清华大学数据库技术与应用

实用机器学习 II

授课教师：计算机系王健楠

授课学期：2026年（春季）

清華大學 Tsinghua University

课程大纲

01
异常检测

02
自动机器学习

- AutoML 动机与概述
- 自动特征生成
- Featuretools

03
可解释机器学习

AutoML 的动机

清華大學 Tsinghua University

1. 机器学习非常成功！

2. 但构建一个ML Pipeline需要：

- 具有长期经验的领域专家
- 专业的数据预处理
- 领域驱动的有意义的特征工程
- 选择正确的模型
- 超参数调优
- …… ……

H2O Driverless AI 演示

清華大學 Tsinghua University

思考问题：

- AutoML 软件会取代数据科学家吗？
- 作为数据科学家，如何看待 AutoML？

AI Overview

H2O.ai's revenue reached approximately **$75 million in 2024**, up from $67.8 million in 2023. The Mountain View-based AI company has shown consistent growth, specializing in open-source machine learning and AutoML, with a focus on enterprise AI adoption, serving over 20,000 organizations. GetLatka +4

Key revenue and company figures include:

- **2024 Revenue:** $75M
- **2023 Revenue:** $67.8M
- **Growth:** ~10.65% year-over-year from 2023 to 2024
- **Valuation:** The company reached a $1.7B valuation following a Series E round in November 2021
- **Total Funding:** H2O.ai has raised over $250M from investors, including NVIDIA, Goldman Sachs, and Wells Fargo GetLatka +3

H2O.ai serves a significant portion of the Fortune 500, focusing on automating data science processes. H2O.ai +1

演示视频：https://www.youtube.com/watch?v=ZqCoFp3-rGc

AutoML 的愿景

清華大學 Tsinghua University

对非专家：

- AutoML 使非专家也能使用机器学习模型和技术
- 无需先成为该领域的专家

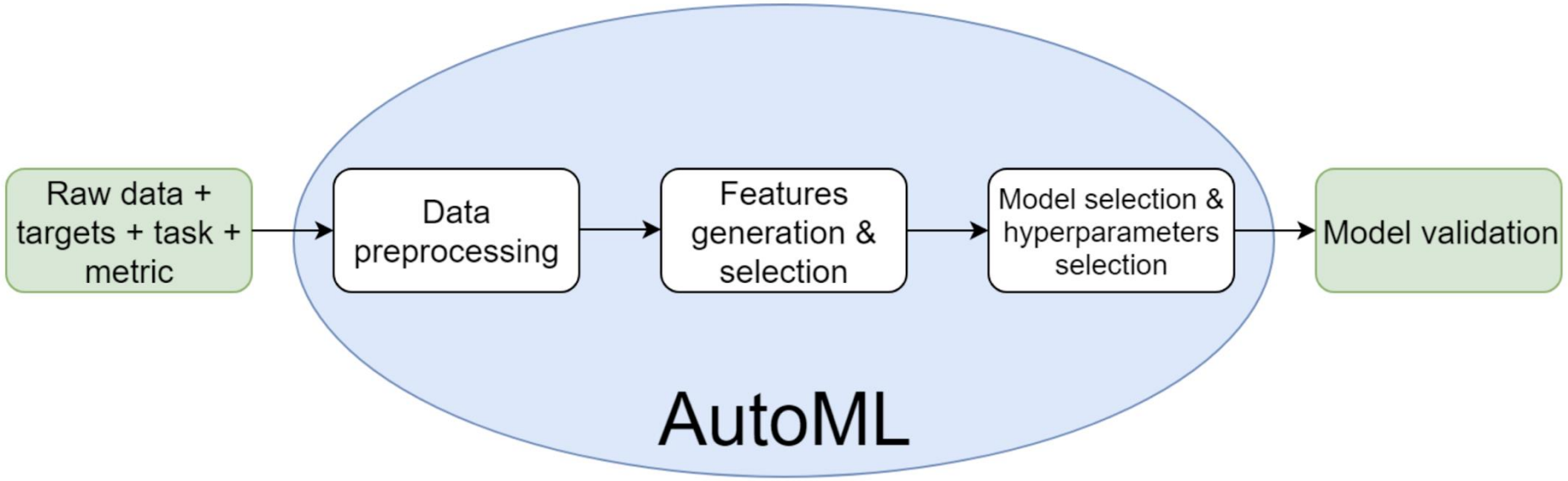
对数据科学家：

- AutoML 旨在增强（augment）而非自动化（automate）
- 数据科学团队的工作和工作实践

Wang et al. "Human-AI Collaboration in Data Science." CSCW (2019)

什么是 AutoML?

