

Causal Inference ML & Application in Marketing Uplift Modeling

Writer: Nguyen Tiet Nguyen Khoi @ Data Scientist Intern

Table of Content:

- 1. Introduction
 - 1.1. Abstract
 - 1.2. Challenges
 - 1.3. Problem statement
 - 1.4. Solution: Uplift Modeling
 - Introduction & usage
 - Inputs
- 2. Theories
 - 2.1. Concept & algorithms
 - The conditional average treatment effect (CATE)
 - Two major approaches for Uplift Modeling
 - Meta-learner: Solo-model (S-learner)
 - Meta-learner: Two-model (T-learner)
 - Meta-learner: X-learner
 - Key ideas of Meta-learner
 - Selection Bias Correction within Neural Networks (DragonNet)
 - Direct Uplift Model: Tree-based method
 - 2.2. Evaluation
 - The Bins of Uplift: Uplift on the top k percent users in each group.
 - The Uplift Curve: a function of accumulated uplift up to the number of users.
 - Qini curve: a function of incremental uplift up to the number of users.
 - The probabilities of causation
- 3. Challenges of Uplift Modeling in practice & Conclusion
 - 3.1. Challenges of Uplift Modeling in practice
 - 3.2. Conclusion
- 4. Demo
 - 4.1. Dataset
 - 4.2. Demo description
- 5. References

1. Introduction

1.1. Abstract

Uplift Modeling is a marketing modeling technique that directly models the incremental impact of a treatment (marketing action, campaign, strategy) on an individual's behaviors.

From those predictions (by using previous & historical data of all users and applying Uplift Modeling), marketers can understand customers' reaction and behavior on a treatment, then decide the most suitable strategy or evaluate the effect of the treatment directly, and put more effort on the well-performed treatment to maximize our ROI.

It is hard to define the correct group of users for treatment. Therefore, it is necessary to apply Machine Learning to **learn** whether the group we chose **truly response positive** to our treatment.

In this article, I will give a brief introduction about Machine Learning for Uplift Modeling, different types of applicable ML models for this task, and do a demo using a public dataset from the Criteo AI Lab (Kaggle).

1.2. Challenges

Let's consider a situation in which we need to decide influence marketing strategies for our users. Specifically:

- Increase users' engagement
- Retain our 'churn' users.

Increase users' engagement: A normal task for marketers. However, saying is a lot easier than doing: It is hard to decide how to do so, as well as necessary treatments and whether the users (who we are trying to impact) really needs the treatments or not. How do we know that the treatments (marketing campaigns, strategies, etc.) truly increase users' engagement?

Retain our churn users: In marketing, users who *previously used our product but has stop using it now* are called 'churn' users. It is more costly for marketer to attract new users rather than to retain existing 'churn' users. How could we 'predict' whether certain treatments could indeed retain soon-to-churn users?

1.3. Problem statement

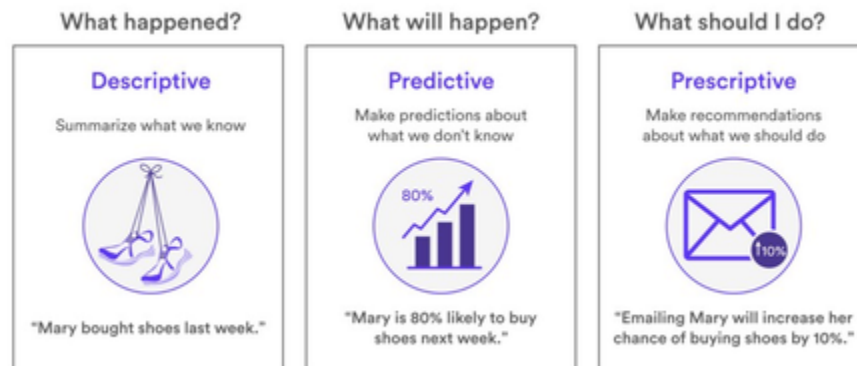
How different marketing treatments influence customer behaviors is important with marketers and the business. Therefore, it raises 2 critical questions:

- How to define correct group of users for treatment?
- How to evaluate whether users positively response to a specific treatment?

1.4. Solution: Uplift Modeling

Introduction & usage

Modern marketing involves applications of a variety of data science and machine learning techniques. These techniques can be generally grouped into three buckets: descriptive, predictive, or prescriptive. Uplift models fall into the Prescriptive category.



