An Introduction to Machine Learning for Economics

Quentin Batista

January 28, 2019

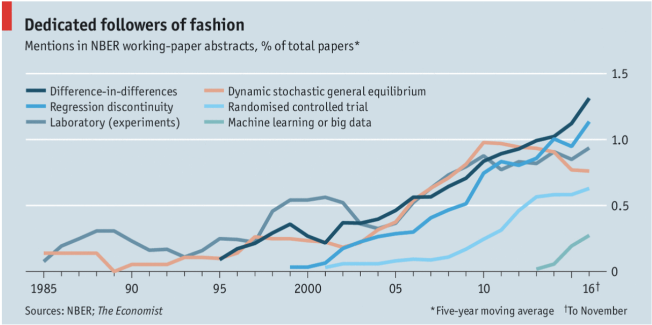
Machine Learning is the Latest Fad?

Economists are prone to fads, and the latest is machine learning

Big data have led to the latest craze in economic research

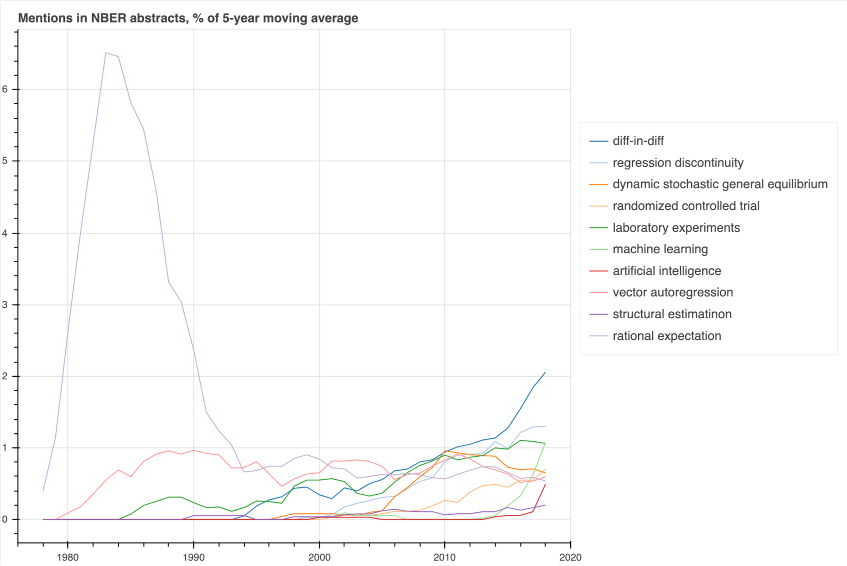
Print edition | Finance and economics ›

Nov 24th 2016



Economist.com

Machine Learning is the Latest Fad?



Application of Machine Learning: Predicting Demographic Makeup

Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US

Timnit Gebru[1], Jonathan Krause[1], Yilun Wang[1], Duyun Chen[1], Jia Deng[2], Erez Lieberman Aiden[3,4], Li Fei-Fei[1]

[1] *Stanford University, 353 Serra Mall, Stanford, CA 94305.*

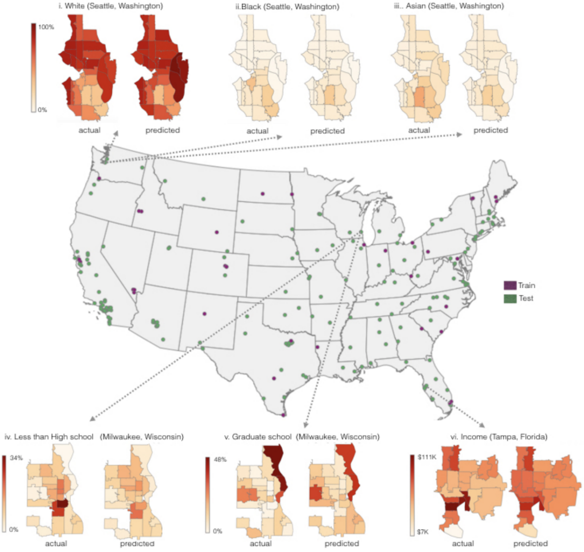
[2] *University of Michigan, 2260 Hayward Street, Ann Arbor, MI 48109.*

[3] *Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030.*

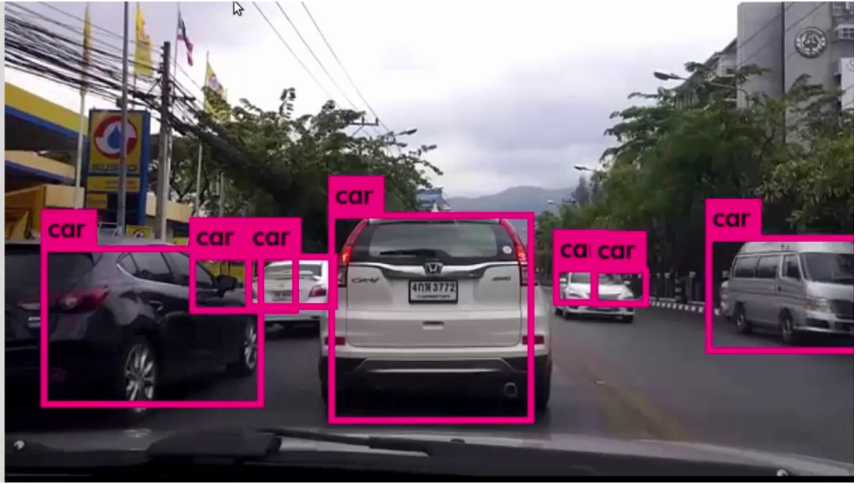
[4] *Rice University, 6100 Main Street, Houston, TX 77005.*

