Introduction to Uplift Modeling

Dr. Juan Orduz

PyConDE & PyData Berlin 2022

Motivation

How can we optimally select customers to be treated by marketing incentives?

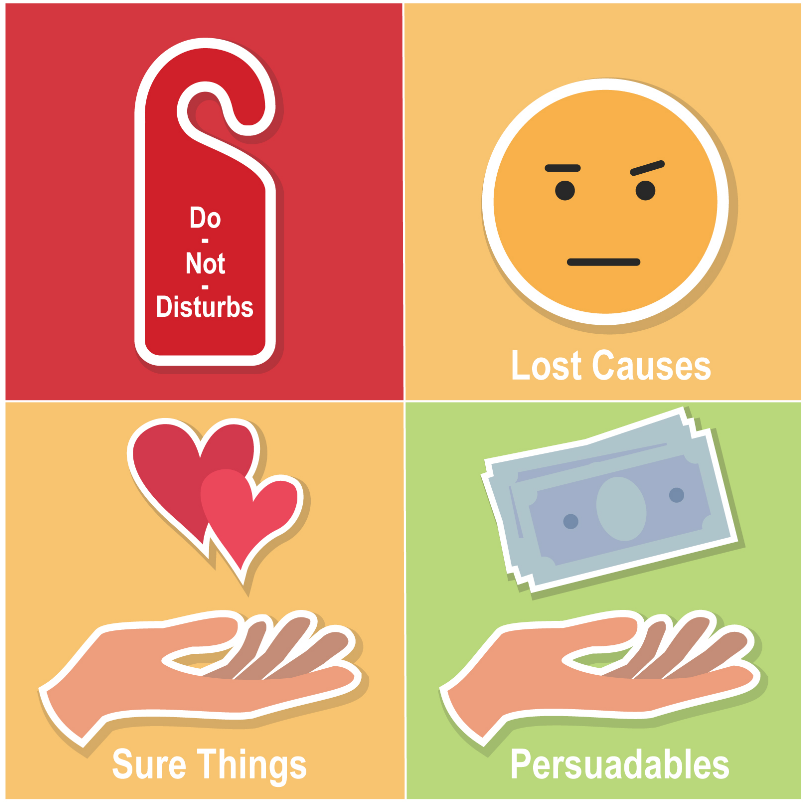


Image taken from https://www.uplift-modeling.com/en/latest/user_guide/introduction/clients.html

We can not send and not send incentives to the same customers at the same time

What is Uplift Modeling?

From Gutierrez, P., & Gérardy, J. Y. (2017). "Causal Inference and Uplift Modelling: A Review of the Literature"

> - Uplift modeling refers to the set of techniques used to model the incremental impact of an action or treatment on a customer outcome.
> - Uplift modeling is therefore both a Causal Inference problem and a Machine Learning one.

Conditional Average Treatment Effect

- Let Y_i^1 denote person i's outcome when it receives the treatment and Y_i^0 when it does not receive the treatment.
- We are interested in understanding the *causal effect* $Y_i^1 - Y_i^0$ and the *conditional average treatment effect* $CATE = E[Y_i^1|X_i] - E[Y_i^0|X_i]$, where X_i is a feature vector of the i-th person.
- **However, we can not observe them!** 🙁

Taken from Gutierrez, P., & Gérardy, J. Y. (2017). "Causal Inference and Uplift Modelling: A Review of the Literature"

Uplift

Let W_i is a binary variable indicating whether person i received the treatment, so that

$$Y_i^{obs} = Y_i^1 W_i + (1 - W_i) Y_i^0$$

Unconfoundedness Assumption

If we **assume** that the treatment assignment W_i is independent of Y_i^1 and Y_i^0 conditional on X_i, then we can estimate the $CATE$ from observational data by computing the empirical counterpart:

$$\mathbf{uplift} = \widehat{CATE} = E[Y_i | X_i, W_i = 1] - E[Y_i | X_i, W_i = 0]$$

Taken from Gutierrez, P., & Gérardy, J. Y. (2017). "Causal Inference and Uplift Modelling: A Review of the Literature"

Estimating UpLift

- Meta algorithms
- Direct measurements (trees)

S-Learner

Step 1: Training

$$\underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} & w_1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{11} & \cdots & x_{nk} & w_n \end{pmatrix}}_{X \bigoplus W} \xrightarrow{\mu} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Step 2: Uplift Prediction

$$\widehat{\textbf{uplift}} = \hat{\mu}\begin{pmatrix} x_{11} & \cdots & x_{1k} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{11} & \cdots & x_{mk} & 1 \end{pmatrix} - \hat{\mu}\begin{pmatrix} x_{11} & \cdots & x_{1k} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ x_{11} & \cdots & x_{mk} & 0 \end{pmatrix}$$

Taken from https://www.uplift-modeling.com/en/latest/user_guide/models/index.html

T-Learner

Step 1: Training

$$\underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{n_C k} \end{pmatrix}}_{X|_{\text{control}}} \xrightarrow{\mu_C} \begin{pmatrix} y_1 \\ \vdots \\ y_{n_C} \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{n_T k} \end{pmatrix}}_{X|_{\text{treatment}}} \xrightarrow{\mu_T} \begin{pmatrix} y_1 \\ \vdots \\ y_{n_T} \end{pmatrix}$$

Taken from https://causalml.readthedocs.io/en/latest/methodology.html#meta-learner-algorithms

T-Learner

Step 2: Uplift Prediction

$$\widehat{\textbf{uplift}} = \hat{\mu}_T \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{mk} \end{pmatrix} - \hat{\mu}_C \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{mk} \end{pmatrix}$$