Common approaches in marketing (Adopted from Ms. linhna's slides):

LookAlike Model	Identify groups of users with similar behaviors	It is straightforward to classify existing users for specific feature (i.e., churn, spending, ...) based on historical data.	We have not addressed the subsequent important questions: 1. Which marketing 'treatment' we should implement? 2. How will targeted users react then?
(Predictive) Response Model	Measure probability of users to perform action(s) (<i>visit, make purchase, ...</i>) given being intervened (<i>notifications, emails, vouchers, ...</i>)	Behaviors of users being targeted in marketing campaigns can be predicted directly from historical data on previous campaigns.	It is unclear to measure potential outcomes given the absence of treatment -> it is hard to distinguish: 1. Users who will not perform if not being targeted -> favoured targets. 2. Users who will still perform even if not being targeted additional cost
Uplift Model	Measure the difference in user behaviors/ 'incremental' probability of performing action(s) with and without intervention(s)	Directly access the change (increase) in user performances subject to marketing campaigns.	This approach requires valid control groups of reasonable sizes. Moreover, it is easy to misinterpret campaign results. 1. First adoption of uplift models may lead to revenue reduction, 2. Uplift models may need to be refreshed frequently, 3. Results may not be significantly improved from those of conventional models.

The user's response to a corresponding marketing treatment can be divided into 4 groups:



- **Persuadables:** Customers who will convert if we apply treatments on them => generate additional ROI.
- **Sure Thing:** Customers who will always convert with or without treatment => it's not necessary to apply treatment on them.
- **Lost Cause:** Customers who will always not convert => it's a waste to apply treatment on them.
- **Sleeping Dogs:** Customers who will convert only without treatment as well as no convert with treatment => Apply treatment on them will only have reverse effect.

We surely want to target on the **persuadables** group. And we also want to make sure that we will not waste our time and resource on the other 3 groups as well.

Note that, **user's responses are vary by treatments**. In one campaign they can be in the persuadables group, however, in another campaign, they can be in the sleeping dogs group. For example, It is possible for a customer to be a Lost Cause if a campaign offers a 5% discount on the next purchase while being a Persuadables if offered a 20% discount.

Inputs

Defining an Uplift prediction involves specifying 2 inputs:

- Intervention: Which treatment(s) would you like to measure the impact of?
- Outcome: What is the conversion event that your intervention is meant to influence?

For example, if we are considering offering a discount (the intervention) with the goal of driving customer transactions (the outcome), our Uplift prediction for Customer A might tell us: "A discount will increase Customer A's probability of purchasing by 10% compared to if we did not offer the discount."

2. Theories

2.1. Concept & algorithms

Assume that we have 2 groups of users with the same distributions of features:

- Treatment group: We apply treatment for this group ('treatment' column with treatment = 1)
- Control group: We don't apply treatment for this group ('treatment' column with treatment = 0)

The conditional average treatment effect (CATE)

Let $Y_i(t)$ represents some outcome of interest Y for individual i under treatment condition t .

Let $t = 0$ is the control condition and $t = 1$ is the treatment condition.

The potential outcome framework as the causal effect of treatment is:

$$Y_i(1) - Y_i(0)$$

Let X denotes a vector features, x_i denotes feature values of an individual i . Then, the CATE:

$$\tau(x_i) = E[Y_i(1) - Y_i(0) | X = x_i]$$

The CATE allows us to understand how treatment effects vary depending on the observed characteristics of the population of interest => Uplift Modeling can be treated as a ML task which goal is to estimate the CATE.

Note that:

ATE = Average treatment effect: average effect over entire population of interest

CATE = Conditional average treatment effect: average treatment for the population given some condition (e.g. Age is 30 - 40 years) - I. E for a subset of the total population.

ITE = individual treatment effect: treatment effect for a specific individual.

Two major approaches for Uplift Modeling

There are 2 common approaches as follow:

- Indirect uplift model (Meta-learner, Selection Bias Correction within Neural Networks, etc.): Combine standard ML algorithms in various ways to estimate the CATE. In meta-learner, we let treatment as an additional feature.
- Direct uplift model: Modify existing ML algorithms to infer a treatment effect (Causal Tree-based model, Neural networks, etc.)

Two types of techniques:

1. Meta-learners

- i. Class transformation: four groups of users.
- ii. Treatment as an additional feature.

2. Direct Uplift Models

- i. Indirect = two models: one for Control and one for Target
→ Uplift = difference in predictions.
- ii. Direct: modify existing ML algorithms to infer a treatment effect.

Meta-learner are easier to implement since the standard base models can be used in the framework without changing the underlying algorithm. Also, meta-learner are computationally efficient since they can leverage the existing base model implementation that have already been optimized for code efficiency.

However, empirically, the meta-learner models have shown poor performance and only work well for certain conditions.

