Machine Learning for Economists and Policymakers: Prediction, Classification and Causal Effects

Online Course

May 2020

Dr. M. J. Weeks

Associate Professor of Economics
University of Cambridge

This course, presented in a two part sequence, will review the application of machine learning techniques to both prediction problems and so-called causal problems where a firm or policy maker needs to understand the impact of some form of intervention on a heterogeneous population.

One example, is a firm that wishes to understand how the introduction of a change in pricing impacts both aggregate demand, and the demand on different segments of the population. In another example, a policymaker seeks to understand the impact of an intervention both in terms of some form of average effect, but also how individuals differ in the magnitude of the effect. Examples include the impact of job training programmes, the impact of education policies in developing economies, and the differential impact of drugs on survival and recovery.

In this context we make the distinction between the ex post assessment of a change and the ex ante identification of characteristics of individuals that are predictive of the likely impact of such a change.

Using Breiman's (2001) notion of two cultures in the use of statistical modelling, the course begins with a review of the fundamental differences between machine learning and econometrics.

> *There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown.*
>
> Breiman [2001], p199.

We contrast a modelling approach where the analyst makes certain assumption on model specification, including functional form, with an approach where the data mechanism is presumed unknown. In this context we consider the econometrician's concern for internal validity, alongside the focus within machine learning of ensuring that a model is robust in the sense of generalising to unseen data (external validity).

The course will focus upon topics at the intersection of machine learning and econometrics, covering a mix of theory and applications. In making the distinction between models which are used to solve a prediction problem and models which are used to estimate some form of causal effect, we introduce participants to identification strategies in econometrics. Here it is important to demonstrate how empirical strategies such

as unconfoundedness, instrumental variables, and difference-in-difference can be used alongside machine learning methods for prediction.

As a point of departure we make reference to the two broad types of machine learning in terms of supervised and unsupervised learning, making the link to nonparametric regression. We then consider a number of fundamental building blocks, starting with error decomposition in terms of bias and variance, the role of training, estimation and test samples, and the role of regularization as a means to avoid overfitting.

In covering the three broad areas where machine learning is used, namely prediction, classification and causal effects, for each case we link the exposition to a parametric benchmark. For prediction we consider the piecewise nonlinear regression model, for classification we review the fundamentals of parametric binary choice models, and for causal effects we consider the specification of models with instrumental variables and treatment effects.

Participants will also be introduced to the use of ensemble methods as an averaging and regularization device. In this context we will explore a number of general methods for model averaging including bootstrap sampling (so-called bagging) and random forests.

For Machine Learning models in prediction, classification and causal effects we provide examples using Stata and `Python`.

Application
Causal Forest Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes

The introduction of time-of-use electricity prices is an example of a policy with heterogeneous effects. Consumers in different socioeconomic groups and with distinct historical intra-day load profiles and behavioural characteristics, may respond differently to the introduction of tariffs that charge different prices for electricity at different times of the day. Customers who can (cannot) adapt their consumption profile to TOU tariffs will accrue a benefit (cost). Those who consume electricity at more expensive peak periods, and who are unable to change their consumption patterns, could end up paying significantly more.

Analysts often describe subpopulations that are of interest a priori, and which can be defined by a known combination of covariates. However, increasingly researchers face a selection problem given a large number of possible covariates alongside uncertainty as to which covariates are important for heterogeneity, and what functional form best describes the association between these covariates and treatment effects.

In assessing whether demographic variables are informative in terms of the impact of TOU tariffs on load profiles, the Customer-Led Network Revolution project noted

> .. a relatively consistent average demand profile across the different demographic groups, with much higher variability *within* groups than *between* them. This high variability is seen both in total consumption and in peak demand.

In addition, the question of which demographic variables are important when considering the impact of energy policies ignores the fact that many of these variables should be considered together, in a multiplicative fashion. One reason for this finding might be that, for example, it is the (unknown) combination of income, household size, education, and daily usage patterns that describes a particularly responsive or unresponsive group.

Throughout the course we make reference to the problem of identifying the distributional effects of some intervention, without succumbing to the problems of data mining (multiplicity). Here we examine the empirical problem of identifying the characteristics of winners and losers subsequent to the introduction of TOU tariffs following the introduction of a Time-of-Use (TOU) pricing scheme where the price per kWh of electricity usage depends on the time of consumption. The pricing scheme is enabled by smart meters, which records consumption every half-hour.

