

机器学习2024

第1节

涉及知识点：

人工智能、数据科学、机器学习简介、机器学习应用、机器学习基本思想、模型选择、模型泛化性、判别模型与生成模型



机器学习简介

张伟楠 - [上海交通大学](#)

课程安排

参数化有监督学习

1. 机器学习概述
2. 线性模型
3. 双线性模型
4. 神经网络

非参数化有监督学习

5. 支持向量机
6. 决策树
7. 集成学习与森林模型

无监督学习部分

8. 概率图模型
9. 无监督学习

学习理论部分

10. 学习理论与模型选择

前沿话题部分

11. 迁移、多任务、元学习
12. System 1&2 机器意识



人工智能与数据科学

张伟楠 - [上海交通大学](#)

人工智能

定义

- 智能是实现目标的计算能力部分
- 人工智能 (AI) 是由机器展示的智能
- 人工智能学科探讨对机器进行设计的方法论使其可以去完成基于智能的任务



人工智能方法

基于规则的方法

- 直接编程实现
- 借鉴人类启发式学习的思想

基于数据的方法

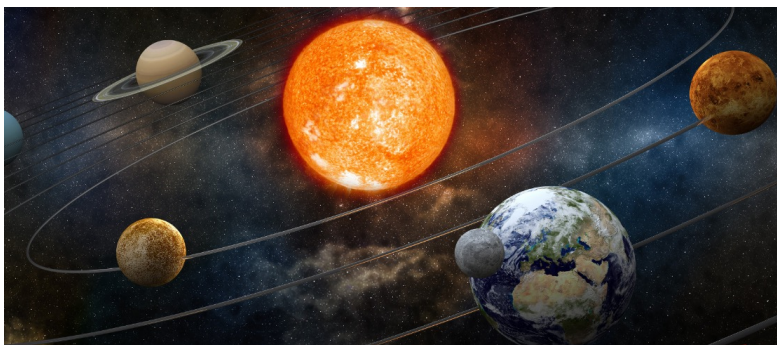
- 专家系统
 - 专家或者数据科学家基于数据创造用于预测或决策的规则
- 机器学习
 - 直接基于数据进行预测或决策
 - 数据科学

什么是数据科学？

- 数据科学其实比自然科学更本质、宽泛

物理学

- 目标：探究世界的基本原理



- 解决方法：从观测结果中构建世界模型

$$F = G \frac{m_1 m_2}{r^2}$$

数据科学

- 目标：发现数据的基本原理



- 解决方法：从观测结果中构建数据模型

$$p(x) = \frac{e^{f(x)}}{\sum_{x'} e^{f(x')}}$$

数据科学

数学上

- 找到联合数据分布 $p(x)$
- 找到条件分布 $p(x_2|x_1)$

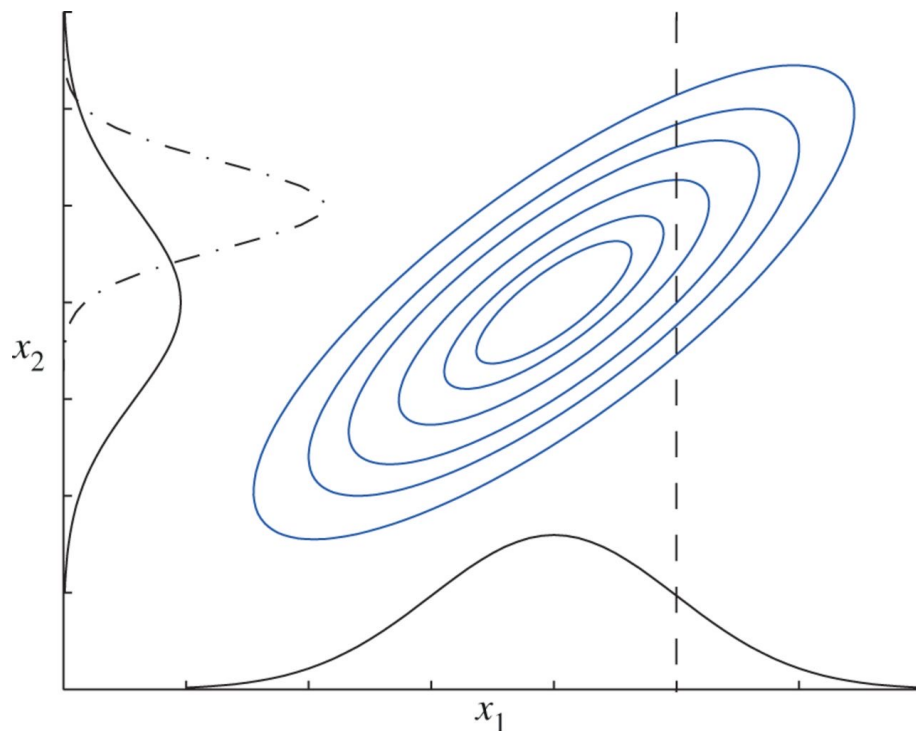
高斯分布

- 多元分布

$$p(x) = \frac{e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)}}{\sqrt{|2\pi \Sigma|}}$$

- 一元分布

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



一个用户行为建模的简单例子

Interest	Gender	Age	BBC Sports	PubMed	Bloomberg Business	Spotify
Finance	Male	29	Yes	No	Yes	No
Sports	Male	21	Yes	No	No	Yes
Medicine	Female	32	No	Yes	No	No
Music	Female	25	No	No	No	Yes
Medicine	Male	40	Yes	Yes	Yes	No

□ 联合数据分布

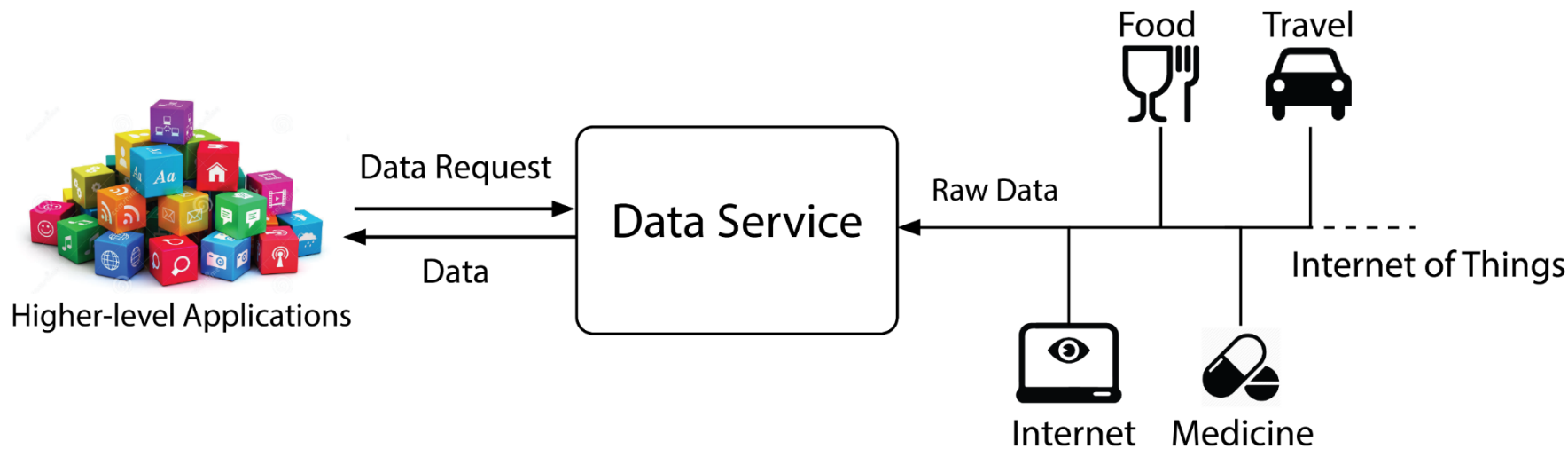
$p(\text{Interest}=\text{Finance}, \text{Gender}=\text{Male}, \text{Age}=29, \text{Browsing}=\text{BBC Sports}, \text{Bloomberg Business})$

□ 条件数据分布

$p(\text{Interest}=\text{Finance} \mid \text{Browsing}=\text{BBC Sports}, \text{Bloomberg Business})$

$p(\text{Gender}=\text{Male} \mid \text{Browsing}=\text{BBC Sports}, \text{Bloomberg Business})$

数据处理技术



数据本身没有价值，数据产生的服务才有价值



机器学习简介

张伟楠 - [上海交通大学](#)

目录

Contents

01 机器学习定义

02 机器学习历史

01

机器学习定义

什么是学习

学习

“学习是系统通过经验提升性能的过程。”

--- Herbert Simon

卡内基·梅隆大学

图灵奖(1975)

人工智能，人类认知心理学

诺贝尔经济学奖(1978)