Meta-learner: Solo-model (S-learner)

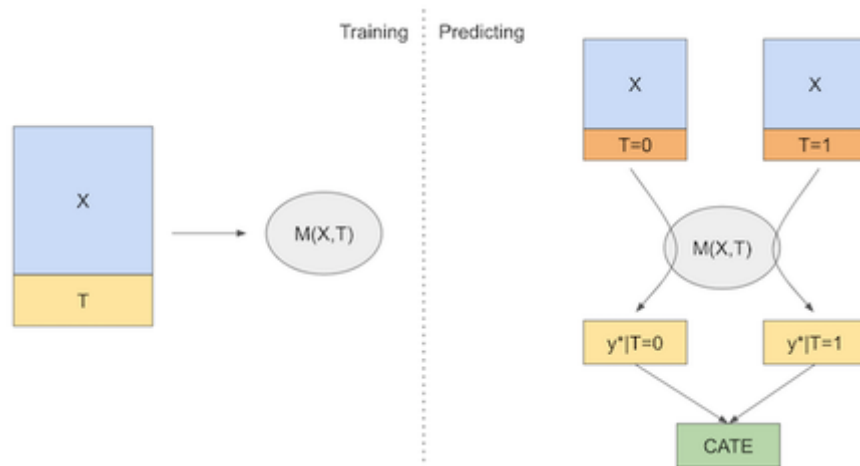
The S-Learner is the simplest learner we can think of. We will use a single machine learning model M_s to estimate:

$$\mu = E[Y | T, X]$$

To do so, we will simply include the treatment as a feature in the model that tries to predict the outcome Y . Then, we can make predictions under different treatment regimes. The difference in predictions between the test and control will be our CATE estimate:

$$\hat{\tau}(x) = M_s(X_i, T = 1) - M_s(X_i, T = 0)$$

We can visualize it as below:



Advantages: The S-learner can be treated as a good baseline for any causal problem due to its simplicity. Not only that, the S-learner can handle both continuous and discrete treatments, while the rest of the meta-learners can only deal with discrete treatments.

Disadvantage : The S-learner tends to bias the treatment effect towards zero. Since the S-learner employs what is usually a regularized machine learning model, that regularization can restrict the estimated treatment effect. If the treatment is very weak relative to the impact other co-variates play in explaining the outcome, the S-learner can discard the treatment variable completely. Notice that this is highly related to the chosen ML model you employ. The greater the regularization, the greater the problem.

Meta-learner: Two-model (T-learner)

The T-learner tries to solve the problem of the S-Learner by forcing the learner to first split on it. Instead of using a single model, we will use one model per treatment variable. In the binary case, there are only two models that we need to estimate:

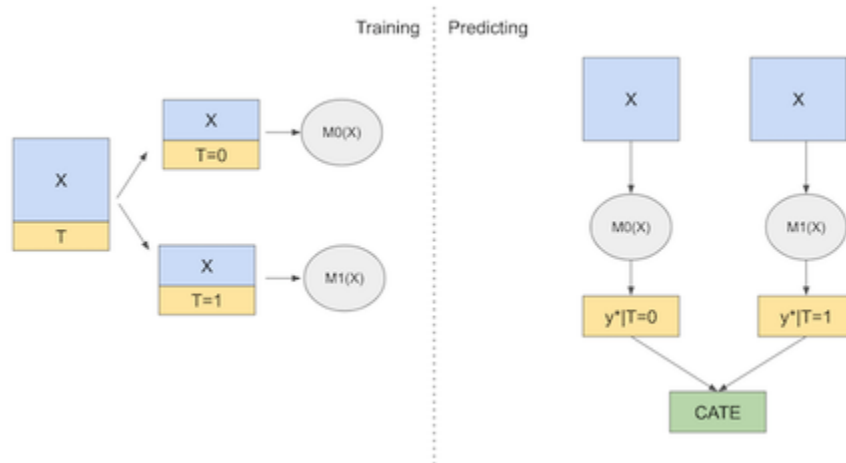
$$\mu_1(x) = E[Y|T = 1, X]$$

$$\mu_0(x) = E[Y|T = 0, X]$$

We estimate the CATE by:

$$\hat{\tau}(x) = M_1(X_i) - M_0(X_i)$$

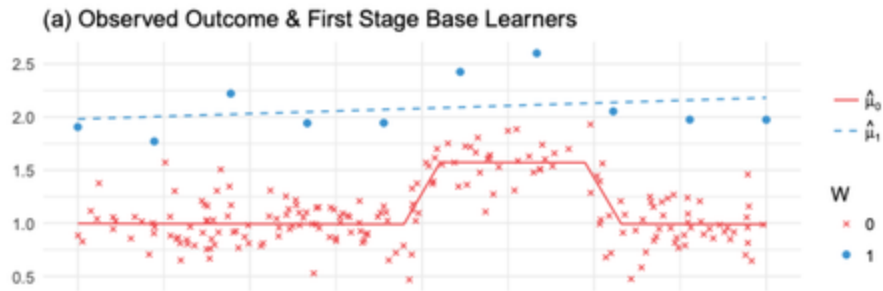
We can visualize it as below:



Advantage: The T-Learner avoids the problem of not picking up on a weak treatment variable

Disadvantage: It can still suffer from regularization bias. Consider the following situation (Kunzela et al, 2019):

We have *lots of data for the untreated* and *very few data for the treated* (a pretty common case in many applications, as treatment is often expensive). Now suppose you have some non linearity in the outcome Y , but the treatment effect is constant. We can see what happens in the following image:



We can see that since the treated group has small sample size, then to prevent over-fitting, the model can look very simple (only a linear line). However, with the control group, then the model is more complex since we have more data. What happens here is that the model for the untreated can pick up the non linearity, but the model for the treated cannot, because we've used regularization to deal with a small sample size.

