同濟大學
TONGJI UNIVERSITY

王睿智 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 机器学习范式

1.5 机器学习工作流程

1.6 获取开放数据

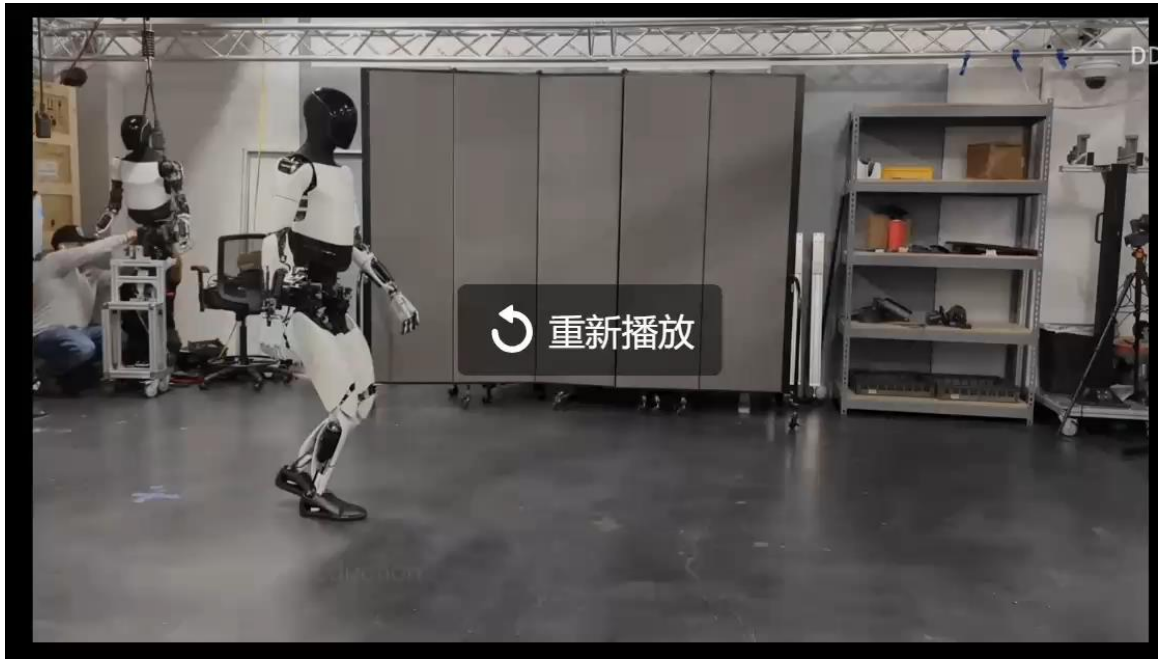
2025年初，火爆全网的DeepSeek-R1

1月27日，英伟达公司股票暴跌近17%，市值蒸发近6000亿美元。美媒称，这是受到中国AI初创公司DeepSeek人工智能模型的冲击。

1月17日，马斯克旗下人工智能公司XAI发布Grok-3，他称其为地球上最聪明的人工智能。靠巨大算力：20万张H100GPU

来源：央视新闻 https://v.cctv.com/2025/01/28/VIDEJJyBUWKf383Yb8pRoePS250128.shtml

人形机器人



2024年，特斯拉人形机器人-Optimus Gen-2

2025年初，宇树科技人形机器人G1