经济组织内的决策过程



什么是机器学习

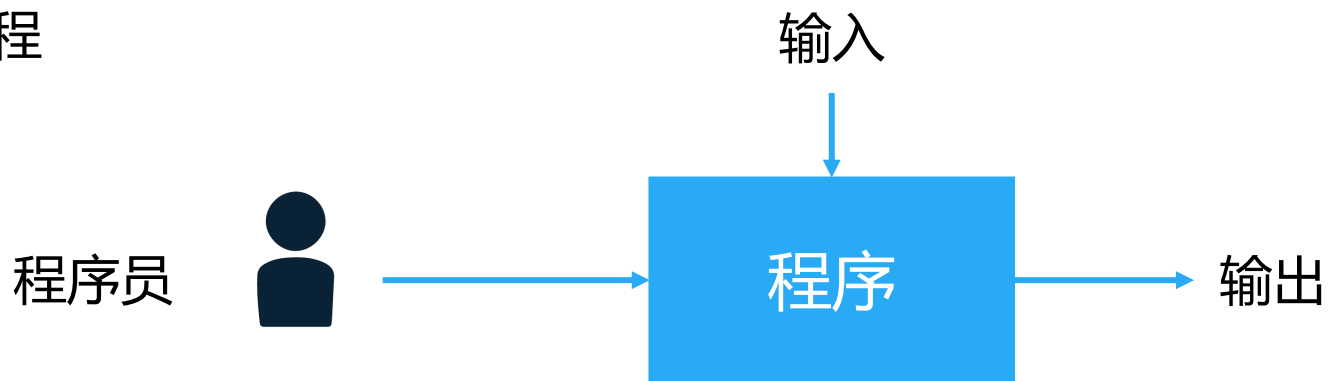
由Tom Mitchell 给出的更加数学化的定义

- 机器学习是一门研究学习算法的学科，这些算法能够：
 - 在某些任务 T 上
 - 通过经验 E
 - 提升性能 P
 - 非显式编程

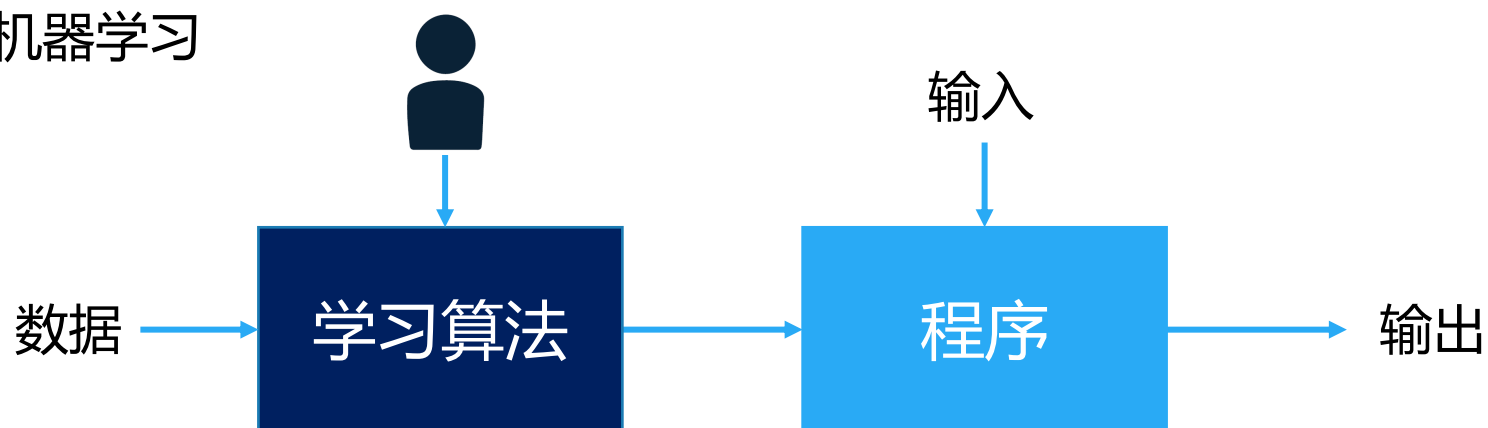
- 一个学习任务可以由三元组 $\langle T, P, E \rangle$ 明确定义

编程 vs. 机器学习

传统编程

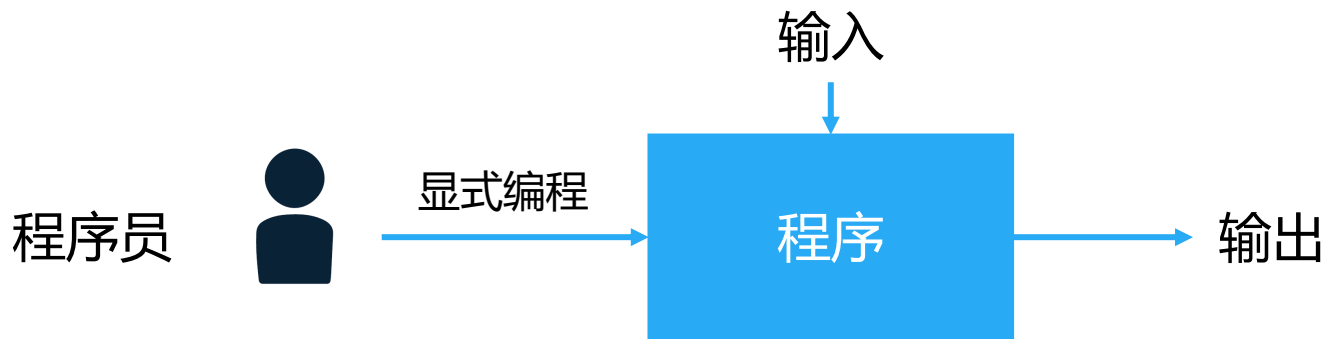


机器学习

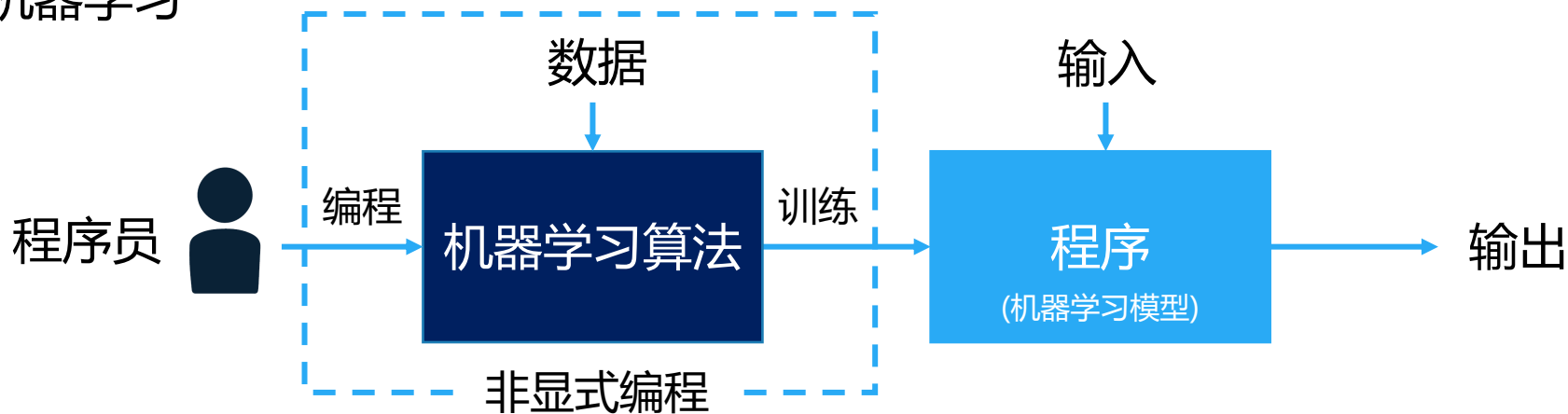


编程 vs. 机器学习

传统编程



机器学习



机器学习在什么情况下具有优势

应用情形：

- 模型基于大量数据
 - 例子：Google 网络搜索，Facebook 新闻提要
- 输出必须是个性化的
 - 例子：新闻/物品/广告推荐
- 人类不能解释专业知识
 - 例子：语音/人脸识别，围棋游戏
- 人类的专业知识不存在
 - 例子：在火星上导航

两种机器学习类型

□ 预测

- 根据数据预测所需的输出（监督学习）
- 生成数据实例（无监督学习）

□ 决策

- 在动态环境中采取行动（强化学习）
 - 转变到新的状态
 - 获得即时奖励
 - 随着时间的推移最大化累积奖励

02

机器学习历史

机器学习的历史

□ 1950年代：

- Samuel 的跳棋程序
- 创建机器学习术语

□ 1960年代：

- 神经网络，感知机
- 模式识别
- Minsky和Papert证明感知机的局限性

□ 1970年代：

- 符号概念归纳
- Winston的结构学习系统
- 专家系统和知识获取瓶颈
- Quinlan的ID3算法
- 使用 AM 的数学发现



Arthur Samuel在1959年创造了"机器学习"这个词

□ 1980年代：

- 高级决策树和规则学习
- 基于解释的学习 (EBL)
- 学习、规划、解决问题
- 三间小屋问题
- 类比
- 认知架构
- 神经网络复苏 (反向传播)
- Valiant的PAC学习理论
- 注重实验方法

机器学习的历史

□ 1990年代

- 数据挖掘
- 自适应软件代理和网络应用程序
- 文本学习
- 强化学习 (RL)
- 归纳逻辑编程 (ILP)
- 组合: 装袋、提升和堆叠
- 贝叶斯网络学习
- 支持向量机
- 核方法

□ 2000年代

- 图模型
- 变分推理
- 统计关系学习
- 迁移学习
- 序列标记
- 集体分类和结构化输出
- 计算机系统应用
 - 编译器
 - 调试
 - 图形
 - 安全性 (入侵、病毒和蠕虫检测)
- 电子邮件管理
- 学习的个性化助手
- 机器人和视觉的学习

机器学习的历史

□ 2010年代

- 深度学习
- 从大数据中学习
- 结合知识图谱的机器学习
- 使用 GPU 或 HPC 学习
- 多任务 + 终身学习
- 深度强化学习, AlphaGo
- 视觉、语音、文本、网络、行为等的庞大应用
- ...

□ 2020年代

- 结合逻辑推理的深度学习
- 超大规模预训练模型 (GPT-3等)
- 无人驾驶
- AI生物制药技术
- 生成式模型的爆发 (ChatGPT、DALL-E、Diffusion、Sora)
- ...



机器学习应用

张伟楠 - [上海交通大学](#)

目录

Contents

01 机器学习分类

02 面向预测问题的机器学习

03 面向决策问题的机器学习

01

机器学习分类

机器学习的三大类

□ 面向预测任务的机器学习

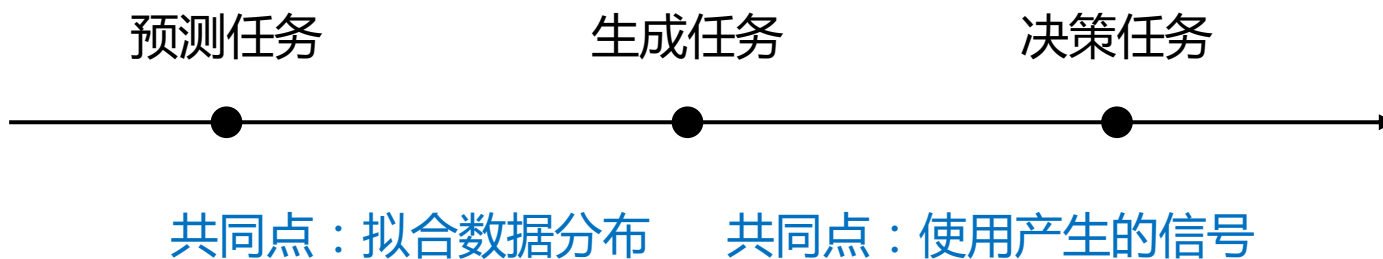
- 给予特征数据，预测目标概率分布（监督学习）

□ 面向生成任务的机器学习

- 生成数据样例（无监督学习）

□ 面向决策任务的机器学习

- 在动态系统中采取序列行动决策



机器学习应用1：网页搜索



- 生成查询建议
- 相关网页排序
- 新一代搜索？



机器学习应用2：推荐和广告

17:45 zmi无线

全部 6.1狂欢 直播 便宜好货 全球 生

猜你喜欢 疯狂抢购 网红推荐 低价抢购 进口好货 享受

00:11

新款潮牌小包男女斜挎包

满98减10·店铺券

¥168 7人付款

HOT 态极自适应科技跑鞋

匹克旗舰店

STABILO

5代原装触控笔

德国Stabilo on思笔乐ipad平板电脑触控笔套苹果安

¥15 38人付款

Microsoft 微软数码专营店

微软surface pen触控笔Pro 5 4笔新款原装4096压

天猫 6.1狂欢

¥68 29人付款

Google iPhone 6s case

Web Shopping News Images Videos More Search tools

Weinan

About 16,900,000 results (0.33 seconds)

iPhone 6s Cases - case-mate.com

Ad www.case-mate.com/iPhone-6s-Cases

4.6 ★★★★★ rating for case-mate.com

Shop The iPhone 6s Case Collection. Free Standard Shipping!
Refined Protection · Slim & Tough · Genuinely Crafted · Premium Designs

iPhone 6s

Ad www.apple.com/

The only thing that's changed is everything. Learn more.
A9 chip · Two sizes · Now in rose gold

Pre-order 9.12 - iPhone Upgrade Program - 3D Touch - Cameras

In the news



Speck's iPhone 6s CandyShell + MightyShell cases bring best-of-breed protection to Apple's latest iPhones

9 to 5 Mac - 1 day ago

With the iPhone 6s and iPhone 6s Plus debuting next week, it's important to start thinking ...

