

# 广义因果森林的构造 以及在在线交易市场的 应用

宛舒 亚利桑那州立大学 博士研究生



# 目录 CONTENT

## 01 背景介绍

1. 在线交易市场
2. 价格需求曲线

## 03 广义因果森林的构造

1. 剂效函数的非线性估计
2. 剂效函数的异质性

## 02 现有算法

1. 业界流行算法
2. 因果森林简介

## 04 实验和部署

1. 模拟结果
2. 在线部署

# 01

## 背景介绍

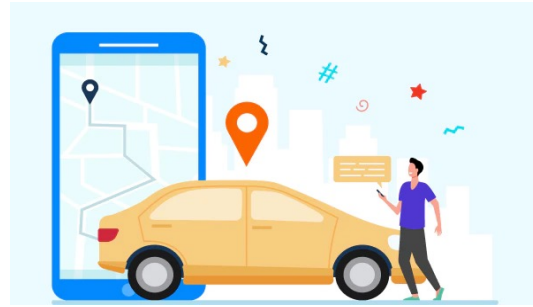
1. 在线交易市场
2. 价格需求曲线



# 在线交易市场蓬勃发展



在线旅游平台



网约车

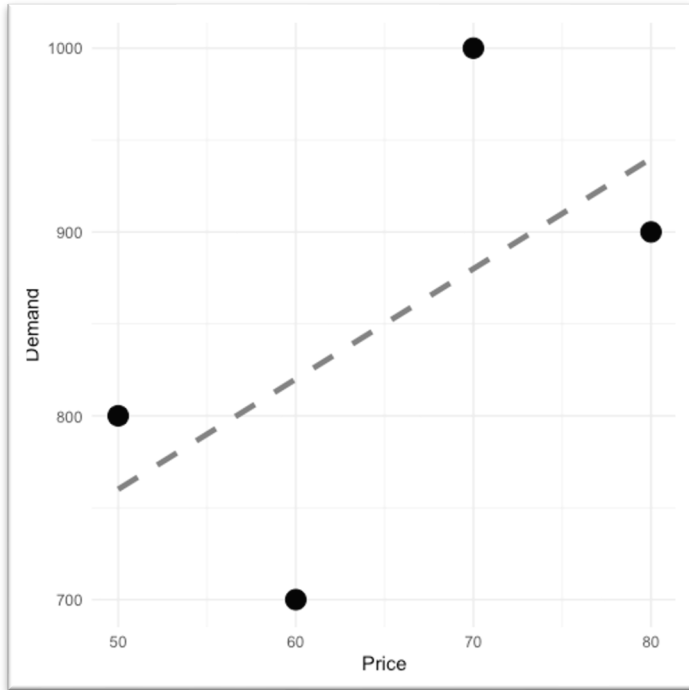


电商



外卖

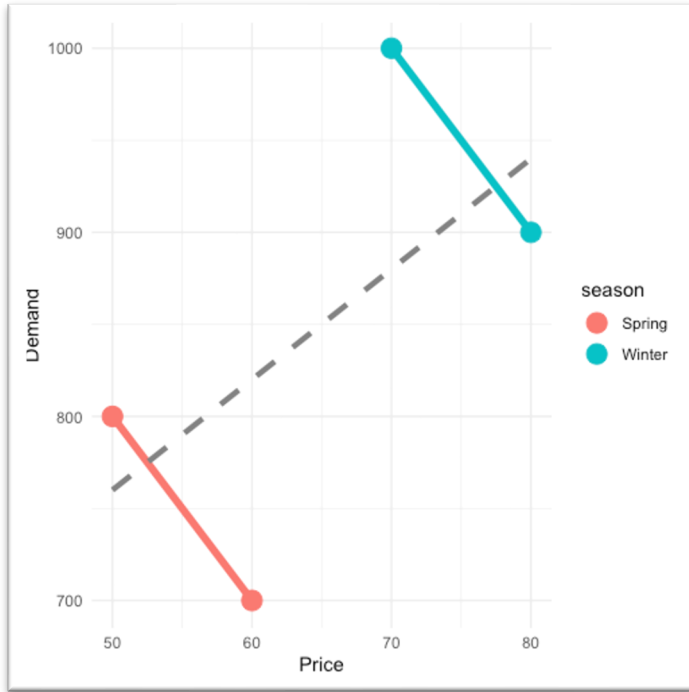
# 价格需求曲线



价格 (w)	需求 (Y)
50	800
60	700
70	1000
80	900

$$Y = \beta_0 + \beta_1 W$$

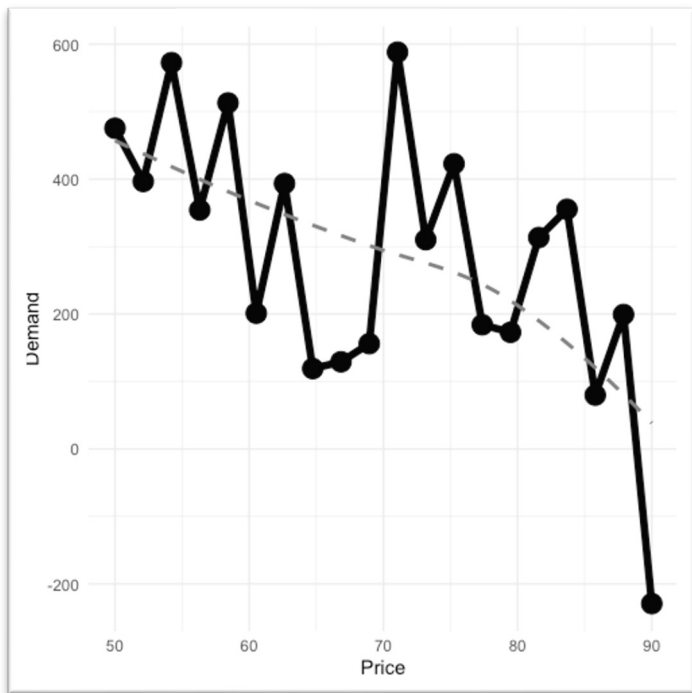
# 价格需求曲线



价格 (W)	需求 (Y)	季节 (X)
50	800	春季
60	700	春季
70	1000	冬季
80	900	冬季

$$Y = \beta_0 + \beta_1 W + \beta_2 X$$

# 在线交易市场下的价格需求曲线



## 价格需求曲线非线性

- 条件处理效应  
CATE, Conditional Average Treatment Effect

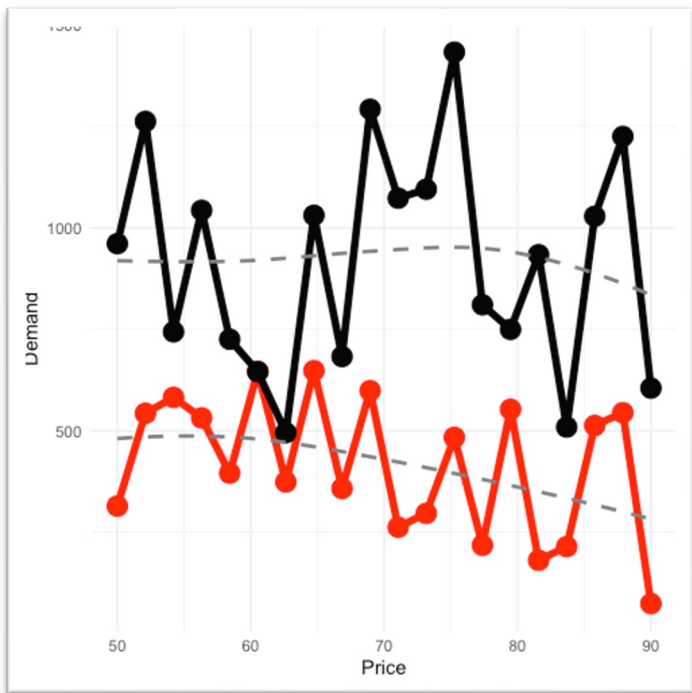
$$\tau = E(Y(W_1) - Y(W_0)|X)$$

- 条件剂量效应函数  
