Heterogeneous Treatment Effects

Michael Thomas

Recap of Experiments

- Previously we saw that experiments can be used to estimate **average treatment effects**:

$$\text{ATE}=\mathbb{E}[Y_i(W_i=1)-Y_i(W_i=0)]$$

- However, much our focus in this class has been on how to prioritize customers for targeting.
- Ultimately, we are interested in which customers would be most likely to respond to our marketing campaign.
  - **Average treatment effects** do not help us with this.
  - This requires estimates of **heterogeneous treatment effects**

Estimating Average Treatment Effects

- We estimated average treatment effects with:

$$Y_i = \beta_0 + \tau W_i + \varepsilon_i$$

- Where $\hat{\tau}$ provided the estimated ATE
- Alternatively, we could run separate regressions for different population groups.
- For example, we could estimate the average response for men ($\hat{\tau}_M$) by running a regression on just the men in the data.
- Similarly, we could estimate $\hat{\tau}_W$ with a regression on just women.

Conditional Average Treatment Effects

- Another way to estimate separate effects for men and women: interact gender with the treatment:

$$Y_i = \beta_0 + \beta_m x_i + \delta_0 W_i + \delta_m x_i W_i + \varepsilon_i$$

- where x_i is a dummy equal to one for males.
- This allows us to estimate the **conditional average treatment effect** (CATE):

$$\begin{aligned}\tau_i &= \mathbb{E}\left[Y_i\left(W_i = 1\right) - Y_i\left(W_i = 0\right)\middle|\mathbf{x_i}\right] \\ &= \mathbb{E}\left[Y_i|\mathbf{x_i}, W_i = 1\right] - \mathbb{E}\left[Y_i|\mathbf{x_i}, W_i = 0\right]\end{aligned}$$

Conditional Average Treatment Effects

- More generally, we could estimate the effect of many x values:

$$Y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \delta_0 W_i + \sum_{k=1}^{p} \delta_k (x_{ik} W_i) + \varepsilon_i$$

- This involves interacting all of the x_k variables with the treatment assignment, W_i.
- Notice that to correctly specify this regression, you need to include all of the x_{ik} terms without interacting them with W_i.
- More generally, you could think about estimating

$$Y_i = \beta_0 + f(x_i) + \delta_0 W_i + g(W_i, x_i) + \varepsilon_i$$

- where $f()$ and $g()$ are estimated using something like trees, or forests...

Conditional Average Treatment Effects

- More generally, we could estimate the effect of many x values:

$$Y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \delta_0 W_i + \sum_{k=1}^{p} \delta_k (x_{ik} W_i) + \varepsilon_i$$

- This involves interacting all of the x_k variables with the treatment assignment, W_i.
- Notice that to correctly specify this regression, you need to include all of the x_{ik} terms without interacting them with W_i.
- More generally, you could think about estimating

$$Y_i = \beta_0 + f(x_i) + \delta_0 W_i + g(W_i, x_i) + \varepsilon_i$$

- where $f()$ and $g()$ are estimated using something like trees, or forests...

Causal Forest

- Recently introduced method in the machine learning/treatment effects literature
  - Wager and Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests" (Journal of the American Statistical Association)
- Causal forest extends the random forest algorithm to directly predict heterogenous treatment effects.
- Potential benefits over regression approach with treatment-interactions:
  - Non-parametric method, designed to automatically detect non-linear relationships between the inputs and the treatment effects and interactions between the inputs.
  - The causal forest is directly trained to predict the CATE, τ_i, not the outcome level, Y_i.

Causal Forest

- Like random forest, causal forest is based on an ensemble of trees.
- Ideally, the leaves of the trees include (almost) homogeneous units i and the treatment assignment in each leaf is random or as good as random.
- In a standard regression tree, the predicted outcome for observation i in leaf l is given by:

$$\hat{Y}_i = \hat{Y}_{R_l} = \frac{1}{N_l} \sum_{j \in R_l} Y_j$$

Causal Forest Predictions

- In a causal tree the predicted CATE for observation i in leaf l is:

$$\hat{\tau}_i = \hat{\tau}_{R_l} = \frac{1}{N_{1l}} \sum_{j \in R_{1l}} Y_j - \frac{1}{N_{0l}} \sum_{j \in R_{0l}} Y_j$$