Moshi's iPhone 6s and 6s Plus cases offer premium protection

iMore - 23 hours ago

Top 5 Best Leather iPhone 6s Cases

Heavy.com - 12 hours ago

More news for iPhone 6s case

iPhone 6s Cases & Covers from OtterBox

Shop for iPhone 6s case on Google

Sponsored

<p>Case-mate - Karat Case Fo...</p> <p>\$49.99</p> <p>Best Buy</p> <p>★★★★★ (163)</p>	<p>Moshi - Iglaze Armour Case...</p> <p>\$39.99</p> <p>Best Buy</p> <p>★★★★★ (161)</p>	<p>Logitech - Protection...</p> <p>\$21.99</p> <p>Best Buy</p> <p>★★★★★ (90)</p>	<p>Moshi - Overture Wall...</p> <p>\$49.99</p> <p>Best Buy</p> <p>★★★★★ (18)</p>
<p>Case-mate - Brilliance Cas...</p> <p>\$44.99</p> <p>Best Buy</p> <p>★★★★★ (294)</p>	<p>Case-mate - Wallet Folio C...</p> <p>\$54.99</p> <p>Best Buy</p> <p>★★★★★ (173)</p>	<p>Marc by Marc Jacobs Metall...</p> <p>\$38.00</p> <p>shopbop</p>	<p>Case-mate - Karat Hard Sh...</p> <p>\$49.99</p> <p>Best Buy</p> <p>★★★★★ (34)</p>

机器学习应用3：人脸识别 - 寻找走失儿童



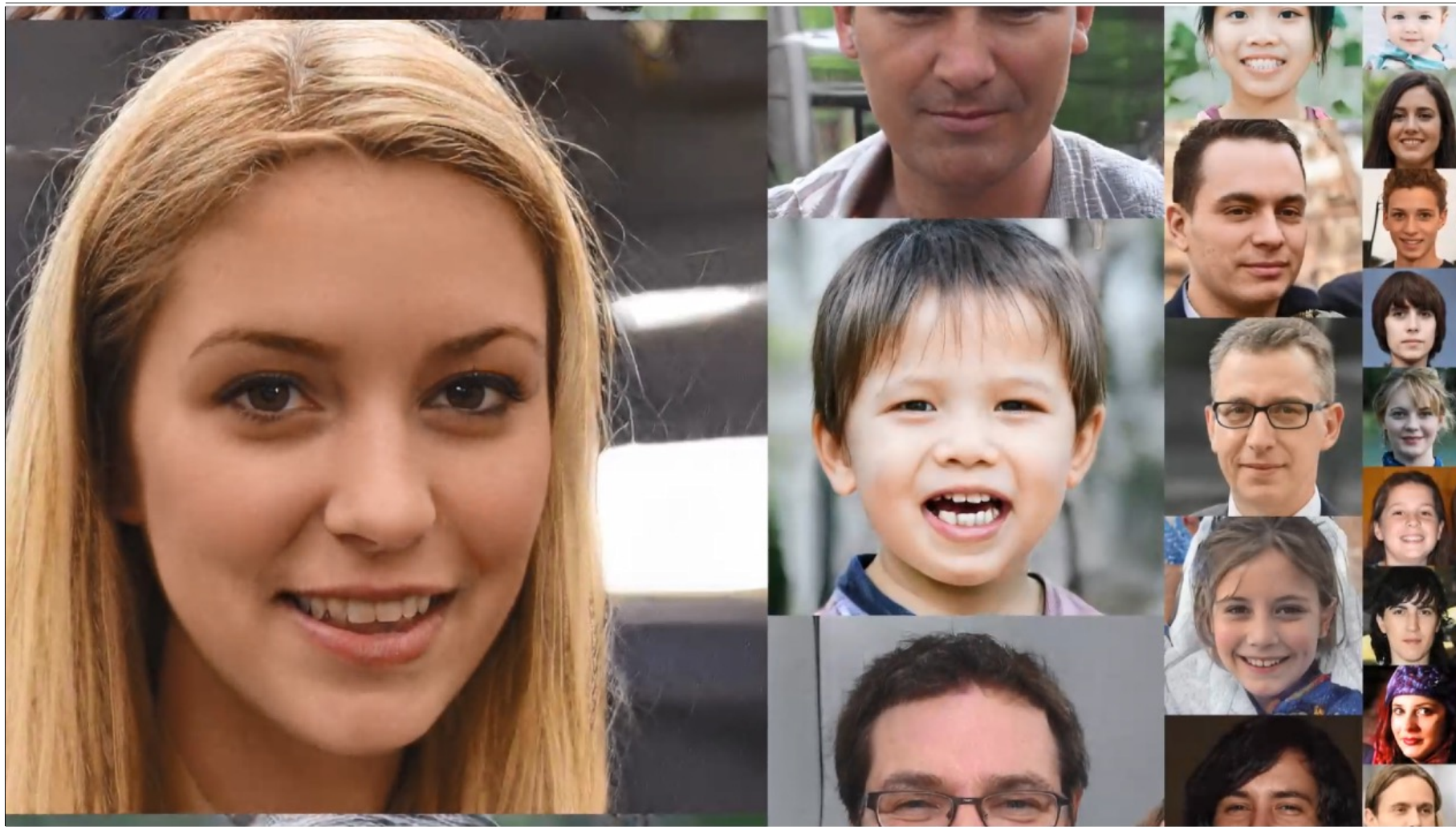
腾讯QQ Alert：寻找走失儿童

机器学习应用4：医疗图像分析 – 新冠影像

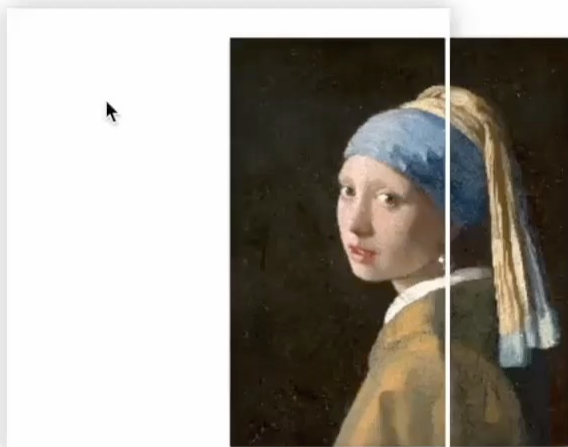


依图胸部CT新型冠状病毒肺炎智能评价系统

机器学习应用6：人脸生成



机器学习应用7：艺术图像生成（DALL-E 2）




机器学习应用8：音乐生成

Dreack[★]

机器学习应用9：对话生成（ChatGPT）

ST

Can you write out an Adobe After Effects expression to make a shape layer wiggle when a null object is within 50 pixels of the shape's anchor point. 



机器学习应用10：艺术图像生成 (DALL-E 3)

ChatGPT●



机器学习应用11：视频生成（ChatGPT）



机器学习应用12：机器人控制

Cook Shrimp
(autonomous)



3x speed

Stanford
University

机器学习应用13：工厂节能



机器学习应用14：自动驾驶汽车

In this experiment, we are going to demonstrate a reinforcement learning algorithm learning to drive a car.

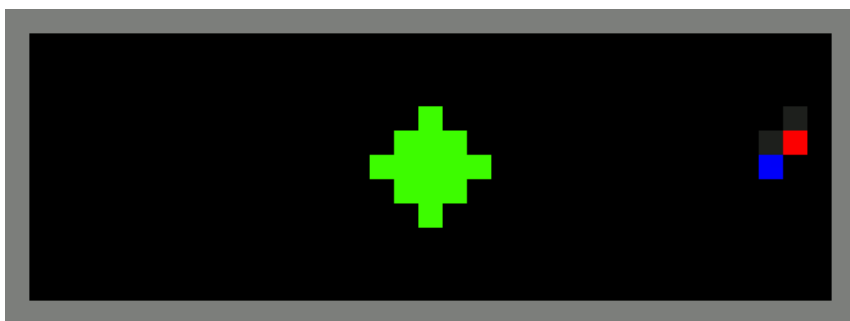
机器学习应用15：游戏智能

- 执行给定屏幕像素的操作
 - <https://gym.openai.com/envs#atari>

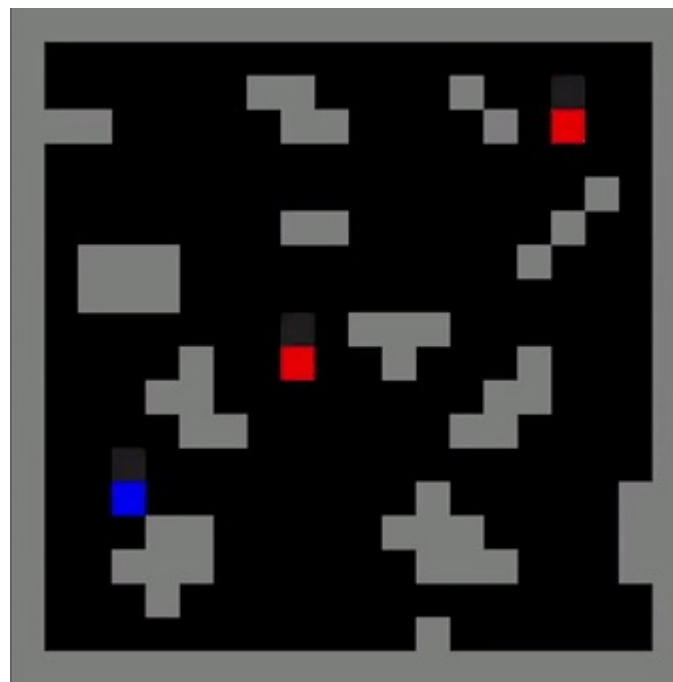


机器学习应用16：多智能体对抗与协作

机器学习最新热点：多智能体行动



双智能体食物争夺



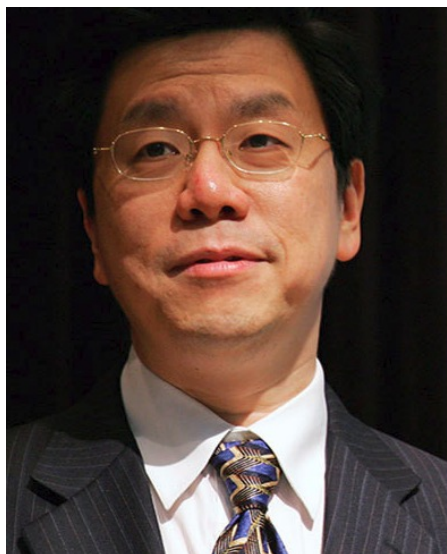
双智能体合作围剿猎物

关于人工智能的重要性

Bill Gates says A.I. like ChatGPT is 'every bit as important as the PC, as the internet'

