Prediction and inference

Seminar Data Science for Economics

Madina Kurmangaliyeva
m.kurmangaliyeva@uvt.nl

Spring 2021

Tilburg University

https://madina-k.github.io/dse_mk2021

Prediction vs. (causal) inference

Causal inference

Example:

What is the effect of years of schooling on income (at some adult age)?

Causal inference

1 more year of schooling → Income (Y)

Causal inference

1 more year of schooling → Income (Y)

$$Y = \beta s + g(X) + \epsilon$$

Causal inference

1 more year of schooling → Income (Y)

$$Y = \beta s + g(X) + \epsilon$$

We are interested in estimating $\hat{\beta}$

Prediction

Example:

What is the predicted income of someone with a certain years of schooling and other characteristics X?

Prediction

$$Y = f(s, X) + \epsilon$$

We are interested in predicting $\hat{Y}$ given s and X

Prediction

$$Y = f(s, X) + \epsilon$$

We are interested in finding best possible $\hat{f}$ that maps s and X into Y

Prediction vs. Inference

$$\boxed{Y} = f(s, X) + \epsilon$$

$$Y = \boxed{\beta} s + g(X) + \epsilon$$

Don't care about the functional form as long as it delivers best predictions

Do care about one particular parameter

Prediction vs. Inference

$$Y = f(s, X) + \epsilon$$

$$Y = \beta s + g(X) + \epsilon$$

Want prediction errors $\hat{\epsilon}$ to be small in expectation (i.e., precisely predicted C)

Want expected standard errors of $\hat{\beta}$ to be small (i.e., precisely estimated)

And $E(\hat{\beta}) = \beta$

Prediction vs. Inference

$$Y = f(s, X) + \epsilon$$

$$Y = \beta s + g(X) + \epsilon$$

Use statistical learning tools based on cross-validation

Use inference tools

Prediction vs. Inference

$$Y = f(s, X) + \epsilon$$

$$Y = \beta s + g(X) + \epsilon$$

Ground truth Y is known

Ground truth β is unknown

Prediction vs. Inference

$$Y = f(s, X) + \epsilon$$

$$Y = \beta s + g(X) + \epsilon$$

Ground truth Y is known

Do not need complicated theory because can always train on one sample and test how it performs out-of-sample

Ground truth β is unknown

Need asymptotic theory, requires quasi-random variation in s

$1mln question: Can you estimate $\widehat{\beta}$ by taking differences in predicted Y at some s and $s-1$?

- train the prediction model (e.g., neural net) $\widehat{f}(s, X)$
- It predicts that $\widehat{f}(s=11, X) = €2,000$ per month, same as $\widehat{f}(s=12, X) = €2,000$ per month.
- Does it mean that one more year of education for someone with $s=11$ and X brings no increase in income?

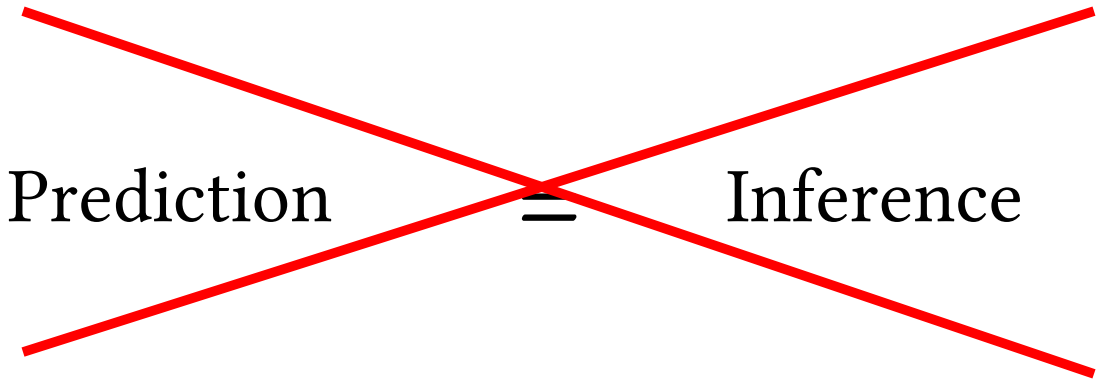
$1mln question: Can you estimate $\hat{\beta}$ by taking differences in predicted Y at some s and $s-1$?

- train the prediction model (e.g., neural net) $\hat{f}(s, X)$
- It predicts that $\hat{f}(s = 11, X) = €2,000$ per month, same as $\hat{f}(s = 12, X) = €2,000$ per month.
- Does it mean that one more year of education for someone with $s = 11$ and X brings no increase in income?

Short answer: No

https://madina-k.github.io/dse_mk2021

Prediction = Inference


Prediction
=
Inference

But can prediction tools be usefull for inference tasks?

But can prediction tools be usefull for inference tasks?

Short answer: Yes