Building a Robot Judge

Data Science for Decision-Making
ETH Zurich, Fall 2020

Welcome to the course!

Instructions before we begin:
(1) Turn on video and set audio to mute
(2) In Participants panel, set zoom name to "Full Name, School, Dept/Major"
(ex: "Leon Smith, ETH Computer Science")
(3) Say "hi" in the chat

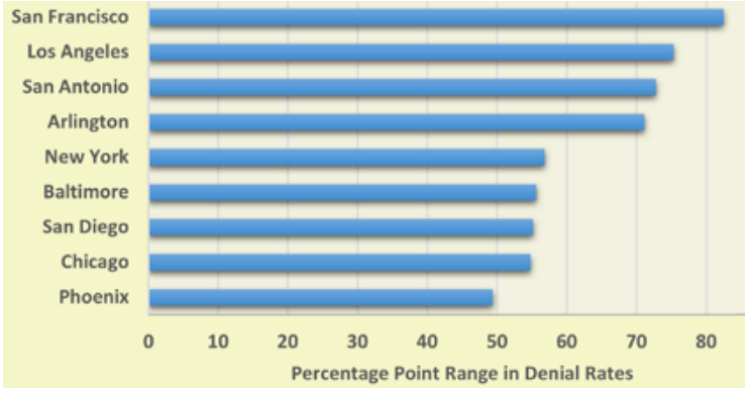
Building a Robot Judge

Data Science for Decision-Making

1. Course Overview and Introduction

What's the matter with human decision-making?

U.S. Asylum Courts: Disparities in Grant Rates



- In San Francisco, one judge grants 90.6% of asylum requests, while another judge grants just 2.9%!

Jailing Decisions Before/After Lunch Breaks

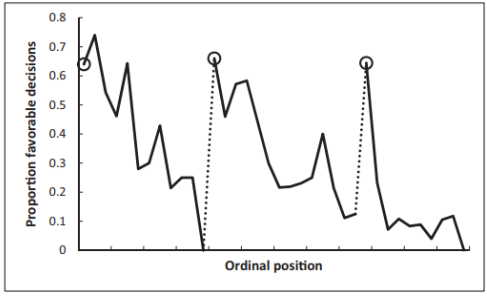
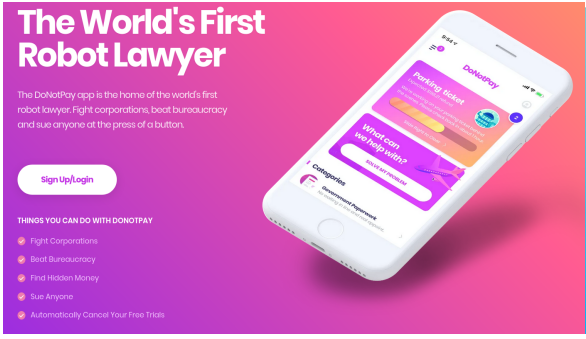


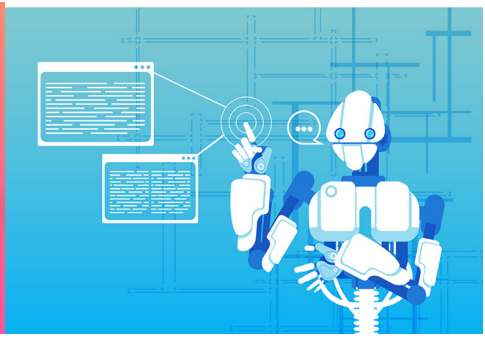
Fig. 1. Proportion of rulings in favor of the prisoners by ordinal position. Circled points indicate the first decision in each of the three decision sessions; tick marks on *x* axis denote every third case; dotted line denotes food break. Because unequal session lengths resulted in a low number of cases for some of the later ordinal positions, the graph is based on the first 95% of the data from each session.

Source: Danziger et al, PNAS 2011, Israel judges deciding on parole.