BY STEVE MOLLMAN

February 3, 2023 at 8:14 AM GMT+8



我认为，人工智能是人类有史以来面临的最大机遇。

李开复

创新工场董事长兼首席执行官



机器学习基本思想

张伟楠 - [上海交通大学](#)

机器学习类别

□ 监督学习

- 给定数据和标签，预测所需的输出

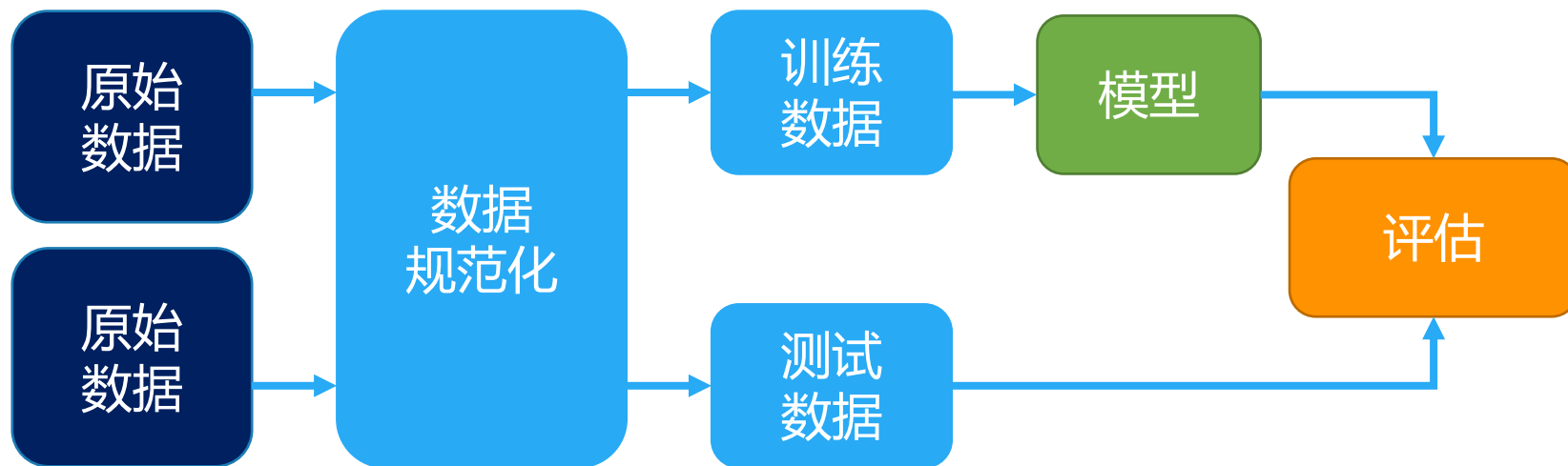
□ 无监督学习

- 分析和利用隐式数据模式/结构

□ 强化学习

- 学习在动态环境中动作执行的决策，并获得尽可能多的奖励值

机器学习过程



- 基本假设：在训练和测试数据中存在相同的模式（[pattern](#)）

监督学习

定义

- 给定带标签的训练数据集

$$D = \{(x_i, y_i)\}_{i=1,2,\dots,N}$$

其中 x_i 为特征数据， y_i 为其对应的标签，让机器学习一个从特征数据映射到标签的函数映射

$$y_i \simeq f_{\theta}(x_i)$$

- 函数集 $\{f_{\theta}(\cdot)\}$ 被称为假设空间
- 学习的过程即为参数 θ 的更新

如何学习？

- 更新参数以使预测结果接近真实的标签
 - 学习目标是什么？
 - 如何更新参数？

学习目标

- 使预测结果接近真实的标签

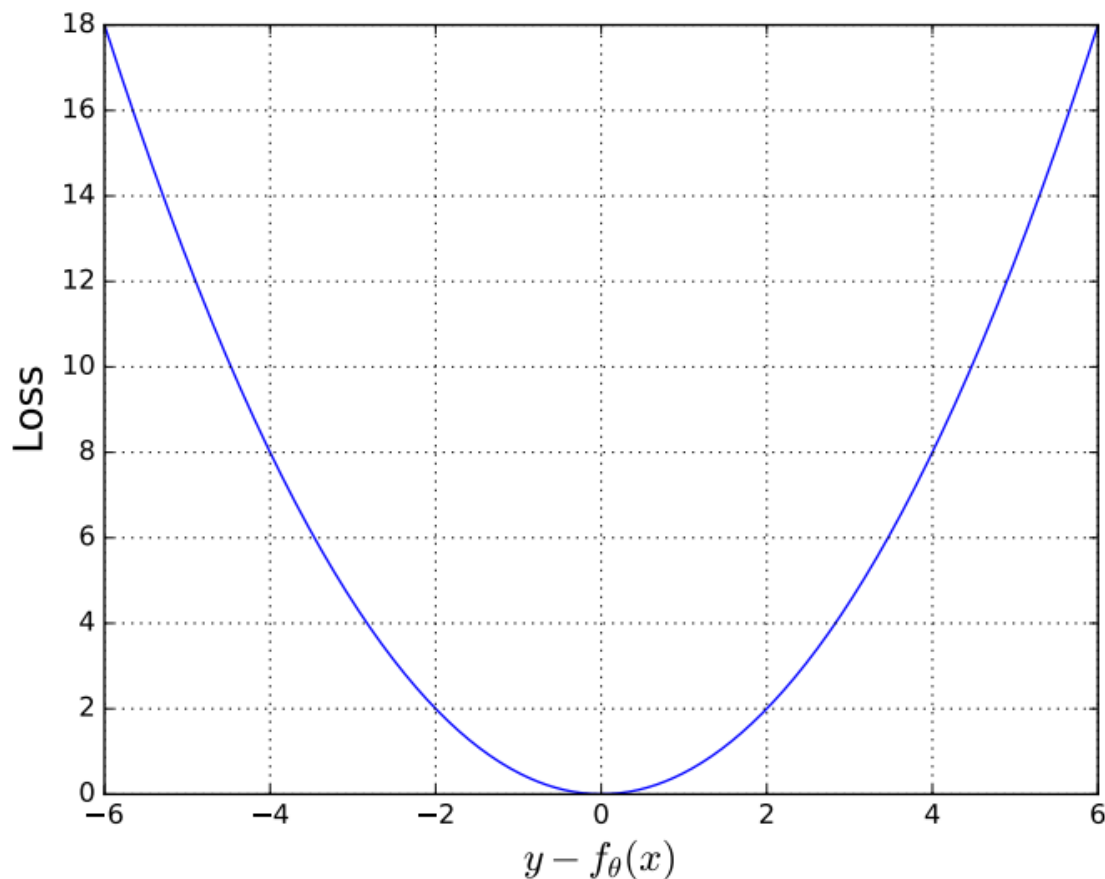
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f_{\theta}(x_i))$$

- 损失函数 $\mathcal{L}(y_i, f_{\theta}(x_i))$ 用来衡量标签和预测结果之间的误差
- 损失函数的定义取决于数据和任务
- 最常见的损失函数：平方误差 (squared loss)