AI正在带来科学发展的新纪元

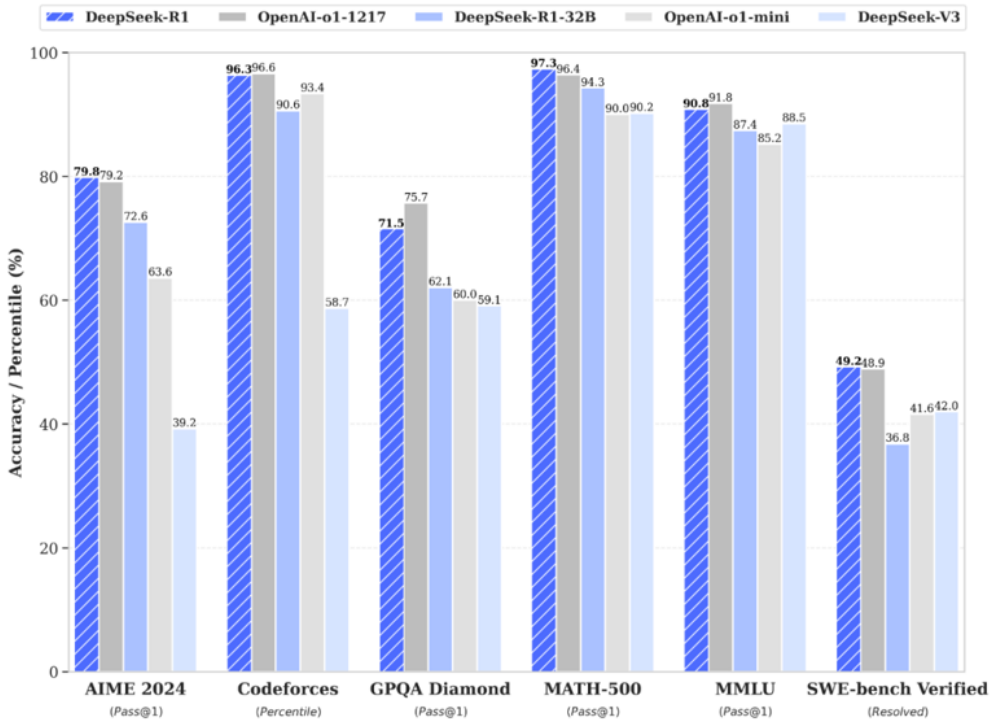
2024年诺贝尔物理学奖和化学奖

表彰他们“基于**人工神经网络实现机器学习**的基础性发现和发明深远影响”。

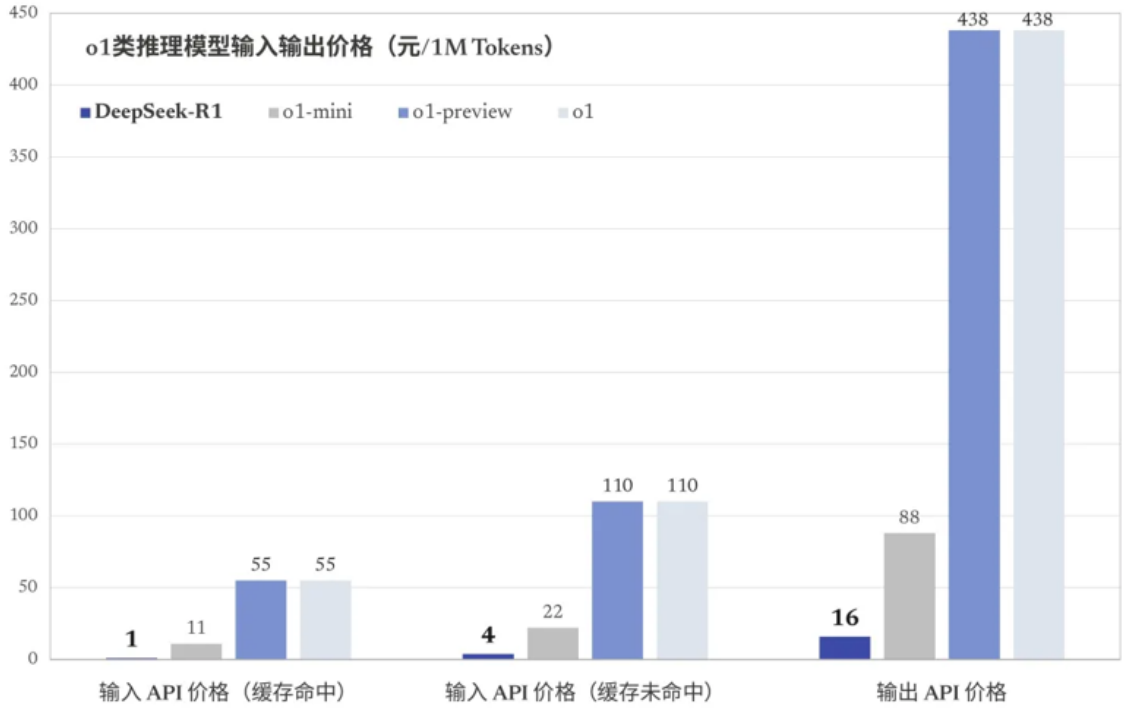
AlphaFold利用了**深度学习**技术，显著提高了蛋白质结构预测的准确性。

DeepSeek-R1爆火原因

DeepSeek性能对齐OpenAI-o1正式版



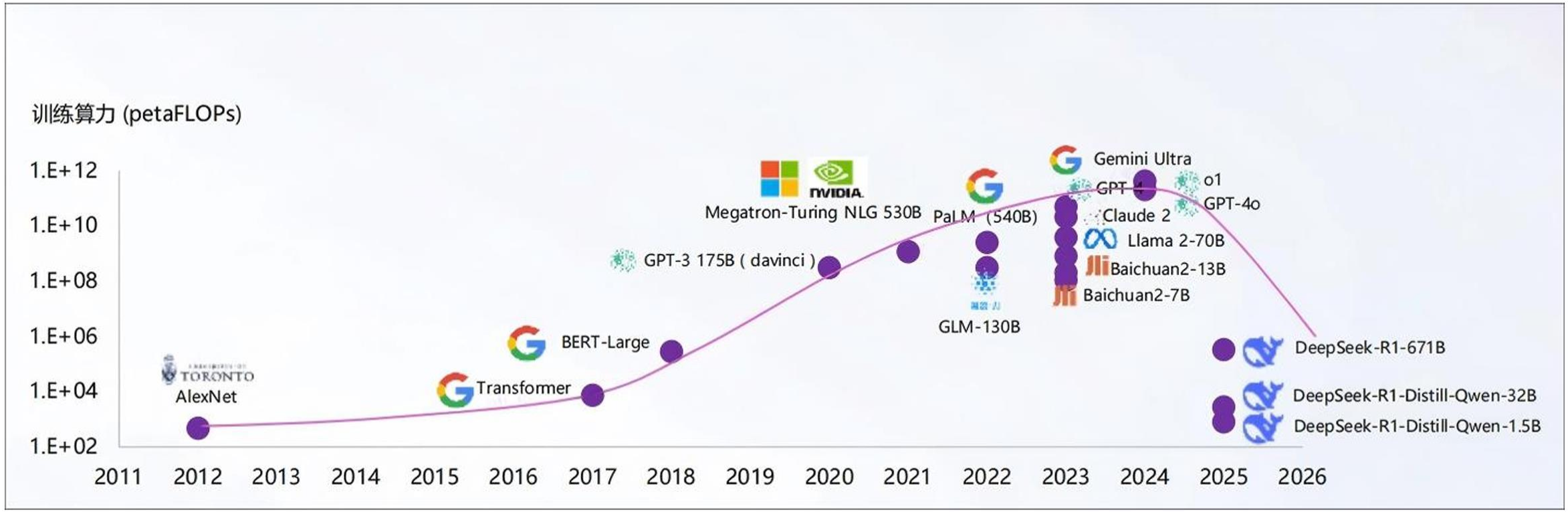
推理成本低至每百万Token0.14美元



来源：DeepSeek 官网，中国银行证券研究院

- DeepSeek-R1的推理能力进入了第一梯队（媲美OpenAI o1），但训练和推理成本低、速度快、全部开源。
- DeepSeek打破了硅谷传统的“堆算力、拼资本”的大模型发展路径。

