Homework 2 & 3

Question 1

- Estimate a prediction model of BuildYear using

  - 80% of "Example.csv" (training data)

  - Stacking model including OLS and RandomForest

- Evaluate it's performance using the remainings 20% of "Example.csv" (test data)

Question 2

- Estimate summary parameters of $\tau(X) = E[Size|Reform = 1, X] - E[Size|Reform = 0, X]$ where $X = [Distance, BuildYear]$ as

  - $E[\tau(X)]$

  - $\tau(X) \simeq \beta_0 + \beta_1 \times Distance + \beta_2 \times BuildYear$

- Use

  - All sample of of "Example.csv"

  - Double Machine Learning (can use DoubleML)

  - Random forest to estimate nuisance functions

Question 3

- Estimate $\tau(X) = E[Size|Reform = 1, X] - E[Size|Reform = 0, X]$ using

  - grf package

  - all sample of of "Example.csv"

Question 4

- Estimate $E[\tau(x)|s(x) \leq median(s(x))]$ as follows;

    - $s(x)$ is a "signal" of conditional average difference

    - estimate $s(x)$ with grf and 50% of data

    - given estimated $s(x)$, estimate $E[\tau(x)|s(x) \leq median(s(x))]$ with AIPW and remaining 50% of data

Qeustion 5

- In generally,

    - When $\tau(x)$ can be interpreted as the conditional average effect?

    - Why AIPW estimator may converge even with slower convergence of estiamted nuisance functions?

Details

- Submit script-file (Code for Question 1-4, and comment for Question 5) on ITC-LMS

- DeadLine: Feb 6