$$\mathcal{L}(y_i, f_{\theta}(x_i)) = \frac{1}{2} (y_i - f_{\theta}(x_i))^2$$

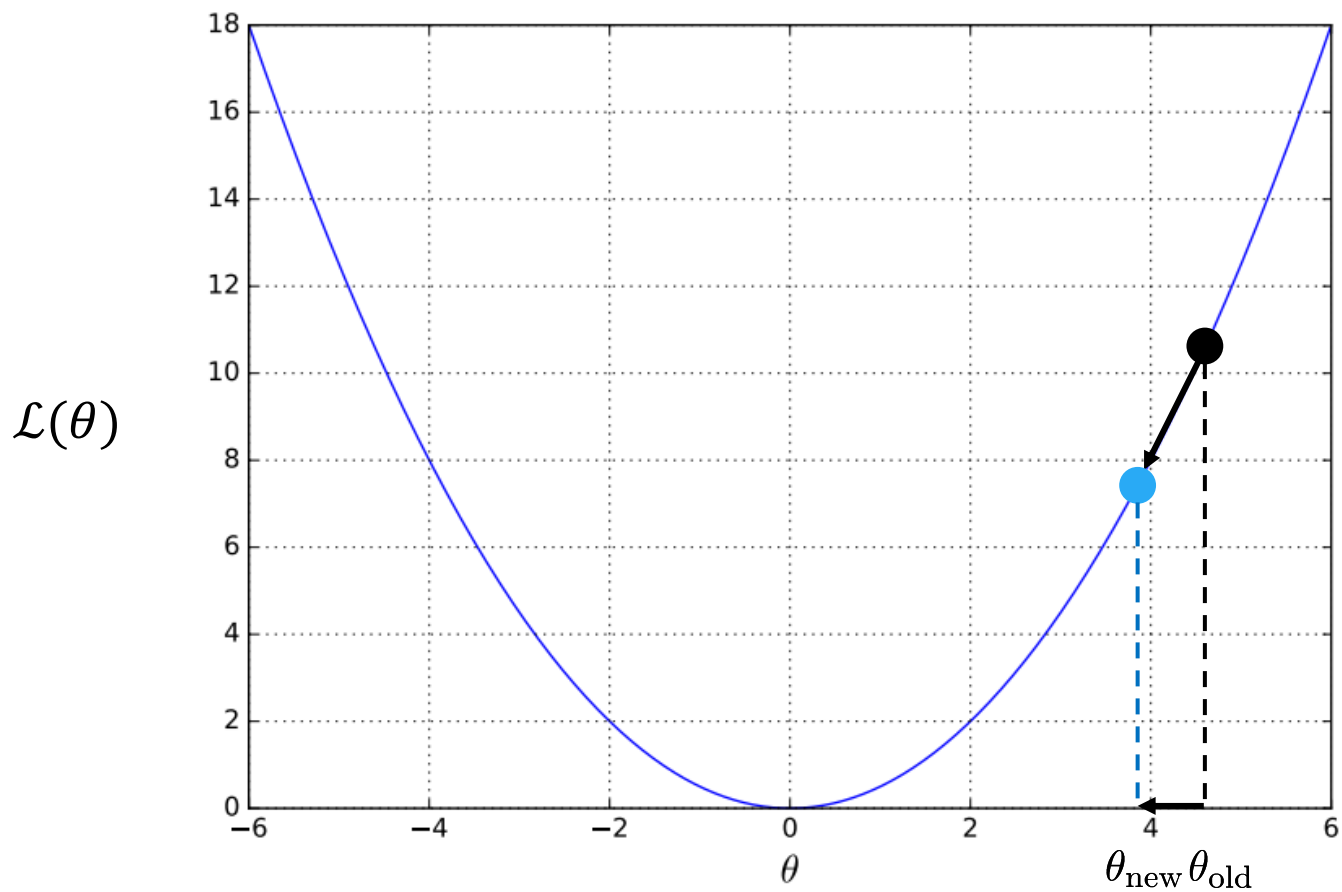
平方误差

$$\mathcal{L}(y_i, f_{\theta}(x_i)) = \frac{1}{2}(y_i - f_{\theta}(x_i))^2$$



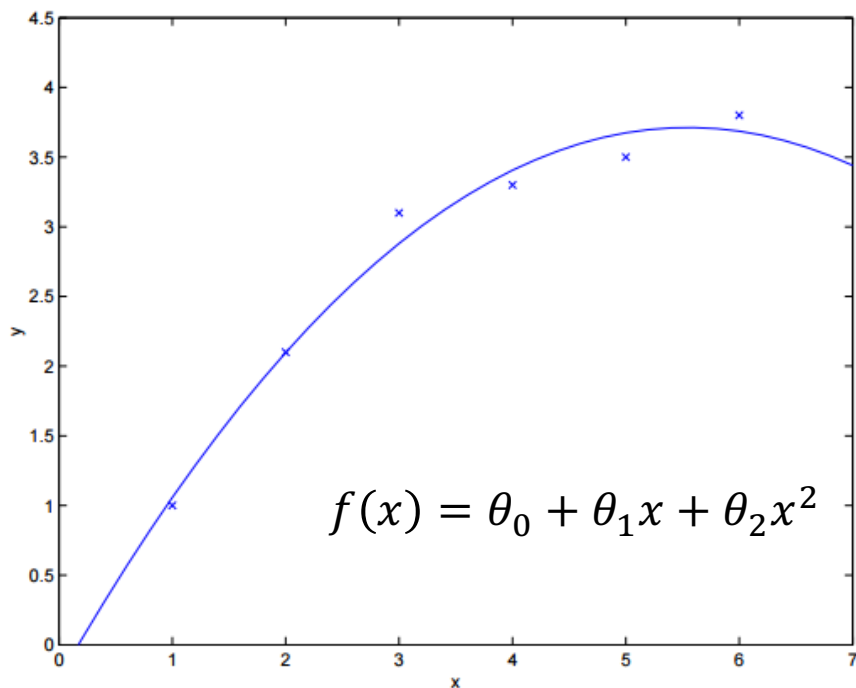
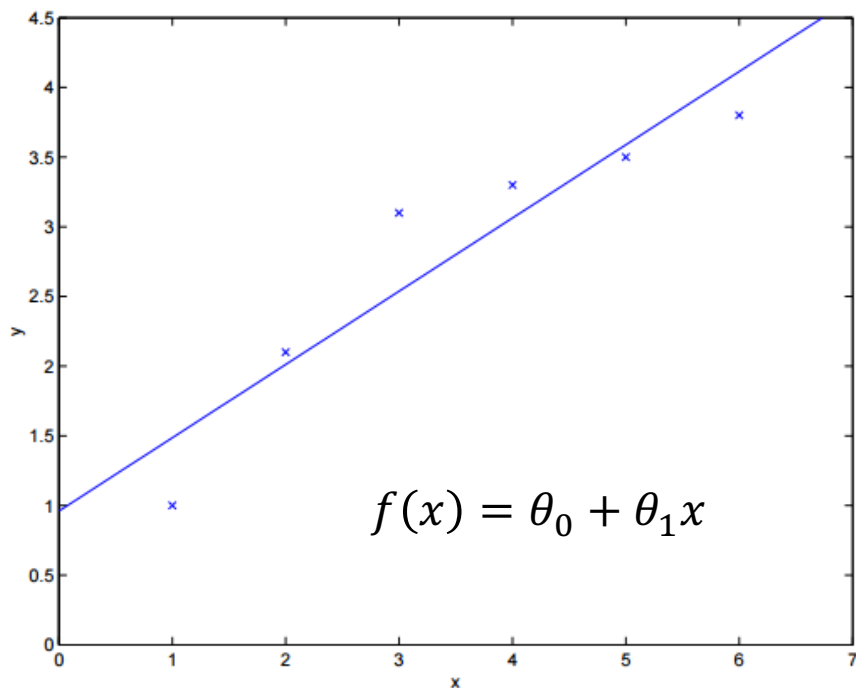
- 距离越远，得到的惩罚更多
- 容忍小距离（误差）
 - 观察噪声等
 - 泛化性

梯度学习方法



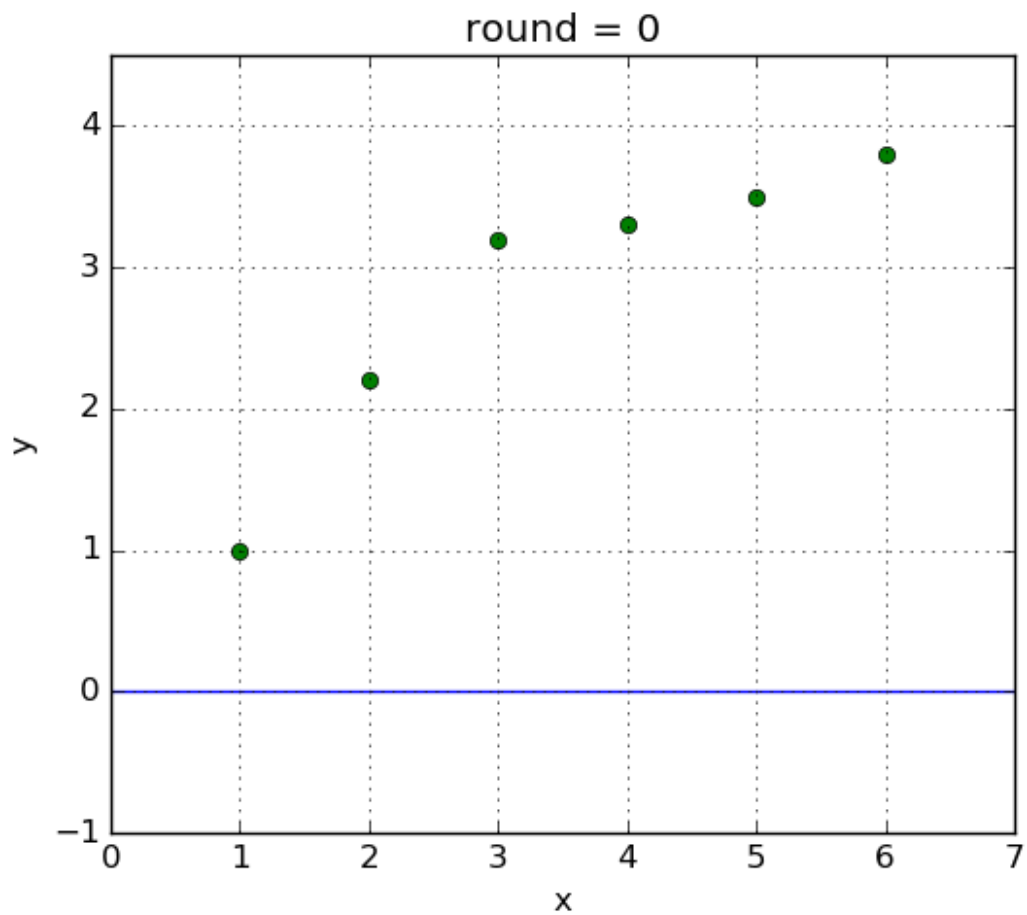
$$\theta_{new} \leftarrow \theta_{old} - \eta \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$$

一个简单例子



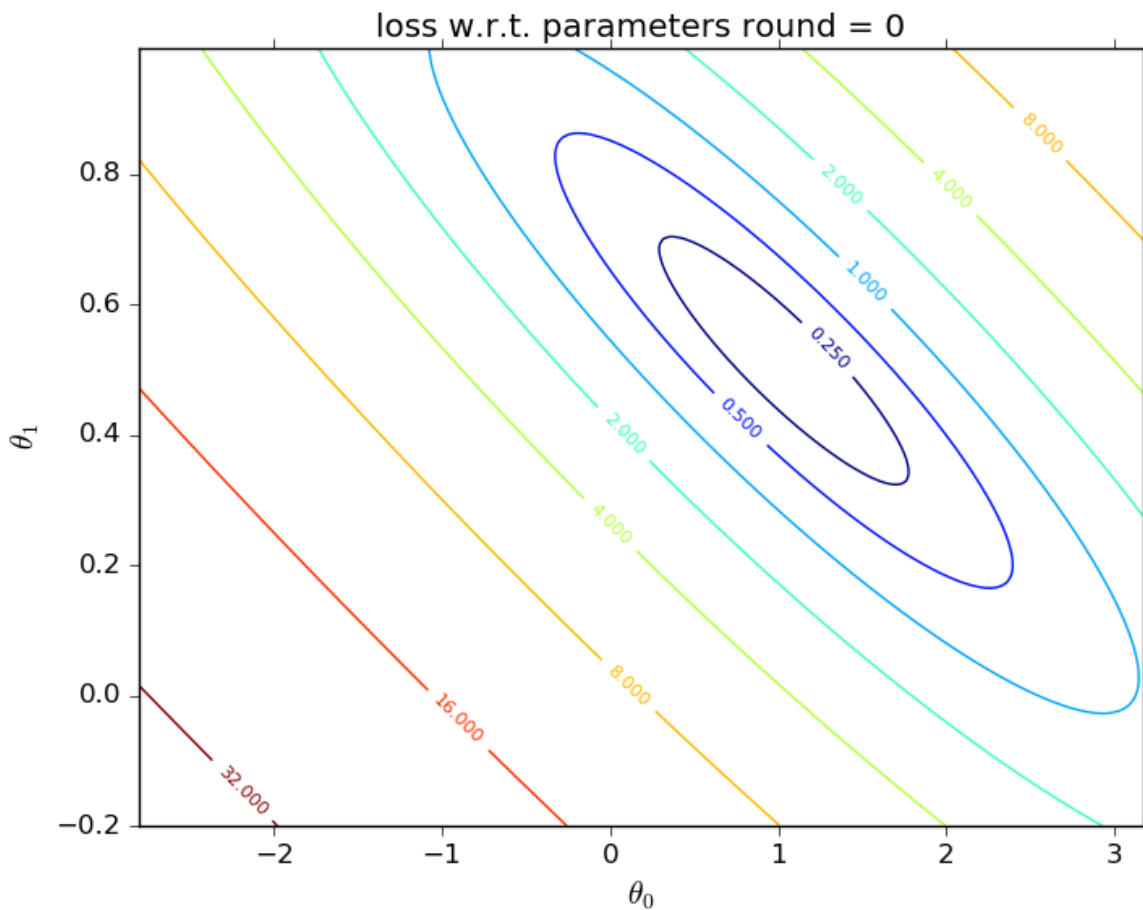
- 观察数据 $\{(x_i, y_i)\}_{i=1,2,\dots,N}$ ，我们可以使用不同的模型(假设空间)来学习
 - 模型选择(线性或二次)
 - 参数学习