Meta-learner: X-learner

The X-learner is more complex than the 2 learners explained above. First, we start by estimating the functions $\mu_0(x)$ and $\mu_1(x)$ using any suitable regression methods and data from the control and treatment group.

First, we estimate $\hat{M}_0(X) \approx E[Y|T=0, X]$; $\hat{M}_1(X) \approx E[Y|T=1, X]$

We estimate the "pseudo-effect" (the imputed treatment effect on the untreated) $\hat{\tau}$ for the observations in the control and treatment groups as:

$$\hat{\tau}(X, T=0) = \hat{M}_1(X, T=0) - Y_{T=0}$$

$$\hat{\tau}(X, T=1) = Y_{T=1} - \hat{M}_0(X, T=1)$$

Then, we fit 2 more models to predict those effects:

$$\hat{M}_{\tau 0}(X) \approx E[\hat{\tau}(X)|T=0]$$

$$\hat{M}_{\tau 1}(X) \approx E[\hat{\tau}(X)|T=1]$$

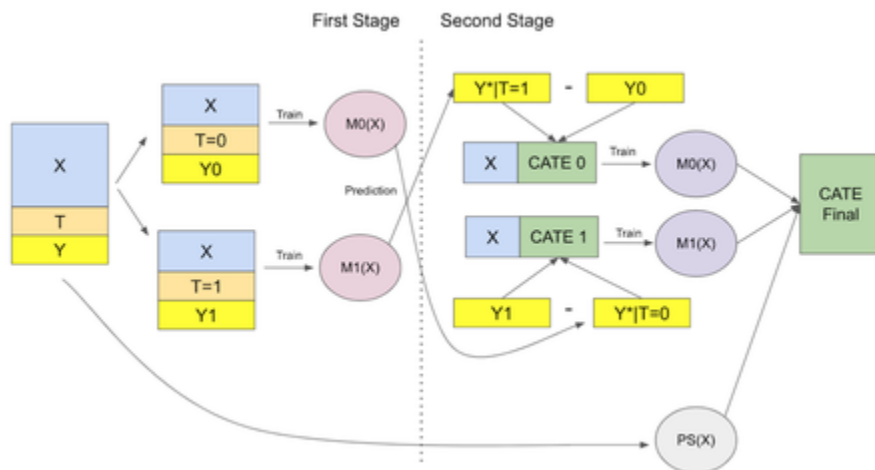
Finally, we combine those 2 estimators to obtain the estimated CATE:

$$\hat{\tau}(x) = \hat{e}(x)\hat{M}_{\tau 0}(x) + [1 - \hat{e}(x)]\hat{M}_{\tau 1}(x)$$

where $\hat{e}(x)$ is the propensity score $P[W_i = 1|X_i = x]$ with W indicates the treatment assignment.

Since there are very few treated units, $\hat{e}(x)$ is very small and thus, the $\hat{M}_{\tau 1}(x)$ will have greater weight than the $\hat{M}_{\tau 0}(x)$. This is reasonable since $\hat{M}_{\tau 0}(x)$ use $\hat{M}_1(X)$ (model for the treatment group) to estimate for the untreated group, and since we use very few data to train the $\hat{M}_1(X)$, it will be more simple. In contrast, since the $\hat{M}_{\tau 1}(x)$ use $\hat{M}_0(X)$, which was trained on a larger data sample, will have complex model than the $\hat{M}_1(X)$.

We can visualize it as below:



Advantage: Compared to the T-learner or S-learner, the X-learner does a much better job in correcting the wrong CATE estimated at the non linearity.

Disadvantage: Meta-learner in general, have poorer performance empirically compared to other methods such as Direct Uplift.

Key ideas of Meta-learner

The S-learner tends to work well when the treatment is not a weak predictor of the outcome. But if that's not the case, the S-learner tends to be biased towards zero or even drop the treatment entirely. Due to its simplicity (and in fact do not have high performance in practice), the S-learner can be used as a baseline.

The T-learner fits one Machine Learning model per treatment level. This works fine when there are enough samples for all treatment levels, but it can fail when one treatment level has a small sample size, forcing a model to be heavily regularized.

We can add another level of complexity using an X-learner, where we have two fitting stages and we use a propensity score model to correct potential mistakes from models estimated with very few data points.

One big problem of these learners (except the S-learner) is that they **assume a binary or categorical treatment**.

Read more about Meta Learners here: <https://matheusfacure.github.io/python-causality-handbook/21-Meta-Learners.html>

Selection Bias Correction within Neural Networks (DragonNet)

We can of course use **two neural networks** as outcome models to estimate the CATE. The two outcome models are likely very similar, since both approximate to a large extent the outcome process without treatment. We may be able to gain efficiency and improve calibration through parameter sharing in the lower hidden layers. The architecture is then best understood as a single multi-task network, with one loss calculated on the control group observations and one (or more) loss calculated on the treatment group observations.