自动化将机器学习应用于实际问题的过程



AutoML 内容概览

- **自动特征生成**
- 自动特征选择 （前面已介绍）
- 自动超参数调优 （前面已介绍）

自动特征生成的动机

- 模型性能高度依赖数据集中特征的质量
- 领域专家生成足够有用的特征非常耗时

特征生成方法

一元算子（作用于单个特征）：

- 离散化数值特征
- 日期的规则展开
- 数学算子（如 Log 函数）

高阶算子（作用于2+个特征）：

- 基本算术运算（如 +, -, ×, ÷）
- 分组聚合（如 GroupByThenAvg, GroupByThenMax）

Katz et al. "ExploreKit: Automatic feature generation and selection." ICDM 2016.

Featuretools

- 开源的自动特征工程库
- 专为跨多关系表的特征生成设计

Featuretools 核心概念

清華大學 Tsinghua University

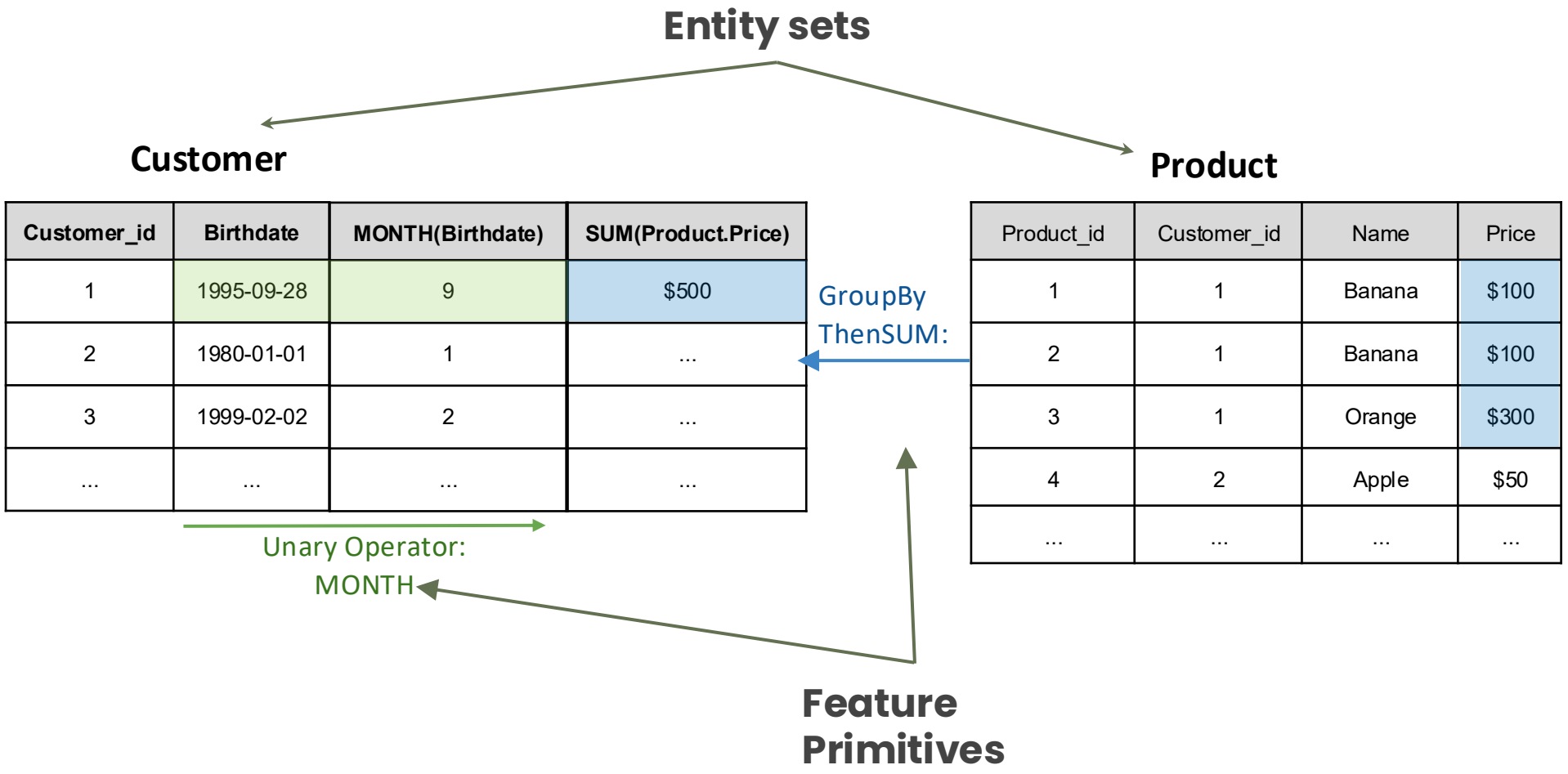
Entity：实体表

EntitySet：实体集合及其之间的关系

Feature Primitives（特征原语）：

- 一元算子：如 MONTH
- 高阶算子：如 GroupByThenSUM

Featuretools 示例



课程大纲

01
异常检测

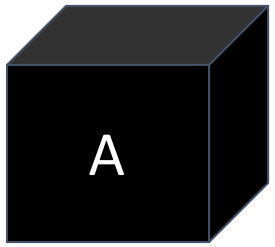
02
自动机器学习

03
可解释机器学习

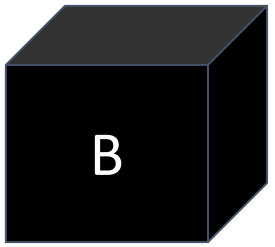
- 研究动机
- 全局视角
- 主流技术

更好的评估

你会选择哪个模型?