利用线性模型学习-曲线拟合



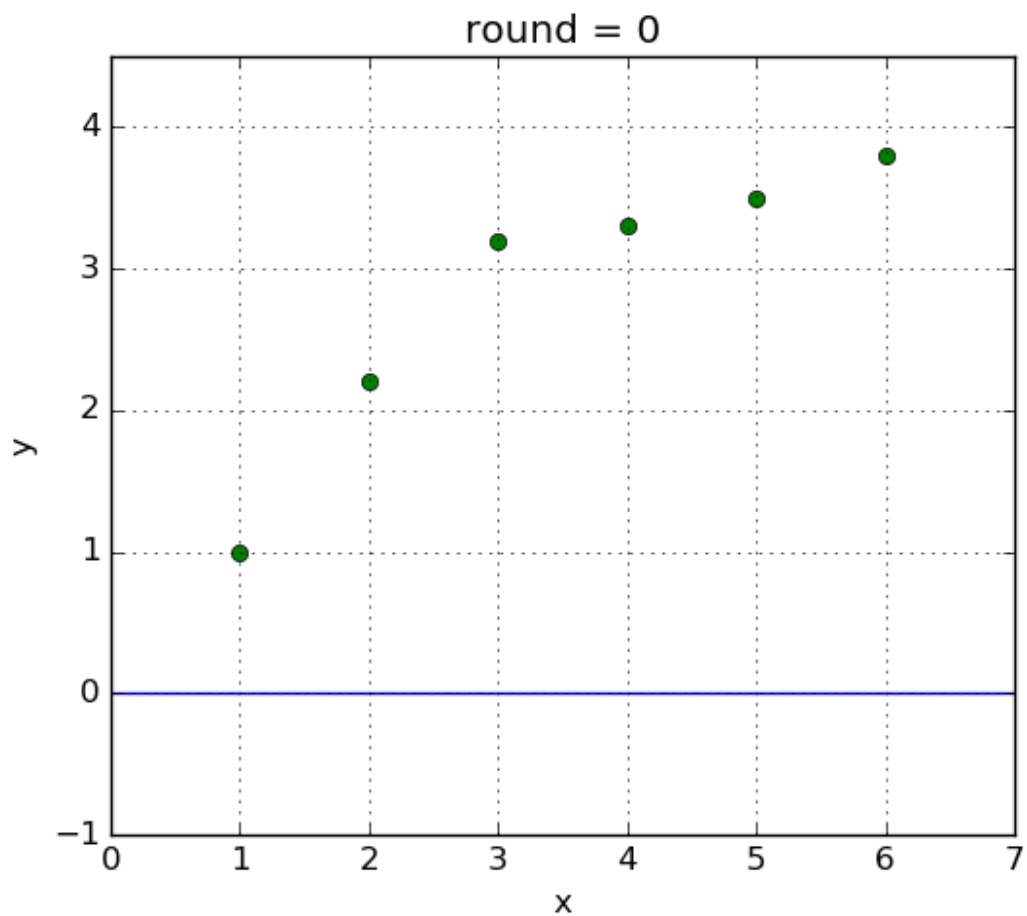
$$f(x) = \theta_0 + \theta_1 x$$

利用线性模型学习-权重分布



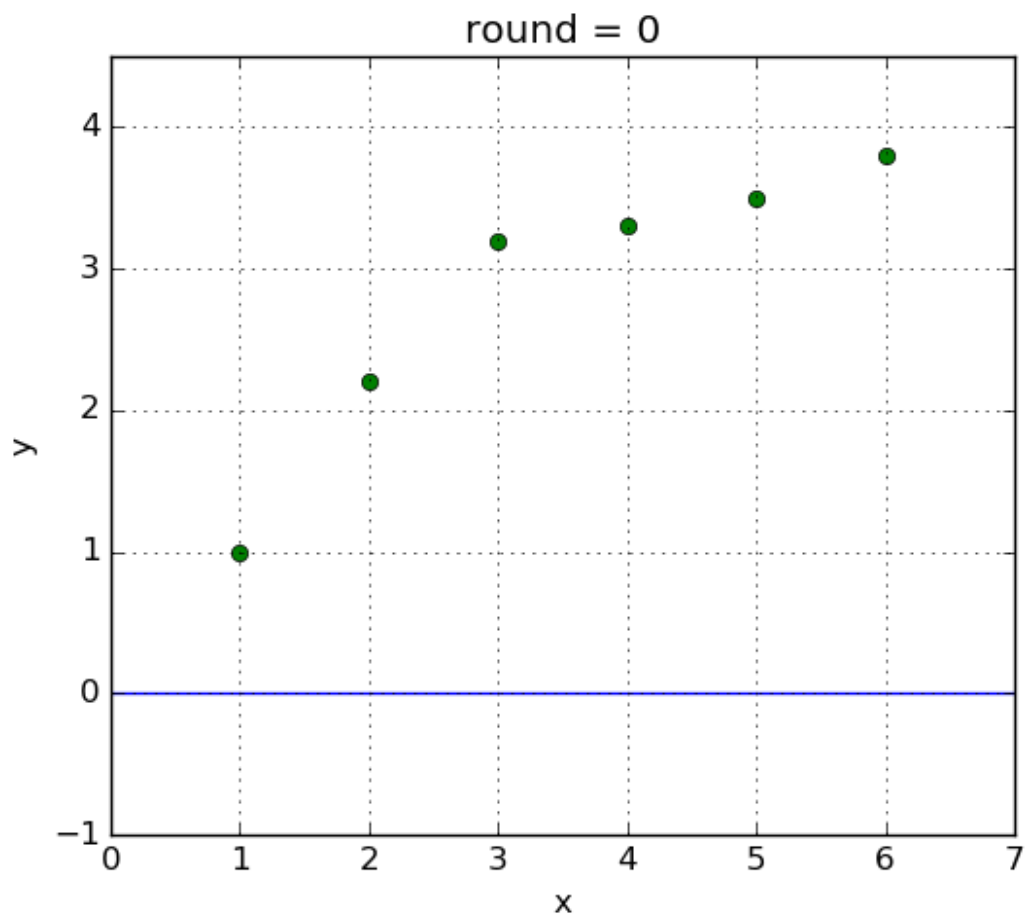
$$f(x) = \theta_0 + \theta_1 x$$

利用二次模型学习



$$f(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

利用三次模型学习



$$f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$



模型选择

张伟楠 - [上海交通大学](#)

目录

Contents

01 欠拟合与过拟合

02 正则化

03 奥卡姆剃刀原则

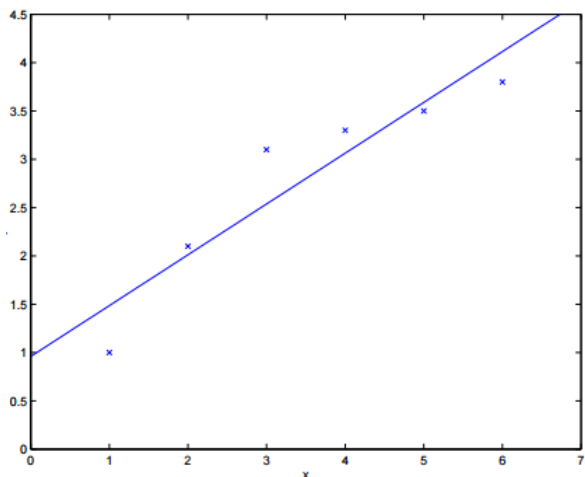
04 交叉验证



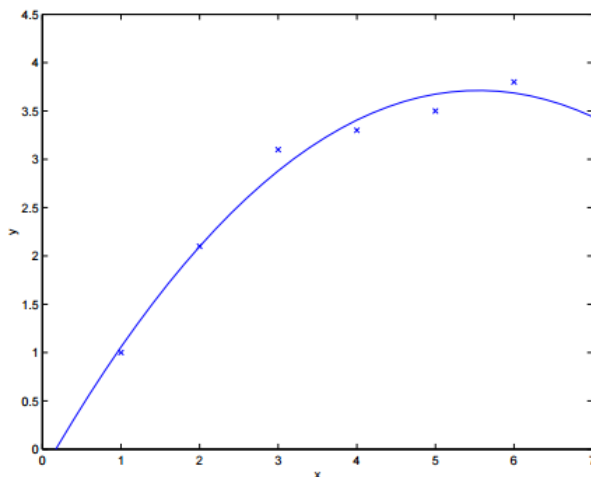
01
欠拟合
与过拟合

欠拟合与过拟合

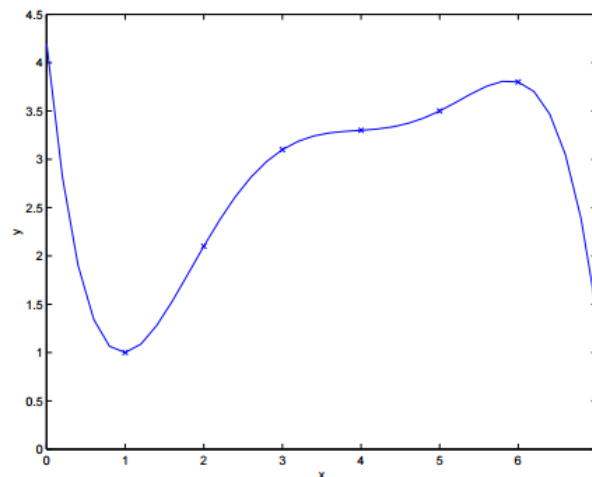
□ 下面哪个模型是最好的？



线性模型：欠拟合



二次模型：合适

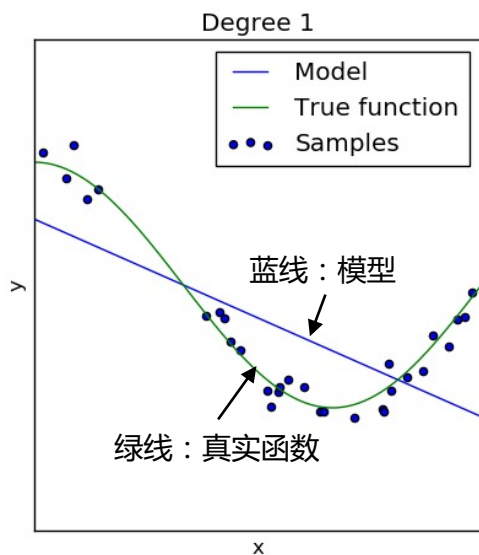


五阶模型：过拟合

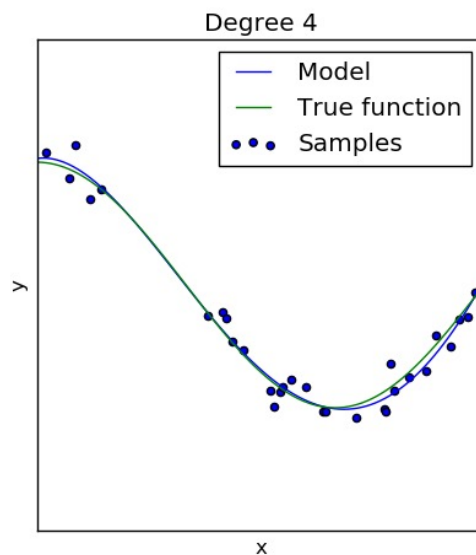
- 当统计模型或机器学习算法无法捕捉数据的基本变化趋势时,就会出现**欠拟合**。
- 当统计模型把随机误差和噪声也考虑进去而不仅仅是考虑数据的基础关联时,就会出现**过拟合**。

