Machine learning for health

Instructors: Jon Kolstad, Ziad Obermeyer (UC Berkeley)

Learning Objectives: This course will prepare students to:

- Describe intuitively how and why a few basic machine learning algorithms work
- Understand what is new about data science and how it differs from traditional estimation (regression). For example, distinguish prediction problems ('can I predict y with x') from estimation problems ('does x cause y').
- Identify major methodological pitfalls encountered in answering these kinds of questions
- Develop a research question around a prediction policy problem in health, and make a research plan that avoids key methodological problems

Evaluation: Grades will be based on in-class participation, as well as a *short* research proposal that applies machine learning methods to a problem in health. Proposals will be evaluated on the basis of how well students incorporate methodological insights from the course, their ability to recognize and work around pitfalls, and the originality of the proposal.

Schedule

Day 1

- **Lecture 1 (JK): Course introduction and prediction vs estimation**
- **Lecture 2 (ZO): What's new about machine learning**
- **Readings**
    - Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), pp.87-106.
    - James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An introduction to statistical learning. New York: Springer. Available here.
        - Ch. 1, 2, 6
    - Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D. and Weinstein, J., 2018. Improving refugee integration through data-driven algorithmic assignment. Science, 359(6373), pp.325-329.
    - Varian HR. Big data: New tricks for econometrics. Journal of Economic Perspectives. 2014 May;28(2):3-28.
    - Blake T, Nosko C, Tadelis S. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. Econometrica. 2015 Jan;83(1):155-74.

Day 2

- **Lecture 3 (JK): Introduction to machine learning methods**
- **Lecture 4 (ZO): Prediction policy problems**

- **Readings**
  - Athey, S., 2017. Beyond prediction: Using big data for policy problems. Science, 355(6324), pp.483-485.
  - Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z., 2015. Prediction policy problems. American Economic Review P&P, 105(5), pp.491-95.
  - Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017. Human decisions and machine predictions. The quarterly journal of economics, 133(1), pp.237-293.
  - Einav, L., Finkelstein, A., Mullainathan, S. and Obermeyer, Z., 2018. Predictive modeling of US health care spending in late life. Science, 360(6396), pp.1462-1465.
  - Anderson A, Kleinberg J, Mullainathan S. Assessing human error against a benchmark of perfection. ACM Transactions on Knowledge Discovery from Data (TKDD). 2017 Aug 21;11(4):45.

Day 3

- **Lecture 5 (ZO): Pitfalls**
  - Mullainathan, S. and Obermeyer, Z., 2017. Does machine learning automate moral hazard and error?. American Economic Review P&P, 107(5), pp.476-80.
  - Corbett-Davies, S., Pierson, E., Feller, A., Goel, S. and Huq, A., 2017, August. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 797-806). ACM.
  - Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017, August. The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 275-284). ACM.
  - Henry, K.E., Hager, D.N., Pronovost, P.J. and Saria, S., 2015. A targeted real-time early warning score (TREWScore) for septic shock. Science translational medicine, 7(299), pp.299ra122-299ra122.
- **Student proposal presentations (PM)**