散模型和Transformer结构的强大能力，以及可能采用的VQ-VAE-2架构和视频压缩技术，能够生成高质量的视频内容。这些技术的融合使得Sora能够在压缩的潜在空间上有效地训练和生成视频，从而创造出逼真的视觉效果。

## 算力、算法、数据



GTP3模型参数有1750亿。  
GPT4参数高达1.8万亿，  
训练一次6300万美元！



### Artificial Intelligence (AI) 官方誕生日 达特茅斯会议, 1956

1956 Dartmouth Conference:  
The Founding Fathers of AI



John McCarthy  
麦卡锡



Marvin Minsky  
明斯基



Claude Shannon  
香农



Nathaniel S. Soltz  
所罗门诺夫



Allen Newell  
纽厄尔



Herbert A. Simon  
西蒙



Arthur Samuel  
塞缪尔



Oliver Selfridge  
塞弗里奇



Nathaniel Rochester  
罗切斯特



Terry A. Moore  
摩尔

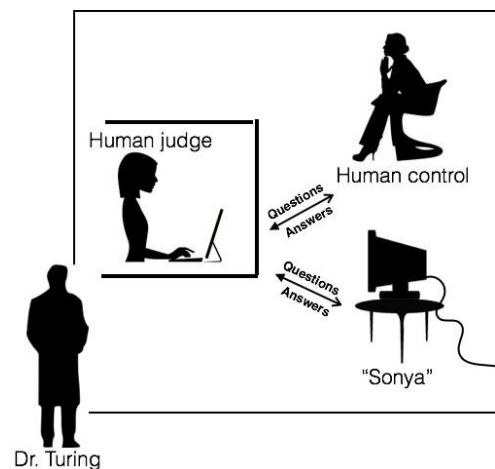
1956年夏，10名年轻学者在达特茅斯大学召开了两个月的学术研讨会，**讨论机器智能问题。**

会上经**麦卡锡**提议正式采用“Artificial Intelligence”这一术语，标志着人工智能学科正式诞生。

1950年 图灵发表的《计算机与智能》中提出“机器能否思考”问题，并设计了一种用以测试机器是否具有智能的方法，即**图灵测试**。



(Alan Mathison Turing, 1912-1954)