The multi-task architecture has an additional advantage when working with observational data. When we cannot conduct an experiment and treatment assignment is not random, we can correct for variables that impact the treatment assignment to still make unbiased estimates. It is in fact sufficient to correct only for the variables that impact treatment assignment (propensity weighting).

An efficient way to filter the relevant information in the multi-task neural network is to correct the shared hidden layers. We correct the last shared layer, for example, by adding the treatment probability as an additional output. Predicting the treatment probability forces the hidden layers to distill the information that is necessary to predict treatment assignment and focus less on the information that is relevant only for outcome prediction, but doesn't differ between the control and treatment group.

Read more here:

Shalit, U., Johansson, F. D., & Sontag, D. (2017). [Estimating individual treatment effect: generalization bounds and algorithms](#). Proceedings of the 34th International Conference on Machine Learning (ICML 2017).

Shi, C., Blei, D. M., & Veitch, V. (2019). [Adapting Neural Networks for the Estimation of Treatment Effects](#). ArXiv:1906.02120 [Cs, Stat].

Alaa, A. M., Weisz, M., & van der Schaar, M. (2017). [Deep Counterfactual Networks with Propensity-Dropout](#). ArXiv E-Prints, arXiv:1706.05966.

Direct Uplift Model: Tree-based method

We can use a modified version of Decision Tree and Random Forest algorithms to handle causal inference task such as uplift modeling. Those modified methods are called Causal Decision Tree/Random Forest a.k.a Uplift Tree.

Basically, the Uplift Tree approach consists of a set of methods that use a tree-based algorithm where the splitting criterion is based on differences in uplift. We can quantify the gain in divergence as the result of splitting as follow:

$$D_{gain} = D_{after_split}(P^T, P^C) - D_{before_split}(P^T, P^C)$$

where D measures the divergence and P^T, P^C refer to the probability distribution of the outcome of interest in the treatment and control groups, respectively. Three different ways to quantify the divergence, KL, ED and Chi.

Read more here: <https://causalml.readthedocs.io/en/latest/methodology.html#tree-based-algorithms>

2.2. Evaluation

The nature of Causal Inference make it hard to find a loss measure of each observations (they can happen either way only) no ground truth.

There are 3 methods to evaluate Uplift Model, including the Bins of Uplift, Uplift curve, and Qini curve.

The Bins of Uplift: Uplift on the top k percent users in each group.

Bins of uplift takes the uplift score and sort the individuals in descending order for both the treatment group and control group separately and then split them up into k segments.

Uplift is evaluated by subtracting purchase rate in treatment group and control group per segment k :

$$u_{kp} = \frac{r_k^t}{n_k^t} - \frac{r_k^c}{n_k^c} \quad (4.10)$$

where u_{kp} is the predicted uplift per segment, r_k^t and r_k^c is the number of purchases in treatment group and control group respectively per segment, n_k^t and n_k^c is the number of individuals in treatment group and control group respectively per segment.

The bins of uplift methodology do not provide any metrics to compare different uplift models to each other; rather, it only visualizes the uplift per segment k .

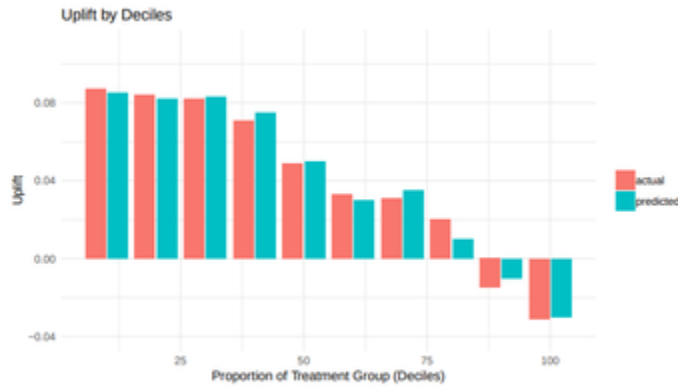


Figure 4.1: Example of Bin of Uplift Graph

The blue bars represent the predicted uplift for each segment, and the red bars represent the actual uplift per segment in the data set.

The Uplift Curve: a function of *accumulated* uplift up to the number of users.

The uplift curve requires a model with a binary target feature, and an uplift score for each individual where a higher value implies a higher chance of purchase given treatment. The individuals are sorted by the uplift score in descending order, and cumulative sum of purchase is computed. The uplift curve assumes that the individual with the highest score is contacted first.