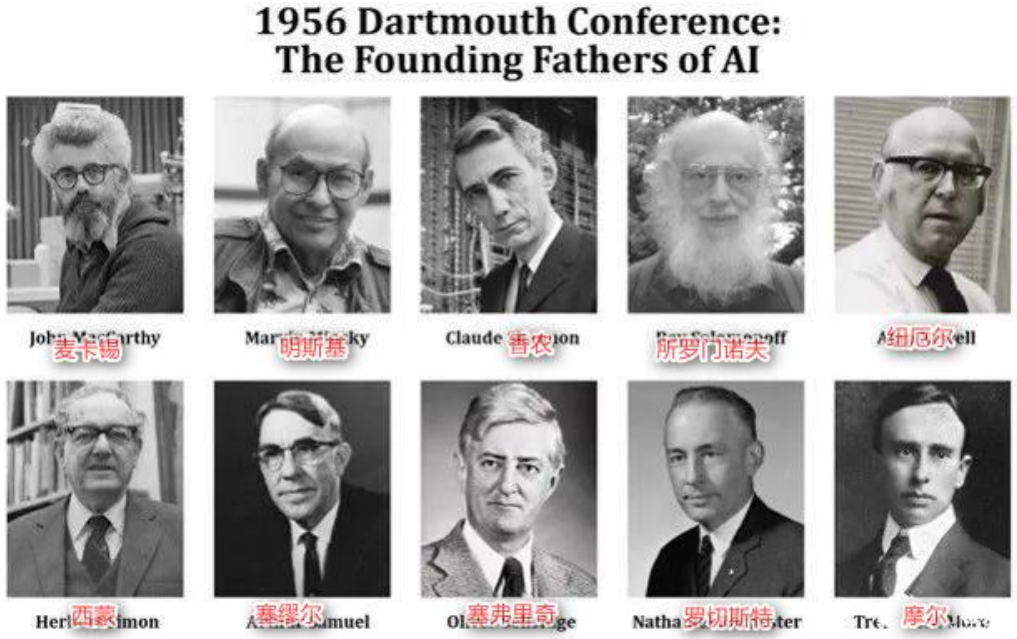
大模型发展对算力的需求演变



大模型相关术语

- **多模态**
  文本、图片、音频、视频
- **AI工具（国内）**
  DeepSeek、豆包、Kimi、腾讯元宝、智谱清言、秘塔搜索 ...
- **通用模型：**
  - 大语言模型
    - 通用大模型、
    - 推理大模型（思维链）、
    - 混合推理
  - 视频模型
  - 多模态模型
- **按应用领域可分为：**
  - 通用大模型
  - 行业模型
    教育、医疗、金融等
  - 垂直模型（垂类模型）

Artificial Intelligence (AI) 官方诞生日　达特茅斯会议, 1956

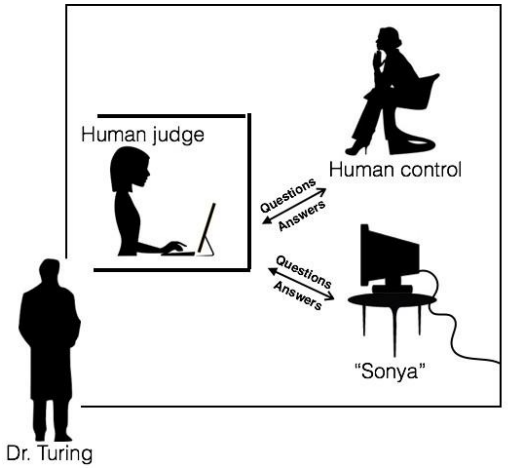


1956年夏，10名年轻学者在达特茅斯大学召开了两个月的学术研讨会，**讨论机器智能问题。**

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《计算机器与智能》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

1.1 人工智能定义

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。

马文•明斯基(1969图灵奖)

