Personalization at Uber Scale

Via Causal Machine Learning

Okke van der Wal

Machine Learning Manager

okke@uber.com

Uber

Agenda

- Uber in Amsterdam
- Causal ML
  - Introduction
  - Case Study

Uber in Amsterdam

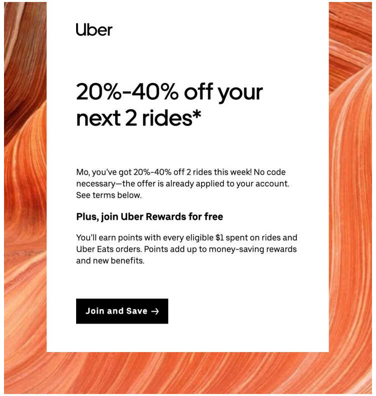
Uber in Amsterdam

- **Amsterdam is our European Headquarters**
- **Recently-opened campus at Tripolis**
- **Two organizations:**
    1. **EMEA Operations**
    2. **Tech Hub for Global Payments & various other smaller teams** ±300 people across data science, engineering & product
- We're hiring! uber.com/careers

The ML team in Amsterdam

Incentives

Optimize incentive spend for attracting users to the platform.

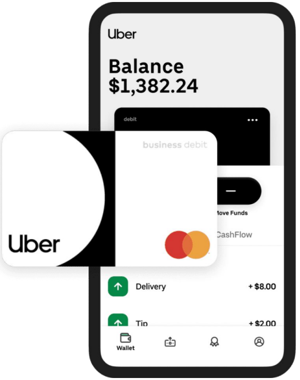


Payments Fraud

Detect fraudulent payment methods on the platform and prevent the risk of future chargebacks.

Financial Products

Apply machine learning to help improve decision-making for our financial products, such as underwriting & offering loans.



CausalML Intro

Classic ML focuses on 'Association', which features correlate with a specific response variable

Classic ML Models

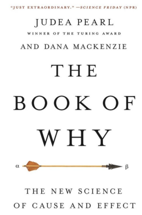
1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

Source: Pearl, J., & Mackenzie, D. (2018)

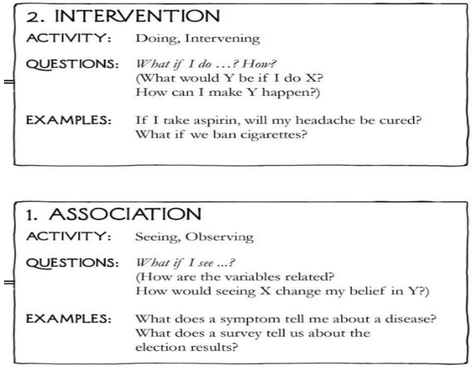


Uber

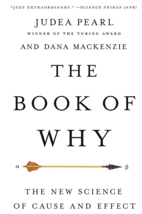
Causal ML focuses on ‘the Intervention’: What happens if we take a certain action

Causal ML Models

Classic ML Models



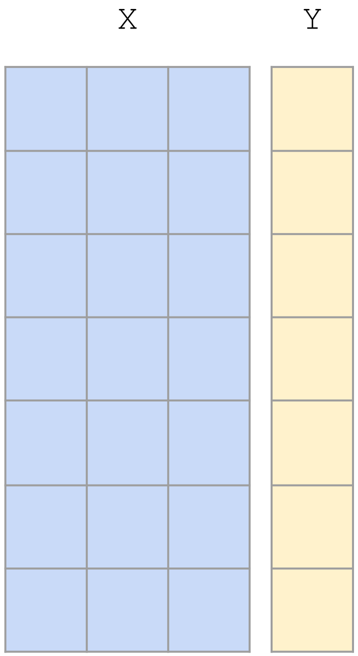
Source: Pearl, J., & Mackenzie, D. (2018)