CDRF, Conditional dose-response function

$$\theta = E(Y(W)|X)$$

- 非参数回归

## 在线交易市场下的价格需求曲线



### 异质效应显著

- 淡季、旺季
  - 地区、国家
  - 用户群体
- 
- 控制所有混淆变量
  - 精细化策略



## 在线交易市场下的价格需求曲线



### 海量、高维数据

- 千万级用户体量
- 特征高达上百维
- 多个策略模型同时在线
  
- 适配大数据机器学习生态
- 提供灵活接口

# 在线交易市场下的价格需求曲线

## 价格需求曲线非线性

- 条件处理效应

CATE, Conditional Average Treatment Effect

$$\tau = E(Y(W_1) - Y(W_0)|X)$$

- 条件剂量效应函数

CDRF, Conditional dose-response function

$$\theta = E(Y(W)|X)$$

- 非参数回归

## 异质效应显著

- 淡季、旺季
- 地区、国家
- 用户群体

- 控制所有混淆变量
- 精细化策略

## 海量、高维数据

- 千万级用户体量
- 特征高达上百维
- 多个策略模型同时在线

- 适配大数据机器学习生态
- 提供灵活接口

## 02 现有算法

1. 业界流行算法
2. 因果森林简介



## 业界流行算法



**GRF**

Causal Forest



Orthogonal Random Forest



**CausalML**

Meta Learner (S,T,X)

$$Y = \theta(x)W + f(X = x, Z)$$

可以处理海量数据  
有效对异质效应进行预估  
**部分线性假设**

因果森林 = 因果推断 + 随机森林