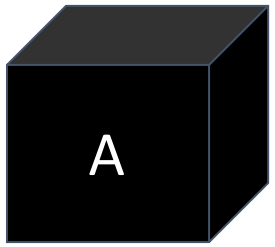
Bird: 99.0%

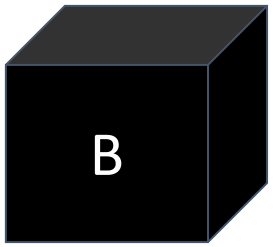


Bird: 99.9%

更好的评估

你会选择哪个模型?





A: Bird 99.0%

"因为它有翅膀和喙"

B: Bird 99.9%

"因为它是白色的且背景是蓝色"

法律合规

SR 11-7: Guidance on Model Risk Management

BOARD OF GOVERNORS
OF THE FEDERAL RESERVE SYSTEM
WASHINGTON, D.C. 20551

DIVISION OF BANKING
SUPERVISION AND REGULATION

SR 11-7
April 4, 2011

TO THE OFFICER IN CHARGE OF SUPERVISION AND APPROPRIATE SUPERVISORY AND EXAMINATION STAFF AT EACH FEDERAL RESERVE BANK

SUBJECT: Guidance on Model Risk Management

Art. 22 GDPR

Automated individual decision-making, including profiling

课程大纲

01
异常检测

02
自动机器学习

03
可解释机器学习

- 研究动机
- ✓ **全局视角**
- 主流技术

分类体系

- **透明模型** 线性回归，决策树，KNN，贝叶斯网络 ...
- **事后解释**
  - **全局模型解释** 特征置换重要性, 部分依赖图, 全局代理模型 ...
  - **个体预测解释** 归因分析, 影响实例, 局部代理模型...

分类体系

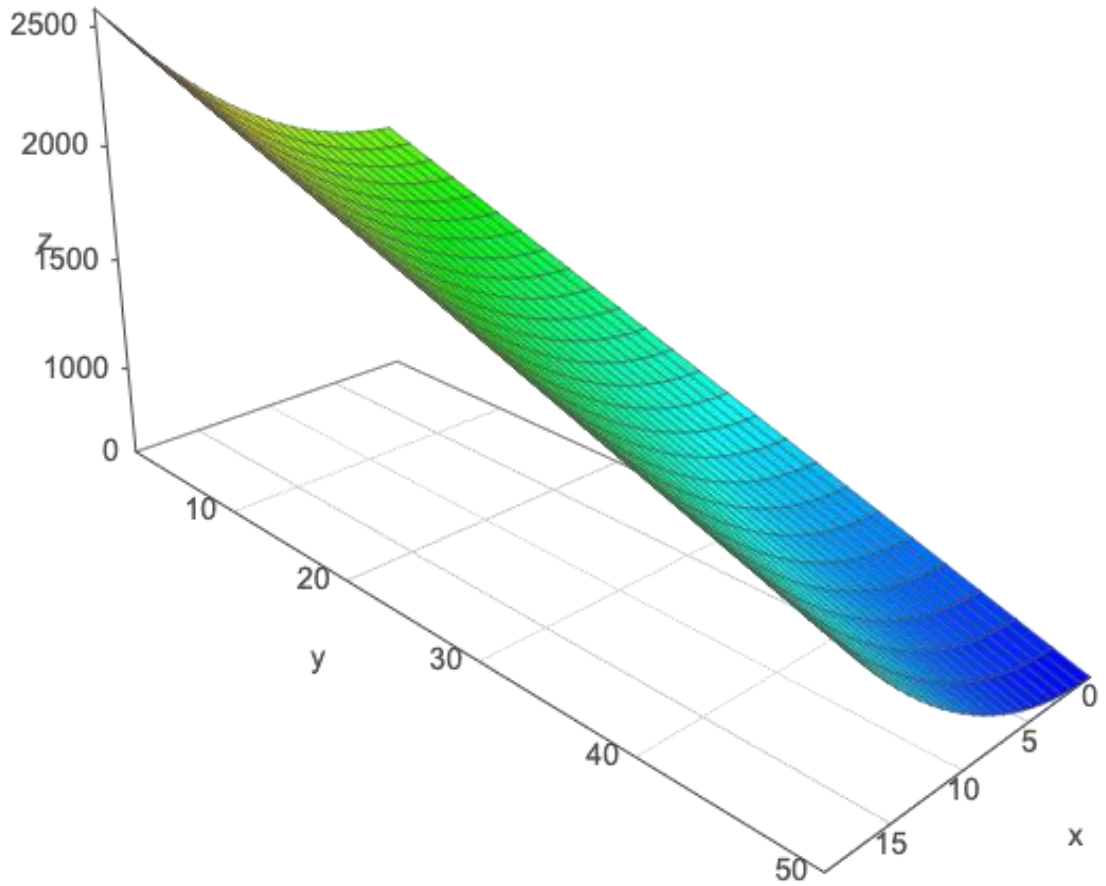
- **透明模型** 线性回归，决策树，KNN，贝叶斯网络 ...
- 事后解释
  - **全局模型解释** 特征置换重要性, 部分依赖图, 全局代理模型 ...
  - **个体预测解释** 归因分析, 影响实例, 局部代理模型...