Taken from https://causalml.readthedocs.io/en/latest/methodology.html#meta-learner-algorithms

X-Learner

Step 1: Training: Same as T-Learner

Step 2: Compute imputed treatment effects

$$\tilde{D}^T := \begin{pmatrix} y_1 \\ \vdots \\ y_{n_T} \end{pmatrix} - \hat{\mu}_C \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{n_T k} \end{pmatrix}$$

$$\tilde{D}^C := \hat{\mu}_T \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{n_C k} \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_{n_C} \end{pmatrix}$$

Taken from https://causalml.readthedocs.io/en/latest/methodology.html#meta-learner-algorithms

X-Learner

Step 3: Train with different targets

$$\underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{n_C k} \end{pmatrix}}_{X|_{\text{control}}} \xrightarrow{\tau_C} \begin{pmatrix} \tilde{D}_1^C \\ \vdots \\ \tilde{D}_{n_T}^C \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{11} & \cdots & x_{n_C k} \end{pmatrix}}_{X|_{\text{treatment}}} \xrightarrow{\tau_T} \begin{pmatrix} \tilde{D}_1^T \\ \vdots \\ \tilde{D}_{n_T}^T \end{pmatrix}$$

Taken from https://causalml.readthedocs.io/en/latest/methodology.html#meta-learner-algorithms

X-Learner

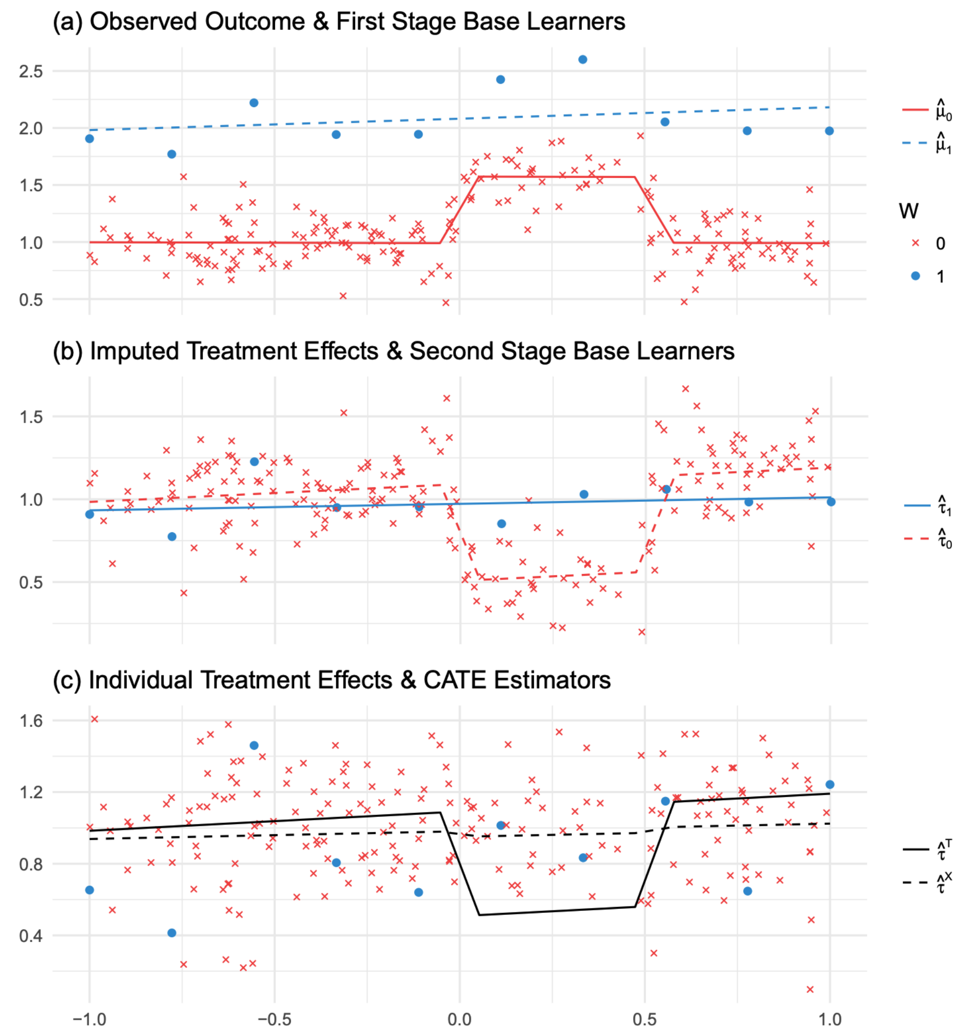
Step 4: Uplift Prediction

$$\widehat{\textbf{uplift}} = g(x)\hat{\tau}_C(x) + (1 - g(x))\hat{\tau}_T(x)$$

where $g(x) \in [0, 1]$ is a weight function.

Taken from https://causalml.readthedocs.io/en/latest/methodology.html#meta-learner-algorithms

Intuition behind the X-Learner



Taken from Sören, R, et.al. (2019) "Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning"

Some python implementations

- `causalml`

CausalML

- `scikit-uplift`

scikit-uplift

uplift modeling in scikit-learn style in python

Demo

Notebook Link

References:

- Gutierrez, P., & Gérardy, J. Y. (2017). "Causal Inference and Uplift Modelling: A Review of the Literature"
- Karlsson, H. (2019) "Uplift Modeling: Identifying Optimal Treatment Group Allocation and Whom to Contact to Maximize Return on Investment"
- Sören, R, et.al. (2019) "Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning"

Thank you!

More Info: juanitorduz.github.io/