人工智能是关于**知识**的学科，即怎样表示知识以及怎样获得知识并使用知识的科学。

尼尔逊（斯坦福AI中心教授）

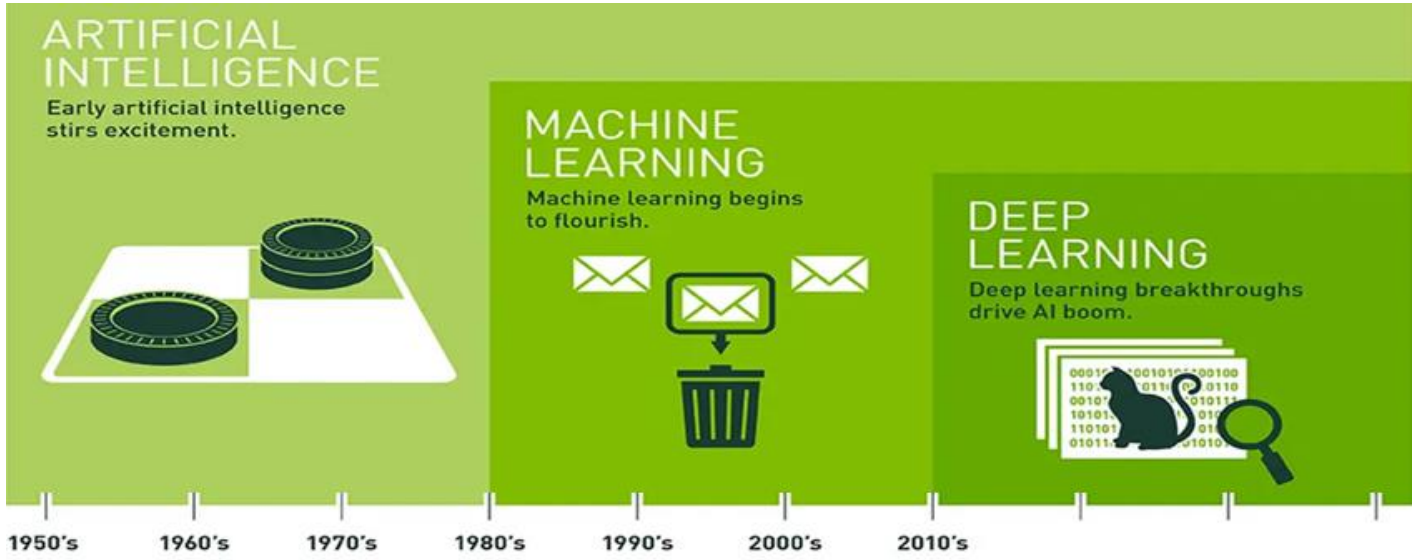
1.1 人工智能定义

人工智能是利用数字**计算机**或者数字计算机控制的**机器**模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

我国《人工智能标准化白皮书》2018

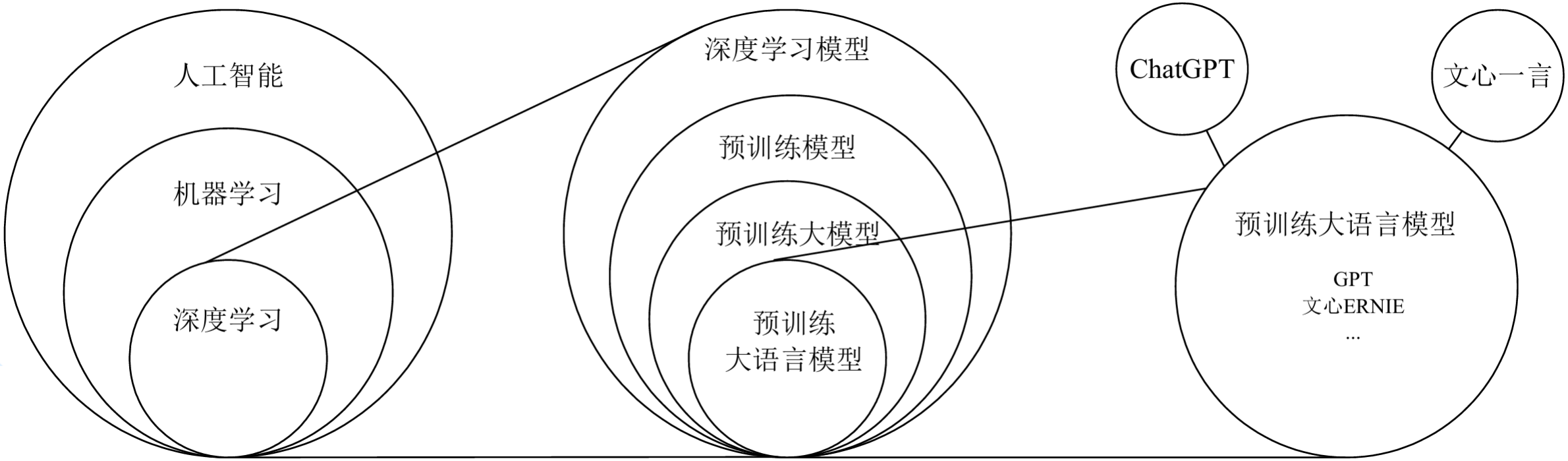
1.2 人工智能、机器学习、深度学习

- 人工智能是一个广泛的领域
  - 包含所有使机器能够模拟或扩展人类智能的技术，如机器学习、深度学习、NLP等。
- 机器学习是人工智能的一个分支
  - 使计算机能通过数据学习并做出预测或决策，而不是通过硬编码的规则。
- 深度学习是机器学习的一个子领域
  - 通过使用具有多层（深层）的神经网络来学习数据的复杂模式和表示。



来源：NVIDIA Blog

人工智能与大模型的关系

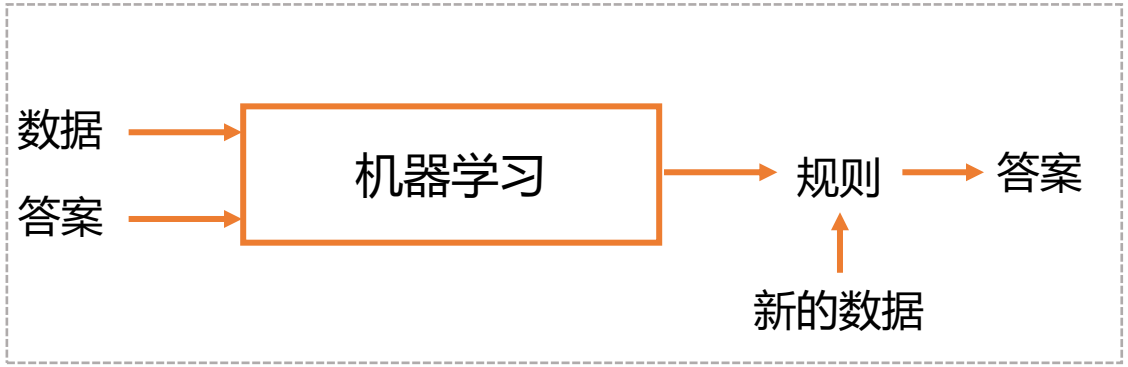


定义1 研究的角度

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 Smuel），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。