线性回归

面积和距离如何影响房租?

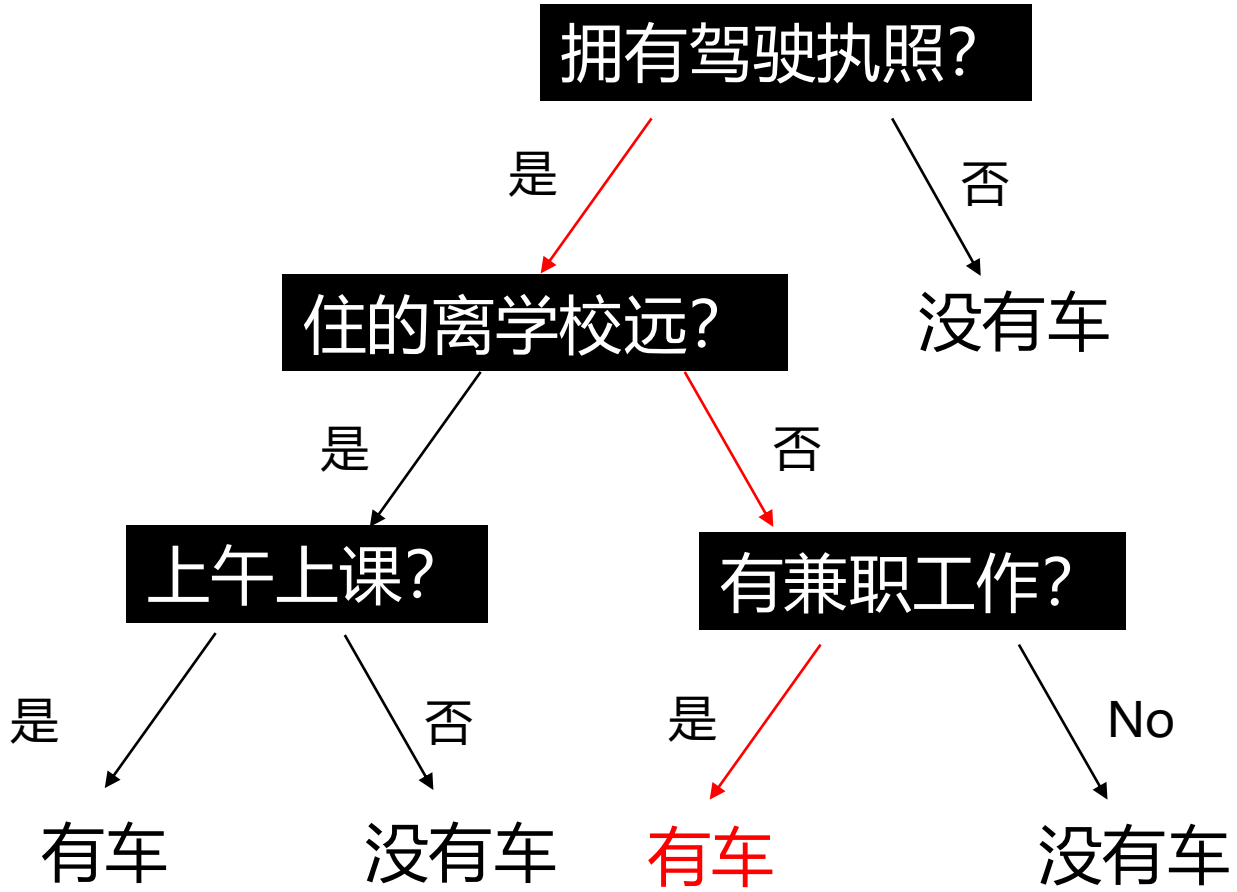
- 房租 (z)
- 面积 (x)
- 距离(y)

$$z = 2.1x - 2.4y + 1800$$



决策树（Decision Tree）

某学生是否拥有汽车?