Uber

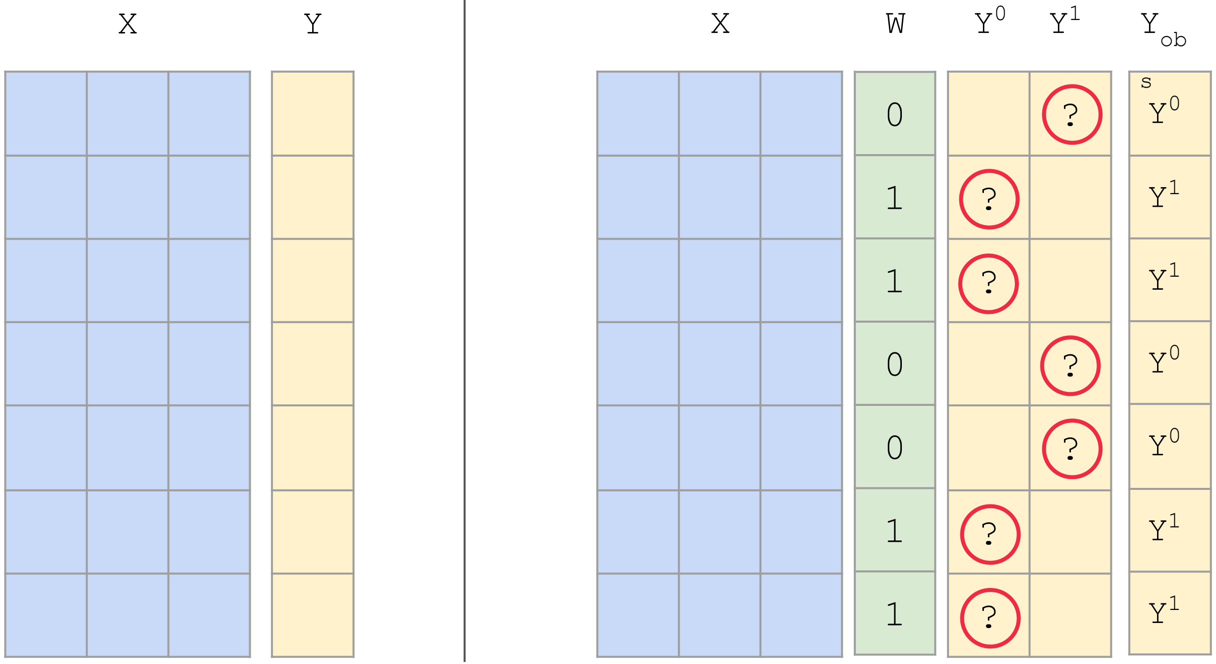
Typical Classic ML Data

In classic machine learning tabular problem settings, the data is structured into two main groups: the features and the response.



Classic ML Dataset Structure

Causal ML Dataset Structure



Predict the counterfactual value

Randomized Experimentation is crucial to effective CausalML

There are three key assumptions that need to be met:

1. Positivity
2. Stable Unit Treatment Value Assumption
3. Common Support

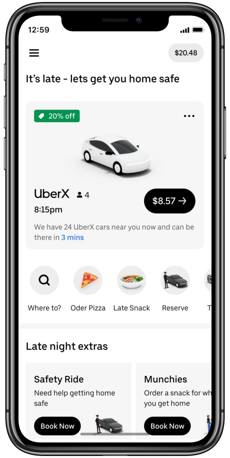
Running A/B experiment allows you to satisfy these assumptions for free!

Experimentation at Uber

“There are over 1,000 experiments running on our in-house platform at any given time.”