Individual	Is purchase	Score given by model	Number of cumulative purchases	Number targeted
x_1	Yes	1.2	1	1
x_2	Yes	1	2	2
x_3	No	0.9	2	3
x_4	Yes	0.8	3	4
x_5	Yes	0.75	4	5
x_6	No	0.7	4	6
x_7	No	0.5	4	7

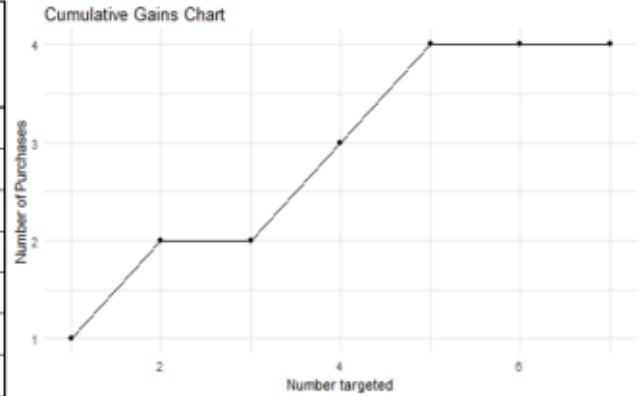


Figure 4.2: Cumulative table of gain and uplift curve

An example of an uplift curve for a sample with a purchase rate of 5 % can be seen. The graph visualizes the number of purchases as a function of the number of individuals treated as follow:

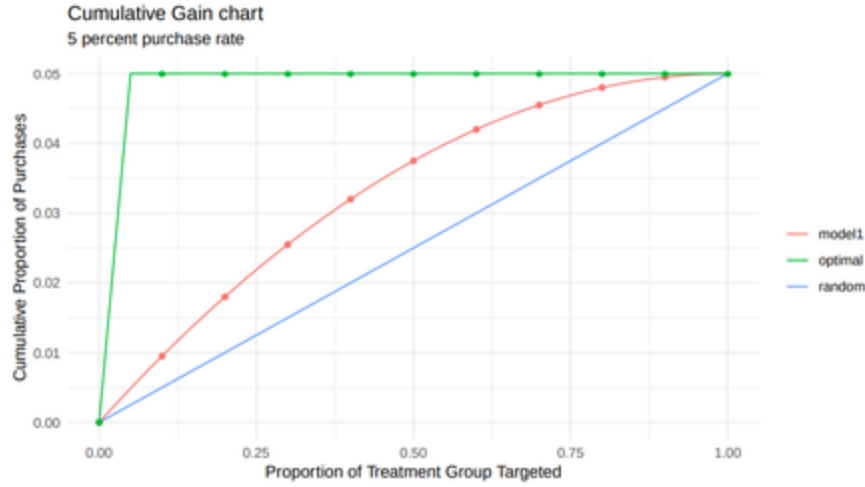


Figure 4.3: Uplift Curve/ Gains Chart

The random line shows the effect of treatment if the selection of whom to treat within the treatment group is random. If the entire sample is targeted, the 5 % that purchases would be found.

In contrast, the optimal model, manages to identify all the 5 % of the sample that will purchase because of treatment. Therefore, the curve has a steep increase until all purchasers due to treatment have been identified, then the curve flattens out horizontally since no other individual in the sample will purchase because of treatment.

A typical uplift model will be somewhere in between the random and optimal curves and an example is visualized as model 1. The closer model 1 is to the optimum, the better model. If model 1 would be below the random curve, it implies that the model has the opposite effect, it captures the individuals that would not purchase because of treatment.

Qini curve: a function of *incremental* uplift up to the number of users.

The Qini measures, Qini coefficient and Qini curve are a generalization of the uplift curve and the Gini coefficient.

The difference between the Qini curve and the gain curve (uplift curve) is that the Qini curve plots the incremental purchases instead of the cumulative number of purchases. The incremental purchase is computed per segment and group:

$$u_{kp} = r_k^t - \frac{r_k^c n_k^t}{n_k^c} \quad (4.11)$$

where u_{kp} is the predicted uplift per segment, r_k^t and r_k^c is the number of purchases in treatment and control per segment respectively, and n_k^t and n_k^c is the group sizes per segment for treatment group and control group. The Qini coefficient is computed in the same manner as the Gini coefficient, but it is computed from the Qini curve instead of the uplift curve.

An example of the Qini curve is as follow:

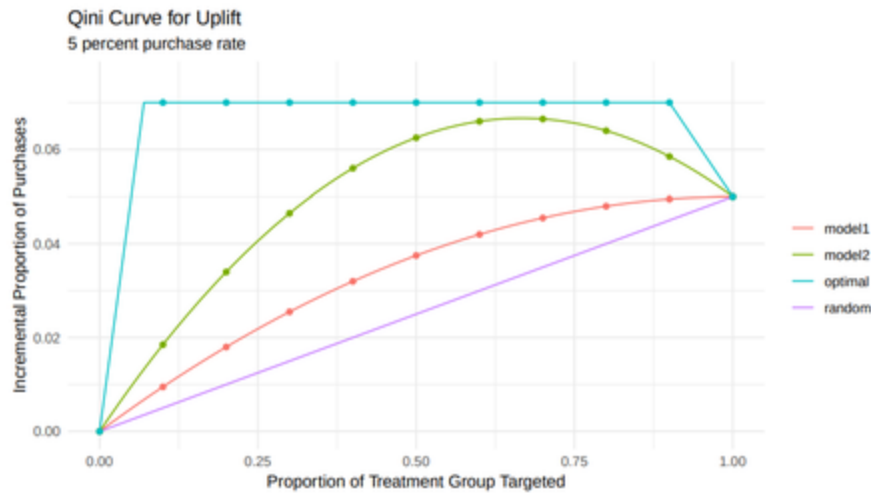


Figure 4.4: Example of Qini curve

The optimal curve represents the theoretically best possible outcome where every individual that can be persuaded by treatment is identified and given treatment first.

The random line represent the outcome if all individuals were given treatment in random order.

Model 1 and Model 2 are two uplift models, where model 2 outperforms model 1.

The highest possible uplift a model can capture is computed by the total uplift minus the negative effect of treatment. So in this example, the model 2 suggest that **approximately 65 % of the treatment group should be given treatment to gain approximately 6.5% uplift.**

The probabilities of causation

Beside those 3 methods above to calculate and evaluate uplift, we can calculate the probability that the treatment are truly **cause** the result.

Then the probability that the treatment is *sufficient* for y to occur can be defined as

$$PS = P(y_t | c, y')$$

This is the probability that the y would occur if the treatment was set to t when in fact the treatment was set to control and the outcome did not occur.

The probability that the treatment is *necessary* for y to occur can be defined as

$$PN = P(y'_c | t, y)$$

This is the probability that y would not occur if the treatment was set to control, while in actuality both y occurs and the treatment takes place.

Finally, the probability that the treatment is both necessary and sufficient is defined as

$$PNS = P(y_t, y'_c)$$

and states that y would occur if the treatment took place; and y would not occur if the treatment did not take place. PNS is related with PN and PS as follows:

$$PNS = P(t, y)PN + P(c, y')PS$$

In bounding the above three quantities, we utilize observational data in addition to experimental data. The observational data is characterized in terms of the joint probabilities:

$$P_{TY} = P(t, y), P(c, y), P(t, y'), P(c, y')$$

Read more about probabilities of causation here: <https://causalml.readthedocs.io/en/latest/methodology.html#probabilities-of-causation>

3. Challenges of Uplift Modeling in practice & Conclusion

3.1. Challenges of Uplift Modeling in practice

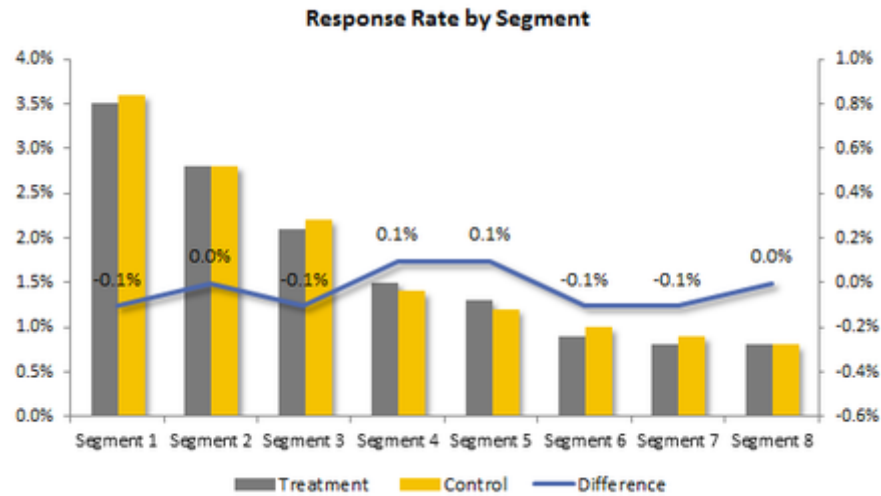
Uplift Modeling, since it's fall to the type of Causal Inference, is hard to setup, as its require deep understanding of experimental testing and randomized controlled trial. Observing causal effect is hard and maybe ambiguous.

Also, in reality, Uplift Modeling is harder to execute than response modeling. While response follows known customer traits (demographics, lifestage, transience, change in circumstances), uplift **can be dependent on variables not commonly used in response modeling**.