—— Tom M. Mitchell（卡内基梅隆大学教授）

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

监督学习、无监督学习、半监督学习、强化学习和自监督学习等。

机器学习			
监督学习	无监督学习	半监督学习	强化学习
线性回归	聚类方法	伪标签方法	值函数方法
逻辑回归	降维方法	一致性正则化方法	策略梯度方法
支持向量机	关联规制	协同训练	演员-评论家
决策树	生成模型	深度生成方法	蒙特卡洛树搜索
随机森林	自编码器	图方法	多智能体强化学习
深度神经网络	……	……	……
……			

- 使用**标记**好的训练数据（一些历史数据）来训练模型，使得模型能够对新的、未标记的数据进行预测或分类。
- 每个训练样本都包含一个输入**特征向量** x和一个对应的输出**标签** y，模型的目标是学习一个映射规则或函数f，这个函数f能够根据输入特征准确预测输出标签。
- **两种主要任务：**
  - 分类，判断新数据所属类别（离散的量） 如，鲑鱼鲈鱼分类。
  - 回归，预测新数据在目标变量上的值（连续的量） 如，房价预测。

分类和回归任务


$x^{(1)}$ $y^{(1)}$: 鲑鱼


$x^{(2)}$ $y^{(2)}$: 鲑鱼


$x^{(3)}$ $y^{(3)}$: 鲈鱼


$x^{(4)}$ $y^{(4)}$: 鲈鱼

分类

f

“鲑鱼”

“鲈鱼”

输出离散量

X 面积（平方米）	y 售价（万元）
$x^{(1)}$ 86	$y^{(1)}$ 380
$x^{(2)}$ 120	$y^{(2)}$ 516
$x^{(3)}$ 130	$y^{(3)}$ 548
$x^{(4)}$ 75	$y^{(4)}$ 338
⋮	⋮

回归

f

预测新发布房屋售价

输出连续值

数据集 (X, y)

X是所有样本构成的特征矩阵；

y是这些样本对应的标签数组。

$x^{(i)}$是第 i个样本的特征向量。

$y^{(i)}$是第 i个样本的标签值。

目标：学习一个映射函数f

$$\hat{y} = f(X)$$

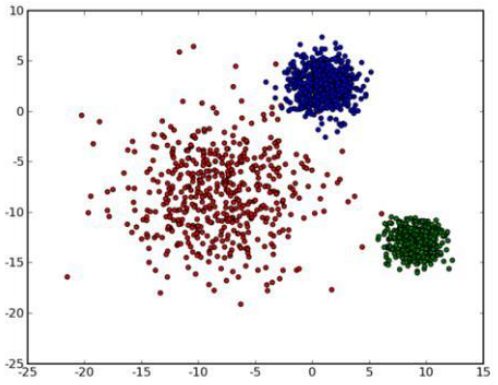
使得 $\hat{y}$ 尽可能接近 y。

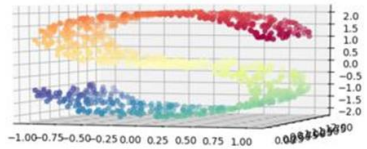
监督学习任务示例:

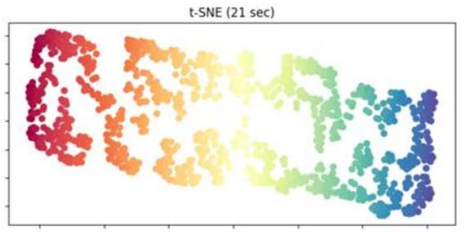
- 识别信封上手写的邮政编码
- 基于医学影像判断肿瘤是否为良性
- 分析生产线上的产品图像来对产品进行自动分类
- 预测海洋波浪的高度和周期
- 识别和分类海洋生物，如鱼类、珊瑚和浮游生物

……

- 在**没有标记**的输入数据上进行训练，即数据没有分类标签或目标输出。在无监督学习中，算法试图自己发现数据中的模式、结构或分布。
- **主要任务：**
  - **聚类**：将数据点分组到不同簇中，使得同一簇内的数据点相似，而不同簇的数据点相异。
  - **降维**：减少数据集的维度数量，同时尽可能保持数据的原有信息。
  - **异常检测**：在数据集中识别出异常或离群点，这些点与大多数其他数据点显著不同。
  - **基于生成模型的数据生成**：学习给定数据的概率分布，然后利用学到的分布生成新数据点。这些新数据点在统计上应该与原始数据相似，具有相同的分布特性。