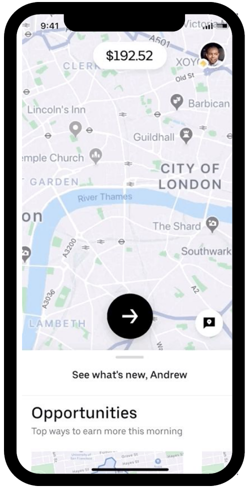
RIDERS

- New home screen



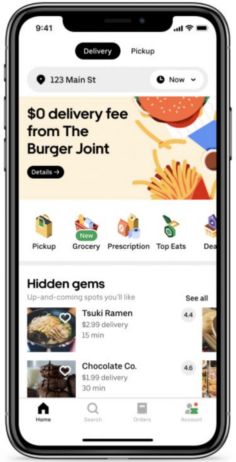
DRIVERS

- New driver experience



EATERS

- New version of the Ranking Algorithm



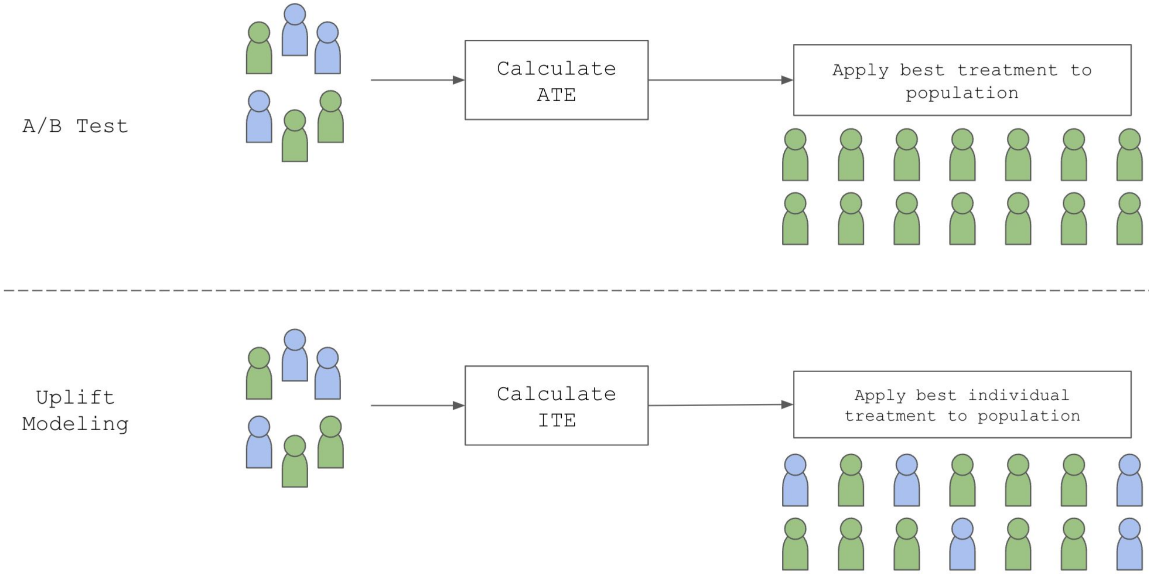
Uber

Modeling the Individual Treatment Effect (ITE) is why Causal ML is important to our Personalization strategy

Causal ML is concerned with the Individual Treatment Effect (ITE).

An A/B Test would assign one treatment to the whole population after the decision making process.

Causal ML can provide personalization and assign the optimal treatment to an unit.



Uber

Case Study at Uber

Uber

20%-40% off your next 2 rides*

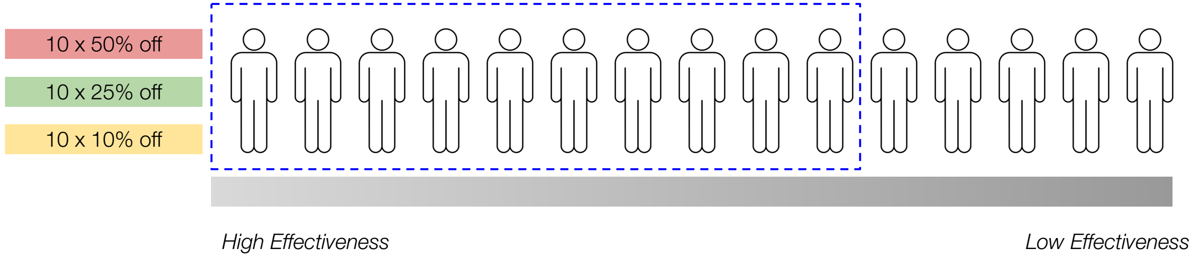
Mo, you've got 20%-40% off 2 rides this week! No code necessary—the offer is already applied to your account. See terms below.

Plus, join Uber Rewards for free

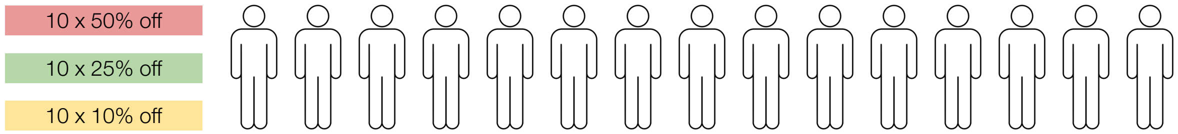
You'll earn points with every eligible $1 spent on rides and Uber Eats orders. Points add up to money-saving rewards and new benefits.

