Prediction:
Causal versus ML Approaches

Nick Eubank

Review

1. Descriptive Questions
    - Description requires summarizing
    - Incumbent on you to ensure summary is representative
2. Causal Questions
    - If treated and untreated units have same potential outcomes, then correlations $\rightsquigarrow$ causal.
    - We can never prove potential outcomes are the same, only argue from case knowledge.
3. Prediction Questions

Prediction

Predict outcomes for observations we have not yet observed

- Use current patient data to predict future complications
- Predict impact of expanded health insurance on wellness

Doesn't have to be predicting the future e.g.

- Using neural network to identify animals in picture. We're *predicting* labels for data whose labels we haven't observed.

Out-of-sample extrapolation might be a better term (but we're stuck with prediction.)

Prediction

Fundamental Problem of Prediction:
Because we are fundamentally interested in predicting outcomes for data *we have not yet observed*, we can never be sure of how well our models will perform.
Way of developing *guesses*:

- Cross-validation; Split samples; statistical model diagnostics; etc.

But by definition, can't run diagnostics on *data we have not yet observed.*
$\Rightarrow$ We can also only argue from theory about prediction accuracy.

Prediction

Not talking about your “test” set, which also has labels. We’re talking about performance on data with no labels / predicted values.

Fundamental Problem of Prediction

Fundamental Problem of Prediction is just like the Fundamental Problem of Causal Inference:
Because accuracy of conclusions depends on something unobservable, we can never *mathematically* prove accuracy.
We *have* to argue from theory.

Fundamental Problem of Prediction

Whether we're using causal inferences or fitting a supervised machine learning model (SML) only your knowledge of the world will tell you whether you can use the model to make predictions on new data!

Fundamental Problem of Prediction

Causal Inference:

- Internal and external validity

Supervised Machine Learning:

- How confident are we the patterns we're detecting in the training data are present in a new context?

Fundamental Problem of Prediction

poll!

Fundamental Problem of Prediction

1. Training data: Duke patient records
2. Trying to predict: Risk of surgical complications
3. Use: Hospitals across the US

Fundamental Problem of Prediction

1. Training data: Duke patient records
2. Trying to predict: Risk of surgical complications
3. Use: Rural hospitals in South Asia

Prediction: Causal versus SML

1. Prediction via Causal Inference
    - Try to model *fundamental causal relationships*
2. Prediction via Supervised Machine Learning (SML)
    - Try to find *correlations* that have predictive power

Prediction: Causal versus SML

What are strengths and weaknesses of each approach?

Prediction: Causal versus SML

Causal:

- + Causal relationships are likely to be more robust, and thus more likely to generalize.
- - Much harder to estimate.
- - Hard to estimate things beyond average, linear relationships (possible, but hard).

Supervised Machine Learning:

- - Simple correlations are much less likely to generalize, and may break down as the world changes.
    - SML can be fragile.
- + Functional form flexibility allows for finding more obscure relationships.

Fragility of SML

Suppose we are interested in explaining cancer survival.

Causal Approach:

- We run an randomized-control trial to test a new drug.
- Internal Validity: We are sure we've measured causal effect of drug.
- External Validity: Might have issues if patients aren't representative.

Fragility of SML

Supervised Machine Learning Approach:
Suppose that we have data on consumer behavior, including what the vending machines people frequent, restaurants they attend, and grocery stores where they shop.

- We feed all the information we can find about patients into a logit regression to predict survival.

Now suppose the office that administers an effective cancer drug is next to the only Coke vending machine in the Duke hospital system.

- Our model now predicts anyone drinking vending machine Coke is likely to survive cancer!

Fragility of SML

Problems:

- If we only want to study Duke patients, this may be fine! It's just proxying for taking this drug.
- But... what if Duke adds more Coke machines? Suddenly our predictions for people who drink Coke in other places will suggest unrealistically high survival rates!
- And *clearly* this doesn't generalize beyond Duke!

Fragility of SML

Supervised machine learning is prone to finding context-specific proxies:

- things we can measure, and
- that are correlated with causal factors, but
- which aren't actually causal, and
- which may not be correlated with the causal factor in other contexts.

e.g. The coke machine, which is correlated with getting an important drug in the data we have, but which obviously isn't causally related to cancer survival, and which isn't likely to be correlated in other contexts.