模型为什么预测学生A有车?

- 沿着决策路径可以直接给出解释

分类体系

- 透明模型：线性回归，决策树，KNN，贝叶斯网络 ...
- 事后解释
  - **全局模型解释**：特征置换重要性, 部分依赖图, 全局代理模型 ...
  - 个体预测解释：归因分析, 影响实例, 局部代理模型...

特征置换重要性（Permutations）

核心思想：

- 通过计算打乱（permuting）某个特征后，模型预测误差的增加量来衡量该特征的重要性

ID	到学校的距离	洗手间个数	面积	最近的公交站	…
1	5.0km	1	$670ft^2$	0.30km	…
2	8.2km	2	$920ft^2$	0.12km	…
3	2.3km	2	$880ft^2$	1.20km	…
…	…	…	…	…	…
9999	10km	1	$680ft^2$	0.05km	…
10000	7.8km	1	$730ft^2$	0.23km	…

特征置换重要性（Permutations）

输入：训练好的模型 + 带标注的评估数据集

输出：每个特征的相对重要性

方法：

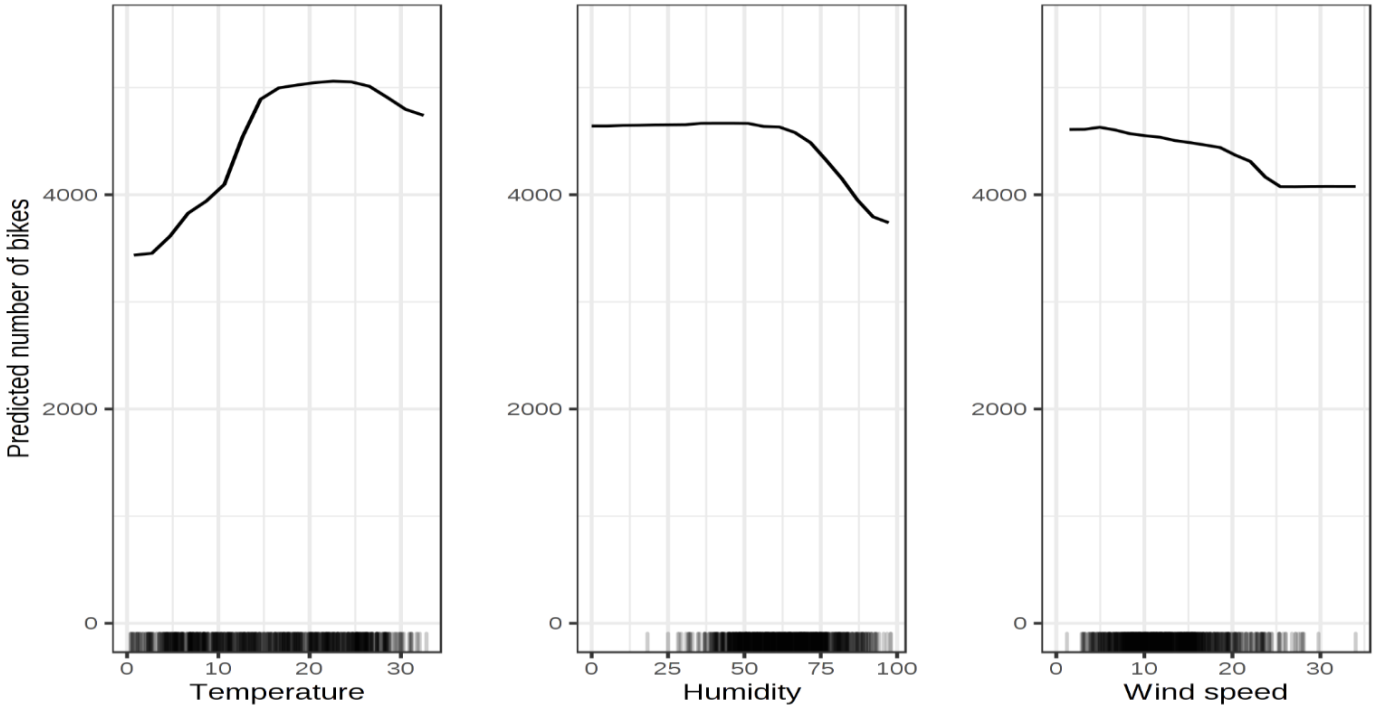
1. 在原始数据集上应用模型，得到估计误差 E
2. 对每个特征：
   - 打乱该特征值，重新应用模型得到新误差 E'
   - 特征重要性 = E' - E 或 E' / E