图灵测试

**定义1 马文·明斯基** 人工智能是一门科学，使得机器做那些人需要通过智能来做的事情。

### 定义2 智能机器

能够在各类环境中自主地或交互地执行各种拟人任务的机器。



IBM 研制的“深蓝”计算机系统



斯坦福大学研制的潜水机器人



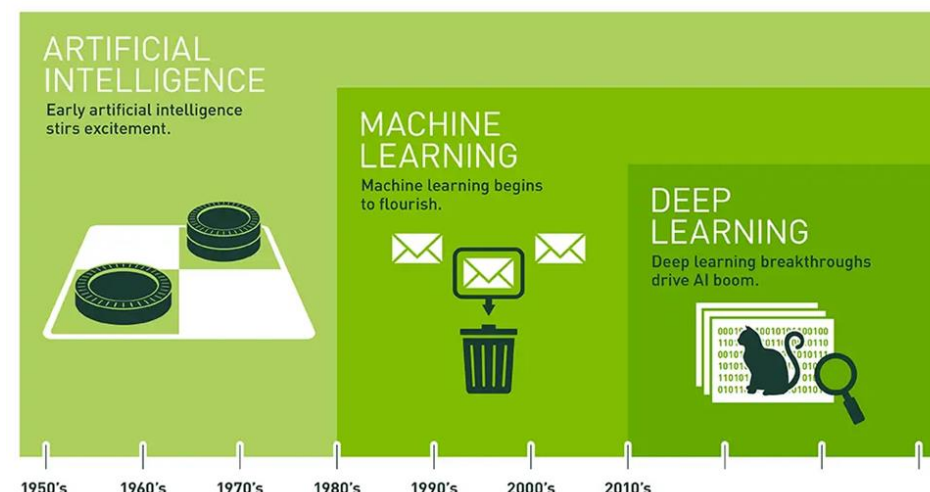
### 定义3 人工智能（能力）

人工智能（能力）是智能机器所执行的通常与人类智能有关的智能行为，如判断、推理、证明、识别、感知、理解、通信、设计、思考、规划、学习和问题求解等思维活动。

### 定义4 人工智能（学科）

人工智能（学科）是计算机科学中涉及研究、设计和应用智能机器的一个分支。它的近期主要目标在于研究用机器来模仿和执行人脑的某些智力功能，并开发相关理论和技术。

- 1950's~1980's 符号主义AI是AI主流。
  - 通过硬编码解决定义明确的逻辑问题，如国际象棋。
  - 难以给出明确的规则来解决更加复杂、模糊的问题，如图像分类、语音识别。
- 1980's~2010's 机器学习为AI主流
  - 机器学习给予计算机从数据中学习规则的能力。
  - 机器学习是数据驱动的AI，是AI的子集。
- 2006年后，深度学习成为AI主流
  - 深度学习是使用深层神经网络从数据中提取模式的技术。
  - 深度学习是机器学习的子集。



来源: [NVIDIA Blog](#)

### 定义1 研究的角度

机器学习是一个研究领域，让计算机无须显式编程就具备学习能力。

—— 亚瑟·塞缪尔 (Arthur Samuel) , 1959

机器学习系统从数据中学习出规则，是数据驱动的AI系统。



机器学习：一种新的编程范式



### 定义2 工程化的角度

一个计算机程序利用**经验E**来学习**任务T**，**性能是P**，如果针对**任务T**的**性能P**随着**经验E**不断增长，则称为机器学习。

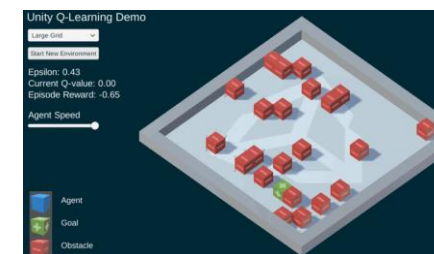
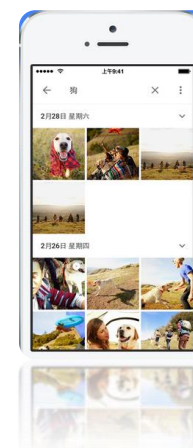
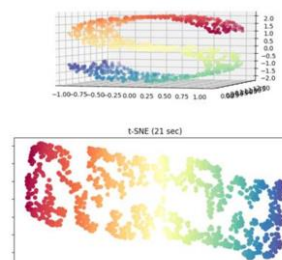
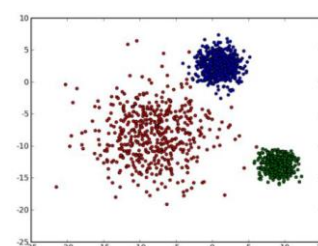
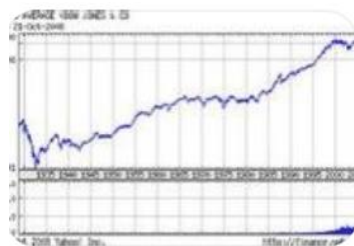
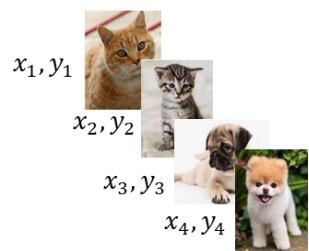
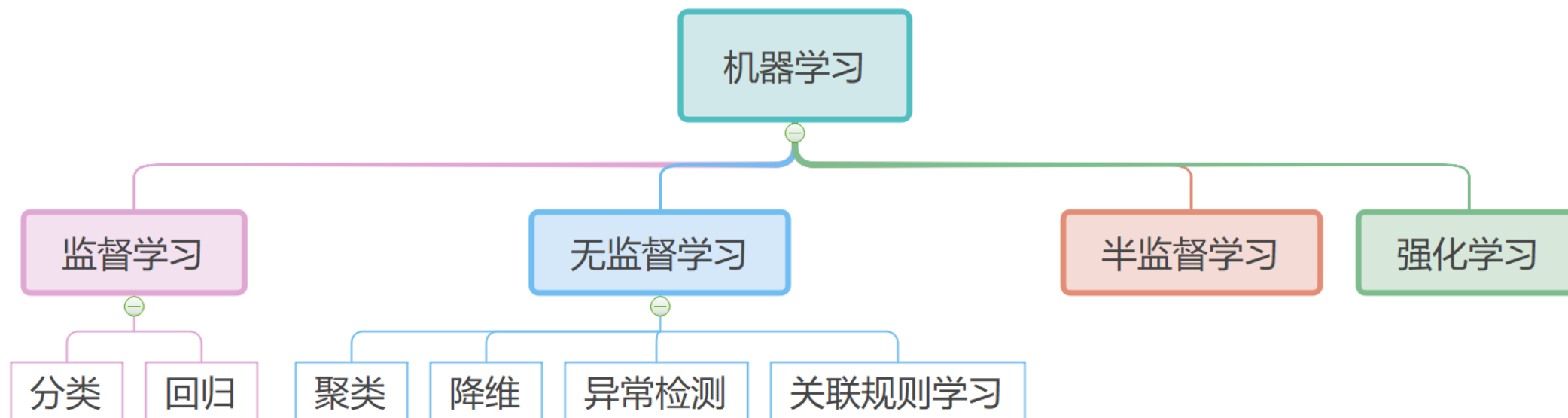
- 如，**垃圾邮件过滤器**是一个机器学习程序，它能从已经人工标注为垃圾邮件或非垃圾邮件的样例中学习判断垃圾邮件。系统用来进行学习的样例称为训练集。每个训练样例称为训练实例。