Join and Save →

The goal is to make sure that the ‘right’ promotion is sent to the ‘right’ user at the lowest cost



1. We run an experiment where we randomly give offers to a group of Riders

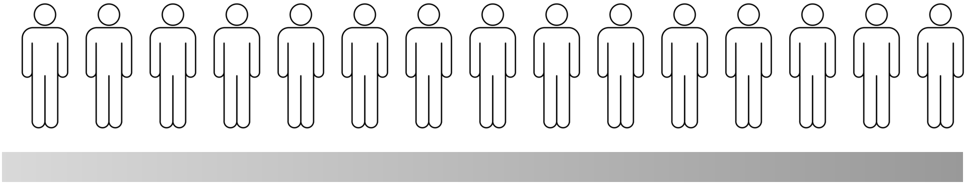


1. Train

Causal learning model is trained based on historical data gathered by targeting riders with random offers (‘Explore’).

For example, 3 offers are available here: 10 x 10%, 25%, 50%.

2. Based on the data from the experiments, combined with a lot of features, we rank the riders based on how likely they are to take trips (ITE)

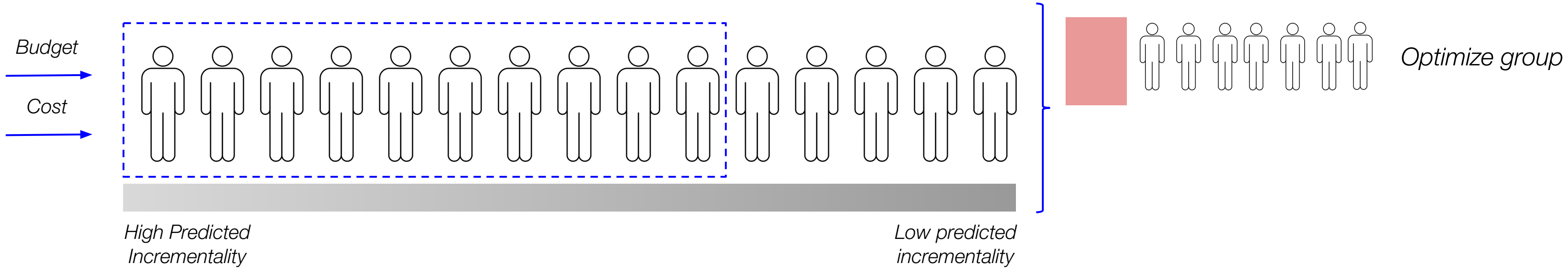


High Predicted Incrementality

Low predicted incrementality

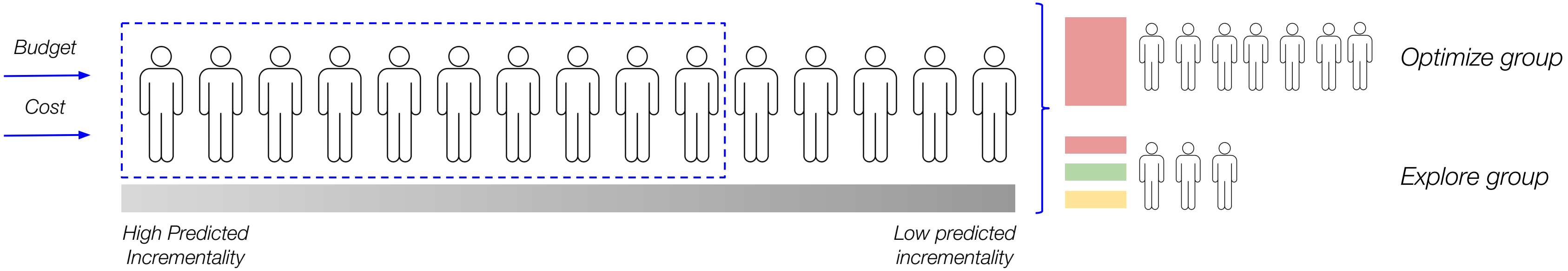
1. Train	**2. Rank**
Causal learning model is trained based on historical data gathered by targeting riders with random offers ('Explore'). For example, 3 offers are available here: 10 x 10%, 25%, 50%.	The model ranks riders by combining the hundreds of rider features to rank riders based on predicted incrementality and looking at the ITE for each of the promotions.