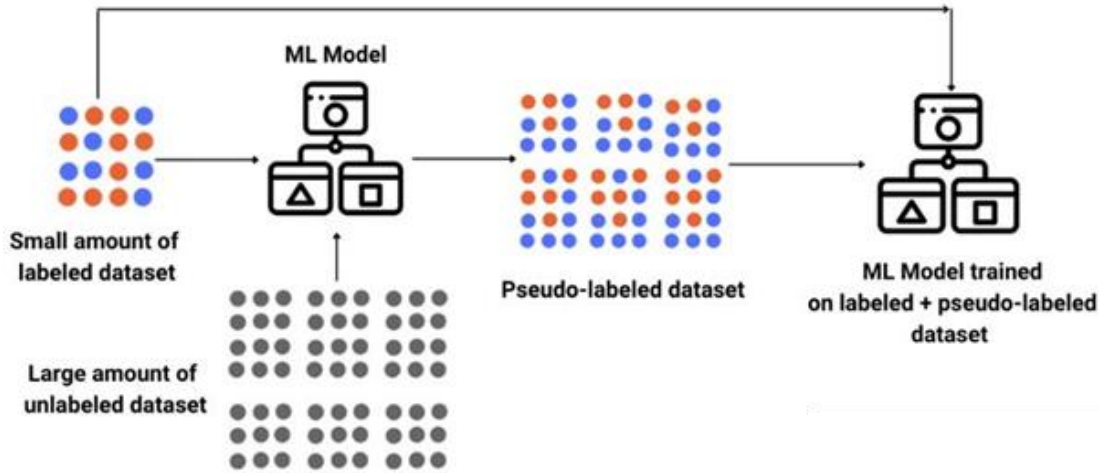


无监督学习任务示例：

- 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
- 在网络中发现社区结构，即社区发现
- 分析海洋传感器数据，识别海水中的异常情况，如污染事件。
- 数据增强，为训练机器学习模型提供更多的训练样本

- 结合**少量有标签数据**和**大量无标签数据**进行学习。特别适用于标签数据稀缺但无标签数据丰富的情况，旨在利用未标记数据来提高模型的泛化能力和预测准确性。
- **方法举例：**
  - **伪标签**：使用模型对未标记数据预测的结果作为新的标记数据，然后重新训练模型。
  - **图方法**：基于数据间的相似性图推断未标注数据的标签。
- **适用场景举例**：

图像分类
文本分类
语音识别
数据融合
……



源自：知乎@deephub

- 关注的是如何让智能体（agent）在**与环境交互**的过程中学习到最优的行为策略，以实现某种目标。在强化学习中，智能体通过执行动作来影响环境的状态，并从环境中获得奖励或惩罚，从而调整自己的行为。
- **方法举例**：