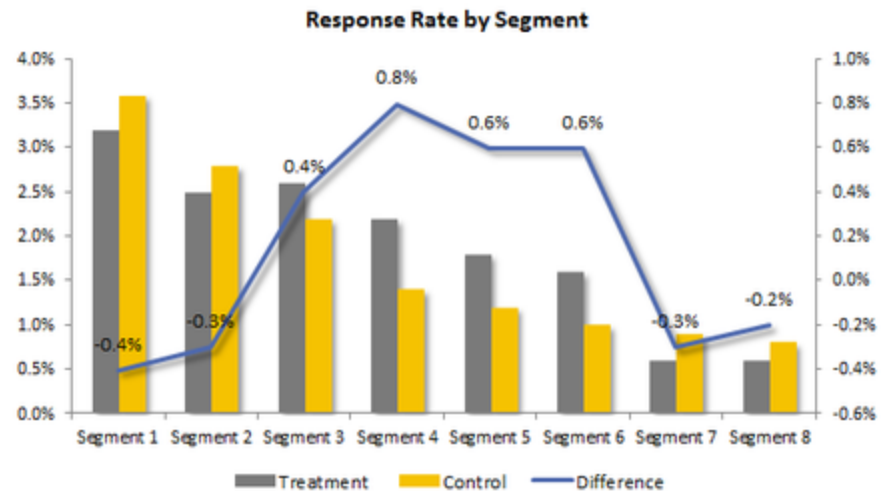
There are many failed attempts to create uplift models, and one common barrier in creating an valid uplift model is ineffective marketing campaign. The dependent variable of the uplift model is the **difference in response** between treatment and control groups. The independent variables can be all sorts of things that impact the effectiveness of the action, from demographics of the segments to variables related to the treatment itself. In order for Uplift Modeling to work, we need to see **variable rates of incremental response for different levels of independent variables**. If there is no variance in response, the model fails because no independent variable impacts the outcome.

Let's consider 2 marketing campaigns as follow:

Campaign 1:



Campaign 2:



Campaign 1 delivers no incremental lift over control. It is a poor campaign, and trying to figure out which factors drive positive change is impossible, because there is no positive change.

Campaign 2 may not be a good fit for some segments, but there are strong positive results in other segments. This suggests that we have an output that not only positive, but has some variability that changes in predictable ways. This campaign is a good candidate for uplift modeling.

Why would a situation where there is no lift over control happen (Campaign 1)? This can happen because:

- The marketing channel we chose is not a good fit with our business.
- The market have already been over-saturated (bão hòa) with our (company) communications.

In large companies, the reliance on direct response gets the market over-saturated to the point of zero marginal sales, so the first attempt at uplift measurement fails miserably.

When faced with a situation like that, the amount of marketing communication (most often – **communication frequency**) **should be taken into account when measuring uplift**, and that's the variable that needs to be optimized first. Unfortunately, most of the time, this variable is not being considered for the analysis or the analytics team has no means to influence this variable.

3.2. Conclusion

Prescriptive analytics, and more specifically uplift modeling, have long been proposed as a paradigm that improves marketing performances. However, there is little empirical evidence of uplift models outperforming predictive models has been presented in the literature. In Vietnam, there is hardly any business knows and applies Uplift Modeling.

However, since Uplift Modeling is a field focuses on experimenting the causal effect, specifically studying customers behavior over different treatments, it is a promising solution for marketers to optimize the marketing ROI, as well as to better understand the effect of their marketing strategy.

4. Demo

4.1. Dataset

The dataset was created by The Criteo AI Lab .The dataset consists of 13M rows, each one representing a user with 12 features, a treatment indicator and 2 binary labels (visits and conversions). Positive labels mean the user visited/converted on the advertiser website during the test period (2 weeks). The global treatment ratio is 84.6%. It is usual that advertisers keep only a small control population as it costs them in potential revenue.

Following is a detailed description of the features:

- f0, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11: *feature values (dense, float)*
- treatment: *treatment group (1 = treated, 0 = control)*
- conversion: *whether a conversion occurred for this user (binary, label)*
- visit: *whether a visit occurred for this user (binary, label)*
- exposure: *treatment effect, whether the user has been effectively exposed (binary)*

4.2. Demo description

In this demo, I will go through several implementations of the Uplift Models, including:

- Meta-learners: T-Learner, S-learner
- Uplift Tree
- Neural Network - the CEVAE

I also plot the Qini curve of each implementation. At the end of the demo, I will compare and analyze the models.

Colab Link: <https://drive.google.com/file/d/1ikF4gBFfe1Tw-qegTGERZm9xc-fEXkJE/view?usp=sharing>

5. References

<http://www.diva-portal.org/smash/get/diva2:1328437/FULLTEXT01.pdf>

<https://matheusfacure.github.io/python-causality-handbook/21-Meta-Learners.html>

https://docs.google.com/presentation/d/1865mkpxUjl9ZQ_03osrWj24tMJr7jiJl/edit?pli=1#slide=id.g17959f4de04_4_8 (Ms. linhna slides)

<https://causalml.readthedocs.io/en/latest/methodology.html#>

<https://pylift.readthedocs.io/en/latest/introduction.html#>

<https://towardsdatascience.com/why-every-marketer-should-consider-uplift-modeling-1090235572ec>

<https://zyabkina.com/challenges-of-uplift-modeling-in-marketing/>

https://johaupt.github.io/blog/Uplift_ITE_summary.html