Computational Methods and Causal Analysis

POLSCI 689L ▪ Fall 2026

Jiawei Fu
Department of Political Science ▪ Duke University

Course overview

1. **Machine learning.** Learn unknown functions for prediction.
2. **Causal inference with ML I.** DML and TMLE.
3. **Causal inference with ML II.** Heterogeneous treatment effects.
4. **Deep learning.** Represent unstructured data: text, images, video, and networks.
5. **Causal inference with unstructured data.** Proximal inference; latent outcomes, treatments, and confounders.
6. **Syllabus.** Grading, midterm, final project, recommended textbooks

Logistics

CLASS

Tuesday + Thursday

3:05–4:20 pm ▪ Languages 114B

LAB

Occasionally, Friday
TA: Trung-Anh Nguyen

Languages 114B

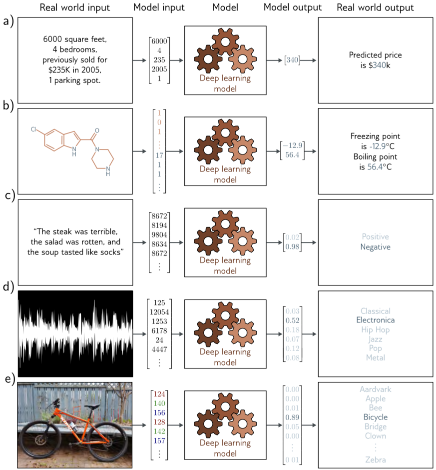
OFFICE HOURS

By appointment

Gross 294A / online ▪ book here

Statistical learning estimates an unknown function

- We observe training examples (X_i, Y_i).
- We estimate $\widehat{f}$ from those labeled examples.
- We use $\widehat{f}(X_{\text{new}})$ to predict the outcome for a new input.



Statistical learning estimates an unknown function

Suppose we observe a response Y and p predictors $X = (X_1, \ldots, X_p)$. A general model is

$$Y = f(X) + \epsilon, \qquad f(X) = \mathbb{E}[Y \mid X].$$

- The function f captures the systematic relationship between X and Y.
- The error ϵ captures variation not explained by X.
- Statistical learning consists of methods for estimating the unknown function f.

OBSERVE

Examples (X_i, Y_i)

Background information paired with an outcome.

LEARN

An estimate $\widehat{f}$

Use data to approximate the unknown relationship.

PREDICT

$\widehat{f}(X_{\text{new}})$

Apply the learned rule to a case not used to fit it.

Flexibility trades bias for variance

VARIANCE

Variance measures how much $\widehat{f}$ would change if it were estimated using a different training set. More flexible methods generally have higher variance.

BIAS

Bias is the error introduced by approximating a complicated relationship with a simpler model. More flexible methods generally have lower bias.

Key idea

As flexibility increases, variance tends to rise while bias tends to fall.

The goal is to generalize, not memorize

TOO SIMPLE

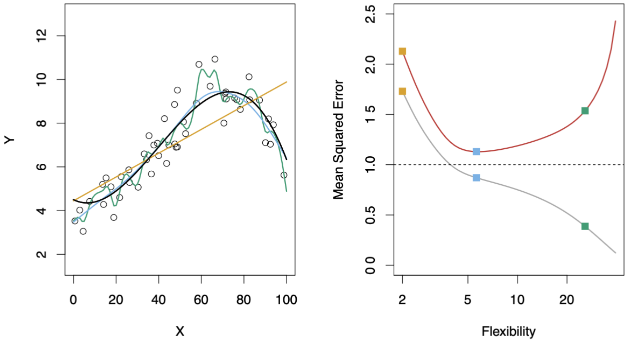
High bias

The rule is too rigid to capture real structure.

TOO SENSITIVE

High variance

The rule follows noise and changes too much across samples.



Key idea

We will learn how validation, tuning, and regularization manage this trade-off.

ML methods in this course

Penalized regression

Ridge, LASSO, adaptive LASSO, and elastic net control complexity in high dimensions.

Nonparametric regression

Kernels, local polynomials, and splines learn flexible smooth relationships.

Trees and ensembles

CART, random forests, and boosting learn nonlinearities and interactions.

Evaluation and tuning

Validation, test sets, and cross-validation assess out-of-sample performance.

Key idea

We first learn these as prediction tools—then reuse them inside causal estimators.

Prediction is not causation

PREDICTION

Who will vote?

Learn patterns in observed turnout and assess accuracy on new people.

$\neq$

CAUSATION

Would a message increase turnout?

Compare the outcome under two interventions for the same target population.

Key idea

Prediction describes the observed world; causation requires a counterfactual and identifying assumptions.

Identification reveals nuisance functions

If consistency, conditional exchangeability, and overlap identify the effect,

$$\begin{aligned} \text{ATE} &= \mathbb{E}[Y_i(1) - Y_i(0)] \\ &= \mathbb{E}[\mu_1(X) - \mu_0(X)] \end{aligned}$$

TREATED OUTCOME

$\mu_1(x)$

$\mathbb{E}[Y \mid D = 1, X = x]$

CONTROL OUTCOME

$\mu_0(x)$

$\mathbb{E}[Y \mid D = 0, X = x]$

TREATMENT ASSIGNMENT

$e(x)$

$\Pr(D = 1 \mid X = x)$, used by robust estimators

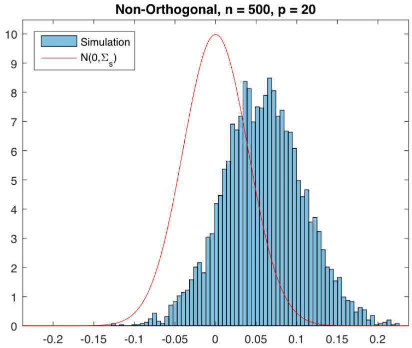
Key idea

Identification gives these nuisance predictions causal meaning; ML estimates them flexibly.