- 任务T，是判断新邮件是否为垃圾邮件；
- 经验E，是训练数据；
- 性能P，可定义为被正确分类的邮件所占比例。



来源: <https://www.mailsafi.com/email-spam-filter>

## 1.4 机器学习系统的类型



- **监督学习：**

已知一些历史样本的集合，其中每个样本有一组**特征**（也称输入），一个或几个表示其自身真实类别或数值的**标签**（也称输出）。将成对的输入和预期输出提供给算法，算法会找到一种方法（也称模型），根据给定输入给出预期输出。对新数据，也能给出可信度较高的预测输出。

- **监督学习算法：**从输入/输出对中进行学习的机器学习算法。

k-近邻

线性回归

逻辑回归

支持向量机

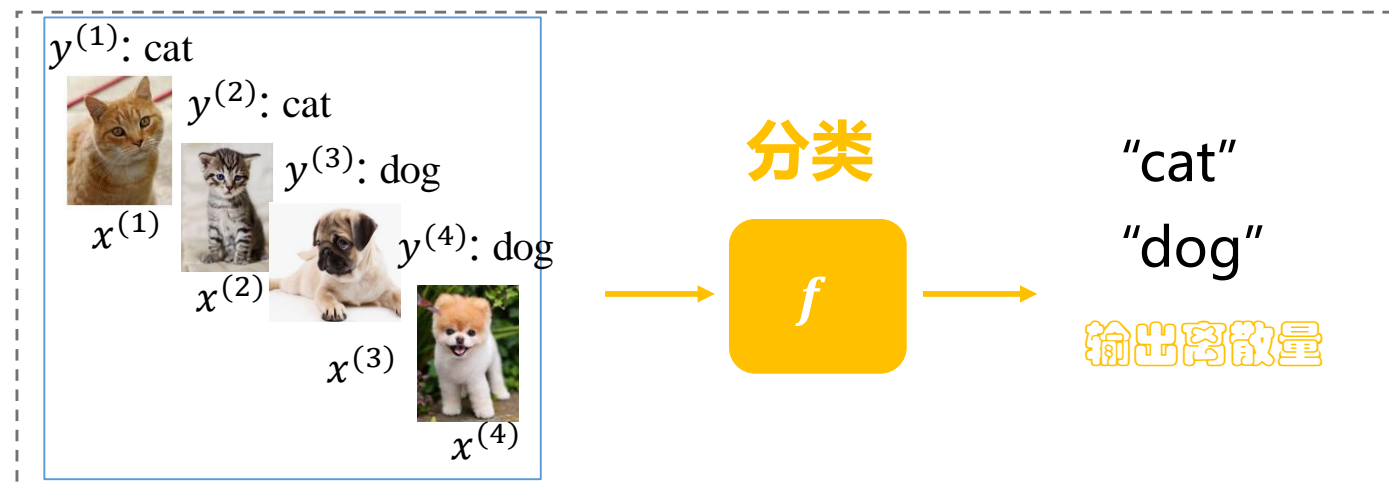
决策树与随机森林

神经网络 ...