3. The riders with highest incrementality (ITE) are targeted first with the cohort size adjusted, depending on the budget



1. Train	2. Rank	**3. Select Riders**
Causal learning model is trained based on historical data gathered by targeting riders with random offers ('Explore'). For example, 3 offers are available here: 10 x 10%, 25%, 50%.	The model ranks riders by combining the hundreds of rider features to rank riders based on predicted incrementality and looking at the ITE for each of the promotions.	Based on the budget and predicted cost per rider , the model selects the best riders to target to maximize efficiency.

4. Not everyone receives their optimal offer to continue to collect randomized training data, especially if it's a continuously-running campaign.



1. Train	2. Rank	3. Select Riders	**4. Offer Execution**
Causal learning model is trained based on historical data gathered by targeting riders with random offers ('Explore'). For example, 3 offers are available here: 10 x 10%, 25%, 50%.	The model ranks riders by combining the hundreds of rider features to rank riders based on predicted incrementality and looking at the ITE for each of the promotions.	Based on the budget and predicted cost per rider , the model selects the best riders to target to maximize efficiency.	Optimize: Near-optimal offers are selected in the ranking process Explore: Small random group is added each week to test all offers.

‘Under the hood’

Problem Statement

We try to make optimal decisions of **who** should get the promo to **maximize the total incremental revenue (GB)** w.r.t a **promo budget constraint**.

$$\max_{w_i} \sum_{i=1}^{n} \underbrace{\boxed{w_i}}_{\text{Decision variable}} \left[\underbrace{\boxed{Y_i(1) - Y_i(0)}}_{\text{Incremental GB}}\right]$$

Subject to

$$\sum_{i=1}^{n} w_i \left[\underbrace{\boxed{C_i(1) - C_i(0)}}_{\text{Incremental Cost}}\right] \leq \underbrace{\boxed{B}}_{\text{Budget}}$$

Y : Revenue for the rider in experiment period

- Y(1) [Y(0)] → the GB that would be observed if the rider is in treatment [control]

w : treatment decision for the rider, e.g., 1 for treatment, 0 for control

C : cost (promo spend) in experiment period

- C(1) [C(0)] → the Cost that would be observed if the rider is in treatment [control]

ITE Estimation framework

We have two key components...

Spend estimator: $c(x) = E[C|W = 1, X = x]$

GB estimator: $y(x, w) = E[Y|W = w, X = x]$

ITE Estimation framework

We have two key components...

Spend estimator: $c(x) = E[C|W = 1, X = x]$

GB estimator: $y(x, w) = E[Y|W = w, X = x]$

→ Based on these, we then need to estimate a single efficiency metric: **iGB per $ spent (IGBS)**

$$Y_i = \begin{cases} y(X_i, 0) + \tau(X_i)c(X_i) + \epsilon_i & W_i = 1 \\ y(X_i, 1) - \tau(X_i)c(X_i) + \epsilon_i & W_i = 0 \end{cases}$$

Observed GB

Predicted Counterfactual GB

IGB = IGBS * Spend

Treatment/Control indicator

Uber

ITE Estimation framework

We have two key components...

Spend estimator: $c(x) = E[C|W=1, X=x]$

GB estimator: $y(x,w) = E[Y|W=w, X=x]$

→ Based on these, we then need to estimate a single efficiency metric: **iGB per $ spent (IGBS)**

$$\underbrace{Y_i}_{\substack{\text{Observed}\\ \text{GB}}} = \begin{cases} y(X_i,0) + \tau(X_i)c(X_i) + \epsilon_i & W_i = 1 \\ \underbrace{y(X_i,1)}_{\substack{\text{Predicted}\\ \text{Counterfactual}\\ \text{GB}}} - \underbrace{\tau(X_i)c(X_i)}_{\substack{\text{IGB}\\ = \text{IGBS * Spend}}} + \epsilon_i & \underbrace{W_i = 0}_{\substack{\text{Treatment/Control}\\ \text{indicator}}} \end{cases}$$

Task:

Need to learn estimators for:

- *Spend*
- *GB in treatment/control*
- *IGBS*

Challenges:

1. Only **one** of [Y(1), Y(0)] can be observed
2. How to do this in an effective and scalable manner since all estimators are interdependent?