Causal Forest

Causal Inference

Random Forest

# 因果推断 随机森林

Causal Inference

Random Forest



# 样本采样

记训练数据集

$$\Omega = \{(X_i, Y_i, W_i)_{i=1}^n\}$$

用于某棵树 $T_b, b \in (1, B)$ 的数据为 $\Omega$ 的随机抽样,  $S \in \Omega$

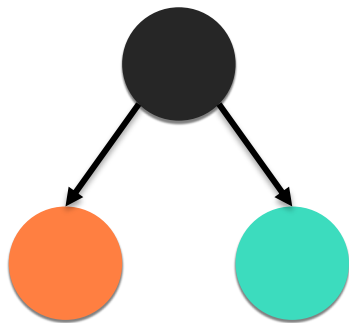
基于这棵树的估计值可以表示为

$$\widehat{\theta}_b(x) = \sum \frac{1_{X_i \in L_b(x)} * \theta_i}{|L_b(x)|}$$

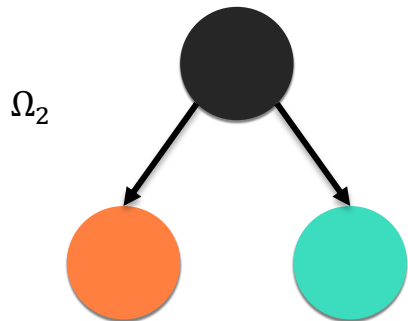
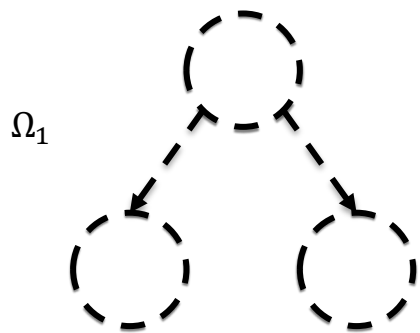
即给予落在同一节点的样本权重1, 否则为0

随机森林的估计器为决策估计值的平均

$$\theta = \frac{1}{B} \sum \hat{\theta}_b(x)$$



# 诚实估计



何为诚实树?

用于决定树结构的数据  $\Omega_1$  和用于决定估计量的数据  $\Omega_2$  不相同

样本抽样 + 诚实估计 + 其他一些条件

$$\frac{\widehat{\theta(x)} - \theta(x)}{\sigma_n(x)} \rightarrow N(0,1), \quad \sigma_n^2(x) \rightarrow_p 0$$

因果森林的估计量是相合的



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/65511222112011101>