Szymon Jaroszewicz[1,2] & Wouter Verbeke[3]

[1]Institute of Computer Science, Polish Academy of Sciences
[2]Faculty of Mathematics and Information Science, Warsaw University of Technology
[3]KU Leuven

September 18, 2022

IPI PAN

Faculty of Mathematics and Information Science
WARSAW UNIVERSITY OF TECHNOLOGY

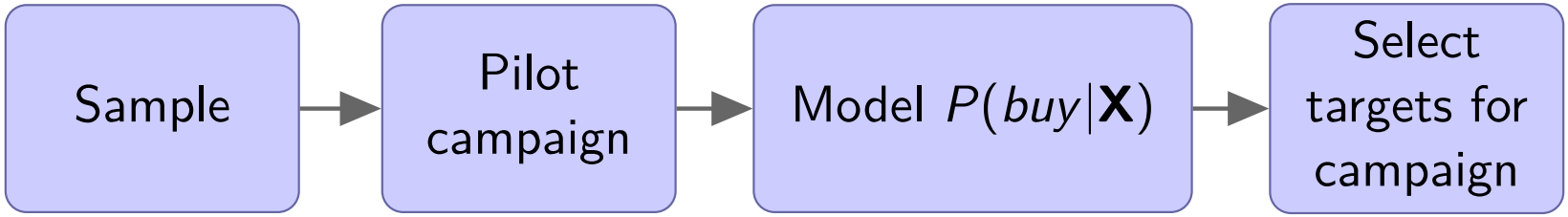
KU LEUVEN

Tutorial overview

- Introduction to uplift modeling (**SJ**)
  - how it differs from other approaches
- Uplift modeling methods 1 (**SJ**)
  - decision trees
  - ensemble methods
  - linear models
- Uplift modeling methods 2 (**VW**)
  - meta-learners
  - deep learning
  - learning2rank
- Evaluation of uplift models (**WV**)
- Implementing uplift models: software packages (**SJ**)
- Open issues (**WV**)

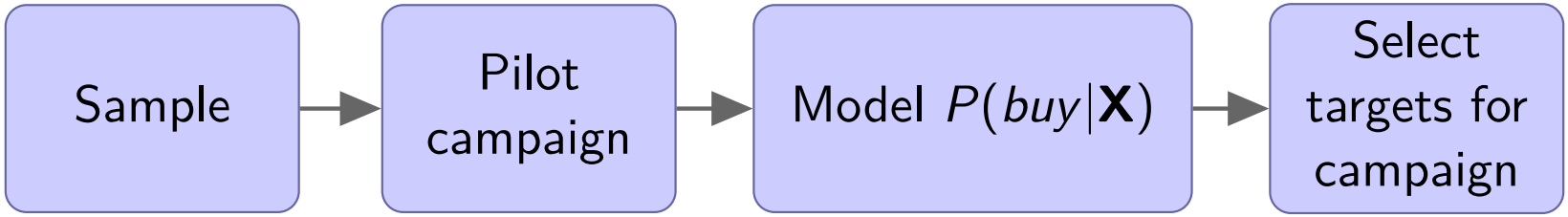
What is uplift modeling?

Old style marketing marketing campaign



What is uplift modeling?

Old style marketing marketing campaign



- But this is not what we need!