部分依赖图

核心思想：

- 展示一个或两个特征对机器学习模型预测结果的边际效应

ID	Temperature	Humidity	Wind Speed	Rental#
1	20	30	20	3000
2	25	35	10	2500
3	22	25	15	3300
..	..	…	…	…



部分依赖图（PDP）示例

清華大學 Tsinghua University

通过PDP可以直观看到：

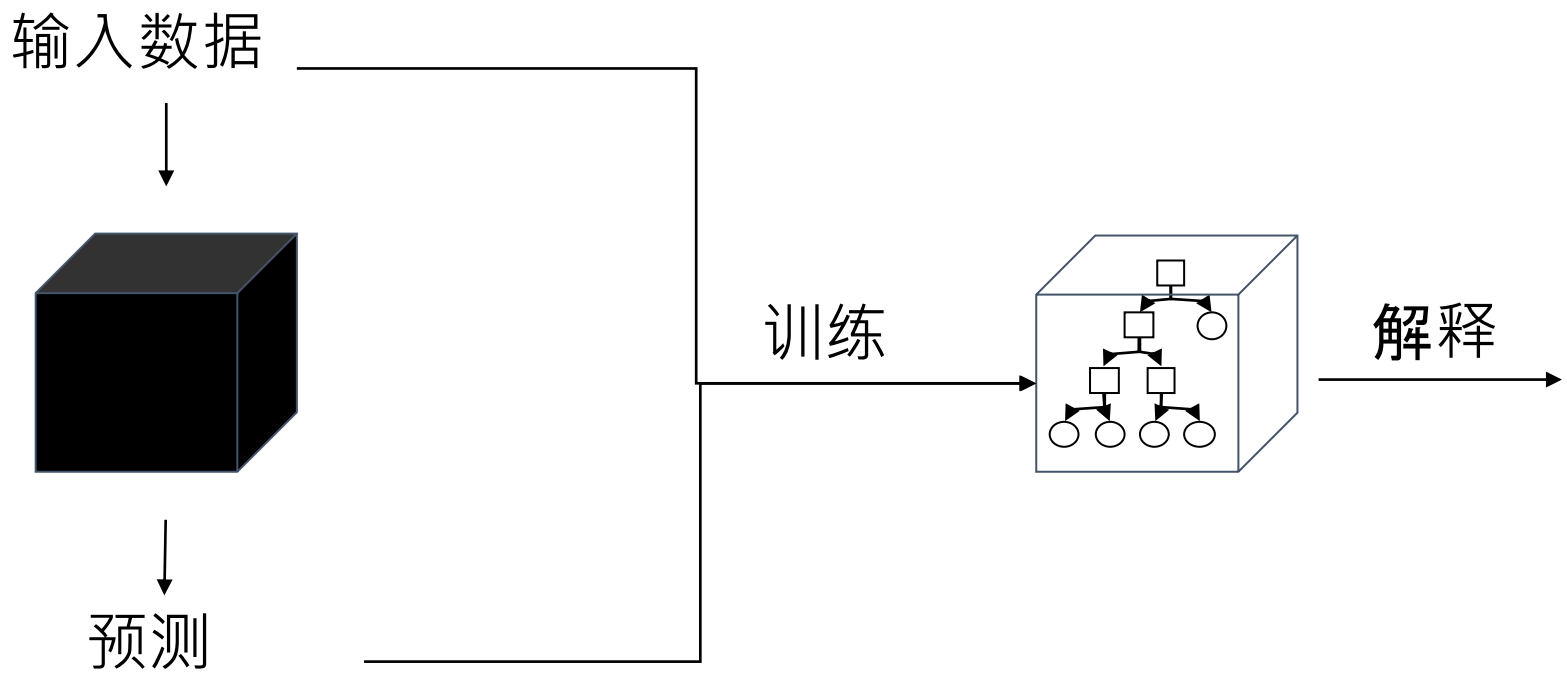
- 某个特征值的变化如何影响预测输出
- 特征与目标变量之间的非线性关系

局限性：假设特征之间相互独立

全局代理模型

核心思想：

- 训练一个透明模型来近似黑盒模型的预测



全局代理模型

步骤：

1. 用黑盒模型对数据进行预测
2. 用透明模型（如决策树）拟合黑盒模型的预测
3. 解释透明模型 → 近似理解黑盒模型

注意：

- 代理模型的保真度很重要
- 代理模型越简单越好解释，但可能不够准确

分类体系

- **透明模型** 线性回归，决策树，KNN，贝叶斯网络 ...
- **事后解释**
  - **全局模型解释** 特征置换重要性, 部分依赖图, 全局代理模型 ...
  - **个体预测解释** 归因分析, 影响实例, 局部代理模型...