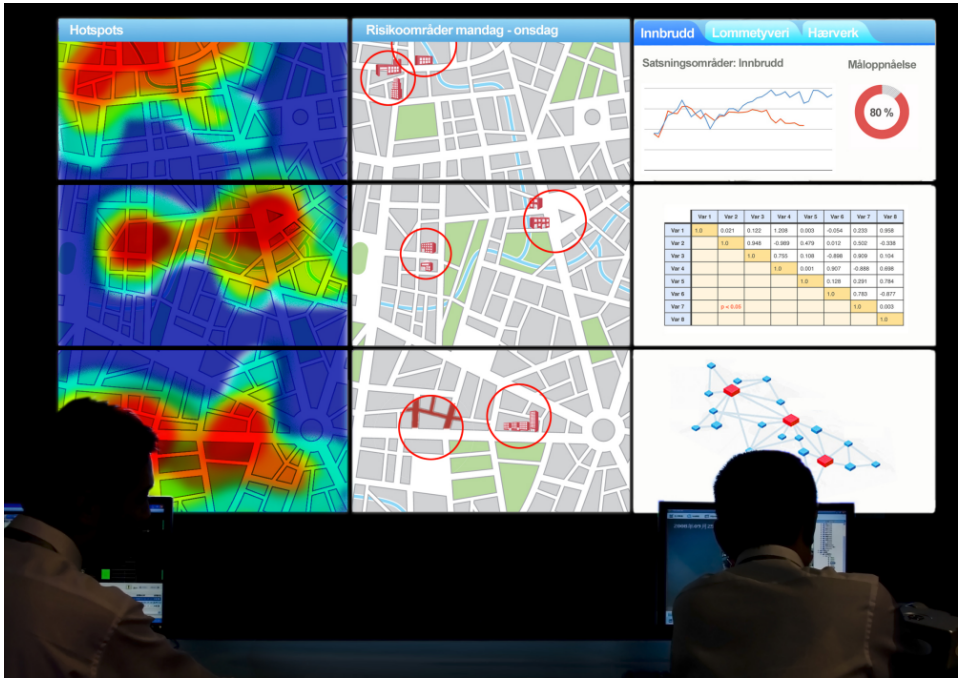
How about robot decision-making?

Your Court-Appointed Chatbot – Is Artificial Intelligence Threatening the Legal Profession?





Predictive Policing



Predictive policing poses discrimination risk, thinktank warns

Machine-learning algorithms could replicate or amplify bias on race, sexuality and age

▲ One officer said human biases including more stop and searches of black men were likely to be introduced into algorithm data sets. Photograph: Carl Court/Getty Images

Zoom Poll 1.1

Welcome to ***Building a Robot Judge***

- This course focuses on **machine learning** and **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.

Welcome to ***Building a Robot Judge***

- This course focuses on **machine learning** and **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.

Welcome to ***Building a Robot Judge***

- This course focuses on **machine learning** and **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.
- Scientific goals:
  - Understand the factors underlying decisions of judges.

Welcome to ***Building a Robot Judge***

- This course focuses on **machine learning** and **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.
- Scientific goals:
  - Understand the factors underlying decisions of judges.
  - Assess the real-world impacts of decisions on society – e.g. defendants, patients.

Learning objectives

Learning objectives

1. **Implement and evaluate machine learning pipelines.**

Learning objectives

1. **Implement and evaluate machine learning pipelines.**
   - Evaluate (find problems in) existing machine learning pipelines.
   - Design a pipeline to solve a given ML problem.
   - Implement some standard pipelines in Python.

Learning objectives

1. **Implement and evaluate machine learning pipelines.**
   - Evaluate (find problems in) existing machine learning pipelines.
   - Design a pipeline to solve a given ML problem.
   - Implement some standard pipelines in Python.
2. **Implement and evaluate causal inference designs.**

Learning objectives

1. **Implement and evaluate machine learning pipelines.**
   - Evaluate (find problems in) existing machine learning pipelines.
   - Design a pipeline to solve a given ML problem.
   - Implement some standard pipelines in Python.
2. **Implement and evaluate causal inference designs.**
   - Evaluate (find problems in) causal claims.
   - Apply the standard research designs to produce causal evidence for a given empirical setting – or articulate why it is not possible.
   - Implement these research designs using Stata regressions.