Naive plug-in can bias causal estimates

WHY ACCURACY IS NOT ENOUGH

- Regularization intentionally trades some bias for better prediction.
- Reusing the same sample can let fitted noise leak into the causal target.



Simulation: a naive non-orthogonal estimate shifts from its target.

Key idea

DML keeps nuisance-model error from dominating the causal estimate.

DML: partial linear model

$$Y = \theta_0 D + g_0(X) + \varepsilon$$

CAUSAL TARGET

θ_0

Program effect under the partially linear model and identifying assumptions.

TREATMENT NUISANCE

$m_0(X)$

$\mathbb{E}[D \mid X]$: what background predicts about participation.

Key idea

Naive plug-in ML estimated $g_0(X)$ will bias the estimator of θ_0

DML through partialling out

$l_0(X) = \mathbb{E}[Y \mid X]$: what background predicts about the outcome.

1. Explain the outcome
$\widetilde{Y} = Y - \widehat{\ell}(X)$

→

2. Explain participation
$\widetilde{D} = D - \widehat{m}(X)$

→

3. Relate what remains
regress $\widetilde{Y}$ on $\widetilde{D}$

In this partially linear example, DML residualizes both the outcome and the treatment.

Two DML guardrails

NEYMAN ORTHOGONALITY

Protect the target from small nuisance errors

Small nuisance-model errors have no first-order push on the effect estimate.

CROSS-FITTING

Use out-of-sample nuisance predictions

Each observation is scored using nuisance models trained without that observation.

WHAT DML CAN BUY

Flexible, high-dimensional nuisance learning plus familiar inference under conditions.

WHAT DML CANNOT DO

It cannot eliminate bias from unmeasured confounding or repair failed identification and overlap.

Key idea

We will learn both tools in detail; and apply them to DID, IV, RDD, etc

TMLE targets the causal quantity

TMLE = targeted maximum likelihood estimation

1. INITIAL LEARNING

Fit flexibly

Estimate outcome and treatment mechanisms with ML.

2. TARGETING STEP

Update deliberately

Make a small update chosen for the causal quantity of interest.

3. ESTIMATE + INFER

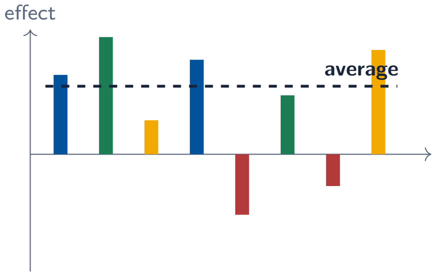
Plug in and quantify

Compute the effect from updated predictions and quantify uncertainty.

Key idea

DML builds from an orthogonal score; TMLE updates predictions to satisfy the causal estimating equation. Both require identification.

Find HTE with ML

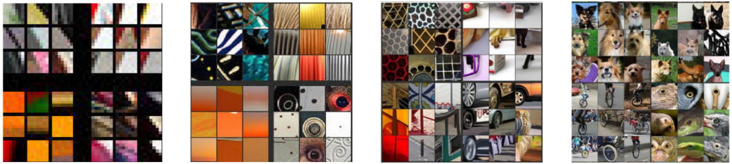


HETEROGENEOUS TREATMENT EFFECTS

- CATE(x)$=E[ITE|X=x]$
- S-, T-, X-, and DR-learners
- causal trees and generalized random forests
- validation of discovered heterogeneity

Deep learning

text, pixels, audio, networks → learned numerical representation → prediction or measurement task



Key idea

Deep learning is a power ML with deep neural networks.

We will cover...

- Basic Structures: neural networks
- Algorithms and properties: Backpropagation Algorithm, double descent, residual learning
- Convolutional Networks for image data
- Attention and Transformers for NLP
- If we have time, we can cover more

Causal Inference with unstructured data

Treatment

A randomized text, image, or video stimulus—or a component deliberately manipulated by design.

Outcome

A post-treatment open response, image, speech, or video.

Confounder

A pre-treatment representation related to both treatment and outcome.

Proxy

A pre-treatment measurement informative about latent confounding under additional assumptions.

If we have time, we will also cover causal inference with interference.

Grading at a glance

50%
PROBLEM SETS
Theory questions plus programming exercises that implement the methods.

25%
MIDTERM
A computational causal-analysis competition estimating ATE and HTE.

25%
FINAL PROJECT
Use a course method in a research-shaped project of your choice.

Key idea

The assessments move from understanding $\rightarrow$ implementation $\rightarrow$ independent research judgment.

Final Projects

Choose one of the following three options

1. apply a method in your own ongoing research;
2. replicate an applied paper and add an extension;
3. propose a methodological research project with preliminary work.