Using machine learning methods we describe the association between the effect of TOU pricing schemes on household electricity demand and a range of variables that are observable before the introduction of the new pricing schemes.

TEXTBOOKS

L. Breiman, J. Freidman, R. Olshen, C. Stone. *Classification and Regression Trees.* Klein-Verlag, 1990.

J. Freidman, T. Hastie, R. Tibshirani. *The Elements of Statistical Learning.* Springer, 2009.

Day 1

Outline:

The course is designed to provide both the tools to undertake projects using machine learning (ML), and critically ensure that participants understand and can communicate how the methods work.

Towards this objective, on Day 1, Session 1 we introduce participants to the vernacular of machine learning tools.

In Session 2 will further explore the links between ML, econometrics and data mining. We also examine how ML utilise data mining tools, suitably adapted to allow inference.

The course is designed in such as way to ensure that participants are given the necessary context to understand the genesis of ML methods. To this end, the first point of departure reviews the ordinary least squares estimator and provides links to ML using kernel density estimation.

We also provide the necessary links to econometrics and nonparametric statistics.

Session 1 Introduction

1. High-level overview of Machine Learning and AI
2. Machine Learning: The Vernacular
3. The Nature of Prediction Problems
4. Prediction, Evaluation and Causal Inference

Session 2 Machine Learning: Tools and Vernacular

1. Econometrics
2. Machine Learning: Tools and Vernacular
   - i Bias Variance Tradeoff
   - ii Regularisation
   - iii Multiplicity and P-values
   - iv Ensemble Learning
3. Point of Departure I: The Ordinary Least Squares Estimator

Day 2

Outline:

Day 2, Session 1 begins with the second point of departure - high dimensional methods in statistics. These methods are used when analysts face a big data problem in terms of which of a large set of explanatory variables to include in a regression model.

We follow this with a practical where participants can explore the use of regularised regression tools with a number of empirical applications.

In session 2 we provide an introduction to a number of machine learning methods including regression trees and forests.

This is then followed by a practical where we examine the use of ML methods for prediction.

Session 1 Point of Departure II: High Dimensional Methods

1. High Dimensional Methods
2. Least absolute Shrinkage and Selection (LASSO)
   - i Choosing λ
   - ii Causal Inference in High-Dimensions
   - iii LASSO For Treatment Models
   - iv Double LASSO
3. Practical: Regularized Regression

Session 2 Machine Learning and Decision Trees

1. Machine Learning and Decision Trees
   - i Machine Learning: Terminology and Concepts
   - ii An Overview of Regression Trees
   - iii The Bias-Variance Tradeoff
   - iv Training, Testing and Cross Validation
   - v Regularization: Variance reduction and Ensemble Learning
2. Practical: Machine Learning for Prediction

Looking ahead

In Part II, a follow-up course, we explore in more detail some of the concepts introduced in Part I. In addition, we extend the coverage to include machine learning methods for both classification problems and for causal effects.

With regards the latter, this material covers a relatively new and rapidly expanding field, where the potential of ML is applied to policy problems. A useful point of departure here is the work of this years Nobel Laureate Esther Duflo.

END

PART II: MACHINE LEARNING FOR CLASSIFICATION AND CAUSAL EFFECTS

SUMMARY AND OBJECTIVES

DAY 1

Outline:

In Part II, the second installment of this two part course, we explore in more detail some of the concepts introduced in Part I, and extend the coverage to include machine learning methods for both classification problems and for causal effects.

The material on causal effects takes as its point of departure the literature on treatment effects and examines the potential of ML to address a number of policy problems. A useful point of reference here is the work of this years Nobel Laureate, Esther Duflo.

Part II of this course is also constructed so that participants can take this module without needing to take Part I. That said, the overall learning experience is greater if participants take Part I and II as a sequence.

On Day 1, Session 1, we review some of the fundamentals of machine learning that were introduced during Part 1. This includes the use of ML for prediction, classification and causal effects, alongside the key methodological concepts such as the bias-variance tradeoff and methods to achieve regularisation.

In Session 2 we examine the use of ML tools applied to so-called classification problems. A useful frame of reference here is the decision to grant a loan or provide some form of service such as insurance. These models utilise characteristics of individuals/firms to understand the determinants of key outcomes such as loan default or an excess number of claims.