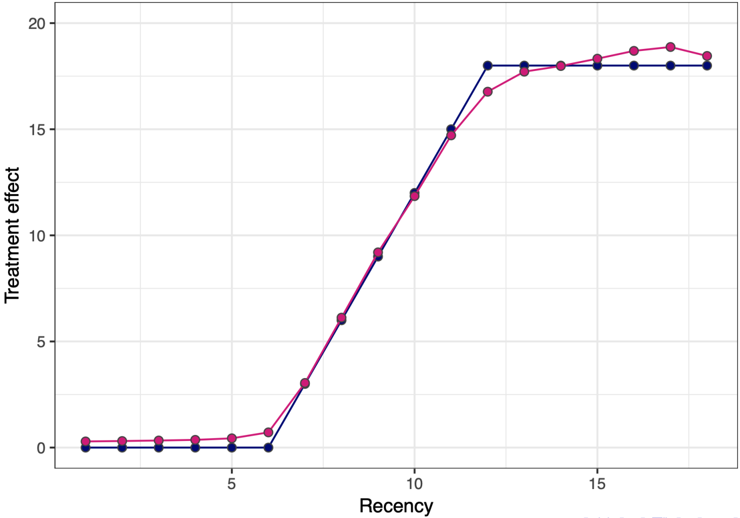
- R_{1l} includes all observations j in leaf l with treatment status $W_j = 1$.
- R_{0l} includes all observations j in leaf l with treatment status $W_j = 0$.
- The predicted CATE is based on the difference in the mean outcomes of the treated and untreated units.
- This can produce an unbiased estimate of the true CATE, τ_i.

Random Forest Illustration

- Imagine that we have a company with both an online and offline presence.
- We ran an experiment in which customers were randomly targeted with emails.
- In reality, customers respond to emails as follows based on how many months have passed since they made a purchase ("recency")
  - If recency is less than 6 months, they do not respond.
  - If recency is between 6 and 12 months, their response increases in recency.
  - The response for offline customers is four times the response for online customers.

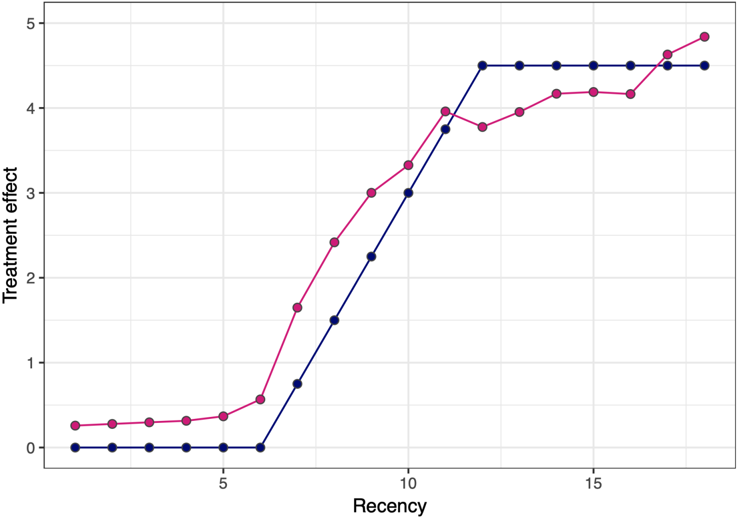
Random Forest Illustration

- Actual and estimated τ for customers who buy offline (segment with large CATE):



Random Forest Capabilities

- Actual and estimated τ for customers who buy online (segment with small CATE):



Model validation: Standard Approach

1. Estimate model in training sample.
2. Use the model to score the customers in the holdout sample.
3. For each customer, calculate the profit we earn from them after removing the cost of our marketing.

Model validation: Heterogenous Treatment Effects

1. Estimate model in training sample.
2. Use the model to score the customers in the holdout sample using the estimated CATE.
3. Divide customers into groups and calculate the incremental profit we obtain from spending advertising on them.

Practice Problem

When we prioritize customers for marketing we ideally would like to prioritize them based on our predictions of:

- How much they are likely to buy.
- How likely they are to buy.
- How likely they are to respond to our marketing instrument.
- Ans: How much incremental profit our marketing instrument will generate from them.

Practice Problem

Which of the following is *false* about heterogeneous treatment effects:

- They can be estimated using an experiment.
- They can be estimated using machine learning methods.
- They can be estimating using standard regression.
- Ans: They are not typically as useful as average treatment effects.

Practice Problem

A benefit of using Causal Forest is

- Ans: Obtaining non-parametric estimates of heterogeneous treatment effects.
- It does not require experimental data.
- It runs faster than Random Forest.
- It does not require as much data as other approaches.