  值函数方法：估计状态-动作值函数，如 Q-learning

  多智能体强化学习 (MARL)：处理多个智能体协作或对抗任务。
- **应用场景**

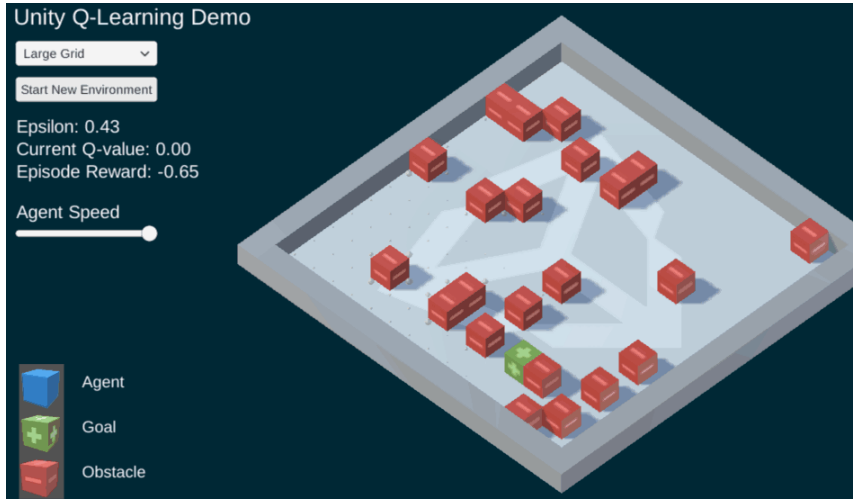
  游戏控制

  路径规划

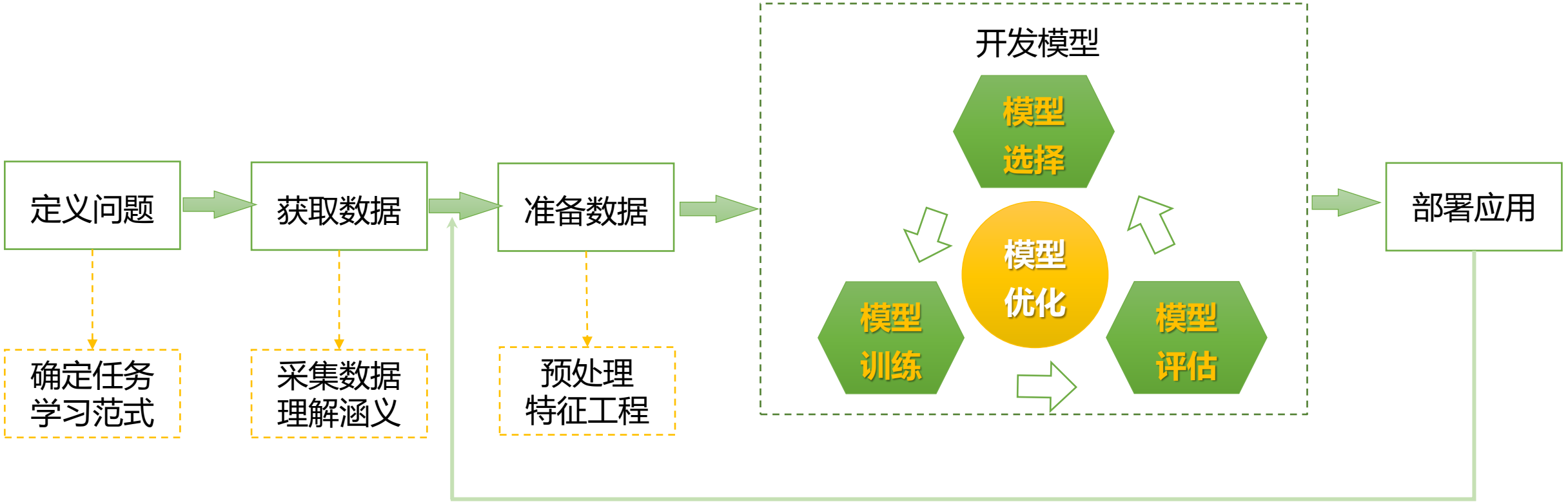
  机器人控制

  自动驾驶

  经济模拟



1.5 机器学习工作流程


定义问题
获取数据
准备数据
开发模型
模型选择
模型优化
模型训练
模型评估
部署应用
确定任务
学习范式
采集数据
理解涵义
预处理
特征工程

案例1：鸢尾花分类

某地区有三种鸢尾花（virginica、Setosa、Versicolor），如果采集一些花的数据，包括花的特征和类别信息。能否通过机器学习得到一个模型，当输入鸢尾花特征时自动识别它的类别？

1. 定义问题

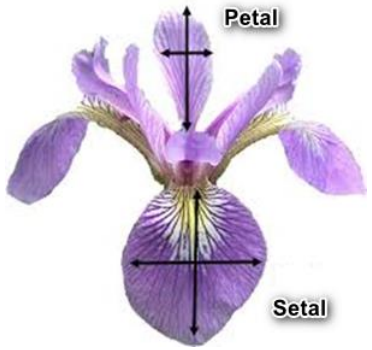
可否用机器学习解决？属于哪类问题？

2. 获取数据

下载iris数据集：https://cloud.tencent.com/developer/article/1869024

iris.csv文件以','作为列间分隔符。

- 150个样本分三类：setosa(山鸢尾),versicolor(杂色鸢尾),virginica(维吉尼亚鸢尾)
- 每个样本由4个特征 sepal length, sepal width, petal length, petal width来描述。



特征提取 $\begin{pmatrix} 5.1 \\ 3.5 \\ 1.4 \\ 0.2 \end{pmatrix}$

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3	1.4	0.1	setosa
4.3	3	1.1	0.1	setosa
5.8	4	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa

3. 准备数据

使数据可以被机器学习模型处理。

数据预处理:

- 数据清洗、
- 类型变换、
- 编码转换、
- 数据合并

……

特征工程:

- 特征选择、
- 特征变换、
- 降维处理

……

4. 选择模型

选择合适的机器学习算法，如决策树、SVM、神经网络等。

本例选择一个简单的模型，如 K近邻法。

K 近邻法

判断一个样本 x 属于哪个类别?

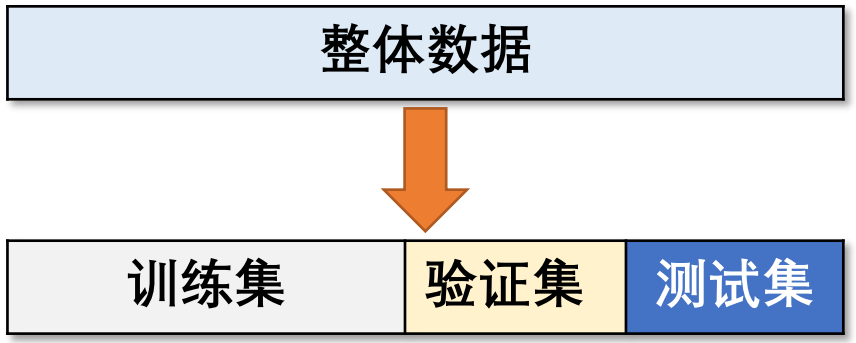
找出（训练集中）距离 x 最近的k个邻居，

看它的k个邻居大多数属于哪个类别，就推测x也属于这个类别。

划分数据集

- 所有机器学习算法都是基于对现有数据集的学习，而要求所学得的模型能够适配于未来的新数据，即所谓的**模型泛化能力**。
- 为了解一个模型在多大程度上适用于新数据，唯一方法是在新数据上**测试**。
- 在模型训练前，把整体数据划分为**训练集**和**测试集**，以保证测试所用数据是模型没有见过的新数据。一般按8:2或7:3来划分。一旦划分无需变动。
- 一般训练多个候选模型，从中选泛化性能最强的。为了评估不同模型性能，一般在训练集中分出一部分做**验证集**。