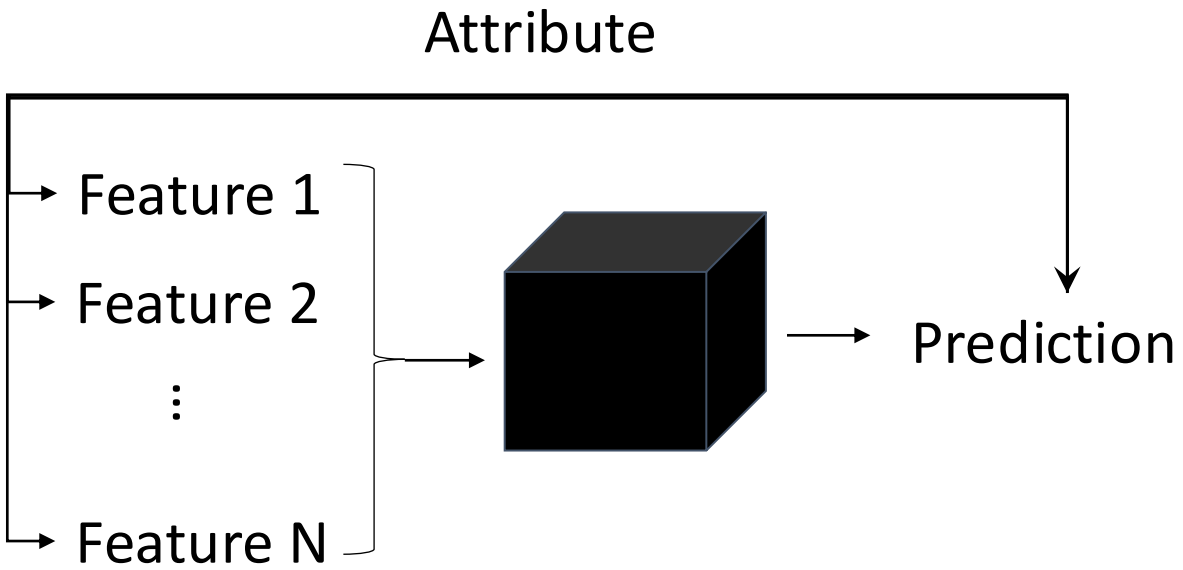
归因分析（Attribution）

核心思想：

- 将模型对某个样本的预测归因到其各个输入特征

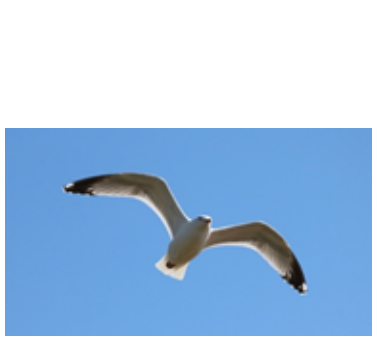
方法：

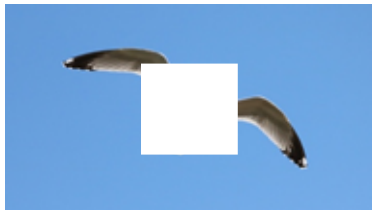
- 消融法
- Shapley 值
- ……

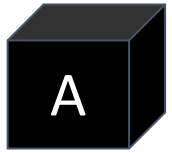


归因分析：消融法（Ablation）

- 逐个遮挡/移除特征，将预测变化归因给该特征



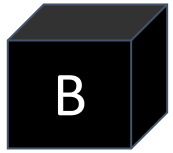




Bird (99%)

Bird (20%)

Bird (96%)



Bird (99%)

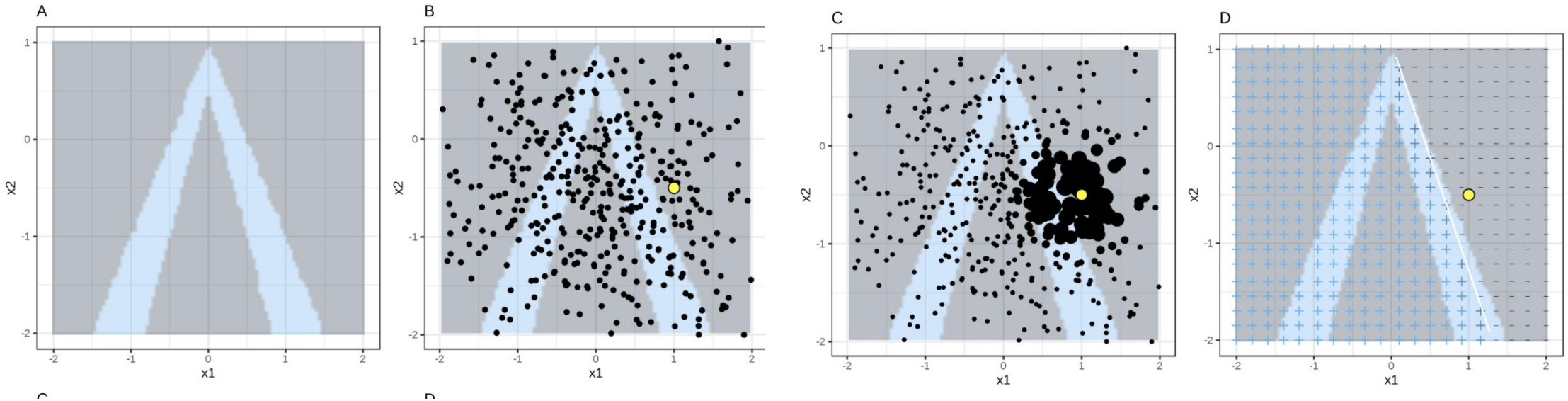
Bird (98%)