欠拟合与过拟合

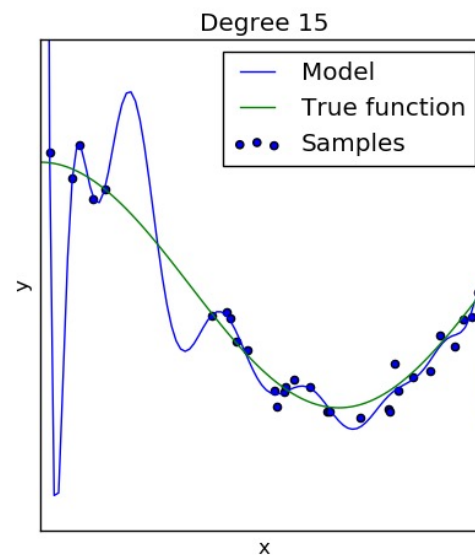
□ 下面哪个模型是最好的？



线性模型：欠拟合



四阶模型：合适



十五阶模型：过拟合

- 当统计模型或机器学习算法无法捕捉数据的基础变化趋势时,就会出现**欠拟合**。
- 当统计模型把随机误差和噪声也考虑进去而不仅仅是考虑数据的基础关联时,就会出现**过拟合**。



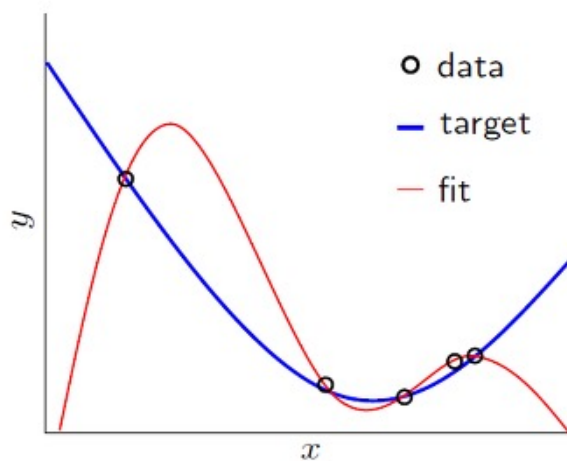
02

正则化

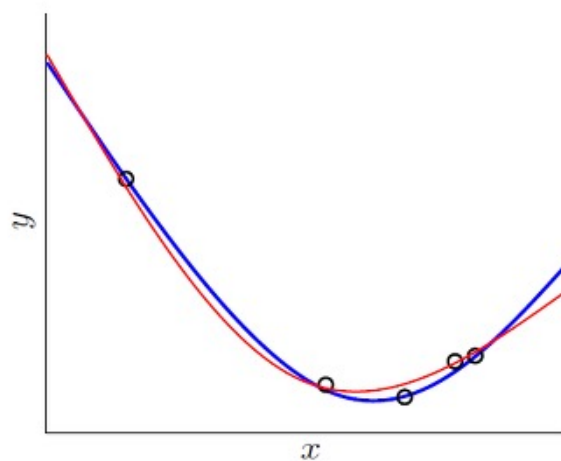
正则化

- 添加参数的惩罚项，防止模型对数据过拟合

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f_{\theta}(x_i)) + \lambda \Omega(\theta)$$



(a) without regularization



(b) with regularization

经典正则化方法

□ L2正则化 (岭回归Ridge)

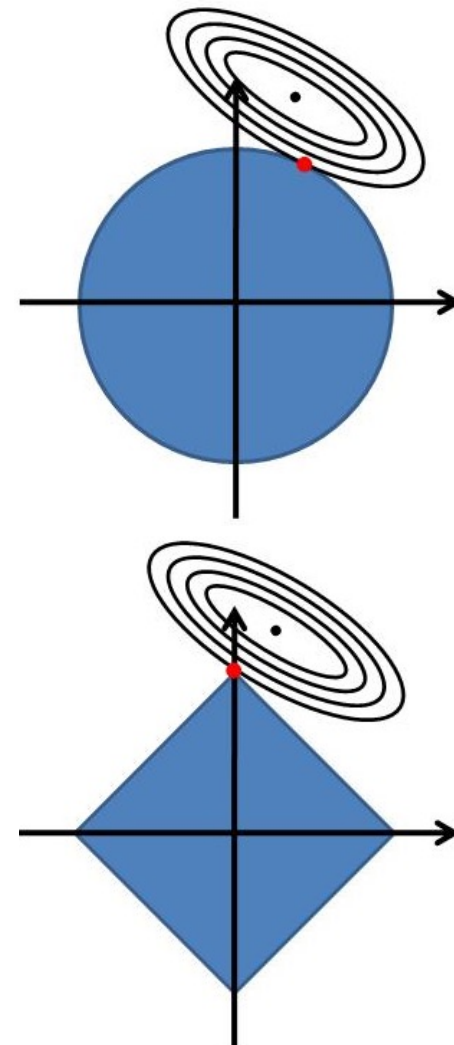
$$\Omega(\theta) = \|\theta\|_2^2 = \sum_{m=1}^M \theta_m^2$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f_{\theta}(x_i)) + \lambda \|\theta\|_2^2$$

□ L1正则化 (拉索回归LASSO)

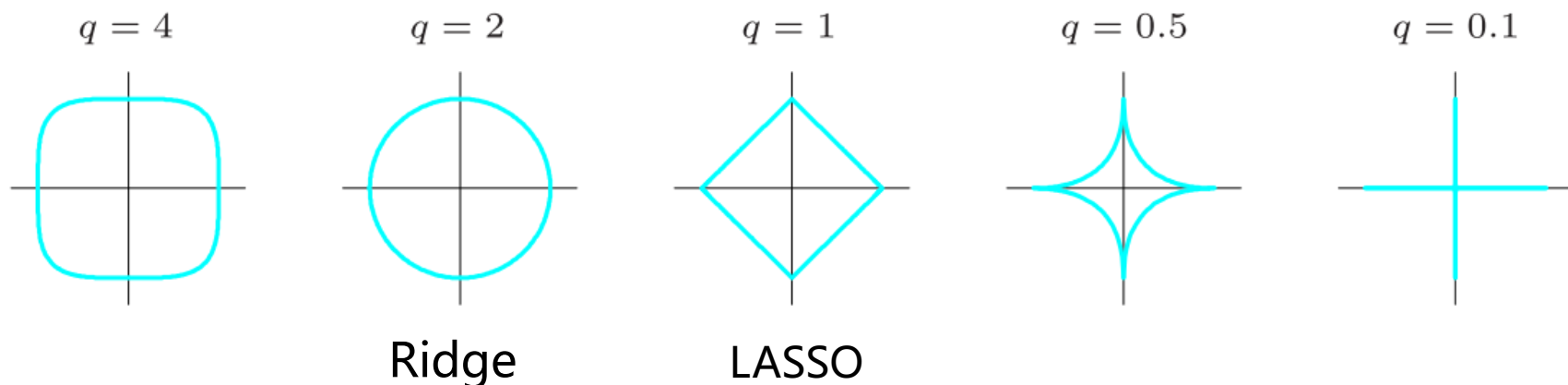
$$\Omega(\theta) = \|\theta\|_1 = \sum_{m=1}^M |\theta_m|$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f_{\theta}(x_i)) + \lambda \|\theta\|_1$$



更常规的正则化方法

□ 常值 $\sum_j |\theta_j|^q$ 的数值分布图



- 当 $q \leq 1$ 的时候，模型进行稀疏性学习
- 很少会用 $q > 2$ 来进行正则化
- 99% 的情形下都取 $q = 1$ 或 2

03

奥卡姆剃刀
原则

奥卡姆剃刀原则

□ 有多个假设模型时，我们应该选择假设条件最少的建模方法。

□ 函数集 $\{f_\theta(\cdot)\}$ 被称作假设空间

$$\min_{\theta} \left[\frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f_\theta(x_i)) \right] + \lambda \Omega(\theta)$$

原始损失 基于假设的罚值

模型选择

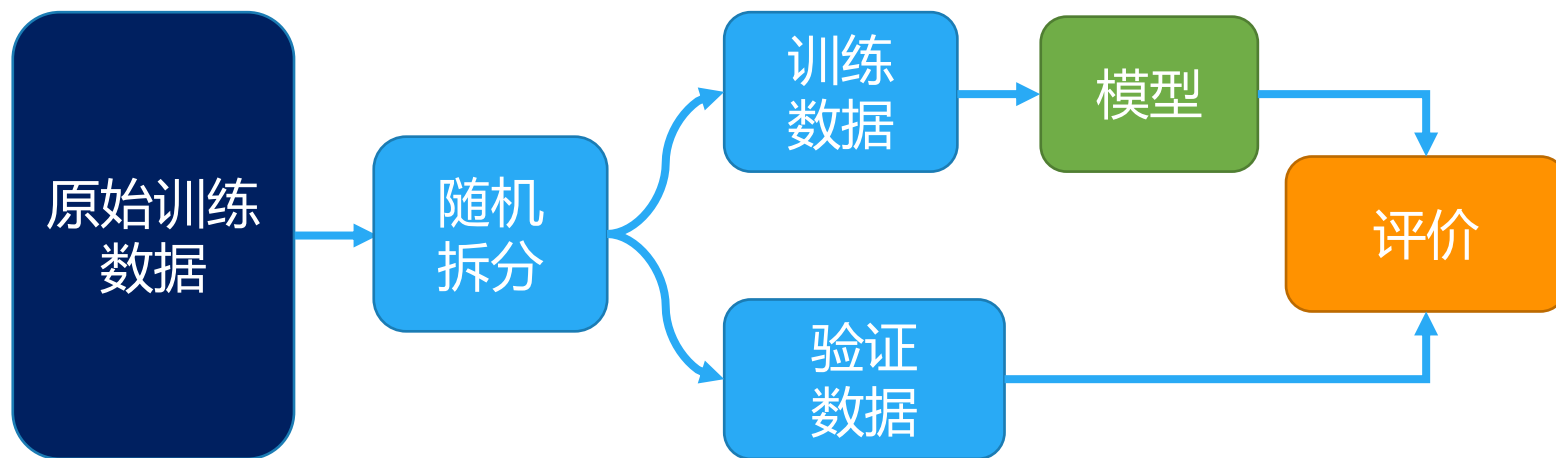
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f_{\theta}(x_i)) + \lambda \|\theta\|_2^2$$