## 监督学习两种主要任务：

- **分类**：标签是离散数据
- **回归**：标签是连续数据

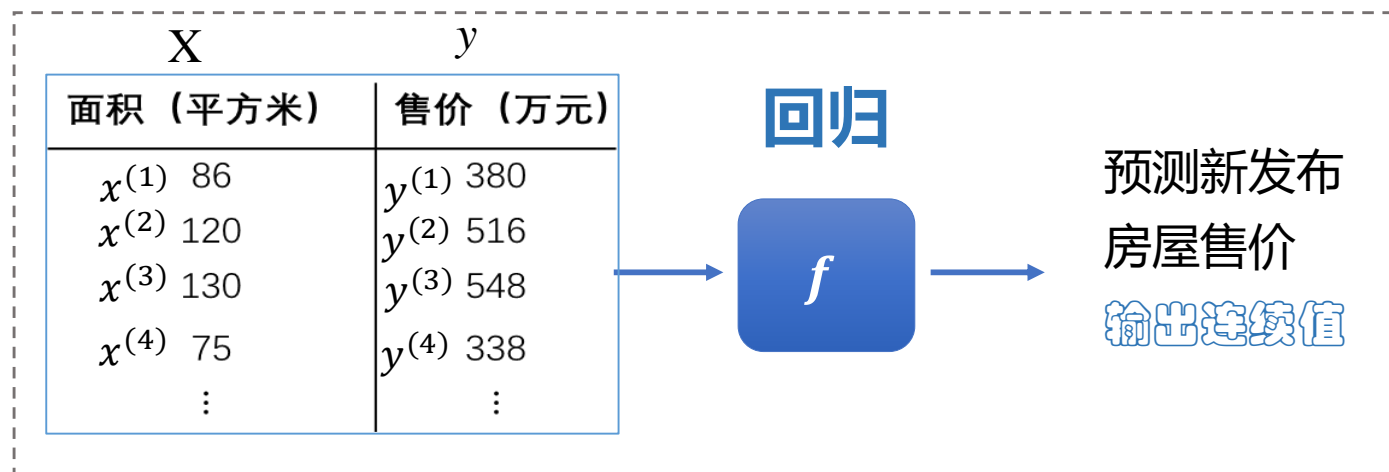


$x^{(i)}$  是第  $i$  个样本的特征向量。

$y^{(i)}$  是第  $i$  个样本的标签值。

$X$  是所有训练样本构成的特征矩阵；

$y$  是这些样本对应的标签数组。



### 监督学习任务示例：

- 识别信封上手写的邮政编码
- 基于医学影像判断肿瘤是否为良性
- 分析生产线上的产品图像来对产品进行自动分类
- 论坛中自动标记恶评
- 预测未来几天的天气或公司销售额的趋势