Bird (35%)

局部代理模型（LIME）

核心思想：

- 向机器学习模型输入数据的不同变体，观察预测如何变化

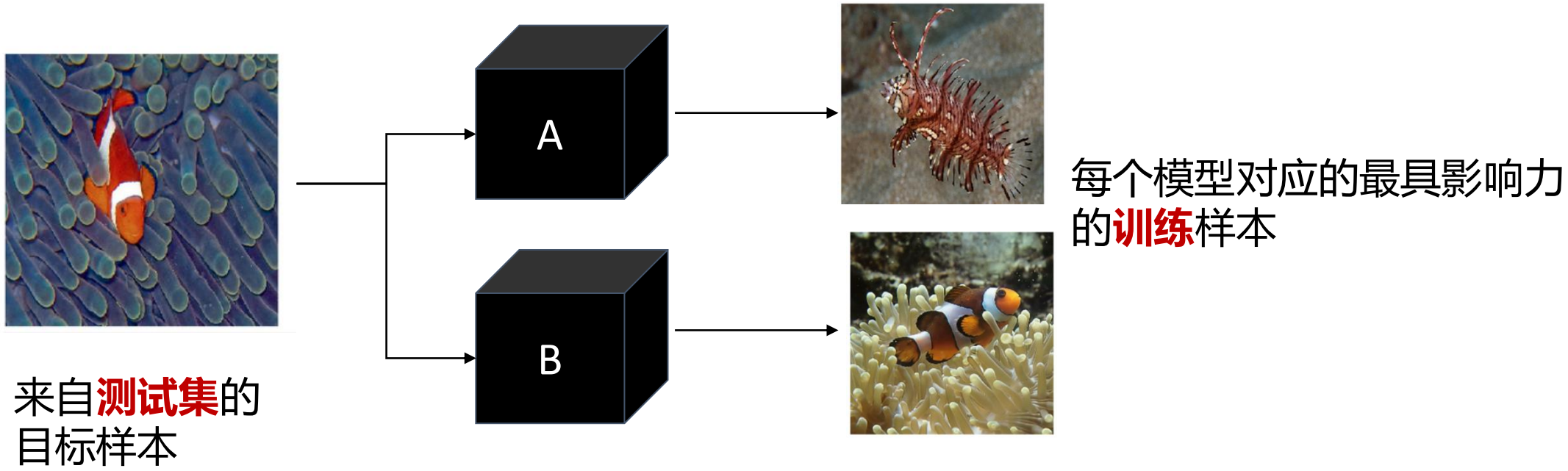


RIBEIRO, MARCO TULIO, SAMEER SINGH, AND CARLOS GUESTRIN. "WHY SHOULD I TRUST YOU?: EXPLAINING THE PREDICTIONS OF ANY CLASSIFIER." PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. ACM (2016)

影响力实例（Influential Instances）

核心思想：

- 通过识别对模型预测影响最大的训练实例来调试机器学习模型
- 当从训练数据中删除某个实例后模型预测显著变化时，该实例具有影响力



Koh & Liang. "Understanding black-box predictions via influence functions." ICML 2017.

可解释机器学习：总结

- **透明模型** 线性回归，决策树，KNN，贝叶斯网络 ...
- **事后解释**
  - **全局模型解释** 特征置换重要性, 部分依赖图, 全局代理模型 ...
  - **个体预测解释** 归因分析, 影响实例, 局部代理模型...

课程总结

第一部分：异常检测与机器学习实战

- 从问题到部署的完整ML解决方案设计流程
- 异常检测、聚类、特征工程、参数调优、模型部署

第二部分：自动化机器学习（AutoML）

- 自动特征选择、超参数调优、特征生成
- 神经架构搜索（NAS）

第三部分：可解释机器学习

- 透明模型 vs 事后解释
- 全局解释与个体预测解释方法

数据驱动 → 模型构建 → 自动化 → 可解释

清华大学数据库技术与应用

谢谢！

授课教师：计算机系 王健宗

授课学期：2026年（春季）

清華大学
Tsinghua University