Challenge: Scalable Modeling Solution

$(X_i, W_i, Y_i, C_i) \rightarrow$

$c(x,1)$

$$\hat{L}(c) = \sum_{\{i:W_i=1\}} (C_i - c(X_i,1))^2$$

Spend

$y(x,1)$

$$\hat{L}(y(1)) = \sum_{\{i:W_i=1\}} (Y_i - y(X_i,1))^2$$

Treatment GB

$y(x,0)$

$$\hat{L}(y(0)) = \sum_{\{i:W_i=0\}} (Y_i - y(X_i,0))^2$$

Control GB

$\tau(x)$

$$\hat{L}(\tau) = \hat{L}(\tau,0) + \hat{L}(\tau,1)$$

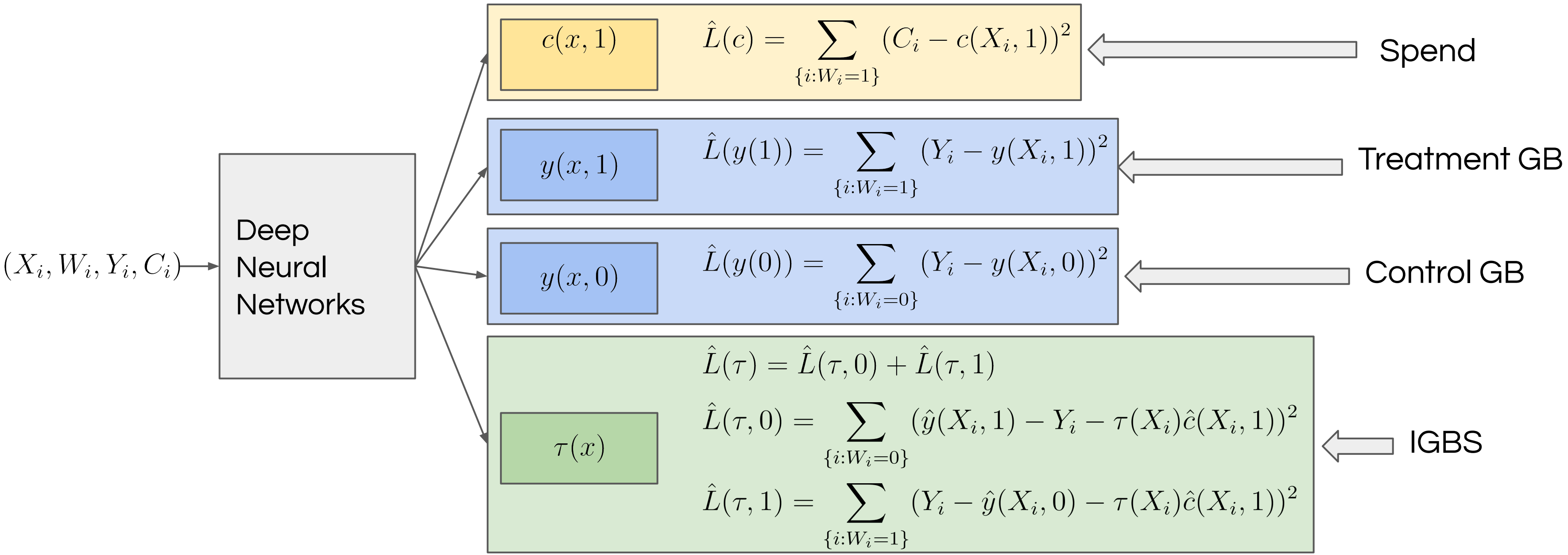
$$\hat{L}(\tau,0) = \sum_{\{i:W_i=0\}} (\hat{y}(X_i,1) - Y_i - \tau(X_i)\hat{c}(X_i,1))^2$$

$$\hat{L}(\tau,1) = \sum_{\{i:W_i=1\}} (Y_i - \hat{y}(X_i,0) - \tau(X_i)\hat{c}(X_i,1))^2$$