- **无监督学习：**

数据中没有标签信息。学习在无监督下进行，尝试在数据中发现模式。

- **无监督学习两种主要任务：**

- **聚类**：将数据集中类似的数据分为一组（称为簇）。
- **降维**：以保持数据间现有距离关系不变为目标，将高维数据转换为低维数据。

- **聚类算法**：k-均值算法、DBSCAN、分层聚类分析（HCA）等

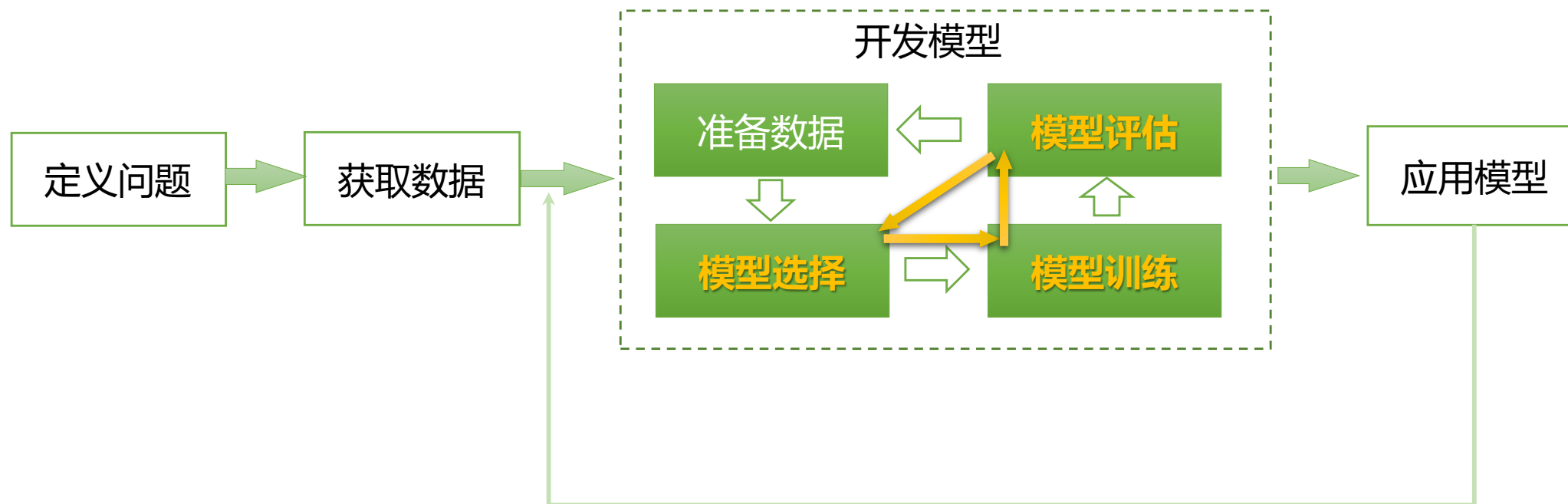
- **降维算法**：主成分分析（PCA）、核主成分分析、  
局部线性嵌入（LLE）、t-分布随机近邻嵌入（t-SNE）等



### 无监督学习任务示例:

- 检测信用卡欺诈
- 基于客户的购买记录对客户进行分组，对每一组客户设计不同的市场策略
- 用清晰而有洞察力的图表来表示复杂的高维数据集
- 在网络中发现社区结构，即社区发现

## 1.5 机器学习工作流程



### (1) 定义问题

- 了解问题所属领域和客户需求背后的业务逻辑
- 确定是否为机器学习的问题。属于哪类机器学习？

### (2) 获取数据

- 应根据问题定义采集相关数据，理解数据所代表的涵义。
- 数据可能来自机构内部、互联网或实时采集等。

### (3) 开发模型

- **准备数据**：使其可以被机器学习模型处理。  
数据清洗、类型变换、编码转换、数据合并等**数据预处理**。  
特征选择、特征变换、降维处理等**特征工程**操作。
- **模型选择**：探索不同机器学习算法、调试各种**超参数**以获得最优模型。  
**超参数**是需预先设定的模型参数，它们不能通过数据学习出来。
- **模型训练**：学习阶段。学出可以表征已知历史数据集的模型。  
在数据集上学习出模型的过程，称为**训练**。
- **模型评估**：衡量模型优劣，需要客观标准 (即评估方法)。  
若当前模型不满足性能要求，则需要调整模型参数、重新训练，再评估。

### (4) 应用模型

保存训练好的模型，将其部署到生产环境中，用于新数据的**预测**。  
开始收集构建下一代模型所需要的数据。



- **常用的开放数据存储库：**

- UC Irvine Machine Learning Repository(<http://archive.ics.uci.edu/ml/>)  
<https://archive-beta.ics.uci.edu/>
- Kaggle datasets(<https://www.Kaggle.com/dataset>)
- Amazon's AWS datasets(<https://aws.amazon.com/fr/datasets>)

- **列出流行的开放数据集的页面：**

- Wikipedia's list of Machine Learning datasets(<https://homl.info/9>)
- Quora.com(<https://homl.info/10>)
- The datasets subreddit(<https://www.reddit.com/r/datasets>)