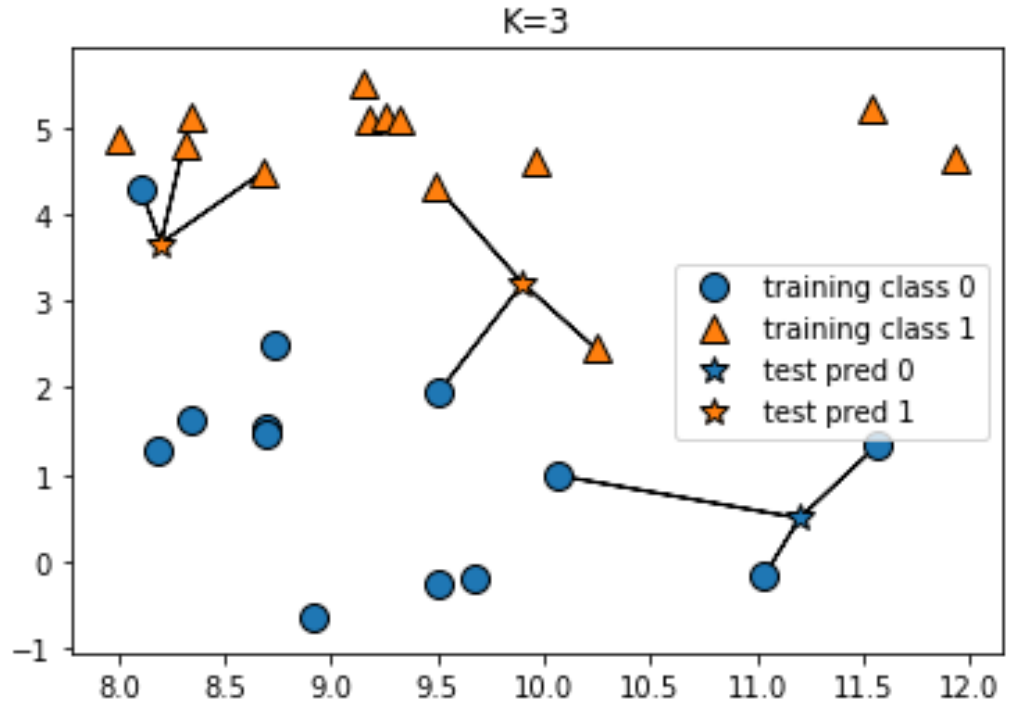
- **训练集**：用于对模型进行训练。
- **验证集**：从多个候选模型中选最佳。
- **测试集**：用来测试模型的效果。



5. 训练模型

使用训练数据集对模型进行训练，即通过学习数据集的特征和标签来调整模型的参数。

K近邻分类算法，基于数据点与邻居之间的距离进行分类，没有显式建立模型并训练。



6. 评估模型

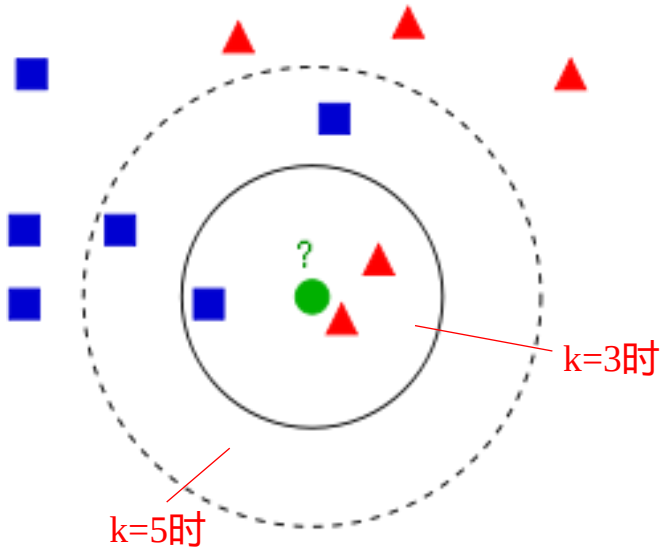
- 模型在新数据上的错误率称为模型的**泛化误差**，在训练数据集上的错误率称为模型的**训练误差**。
- **评估一个模型**，需要通过**在验证集/测试集上测试模型**，获得泛化误差的估计值。该值说明模型在从未见过数据上的性能。
  - 如果训练误差很低，但泛化误差很高，则说明模型对训练数据**过拟合**。
  - 若训练误差就很高，说明模型对训练数据拟合不足，称为**欠拟合**。

模型评估指标

- 实际中，根据任务类型采用相应的评估指标，来评估模型的泛化能力。
  - **分类指标**：分类准确率、精度、召回率、F1-score，ROC 等
  - **回归指标**：决定系数（R^2）、均方误差（MSE）、绝对值误差等。
  - 如果训练指标很高（如F1-score很高），但验证/测试指标很低，则说明模型对训练数据**过拟合**。
  - 若训练指标就很低（如F1-score很低），说明模型对训练数据拟合不足，称为**欠拟合**。

7. 模型优化

- **根据评估结果对模型进行调整**，包括**超参数调优**和特征选择。
- K 近邻法中，k 不能通过训练数据学习出来，一般需要人为指定
- 这类模型需要在训练前人为指定的参数被称为超参数。训练过程中尝试不同的超参数，选择最优超参数，即所谓的超参数调优。



绿圆点 *x*是一个新数据点。

当k=3时，判定它的类别是红三角；

而当k=5时，判定它的类别属于蓝方块。

8. 模型部署

- 将训练好的模型部署到实际应用中，进行预测或决策。
- 常部署在以下几种环境中：
  - **服务器**：通常部署在云服务器或企业内部服务器上，可以通过网络提供API接口，供客户端调用模型的预测服务。
  - **边缘设备**：部署在边缘设备上，如路由器、交换机、物联网设备等，这些设备位于数据产生的源头，可以减少数据传输的延迟。
  - **移动设备**：部署在智能手机、平板电脑等移动设备上。
  - **嵌入式设备**：对于一些对资源限制非常严格的场景，如自动驾驶车辆、无人机、机器人等，机器学习模型可以部署在嵌入式系统中。
  - **集群计算环境**：对于需要大量计算资源的机器学习模型，可能会部署在分布式计算集群中，如Hadoop或Spark集群。

机器学习工作流程小结

(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、编码转换等**数据预处理**；特征选择、特征变换、降维等**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 。
  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)