IGBS

Solution: Neural Networks

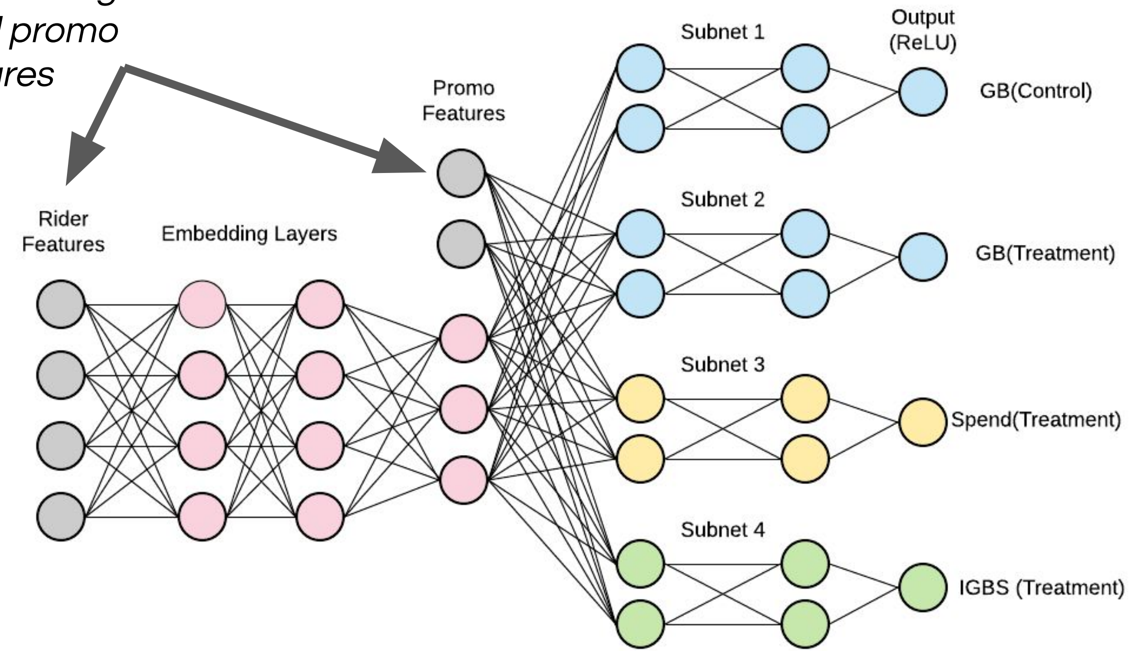
The goal is to jointly learn the estimators for spend, GB, and IGBS, something which Neural Networks excel at.



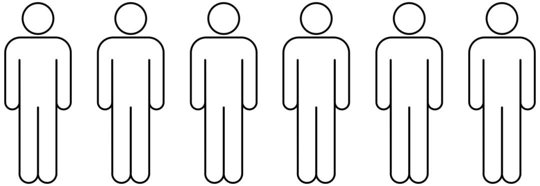
Scalable DL Modeling Solution - MTY

Inspired by the joint-learning idea from [Y-learner], we end up with the following **Multitask Y-learner (MTY)** architecture.

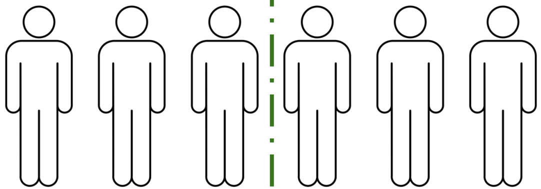
Separate handling of rider and promo features



Coming back to our initial user base….



Based on the budget for the campaign, and possibly our 'minimum efficiency target value', we can choose who we target.



IGBS	3.5	3.0	2.7	2.0	1.6	1.1
Spend	1.5	3.3	2.2	0.4	1.1	0.1

$$B = 7 \Rightarrow \gamma_B = 2.5$$

Plug
CausalML

CausalML

CausalML is a Python package, developed at Uber and made available open-source, that provides a suite of **uplift modeling** methods using machine learning & deep learning algorithms.

Uplift Trees

- KL Divergence
- Chi-Square
- Euclidean Distance
- Contextual Treatment Selection

Meta Learners

- R-Learner
- X-Learner
- T-Learner
- S-Learner

Neural Networks

- DragonNet
- CEVAE

Instrumental Variables

- 2SLS

Thank you!

Questions?

okke@uber.com