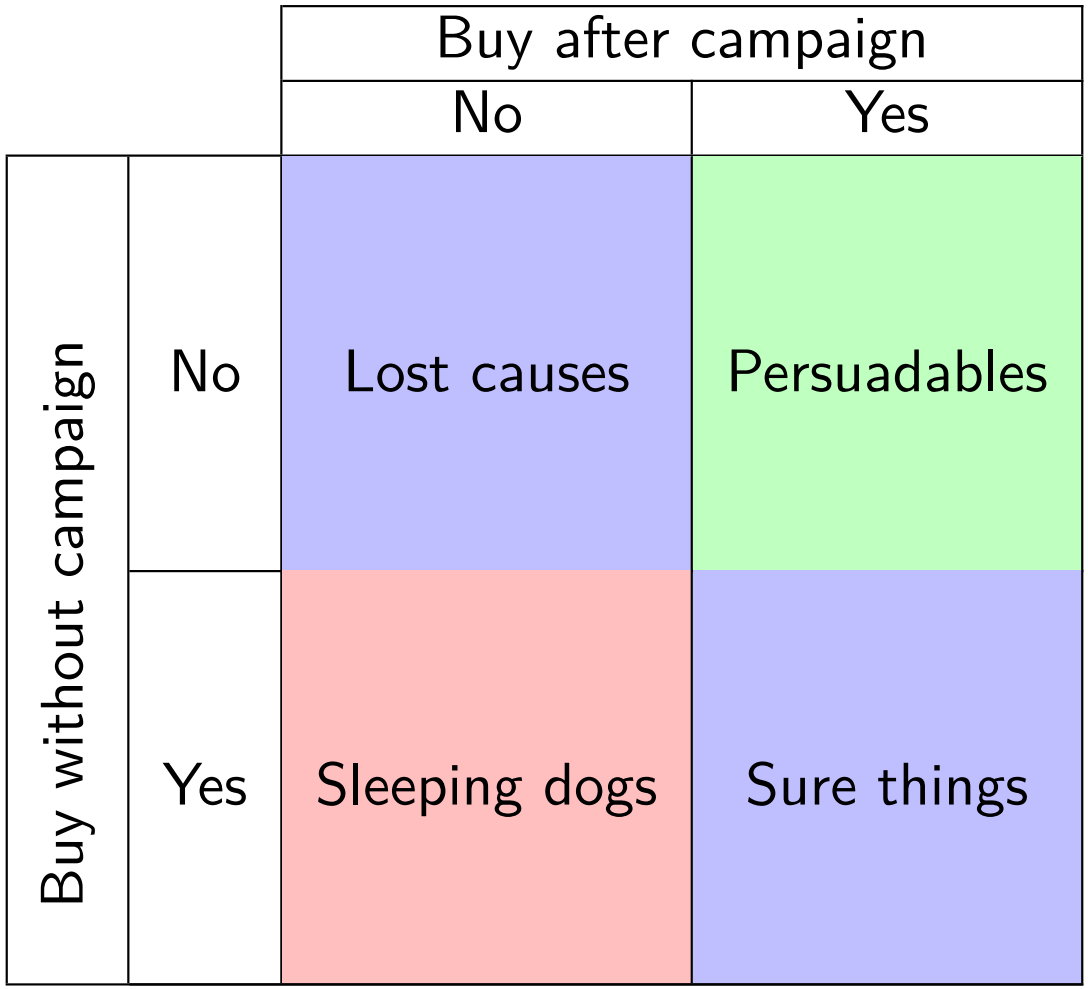
- We want people who bought because of the campaign
- Not people who bought after the campaign

Four groups of customers

We can divide potential customers into four groups

1. Responded **because** of the action
   (**the people we want**)
2. Responded, but would have responded **anyway**
   (**unnecessary costs)**
3. Did not respond and the action had **no impact**
   (**unnecessary costs**)
4. Did not respond **because** the action had a
   (**negative impact**)

Four groups of customers



Solution: Uplift modeling

- Solution: Uplift modeling
- Two training sets:
  1. the treatment group
     on which the action was taken
  2. the control group
     on which no action was taken
     used as background
- Build a model which predicts the difference between class probabilities in the treatment and control groups
- Random assignment to treatment/control groups allows for causal interpretation.
- Similar to a randomized clinical trials in medicine

Difference with traditional classification

Old style models predict the conditional probability

$$P(Y \mid x, Treatment)$$

Uplift models predict change in behaviour resulting from the action

$$P(Y \mid x, \text{Treatment}) - P(Y \mid x, \text{Control})$$

Uplift modeling within causal discovery

- Uplift modeling is part of a broad field of causal discovery
  - most areas have different focus (e.g. causal graph discovery)
- Individual Treatment Effect (ITE) estimation has similar goals
  - estimate effect of an action at the level of individuals

$$CATE(x) = \mathrm{E}(Y \mid x, \text{Treatment}) - \mathrm{E}(Y \mid x, \text{Control})$$

- How is uplift modeling different from ITE?

How is uplift modeling different from ITE?

- Uplift modeling and ITE estimation developed in parallel
  - several ideas rediscovered several times
- Different origins
  - uplift modeling has origins in marketing and ML
  - ITE estimation has origins in social/medical sciences and statistical community
- Result: different focus

How is uplift modeling different from ITE?

Problem setting	
Uplift modeling	**ITE estimation**
Primarily randomized experiments • easier to obtain in marketing (e.g. A/B testing)	Biased treatment assignment • e.g. doctor assigned therapy • RTs expensive/unethical in medical/social domain
Well designed experiment $\Rightarrow$ causal models	Nontrivial assumptions needed, e.g. no unmeasured confounders

How is uplift modeling different from ITE?

Goals & methodology	
Uplift modeling	**ITE estimation**
Obtain the best possible estimator of treatment effect	Unbiased CATE estimation while correcting treatment assignment bias
Focus on prediction: Machine Learning models predicting the effect directly	Sophisticated statistical methods for bias correction • doubly robust methods • joint CATE and propensity score estimation

How is uplift modeling different from ITE?

Evaluation	
Uplift modeling	ITE estimation
• Ranking based methods • Curves	PEHE (MSE of estimated effect)