SESSION 1 Introduction

1. Review of PART I
2. Bias-Variance trade-off, overfitting and prediction
3. Prediction, Evaluation and Causal Inference

SESSION 2

1. Classification problems
2. Parametric Benchmarks: binary choice models
3. Application: Surviving the Titanic with Python integration
4. Application: Credit Card default

Day 2

Outline:

Day 2, Session 1 begins with the third point of departure - programme evaluation and treatment effects. We will review the econometric methodology which includes methods to handle both endogeneity and move away from parametric functional forms. We make reference to the work of the Nobel Laureate Esther Duflo who has made significant contributions to the use of randomised control trials, in addition to the utilisation of machine learning methods in this context.

In Session 2 we examine the use of machine learning methods for causal inference. Relative to many of econometric methods, studies which employ ML techniques have sort to exploit so-called big data to provide a coherent approach to uncover variation in treatment effects without succumbing to the pitfalls of data mining.

This is followed by a practical where we examine the use of ML methods applied to the impact of time-of-use electricity on individual-level demand response. A key question here is whether it is possible to identify characteristics of households that enable policy makers to identify so-called winners and losers once we move to a price system where prices vary throughout the day.

Session 1 Point of Departure III: Programme Evaluation and Treatment Effects

1. Overview
2. Ignorability of Treatment
3. Endogenous Selection
4. Matching Estimators
5. The Difference-in-Difference Estimator
6. Application: Job Training Programs

Session 2: Machine Learning and Causal Inference

1. Causal Trees
2. Honest Estimation
3. Forests and Variance Reduction Methods
4. Testing for Heterogeneity in Treatment Effects
5. Application: Time of Use Tariffs and Smart Meter Data

END

Readings

Machine Learning: Overview

[1] L. Breiman, J. Freidman, R. Olshen, C. Stone. *Classification and Regression Trees.* Klein-Verlag, 1990.

[2] *Random Forests.* `https://en.wikipedia.org/wiki/Random_forest.`

[3] *Training, Validation, and Test sets.* `https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets`

[4] J. Freidman, T. Hastie, R. Tibshirani. *The Elements of Statistical Learning.* Springer, 2009.

[5] G. James, D. Witten, T. Hastie, R. Tibshirani. *An Introduction to Statistical Learning with Applications in R.* Springer, 2013.

[6] S. Russell, P. Norvig Artificial Intelligence: A Modern Approach *3rd edition, 2009*

Machine Learning and Econometrics

[1] L. Breiman *Statistical Modeling: The Two Cultures* Statistical Science, Vol. 16, No. 3. pp. 199-215

[2] S. Athey *The Impact of Machine Learning on Economics.* in, The Economics of Artificial Intelligence: An Agenda, 2018. National Bureau of Economic Research. See `http://bit.ly/2EENtvy` S. Athey,G. Imbens *Machine Learning Methods Economists Should Know About.* Working Paper, 2019, Graduate School of Business, Stanford University.

[3] S. Mullainathan, J. Spiess. *Machine Learning: An Applied Econometric Approach* Journal of Economic Perspectives vol. 31, 2017, pp. 87-106.

[4] A. Belloni, V. Chernozhukov, C. Hansen. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives, 28(2):29-50, 2014(a)*

Policy Problems

[1] Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015), *Prediction policy Problems.* American Economic Review, 105(5): 491-495.

[2] Andini, M., Ciani, C., Blasio, G. and D'Ignazio, A. (2018). *Effective Policy Targeting Machine Learning* `https://voxeu.org/article/effective-policy-targeting-machine-learning`

[3] Andini, M., Ciani, C., Blasio, G., D'Ignazio, A. and Paladini, A. (2018). *Machine learning in the service of policy targeting: the case of public credit guarantees.* Banca d'Italia Temi di discussione.

[4] Chalfin, A, O. Danieli, Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., and S. Mullainathan (2016), *Productivity and selection of human capital with machine learning*, American Economic Review, 106(5): 124-127.

[5] Athey, S. (2017). *Beyond prediction: Using big data for policy problems.* Science 03, Vol. 355, pp. 483-485

Treatment Effect Models

[1] Angrist, J.D. and J.-S. Pischke, (2009), *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

[2] Joshua D. Angrist (2004) *Treatment Effect Heterogeneity in Theory and Practice.* The Economic Journal, Vol. 114.

