

Introduction

Machine Learning for Social Science

What is Machine Learning?

There are many possible definitions.

One definition:

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

– Tom Mitchell (1997)

Examples:

- Self-driving cars
- Facial recognition
- Targeted Ads

What is Statistical Learning?

[Use data] to extract important patterns and trends, and understand “what the data says”. We call this learning from data.

– Hastie, Tibshirani, Friedman (2009)

What is Machine Learning for Social Science?

The application of Machine Learning techniques to answer Social Science questions.

Traditionally, the social sciences have been focused on **inference**.

Example: Is there a statistical significance difference between groups A and B? Is the slope of a linear regression line significantly different from 0?

With ML, we are more concerned with **prediction**.

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. *Potash et al. 2015*

- Lead Poisoning is a big issue for children living in old homes with lead paint.
- Blood tests can identify elevated lead levels ... but by then, it's too late.
- Use Machine Learning to predict risk of a child being poisoned to provide intervention.

Machine Learning Applications

Identifying Police Officers at Risk of Adverse Events. *Carton et al. 2015*

- Early Intervention System in place to flag officers at high risk for involvement in adverse events, but it wasn't data-driven and only based on intuition.
- Use Machine Learning to predict adverse events before they happen and provide intervention (e.g. counseling).
- Identified 12% more high-risk officers and flagged 32% fewer low-risk officers.

Machine Learning Applications

Reducing Incarceration through Prioritized Interventions. *Bauman et al. 2018*

- Programs to be put in place to help people at high risk of incarceration.
- Use Machine Learning to predict risk of incarceration in the next 12 months (in this case, specifically recidivism).
- Improved on baseline heuristic model by 1.5 times and on random model by 4.6 times.

Why Machine Learning?

Goal: Build adaptable and scalable systems that are effective and easy to maintain.

Adaptable: Relationships may change over time.

Easy to Maintain: Rule-based systems may become cumbersome and difficult to update.