- 一个机器学习的解决方案的模型包含参数 θ 和超参数 λ
- 超参数
 - 定义模型的更高层次的概念，如复杂性或学习能力。
 - 在标准模型训练过程中**无法直接从数据中学习**，需要预先定义。
 - 可以通过不同的参数设置、训练不同的模型以及选择最好的测试结果来进行超参数选择
- 模型选择（或超参数优化）关注如何选择最佳超参数

04

交叉验证

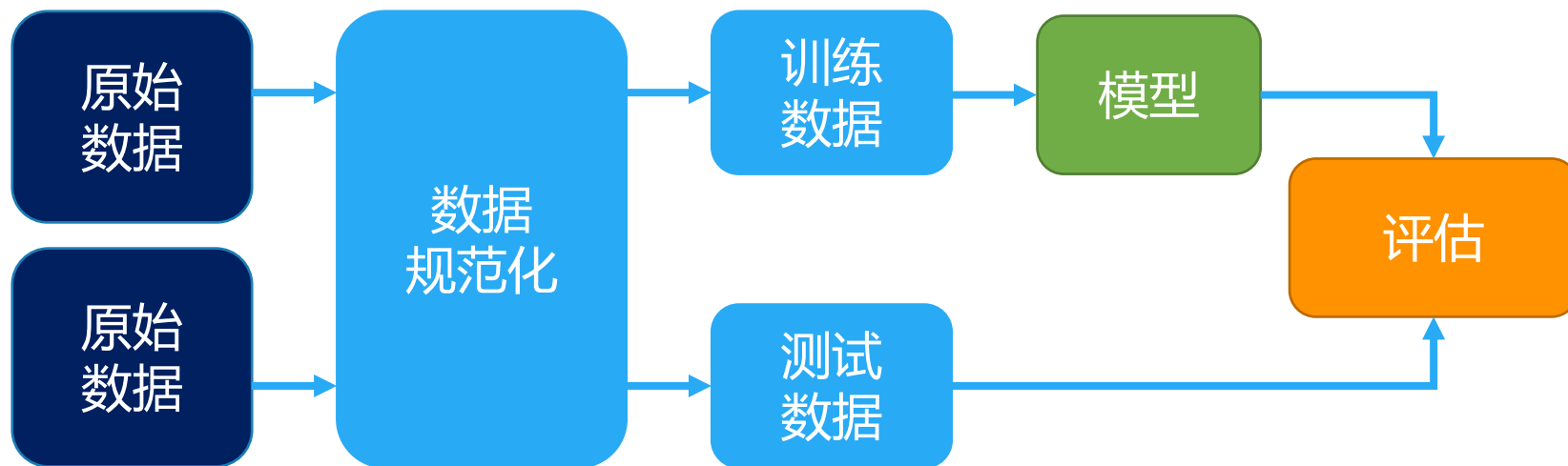
交叉验证



K-折交叉验证

1. 设置超参数
2. 将原始训练数据随机拆分为K份
3. 重复K次:
 - 若当前为第 i 次重复 ($i=1, \dots, K$)，选择第 i 份数据作为验证数据集，其余 $K-1$ 份作为训练数据集
 - 对训练数据进行建模,并在验证数据上对其进行评估,从而获得评估分数
4. 对K个评估分数取平均作为模型性能

机器学习过程



- 选择了“好的”超参数后，对整个训练数据进行模型训练，然后用测试数据对模型进行测试。



模型泛化性

张伟楠 - [上海交通大学](#)

泛化能力

□ 泛化能力指的是模型对未观测数据的预测能力

- 可以通过泛化误差来评估，定义如下：

$$R(f) = \mathbb{E}[\mathcal{L}(Y, f(X))] = \int_{X \times Y} \mathcal{L}(y, f(x)) p(x, y) dx dy$$

- $p(x, y)$ 是潜在的（可能是未知的）联合数据分布

□ 在训练数据集上对泛化能力的经验估计为：

$$\hat{R}(f) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, f(x_i))$$

泛化误差

□ 有限假设集 $\mathcal{F} = \{f_1, f_2, \dots, f_d\}$

□ 泛化误差约束定理：

对任意函数 $f \in \mathcal{F}$ ，以不小于 $1 - \delta$ 的概率满足下式：

$$R(f) \leq \hat{R}(f) + \epsilon(d, N, \delta)$$

其中，

$$\epsilon(d, N, \delta) = \sqrt{\frac{1}{2N} \left(\log d + \log \frac{1}{\delta} \right)}$$

- N ：训练实例个数
- d ：假设集的函数个数

引理：Hoeffding不等式

- 令 X_1, X_2, \dots, X_n 为独立同分布的随机变量，其中 $X_i \in [a, b]$ ，那么平均变量 Z 为：

$$Z = \frac{1}{n} \sum_{i=1}^n X_i$$

那么下述不等式成立：

$$P(Z - \mathbb{E}[Z] \geq t) \leq \exp\left(\frac{-2nt^2}{(b-a)^2}\right)$$

$$P(\mathbb{E}[Z] - Z \geq t) \leq \exp\left(\frac{-2nt^2}{(b-a)^2}\right)$$

泛化误差约束定理证明

□ 假设约束的损失函数 $L(y, f(x)) \in [0,1]$

□ 基于Hoeffding不等式，对 $\epsilon > 0$ ，有

$$P(R(f) - \hat{R}(f) \geq \epsilon) \leq \exp(-2N\epsilon^2)$$

□ 由于 $\mathcal{F} = \{f_1, f_2, \dots, f_d\}$ 为有限集，满足

$$\begin{aligned} P(\exists f \in \mathcal{F}: R(f) - \hat{R}(f) \geq \epsilon) &= P\left(\bigcup_{f \in \mathcal{F}} \{R(f) - \hat{R}(f) \geq \epsilon\}\right) \\ &\leq \sum_{f \in \mathcal{F}} P(R(f) - \hat{R}(f) \geq \epsilon) \\ &\leq d \exp(-2N\epsilon^2) \end{aligned}$$

泛化误差约束定理证明

□ 下述不等式等价：

$$\begin{aligned} P(\exists f \in \mathcal{F}: R(f) - \hat{R}(f) \geq \epsilon) &\leq d \exp(-2N\epsilon^2) \\ &\Downarrow \\ P(\forall f \in \mathcal{F}: R(f) - \hat{R}(f) < \epsilon) &\geq 1 - d \exp(-2N\epsilon^2) \end{aligned}$$

□ δ 参数化方式如下：

$$\delta = d \exp(-2N\epsilon^2) \Leftrightarrow \epsilon = \sqrt{\frac{1}{2N} \log \frac{d}{\delta}}$$

那么泛化误差以 $1 - \delta$ 概率被约束

$$P(R(f) < \hat{R}(f) + \epsilon) \geq 1 - \delta$$



判别模型与生成模型

张伟楠 - [上海交通大学](#)

判别模型与生成模型

□ 判别模型

- 对可观测变量和未知变量的关联性建模
- 也称为条件模型
- 确定性判别： $y = f_{\theta}(x)$
- 随机判别： $p_{\theta}(y|x)$

□ 生成模型

- 建模数据的联合概率分布
- 给定一些隐参数或隐变量 $p_{\theta}(x, y)$
- 进行条件推断

$$p_{\theta}(y|x) = \frac{p_{\theta}(x, y)}{p_{\theta}(x)} = \frac{p_{\theta}(x, y)}{\sum_{y'} p_{\theta}(x, y')}$$

判别模型

- 对可观测变量和未知变量的关联性建模
- 也称为条件模型
- 确定性判别： $y = f_{\theta}(x)$
- 随机判别： $p_{\theta}(y|x)$

- 直接建模预测标签与已知特征的关联
- 易于定义特定依赖的特征和模型
- 实际上产生更高的预测性能

- 线性回归，逻辑回归， k 近邻，支持向量机，（多层）感知机，决策树，随机森林等

生成模型

- 建模数据的联合概率分布
- 给定一些隐参数或隐变量 $p_{\theta}(x, y)$
- 进行条件推断

$$p_{\theta}(y|x) = \frac{p_{\theta}(x, y)}{p_{\theta}(x)} = \frac{p_{\theta}(x, y)}{\sum_{y'} p_{\theta}(x, y')}$$

- 探寻数据分布（数据科学的本质）
- 受益于隐变量建模

- 朴素贝叶斯，隐马尔可夫模型，混合高斯，马尔可夫随机场，隐狄利克雷分布 (LDA)



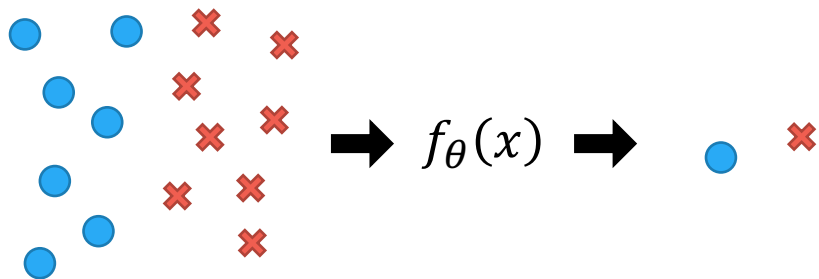
参数化模型与非参数化模型

张伟楠 - [上海交通大学](#)

参数化模型与非参数化模型

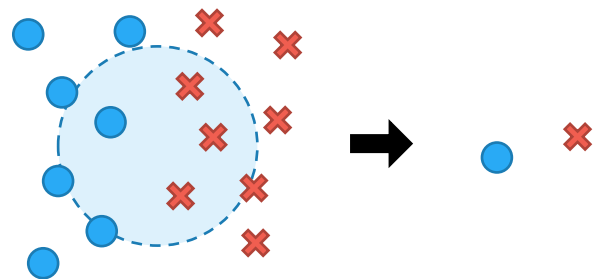
□ 参数化模型

- 对数据有基本假设
- 使用一套带参数的函数 $f_{\theta}(x)$ 来建模，参数空间对应模型空间
- 训练数据的知识注入在参数中
- 参数量不随数据量增加而改变
- 对参数的更新也就是对应模型的学习



□ 非参数化模型

- 对数据没有基本假设
- 没有直接的参数需要学习，可以直接定义模型空间（例如树模型）
- 对目标数据的预测直接来自一部分训练数据及其标签，这些数据直接代表了知识
- 实际模型参数量随数据量增加而增加



课程安排

参数化有监督学习

1. 机器学习概述
2. 线性模型
3. 双线性模型
4. 神经网络

非参数化有监督学习

5. 支持向量机
6. 决策树
7. 集成学习与森林模型

无监督学习部分

9. 概率图模型
10. 无监督学习

学习理论部分

11. 学习理论与模型选择

前沿话题部分

12. 迁移、多任务、元学习
13. System 1&2 机器意识

THANK YOU