[3] LaLonde, R.J. 1986. *Evaluating the econometric evaulations of training programs with experimental data.* American Economic Review, Vol.76, No.4, pp. 604-620.

[4] Dehejia, R., and Wahba, S. (1999). *Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs*, Journal of the American Statistical Association, Vol. 94, No. 448, pp. 1053-1062.

[5] Blundell, R.W., and M. Costa Dias (2002): *Alternative Approaches to Evaluation in Empirical Microeconomics.* Portuguese Economic Journal, 1, 91-115. `http://cemmap.ifs.org.uk/wps/cwp0210.pdf`

[6] Imbens, G., and J. Wooldridge (2009): *Recent Developments in the Econometrics of Program Evaluation.* Journal of Economic Literature, 47(1): pp. 5-86.

[7] Athey, S., and G. Imbens (2017). *The State of Applied Econometrics: Causality and Policy Evaluation.* Journal of Economic Perspectives, 31 (2): 3-32.

[8] Chib, S. and B.H. Hamilton, (2000), Bayesian analysis of cross-section and clustered data treatment models, *Journal of Econometrics*, 97(1), 25-50

[9] Munkin, M.K. and P.K. Trivedi, (2003), Bayesian analysis of a self-selection model with multiple outcomes using simulation-based estimation: An application to the demand for healthcare, *Journal of Econometrics*, 114(2), 197-220

[10] Li, M, and J. Tobias, (200*), Bayesian analysis of Treatment Effects in an Ordered Potential Outcomes Mode1, Advances in Econometris,

LASSO

[1] Efron, B., Hastie, T., Johnstone, I. and R. Tibshirani, (2004). Least angle regression (with discussion), *Annals of Statistics* 32(2): 407- 499

[2] Friedman, Jerome; Hastie, Trevor; Tibshirani, Robert (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition (Springer Series in Statistics) (Kindle Locations 13024-13026). Springer - A. Kindle Edition.

[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, Vol. 58, No. 1, pages 267-288).

[4] Hofmarcher, P., Cuaresma, J., Grun, B., and K. Hornik (2015). Last Night a Shrinkage Saved My Life: Economic Growth, Model Uncertainty and Correlated Regressors, *Journal of Forecasting*, Vol. 34, pages 133-144

[5] Park, T., and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681-686.

[6] O'Hara, R. B., & Sillanpaa, M. J. (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian analysis*, 4(1), 85-117.

[7] Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5(2), 369-411.

[8] Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273-282.

[9] Belloni, A., Chernozhukov, V. and C. Hansen (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects, Journal of Economic Perspectives, Vol. 28, no. 2 pp. 29-50,

[10] Belloni, A., Chernozhukov, V. and C. Hansen (2014). Inference on Treatment Effects after Selection amongst High- Dimensional Controls. Review of Economic Studies ...

[11] Imai, K., and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics* 7(1):443-470.

Machine Learning: Causal Effects and Random Forests

[1] S. Athey, G. Imbens. Recursive partitioning for heterogeneous causal effects *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

[2] E. O'Neill, M. Weeks. Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes arXiv:1810.09179v1, 2018.

[3] S. Athey, G. Imbens, Y. Kong, V. Ramachandra. An Introduction to Recursive Partitioning for Heterogeneous Causal Effects Estimation Using causalTree package. https://github.com/susanathey/causalTree/blob/master/briefintro.pdf, 2016.

[4] S. Athey, S. Wager. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 2017.

[5] Y. Lin, J. Yongho Random forests and adaptive nearest neighbors *Technical Report No. 1055. University of Wisconsin, 2002* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.153.9168

[6] S. Athey, G. Imbens, S. Wager(201x). *Estimating Average Treatment Effects: Supplementary Analyses*

[7] V. Chernozhukov Double/debiased machine learning for treatment and structural parameters *The Econometrics Journal*, 21 (1) pp. C1-C68.

[8] A. Belloni, V. Chernozhukov, C. Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies, 81(2):608-650, 2014(b).*

[9] A. Belloni, V. Chernozhukov, I. Fernandez-Val,C. Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica, 85(1): 233-298, 2017.*

[10] L. Breiman Random Forests *Machine Learning, Vol. 45, pp.5-32*

[11] *Random Forests.* https://en.wikipedia.org/wiki/Random_forest.

[12] *Training, Validation, and Test sets.* https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets