Machine learning in economics

Hannah Bull

November 28, 2019

1 Introduction

The aim of this class is to provide an introduction to machine learning techniques, such as lasso and ridge regression, random forests and support vector machines, and to learn how to apply such techniques to pure prediction problems, as well as to estimate causal effects.

The course is divided into three parts. Each of the parts has a theoretical and a practical component where we will use R. There will be three equally-weighted assignments based on the three practical components.

The practical sessions are available online at: `ml-in-econ.appspot.com`.

2 Part 1: Prediction problems

2.1 Theoretical session: The bias-variance trade-off

Bias-variance trade-off, cross-validation and regularised regression.

Given any dataset (X,Y), it is possible to perfectly fit a function $Y = f(X)$ to all of the observations just by using a high enough degree polynomial. This gives us a perfect predictor of Y given X. However, such a function is undesirable, as f depends strongly on the dataset use to estimate it, and is thus high-variance. On the other hand, a simple linear model $\hat{Y} = X\hat{\beta}$ does not vary much with new data (X',Y') drawn from the same distribution as the initial data. However, such a model might be too simple and have high prediction error. This dilemma is known as the bias-variance trade-off. We discuss how to use cross-validation to balance the bias-variance trade-off and avoid over/under fitting. We also discuss the bias-variance trade-off in the context of regularised regression.

2.1.1 References

- James, Witten, Hastie and Tibshirani (2013). An Introduction to Statistical Learning. Chapters 2.1, 2.2, 5.1, 6.2. Available here: `https://www-bcf.usc.edu/~gareth/ISL/`
- Varian, H. R. (2014). Big data: New tricks for econometrics. Journal of Economic Perspectives 28(2), 3–28
- Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. Journal of Economic Perspectives 31(2), 87–106

2.2 Practical session: Predicting poverty with satellite data

We will try out a number of classic machine learning techniques and ensemble learning using R in order to train a model to estimate poverty from satellite data. We will use the R package SuperLearner.

2.2.1 References

- SuperLearner package: `https://cran.r-project.org/web/packages/SuperLearner/SuperLearner.pdf`
- Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon (2016). Combining Satellite Imagery and Machine Learning to Predict Poverty. Science 353 (6301). American Association for the Advancement of Science: 790–94

3 Part 2: Double machine learning

3.1 Theoretical session: Estimating causal effects using ML

In the first part, we cover prediction methods. Such methods are not developed to estimate *causality*, which as econometricians is what we are most interested in. Nevertheless, there are some links between the prediction methods we have covered and causality. Linear regression is a machine learning method: we are trying to find the *best predictor* of Y given X, assuming that $E[Y|X]$ has a linear form. It just so happens that (under some assumptions) we can call the coefficients β of this linear form 'causal effects'. This leads us to ask whether or not there are other machine learning methods that we can use to estimate a 'causal effect' β.

We can think of the link between prediction and causality in terms of counter-factuals. Assume we have a treatment $X \in \{0, 1\}$ that causes an outcome Y, with controls W. The effect of a treatment on an individual i is $E(Y_i|X_i = 1, W_i) - E(Y_i|X_i = 0, W_i)$, but only one of these terms are observed. The problem is to find a good *predictor* of what would have happened in the counter-factual situation.

Using this idea, we learn how to obtain estimates of a treatment variable using double machine learning.

3.1.1 References

- Belloni, A., V. Chernozhukov, and C. Hansen (2014a). High-dimensional methods and inference on structural and treatment effects. Journal of Economic Perspectives 28(2), 29–50.
- Belloni, A., V. Chernozhukov, and C. Hansen (2014b). Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies 81(2), 608–650.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal 21(1), C1–C68.

3.2 Practical session: Coding up the Double ML estimator in R

We create a function in R to implement the double machine learning estimator. We use simulated data in order to demonstrate unbiasedness and to compare the performance of this estimator with alternative methods.

4 Part 3: Heterogenous treatment effects

4.1 Theoretical session: Estimating and discovering treatment effect heterogeneity

Here we consider the problem of estimating and discovering treatment effect heterogeneity. Technically, this amounts to (non-parametrically) estimating the conditional average treatment effect (CATE). We consider 'causal trees' (Athey and Imbens, 2016) and 'causal forests' (Wager and Athey, 2018). A simple exposition of their methods is given in Davis and Heller (2017). An alternative approach that we will also consider is proposed in Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018).

4.1.1 References

- Athey, S. and G. Imbens (2016). Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences 113(27), 7353–7360.
- Wager, S. and S. Athey (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Journal of the American Statistical Association 113(523), 1228–1242.
- Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogenous treatment effects in randomized experiments. Working paper.

4.2 Practical session: Coding up different heterogenous treatment effects estimators in R

We discuss and implement four different methods to estimate heterogeneous treatment effects:

1. OLS with interaction terms
2. Post-selection Lasso
3. Honest Trees
4. Causal Forests

We compare the heterogeneity identified by each of these methods.
Finally, we compute the Sorted Group Average Treatment Effects (GATES).

5 Exam

The assessment is based on 3 short assignments based on the three practical sessions. The first assignment is a data challenge competition on classifying tweets using machine learning techniques. The second assignment is on implementing an alternative version of the double ML estimator and applying this technique on data from a randomised control trial. The third assignment is on measuring heterogeneity in a large randomised control trial.