Learning objectives

1. **Implement and evaluate machine learning pipelines.**
   - Evaluate (find problems in) existing machine learning pipelines.
   - Design a pipeline to solve a given ML problem.
   - Implement some standard pipelines in Python.
2. **Implement and evaluate causal inference designs.**
   - Evaluate (find problems in) causal claims.
   - Apply the standard research designs to produce causal evidence for a given empirical setting – or articulate why it is not possible.
   - Implement these research designs using Stata regressions.
3. **Understand how (not) to use data science tools (ML and CI) to support expert decision-making.**

Learning objectives

1. **Implement and evaluate machine learning pipelines.**
   - Evaluate (find problems in) existing machine learning pipelines.
   - Design a pipeline to solve a given ML problem.
   - Implement some standard pipelines in Python.
2. **Implement and evaluate causal inference designs.**
   - Evaluate (find problems in) causal claims.
   - Apply the standard research designs to produce causal evidence for a given empirical setting – or articulate why it is not possible.
   - Implement these research designs using Stata regressions.
3. **Understand how (not) to use data science tools (ML and CI) to support expert decision-making.**
   - Explore the connections/distinctions between **prediction**, **inference**, and **decisions**.
   - Evaluate proposed policies/systems that use algorithms for decision support – along accuracy, bias, gaming, and other dimensions.
   - Read and critique research papers reporting on these policies/systems.

Learning objectives

1. **Implement and evaluate machine learning pipelines.**
   - Evaluate (find problems in) existing machine learning pipelines.
   - Design a pipeline to solve a given ML problem.
   - Implement some standard pipelines in Python.
2. **Implement and evaluate causal inference designs.**
   - Evaluate (find problems in) causal claims.
   - Apply the standard research designs to produce causal evidence for a given empirical setting – or articulate why it is not possible.
   - Implement these research designs using Stata regressions.
3. **Understand how (not) to use data science tools (ML and CI) to support expert decision-making.**
   - Explore the connections/distinctions between **prediction**, **inference**, and **decisions**.
   - Evaluate proposed policies/systems that use algorithms for decision support – along accuracy, bias, gaming, and other dimensions.
   - Read and critique research papers reporting on these policies/systems.
   - If you are signed up for the project: Implement/analyze such a system and write a paper about it.

Why did you sign up for this course?

Zoom Poll 1.2

Outline

Lecture Times

- Mondays, 1215h-14h
  - Zoom (Meeting ID 927 5461 2589)
- 10 minute break, 13h-1310h

Online Lecture Norms

Let's make the most of online learning!

- Live attendance at lectures is required.
- Keep video on if connection allows.
- Stay muted when not talking.
- To make questions or comments, type in the chat (private or public) or use the "raise hand" function.

Online Course Materials

- Course Syllabus:
  - `https://bit.ly/BRJ_syll`
- Course Repo (slides, notebooks, and assignments):
  - `https://bit.ly/BRJ_Repo`

Teaching Assistants

Claudia Marangon (claudia.marangon@gess.ethz.ch)
Christoph Goessmann (christoph.goessmann@gess.ethz.ch)

- will hold several TA sessions to go over some more technical/practical issues.
  - e.g. skills needed to do the coding homework assignments.
  - not mandatory – attend if you are new to the tools.
  - we will also post recordings.
- can answer questions about lectures, notebooks, assignments, and projects.

Course Communication

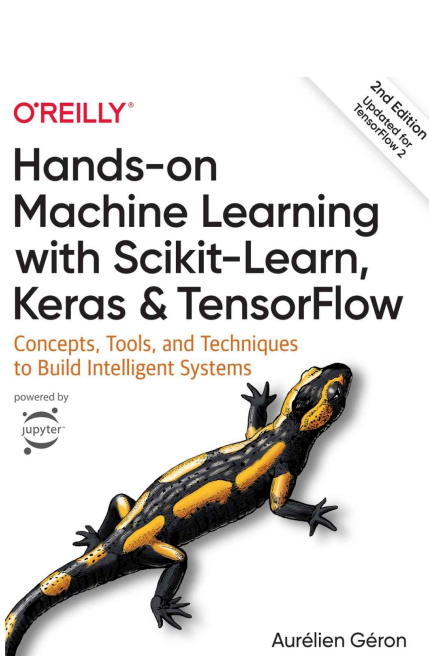
- Course communication will be done through eDoz.
- Questions welcome via email, to me or to the TA's.
- I will be available in the zoom 5 minutes early, during the mid-lecture break, and until 15 minutes after the end of lecture.
- Will schedule meetings with students doing projects.

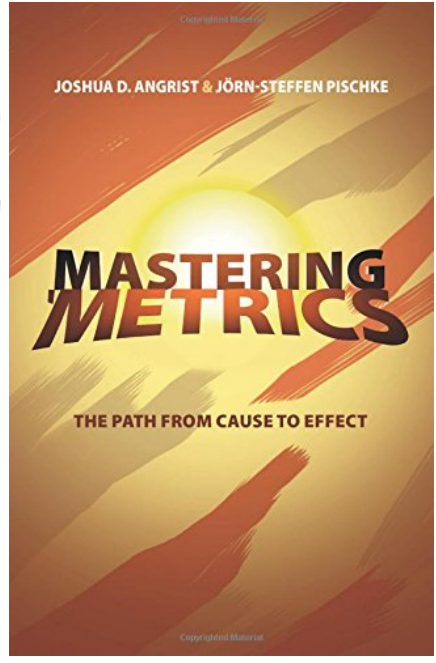
Course Participation

- We will monitor attendance and participation in class activities (e.g. polls, group work).

Syllabus has a long readings list

- There are only a handful of required readings (highlighted).
- Other readings can be used as reference:
  - to complement the slides
  - to be used for reading response essays (more next week)
  - for projects


O'REILLY
2nd Edition
Updated for TensorFlow 2
Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow
Concepts, Tools, and Techniques to Build Intelligent Systems
powered by
jupyter
Aurélien Géron


JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE
MASTERING METRICS
THE PATH FROM CAUSE TO EFFECT

Example Code is in Python and Stata

- Machine learning:
  - Python 3.7 is ideal for machine learning.
  - You can use Anaconda or download the packages we need to a pip environment.

Example Code is in Python and Stata

- Machine learning:
  - Python 3.7 is ideal for machine learning.
  - You can use Anaconda or download the packages we need to a pip environment.
- Econometrics:
  - Stata is closed-source statistical software – we will provide licenses.
  - Unfortunately, there is no good open-source alternative.

Example Code is in Python and Stata

- Machine learning:
  - Python 3.7 is ideal for machine learning.
  - You can use Anaconda or download the packages we need to a pip environment.
- Econometrics:
  - Stata is closed-source statistical software – we will provide licenses.
  - Unfortunately, there is no good open-source alternative.
- See the syllabus for lists of packages.
- If you strongly prefer to use a different programming language, email me about it.

Homework Assignments

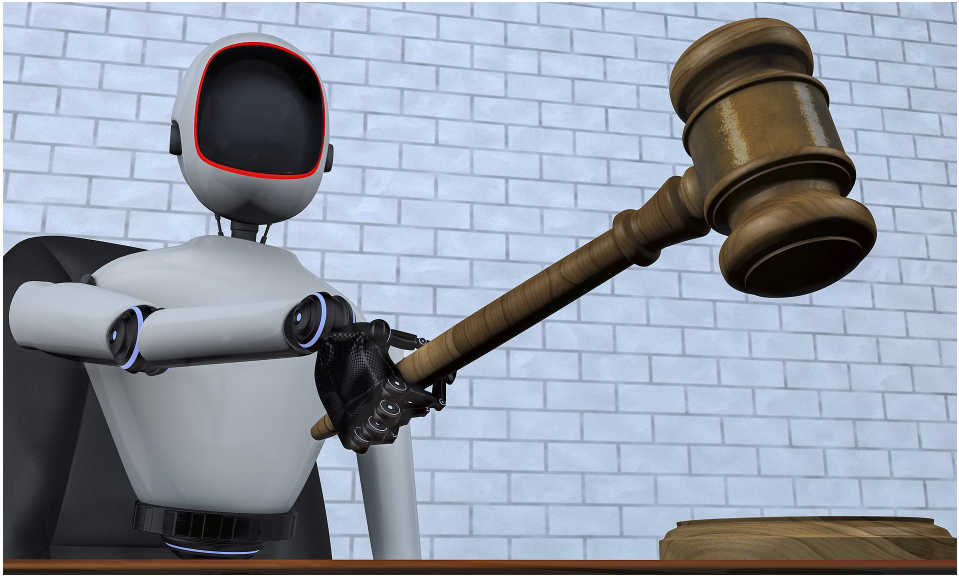
- There is a (small) homework assignment due every week.
  - upload to assignment dropbox (see syllabus)
  - most are completion grades — three are graded.
- More details as we go.

Course Projects

- 2 extra credits available to do a course project (5 credits total):
- About twice as much work expected
  - previous course projects have turned into conference/journal publications, that should be goal.
  - two projects turned into funded Innoisuisse startups.
- Can be done individually or in small groups (preferably 2, up to 4 with good reason).

Course Projects

- 2 extra credits available to do a course project (5 credits total):
- About twice as much work expected
  - previous course projects have turned into conference/journal publications, that should be goal.
  - two projects turned into funded Innoisuisse startups.
- Can be done individually or in small groups (preferably 2, up to 4 with good reason).
- Information session after September 28th (Week 2) lecture (until about 1410h).
  - we have a list of potential topics and postdoc advisors.



Questions

Outline

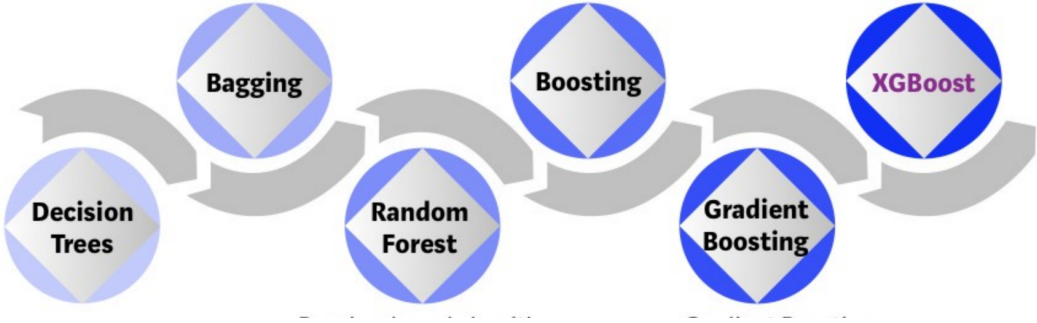
Implement and evaluate machine learning pipelines

Implement and evaluate machine learning pipelines

- Evaluate (find problems in) existing machine learning pipelines.
- Design a pipeline to solve a given ML problem.
- Implement some standard pipelines in Python.

- Week 02 Sept. 28 Machine Learning Essentials
- Week 04 Oct 12 Classification & XGBoost
- Week 06 Oct 19 Deep Learning Essentials
- Week 12: Dec 7 Compression and Explanation

“Extreme Gradient Boosting”: Ingredients



Complicated in theory, easy in practice

```
from xgboost import XGBClassifier
model = XGBClassifier()

model.fit(X_train, y_train,
          early_stopping_rounds=10,
          eval_metric="logloss",
          eval_set=[(X_eval, y_eval)]
          )

y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
```

Predicting U.S. Asylum Court Decisions

Predicting U.S. Asylum Court Decisions

		Predicted	
		Denied	Granted
True	Denied	195,223	65,798
	Granted	73,269	104,406

Accuracy = 68.3%, F1 = 0.60

- Prediction App (Beta): `https://floating-lake-11821.herokuapp.com/`

New Measurements → New Policy Solutions

Source: World Bank.

Implement and evaluate causal inference designs

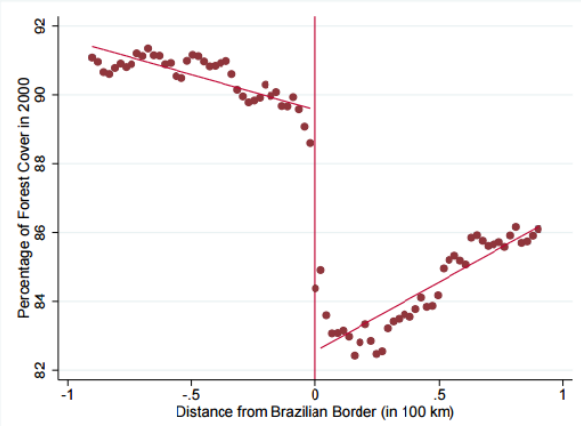
Implement and evaluate causal inference designs

- Evaluate (find problems in) causal claims.
- Apply the standard research designs to produce causal evidence for a given empirical setting – or articulate why it is not possible.
- Implement these research designs using Stata regressions.

- Week 03 Oct 5 Linear Models and Research Design
- Week 05 Oct 26 Machine Learning and Causal Inference
- Week 07 Nov 2 Instrumental Variables

http://www.tylervigen.com/spurious-correlations

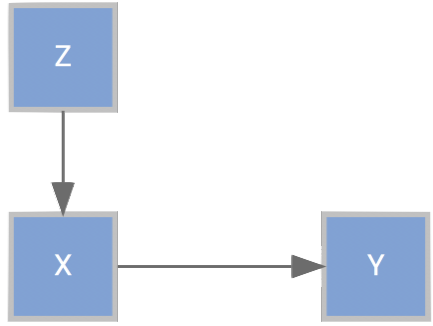
Burgess, Costa, and Olken, "The Brazilian Amazon's Double Reversal of Fortune"



Source: `https://economics.mit.edu/files/12732`

Instrumental Variables

Zoom Poll: Before reading the course syllabus, had you ever heard of instrumental variables?



- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.

- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.
  - estimate a low-dimensional **causal parameter** ρ using

$$y_i = \alpha_i + x_i \cdot \rho + \epsilon_i$$

    where i indexes over documents, α_i includes control variables (and fixed effects), $\cdot$ is dot product, and ϵ_i is the error residual.

- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.
  - estimate a low-dimensional **causal parameter** ρ using

$$y_i = \alpha_i + x_i \cdot \rho + \epsilon_i$$

    where i indexes over documents, α_i includes control variables (and fixed effects), $\cdot$ is dot product, and ϵ_i is the error residual.
  - ρ gives a prediction how outcome y would change if treatment variable x were **exogenously shifted**.
  - useful for policy evaluation.

- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.
  - estimate a low-dimensional **causal parameter** ρ using

$$y_i = \alpha_i + x_i \cdot \rho + \epsilon_i$$

    where i indexes over documents, α_i includes control variables (and fixed effects), $\cdot$ is dot product, and ϵ_i is the error residual.
  - ρ gives a prediction how outcome y would change if treatment variable x were **exogenously shifted**.
  - useful for policy evaluation.

- **Machine learning**:
  - $\boldsymbol{y}$ can be multi-dimensional, $\boldsymbol{x}$ can be high-dimensional.

- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.
  - estimate a low-dimensional **causal parameter** ρ using

    $$y_i = \alpha_i + x_i \cdot \rho + \epsilon_i$$

    where i indexes over documents, α_i includes control variables (and fixed effects), $\cdot$ is dot product, and ϵ_i is the error residual.
  - ρ gives a prediction how outcome y would change if treatment variable x were **exogenously shifted**.
  - useful for policy evaluation.

- **Machine learning**:
  - $\boldsymbol{y}$ can be multi-dimensional, $\boldsymbol{x}$ can be high-dimensional.
  - learn a high-dimensional vector of parameters θ to approximate a (potentially non-linear) function

    $$y_i = h(\boldsymbol{x}; \theta)$$

    that predicts y given covariates $\boldsymbol{x}$.

- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.
  - estimate a low-dimensional **causal parameter** ρ using

$$y_i = \alpha_i + x_i \cdot \rho + \epsilon_i$$

    where i indexes over documents, α_i includes control variables (and fixed effects), $\cdot$ is dot product, and ϵ_i is the error residual.
  - ρ gives a prediction how outcome y would change if treatment variable x were **exogenously shifted**.
  - useful for policy evaluation.

- **Machine learning**:
  - $\boldsymbol{y}$ can be multi-dimensional, $\boldsymbol{x}$ can be high-dimensional.
  - learn a high-dimensional vector of parameters θ to approximate a (potentially non-linear) function

$$y_i = h(\boldsymbol{x}; \theta)$$

    that predicts y given covariates $\boldsymbol{x}$.
  - if we collected more data on $\boldsymbol{x}$, we could predict the associated $\hat{y}$.

- **Econometrics** (applied statistical causal inference):
  - y is one-dimensional, x is low-dimensional.
  - estimate a low-dimensional **causal parameter** ρ using

$$y_i = \alpha_i + x_i \cdot \rho + \epsilon_i$$

    where i indexes over documents, α_i includes control variables (and fixed effects), $\cdot$ is dot product, and ϵ_i is the error residual.
  - ρ gives a prediction how outcome y would change if treatment variable x were **exogenously shifted**.
  - useful for policy evaluation.

- **Machine learning**:
  - $\boldsymbol{y}$ can be multi-dimensional, $\boldsymbol{x}$ can be high-dimensional.
  - learn a high-dimensional vector of parameters θ to approximate a (potentially non-linear) function

$$y_i = h(\boldsymbol{x}; \theta)$$

    that predicts y given covariates $\boldsymbol{x}$.
  - if we collected more data on $\boldsymbol{x}$, we could predict the associated $\hat{y}$.
  - but $h(\cdot)$ does not provide a *counterfactual prediction* – that is, how the outcome would change if $\boldsymbol{x}$'s were exogenously shifted.

Understand how (not) to use data science tools (ML and CI) to support expert decision-making

Understand how (not) to use data science tools (ML and CI) to support expert decision-making

- Appreciate the connections/distinctions between **prediction**, **inference**, and **decisions**.
- Evaluate proposed policies/systems that use algorithms for decision support – along accuracy, bias, gaming, and other dimensions.
- Read and critique research papers reporting on these policies/systems.

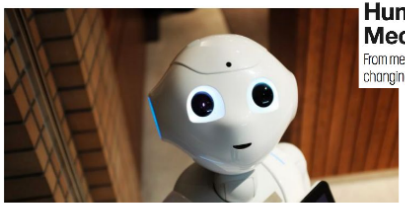
- Week 08 Nov 9 Bias and Discrimination
- Week 09: Nov 16 Algorithms and Decisions 1
- Week 10: Nov 23 Algorithms and Decisions 2
- Week 11: Nov 30 Algorithmic Fairness 1
- Week 12: Dec 7 Compression and Explanation
- Week 13: Dec 14 Algorithmic Fairness 2

Home 〉 Features 〉 Emerging tech & Innovation Features

Researcher explains how algorithms can create a fairer legal system

8 WAYS MACHINE LEARNING WILL IMPROVE EDUCATION

BY MATTHEW LYNCH / JUNE 12, 2018 / 5



Your Future Doctor May Not be Human. This Is the Rise of AI in Medicine.

From mental health apps to robot surgeons, artificial intelligence is already changing the practice of medicine.

PREDICTIVE POLICING: USING MACHINE LEARNING TO DETECT PATTERNS OF CRIME

The New York Times

ROBO RECRUITING

Can an Algorithm Hire Better Than a Human?

By Claire Cain Miller

Source: Hoda Heidari slides.

- **Bansak et al (*Science* 2018):**
  - assign refugees to locations using an algorithm that predicts higher employment, demonstrate large gains relative to random assignment.

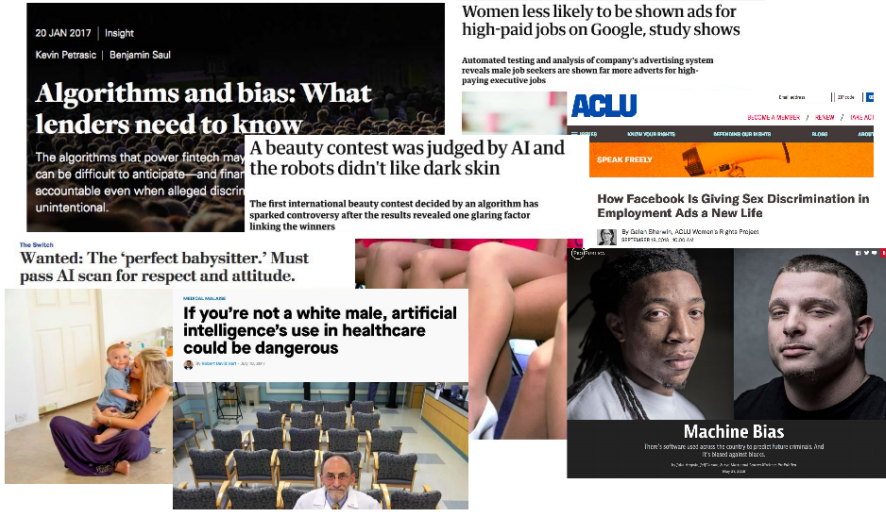
- **Bansak et al (*Science* 2018):**
  - assign refugees to locations using an algorithm that predicts higher employment, demonstrate large gains relative to random assignment.
- **Kleinberg et al (*Quarterly Journal of Economics*, 2018):**
  - decide on bail/parole using an algorithm that predicts recidivism (whether defendant commits another crime), demonstrate that it could reduce both incarceration rates and recidivism.

- **Bansak et al (*Science* 2018):**
  - assign refugees to locations using an algorithm that predicts higher employment, demonstrate large gains relative to random assignment.
- **Kleinberg et al (*Quarterly Journal of Economics*, 2018):**
  - decide on bail/parole using an algorithm that predicts recidivism (whether defendant commits another crime), demonstrate that it could reduce both incarceration rates and recidivism.
- **Ash et al (2020):**
  - show that algorithm can predict fiscal corruption from budget data, could be used to double the detection rate of corruption relative to randomly assigned audits.

