Seminario 27/10

De Predecir a Inferencia Causal

- ML es usado cuando queremos predecir un valor, es decir cuando mapeamos de una entrada conocida X a una salida desconocida pero bien definida (Diagnósticos médicos, detección de objetos, ventas en el próximo mes).
- En ML lo que hacemos es estimar $E[Y|X]$
- Es un proceso “pasivo”. No hacemos nada para cambiar el resultado(Tratamiento)

De Predecir a Inferencia Causal

- ¿Qué pasa si nos planteamos $E[Y|X,T]$?
- Tanto X como T impactan en Y, la diferencia está en que a X no la podemos controlar pero a T si.
- La inferencia causal es el proceso de estimar la relación causal entre T y Y bajo el contexto de X.
- Si sabemos esto podemos plantear $\underset{T}{argmax}\ E[Y|X,T]$ para optimizar Y

De predecir a Inferencia Causal

Supongamos que ahora queremos predecir el salario de un individuo, también tenemos la posibilidad de afectar en la educación del individuo. Podemos plantear:

$$\underset{Educ}{argmax}\ E[Income|X, Educ]$$

Por lo que podemos definir dado el individuo cuánto gastar en la educación.

Ejemplo yo soy malísimo dibujando, por más que invierta mucho dinero en clases de dibujo no podría vivir de dibujante de comics :(

De ATE a CATE

Average Treatment Effect(ATE): $E[Y_1 - Y_0]$

Conditional Average Treatment Effect(CATE): $E[Y_1 - Y_0|X]$

En caso de que el tratamiento sea continuo ...

Average Treatment Effect (ATE): $E[y'(t)]$

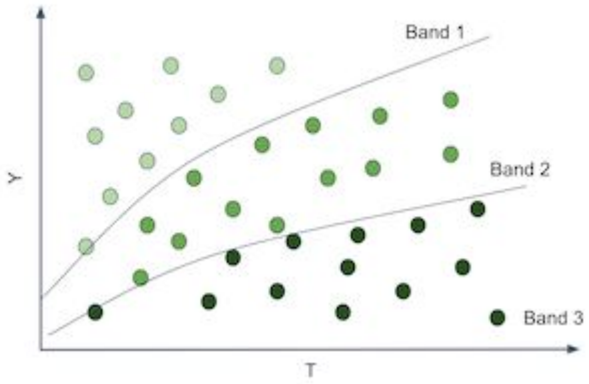
Conditional Average Treatment Effect (CATE): $E[y'(t)|X]$

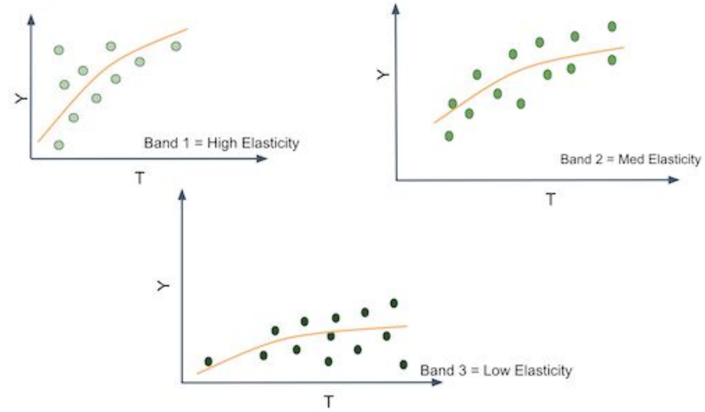
Nota: *Ambas derivadas son respecto a T*

De ATE a CATE

Imaginemos que somos una empresa que quiere aumentar las ventas, tenemos un grupo de clientes y podemos tomar ciertas medidas (bajar los precios, descuentos, cuotas sin interés, …) para fomentar las ventas. Ahora vayamos un paso más e imaginemos que podemos agrupar a nuestros clientes en función de cómo reaccionan al tratamiento, por ejemplo en 3 grupos donde el primero, es en el que pequeños cambios en el tratamiento cambia mucho la salida (Alta elasticidad), y el tercero es el que aunque se hagan en el tratamiento poco va a variar la salida (Baja elasticidad).

De ATE a CATE


Band 1
Band 2
Band 3
Y
T


Band 1 = High Elasticity
Band 2 = Med Elasticity
Band 3 = Low Elasticity
Y
T

De ATE a CATE

Ahora dejamos de querer predecir Y y pasos a querer predecir la derivada de Y respecto a T, $\frac{\delta Y}{\delta T}$ para cada unidad.

Ejemplo motivacional ...

Tenemos una heladería, Y es la venta de helados, T es el precio del helado y cada unidad es el día i. Tenemos la posibilidad de cada día cambiar el precio del helado. Entonces si encontramos los días que $\frac{\delta Sales}{\delta Price}$ es bajo podemos incrementar los precios sin perder mucho las ventas ese día.

De ATE a CATE

No todo es tan fácil ...

¿Cómo puedo predecir la elasticidad $\frac{\delta Sales}{\delta Price}$ si no puedo verla? La elasticidad es esencialmente no observable a nivel de unidad.

Para ver las pendientes individuales, tendríamos que observar cada día bajo dos precios diferentes y calcular cómo cambian las ventas para cada uno de esos precios

$$\frac{\delta Y_i}{\delta T_i} \approx \frac{Y(T_i) - Y(T_i + \epsilon)}{T_i - (T_i + \epsilon)}$$

Otra vez el problema fundamental de la inferencia causal: No podemos ver nunca la misma unidad bajo diferentes condiciones de tratamiento

Prediciendo la elasticidad

¿Cuál es el problema?

Llegamos a que tenemos que predecir $\frac{\delta Y_i}{\delta T_i}$ pero vimos que esto es no observable, por lo que no podemos usar un algoritmo de machine learning y poner esto como target.

Pero tal vez no tenemos que observar a $\frac{\delta Y_i}{\delta T_i}$ para predecirla …

Prediciendo la elasticidad

Planteemos el siguiente modelo lineal:

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + e_i \tag{1}$$

Derivando …

$$\frac{\delta y_i}{\delta t_i} = \beta_1 \tag{2}$$

Y dado que podemos estimar (1) podemos llegar a un $\hat{\beta}_1$

… Entonces ¿Ya está?

Prediciendo la elasticidad

No todavía no está ...

Lo que predijimos fue un $\hat{\beta}_1$ constante para todos los individuos, es decir, predijimos el ATE. Planteemos ahora el siguiente modelo.

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \beta_3 t_i X_i + e_i \quad \textbf{(3)}$$

Que nos termina dando la siguiente predicción:

$$\widehat{\frac{\delta y_i}{\delta t_i}} = \hat{\beta}_1 + \hat{\beta}_3 X_i$$

Donde es un vector de coeficientes para las propiedades de X. Por lo que ahora si cada Xi va a tener una predicción distinta para la elasticidad. Es decir con (3) pudimos estimar el CATE

Evaluando los Modelos Causales

- ¿Cómo sabemos si el modelo que presentamos en la sección anterior es un buen modelo?

  El problema está en que queremos predecir algo que no podemos observar.

  En ML se usa el paradigma de dividir en train-test, donde comparamos los valores predecidos con el modelo en test que se obtuvo en train con los verdaderos valores del test, pero acá no sabemos cuales son los verdaderos valores.

Evaluando los Modelos Causales

Utilizaremos datos no aleatorios para estimar nuestros modelos causales y datos aleatorios para evaluarlos.