Lec12: Causal Machine Learning I

Isidoro Garcia Urquieta

2021

Agenda

- Omitted Variable Bias
- Double Selections LASSO
- Balancing Scores: `e(x)`, IPW, AIPW
- Double Debiased ML
- Residual Balancing
- Causal Trees
- Double Debiased ML for HTE
- Causal forests
- Generalized Random Forests

Set up

Para poder hablar de Causal Machine Learning tenemos que volver brevemente a los basics de inferencia causal

Queremos estimar τ de manera insesgada:

$$Y_i = X\beta + \tau T_i + \epsilon_i$$

Donde: - Y_i es la variable sobre la que queremos evaluar el impacto

- X es una matriz (de alta dimensionalidad) de controles
- T_i es el tratamiento sobre el que nos gustaría evaluar el impacto sobre Y_i ceteris paribus
  - Noten como T_i puede ser dicotómico, categórico o continuo.
- ϵ_i es ruido blanco

Como combinar Inferencia Causal con Machine Learning?

De manera general, la inferencia causal trata de **inferir** sobre **derivadas** o cambios en la métrica Y_i en lugar de la Y_i misma (el nivel)

$$\tau_i = \frac{\partial Y_i}{\partial T_i}$$

Por otro lado, los modelos de Machine Learning que vimos buscan **predecir** Y_i fuera de la muestra:

$$\hat{Y}_i = f(\hat{X}_i) + \epsilon_i$$

Son problemas fundamentalmente distintos!

Un grupo de autores: (Christian Hansen, Susan Athey, Guido Imbens, Stephan Wager, Belloni, Victor Chernozshukov, Matt Taddy, Esther Duflo, Robert Tibshirani) se pusieron a pensar cómo aprovechar ML y aplicarlo a **predecir e inferir impactos causales (derivadas)**

Omitted Variable Bias

Recuerden lo que origina el sesgo en la estimación de τ. Imaginen que tenemos lo siguiente:

- Modelo real: $Y_i = \beta_0 + \beta_1 X_{1i} + \tau T_i + \epsilon_i$
- Modelo estimado: $Y_i = \beta_0 + \tau T_i + \psi_i$
- $\psi_i = \epsilon_i + \beta_1 X_{1i}$

Veamos la $\hat{\tau}$ del modelo estimado:

$$\hat{\tau} = \frac{\Sigma_i^N (T_i - \bar{T})(Y_i)}{\Sigma_i^N (T_i - \bar{T})^2}$$

Sustituyo Y_i

$$\hat{\tau} = \frac{\Sigma_i^N (T_i - \bar{T})(\beta_0 + \tau T_i + \psi_i)}{\Sigma_i^N (T_i - \bar{T})^2}$$

Omitted Variable Bias II

Distribuyo los términos:

$$\hat{\tau} = \frac{\sum_i^N (T_i - \bar{T})\beta_0}{\sum_i^N (T_i - \bar{T})^2} + \frac{\sum_i^N (T_i - \bar{T})\tau T_i}{\sum_i^N (T_i - \bar{T})^2} + \frac{\sum_i^N (T_i - \bar{T})\psi_i}{\sum_i^N (T_i - \bar{T})^2}$$

$$\hat{\tau} = \beta_0 \cancelto{0}{\frac{\sum_i^N (T_i - \bar{T})}{\sum_i^N (T_i - \bar{T})^2}} + \tau \cancelto{1}{\frac{\sum_i^N (T_i - \bar{T}) T_i}{\sum_i^N (T_i - \bar{T})^2}} + \frac{\sum_i^N (T_i - \bar{T})\psi_i}{\sum_i^N (T_i - \bar{T})^2}$$

$$\hat{\tau} = \tau + \frac{\sum_i^N (T_i - \bar{T})\psi_i}{\sum_i^N (T_i - \bar{T})^2}$$

Sustiyo $\psi_i = \beta_1 X_{1i} + \epsilon_i$

$$\hat{\tau} = \tau + \frac{\sum_i^N (T_i - \bar{T})(\beta_1 X_{1i})}{\sum_i^N (T_i - \bar{T})^2} = \tau + \beta_1 \frac{\sum_i^N (T_i - \bar{T})(X_{1i})}{\sum_i^N (T_i - \bar{T})^2}$$

Noten como $\frac{\sum_i^N (T_i - \bar{T})(X_{1i})}{\sum_i^N (T_i - \bar{T})^2}$ es el coeficiente de la regresión $X_{1i} = \delta_o + \delta_1 T_i$.

Omitted Variable Bias III

Por lo tanto:

$$\hat{\tau} = \tau + \beta_1\delta_1$$

El omitted variable bias es el producto de:

- β_1: La relevancia de la variable omitida X_{1i} sobre Y_i
- δ_1: La relevancia de la variable omitida X_{1i} sobre T_i

Si la variable omitida X_{1i} no tiene relación con el tratamiento T_i (δ_1=0), el sesgo es cero y el estimador es causal.

De igual manera, si la variable omitida X_{1i} no tiene relación con Y_i, el sesgo es cero y el estimador es causal

Double LASSO

Los primeros intentos de usar Machine Learning en inferencia causal involucraron LASSO. Veamos cómo:

- Si usamos LASSO sobre $Y_i = X\beta + \tau T_i + \epsilon_i$ vamos a, en principio, a dejar sólo variables que tengan $\beta \neq 0$.

Double LASSO

Los primeros intentos de usar Machine Learning en inferencia causal involucraron LASSO. Veamos cómo:

- Si usamos LASSO sobre $Y_i = X\beta + \tau T_i + \epsilon_i$ vamos a, en principio, a dejar sólo variables que tengan $\beta \neq 0$.
- El problema es que si usamos este LASSO para hacer inferencia causal seguimos con δ_1 grande.

Double LASSO

Los primeros intentos de usar Machine Learning en inferencia causal involucraron LASSO. Veamos cómo:

- Si usamos LASSO sobre $Y_i = X\beta + \tau T_i + \epsilon_i$ vamos a, en principio, a dejar sólo variables que tengan $\beta \neq 0$.
- El problema es que si usamos este LASSO para hacer inferencia causal seguimos con δ_1 grande.
- Belloni, Chernoszhukov y Hansen (2014) argumentaron lo siguiente:

Double LASSO

Los primeros intentos de usar Machine Learning en inferencia causal involucraron LASSO. Veamos cómo:

- Si usamos LASSO sobre $Y_i = X\beta + \tau T_i + \epsilon_i$ vamos a, en principio, a dejar sólo variables que tengan $\beta \neq 0$.
- El problema es que si usamos este LASSO para hacer inferencia causal seguimos con δ_1 grande.
- Belloni, Chernoszhukov y Hansen (2014) argumentaron lo siguiente:
- X puede ser una matriz de alta dimensionalidad por dos razones: 1) Ahora tenemos mucha información (Big data!) y 2) Para hacer inferencia causal, el investigador *necesita* estimar la forma funcional perfectamente $f(X)$