Activity (3 minutes)

Activity (3 minutes)

- Chat **privately** to me on zoom:
  - An example of a decision or judgment that would be difficult to automate, and why.
  - Try to pick one that no one else picks.



Source: Hoda Heidari slides.

The AI Fairness Tradeoff

- Pros:
  - higher accuracy
  - lower cost
  - consistency – all defendants get the same decision for the same evidence.
- Cons:
  - systematic biases – for example those in training data – could be replicated or amplified.
  - lack of transparency / accountability
  - issues of privacy /surveillance
  - risks of gaming the system

The AI Fairness Tradeoff

- Pros:
  - higher accuracy
  - lower cost
  - consistency – all defendants get the same decision for the same evidence.
- Cons:
  - systematic biases – for example those in training data – could be replicated or amplified.
  - lack of transparency / accountability
  - issues of privacy /surveillance
  - risks of gaming the system
- But algorithms can also be used to **detect** systematic bias, to **understand** it – and therefore to help **reduce** it.

Interpretable Machine Learning

- Key point:
  - Standard machine learning techniques cannot be interpreted easily.

Interpretable Machine Learning

- Key point:
  - Standard machine learning techniques cannot be interpreted easily.
- Users and decision subjects want to understand the model

Interpretable Machine Learning

- Key point:
  - Standard machine learning techniques cannot be interpreted easily.
- Users and decision subjects want to understand the model
- Other models/approaches improve interpretability:
  - XGBoost provides feature importance ranking.
  - LIME and related tools can help interpret any model (Ribeiro et al 2016).

Outline

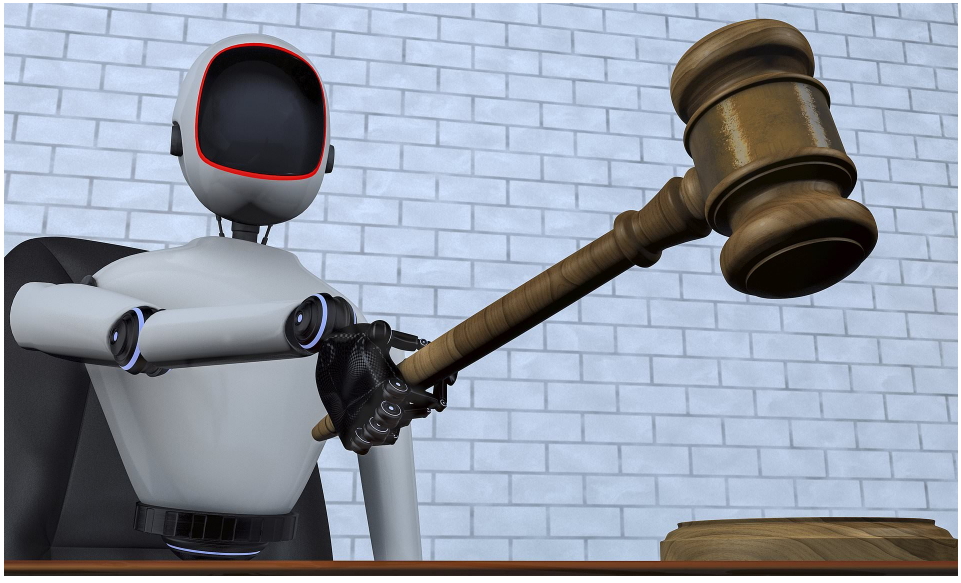
Activity: Breakout Rooms (8 minutes)

- Discuss "Biased algorithms are easier to fix than biased people" by Sendhil Mullainathan in *New York Times* (`bit.ly/nyt-bias`).
  - Think of another task where fixing biases in an algorithm is probably easier than fixing it in humans.
  - Can you think of the opposite case — a task where fixing biases in humans is easier than fixing biases in algorithms?

First Homework Assignment

Homework Assignments Page: `http://bit.ly/BRJ_HW`

- Write a short fake news article (~300 words) about a fake AI technology supporting/replacing expert decisions, such as by doctors or judges.
  - This is a completion grade – have fun with it!
  - See Homework Assignments page for submission instructions (due by 8am next Monday).



Meeting Adjourned!