Thinking of Context-Specific Proxies

Suppose we want to predict customer value

Fragility of SML

Big causal inference / SML trade-off:

- Causal relationships are likely to be much more robust because they reflect causal relationships.
- Supervised Machine Learning models are *just* picking up correlations, and correlations that are predictive in one context may not be predictive in others!

Adversarial Users

An issue related to context-specific proxies are *adversarial users*: users who change their behavior once they learn that they are being observed by a machine learning algorithm.

Adversarial Users

In *many* elementary schools in the US, essays are now being graded by supervised machine learning algorithms. These algorithms were trained by:

- Having humans grade a random sample of essays, then
- Train a model to predict the human grades.

Teachers pay a subscription, submit student essays to this system, and get back a grade.
But what are these algorithms rewarding?

Adversarial Users

Many turn out to rely on *context-specific proxies*.

- In essays written for humans, longer essays tended to be better.
- So model started rewarding length.

The problem is length is just a proxy for what we care about (quality of writing and argumentation).
Had student behavior not changed, that'd be ok. But...

Students quickly realized the algorithm rewarded length, *not* quality, so the started writing very long jibberish essays that no human would ever score well, and... they got As!

Adversarial Users

A machine learning algorithm doesn't actually know what is *important*, it just knows what has predictive power in the training data. But this makes machine learning algorithms manipulable:

- Essay grading
- Computer security
    - Spam
    - Anomaly and fraud detection
    - Malware detection
- Computer vision
- Resume reading

Adversarial Users

For black-box machine learning algorithms (SVMs, neural networks, etc.), the problem isn't just that these algorithms are manipulatable, but also that they're unpredictably manipulable.

One approach is to only use interpretable ML models.

Build models that have only a few parameters, and which are *transparent* and understandable.

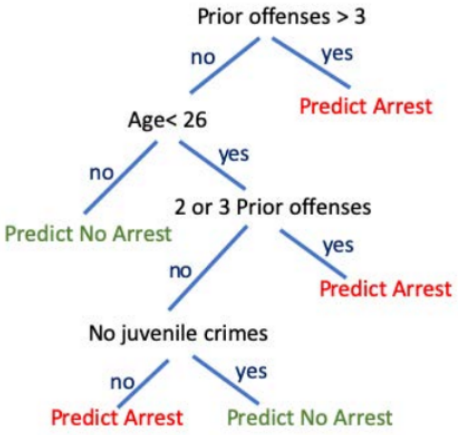
Interpretable Machine Learning (Cynthia Rudin)

1. Any cEEG pattern with Frequency **2** **H**z	1 point	. . .
2. **E**pileptiform Discharges	1 point	+ . . .
3. Patterns include [**L**PD, LRDA, BIPD]	1 point	+ . . .
4. **P**atterns Superimposed with Fast or Sharp Activity	1 point	+ . . .
5. Prior **S**eizure	1 point	+ . . .
6. **B**rief Rhythmic Discharges	**2** points	+ . . .
	Score	= . . .

Score	0	1	2	3	4	5	6+
Risk	<5%	11.9%	26.9%	50.0%	73.1%	88.1%	95.3%

2HELPS2B score for predicting seizures in ICU patients (Struck et al 2017), constructed by the RiskSLIM ML algorithm (Ustun & R 2019).
The factors and point scores were chosen (by an algorithm)

Interpretable Machine Learning (Cynthia Rudin)



An interpretable decision tree to predict whether an individual will be arrested in the future. Hu et al. NeurIPS 2019

Interpretable Machine Learning

Advantages:

- Transparency makes it easy to detect reliance on context-specific proxies
- Also helpful for identifying bias
- Often perform as well as fancier methods

Disadvantages:

- New? Have to go learn them? Honestly, I think we should all be using these.

Summary

- Causal inference tends to be more robust
    - Better when making *bigger* extrapolations
      (e.g. if you plan to deliberately manipulate the world!)
- SML can be more powerful when you're sure your context won't change
    - Training data looks just like data you want to work with
    - World is unlikely to adapt to your model