The United States spends more than $1B each year on initiatives such as the American Community Survey (ACS), a labor-intensive door-to-door study that measures statistics relating to race, gender, education, occupation, unemployment, and other demographic factors [1]. Although a comprehensive source of data, the lag between demographic changes and their appearance in the ACS can exceed half a decade. As

Application of Machine Learning: Predicting Demographic Makeup



Application of Machine Learning: Predicting Demographic Makeup



Application of Machine Learning: Measuring Labor Flows

USING SOCIAL MEDIA TO MEASURE LABOR MARKET FLOWS

Dolan Antenucci
Michael Cafarella
Margaret C. Levenstein
Christopher Ré
Matthew D. Shapiro

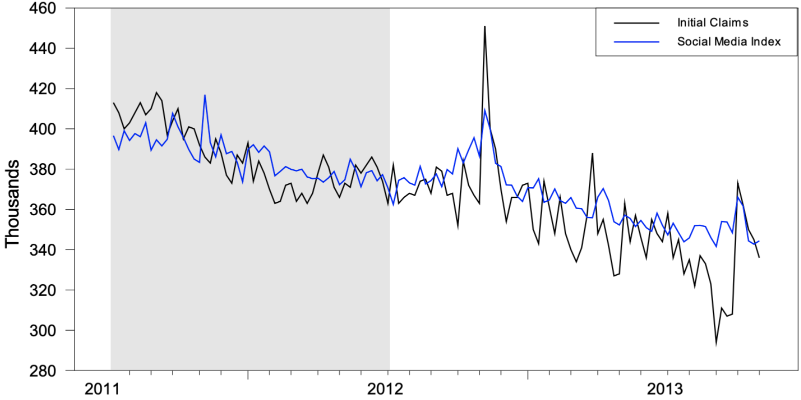
Working Paper 20010
http://www.nber.org/papers/w20010

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2014

This project is part of the University of Michigan node of the NSF-Census Research Network (NCRN) and is supported by the National Science Foundation under Grant No. SES 1131500. Ré is partially supported by the National Science Foundation CAREER Award under No. IIS-1353606. Cafarella and Antenucci are partially supported by the National Science Foundation CAREER Award under No. IIS-1054913. We are grateful to Tomaz Cajner, Michael Elsby, Linda Tesar, Kenneth West, Justin Wolfers, and seminar participants at the NBER Summer Institute, Vanderbilt University, the University of Wisconsin, the University of Michigan, and the Federal Reserve Board for helpful comments and to Mark Fontana for excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Application of Machine Learning: Measuring Labor Flows

Figure 3. Initial Claims for Unemployment Insurance and the Social Media Job Loss Index



Note: Figure shows the Department of Labor's Initial Claims for Unemployment Insurance and the Social Medial Job Loss Index. The Social Media Job Loss Index is estimated in sample in the shaded area and recursively thereafter. See text for details.

Application of Machine Learning: Predicting Income Using Images

BIG DATA AND BIG CITIES:
THE PROMISES AND LIMITATIONS OF IMPROVED MEASURES OF URBAN LIFE

Edward L. Glaeser
Scott Duke Kominers
Michael Luca
Nikhil Naik

Working Paper 21778
http://www.nber.org/papers/w21778

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2015

We would like to acknowledge helpful comments from Andy Caplin, William Kominers, and Mitchell Weiss. E.L.G. acknowledges support from the Taubman Center for State and Local Government; S.D.K. acknowledges support from the National Science Foundation (grants CCF-1216095 and SES-1459912), the Harvard Milton Fund, the Wu Fund for Big Data Analysis, and the Human Capital and Economic Opportunity Working Group (HCEO) sponsored by the Institute for New Economic Thinking (INET); and N.N. acknowledges support from The MIT Media Lab consortia. Comments are welcome and may be sent to eglaeser@harvard.edu, kominers@fas.harvard.edu, mluca@hbs.edu, and naik@mit.edu. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Machine Learning

- “Machine learning is the scientific study of algorithms and statistical models that computer systems use to improve their performance on a specific task.” (Wikipedia)
- “Machine learning is essentially a form of applied statistics with increased emphasis on the use of computers” ([1])
- “The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest”. ([2]).

Machine Learning vs. Econometrics

- Fit and empirical performance vs. asymptotic properties
- Algorithms vs. estimation
- Reference: Susan Athey

Machine Learning vs. Econometrics

- Suppose you have the following model: $y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim N\left(0, \sigma^2\right)$
- You want to solve $\min_\alpha E\left[(y - \hat{\mu})^2\right]$ such that $\hat{\mu} = \alpha \bar{y}$
- The solution is $\alpha^* = \frac{\mu^2}{\mu^2 + \frac{\sigma^2}{n}}$
- The estimator that minimizes mean squared error is **biased**

What is a Learning Algorithm?

Definition

Learning Algorithm (Mitchell 1997)
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.

- The task T is what we want the computer to achieve
- The performance measure P is a quantitative measure of prediction ability
- The experience E is some form of data

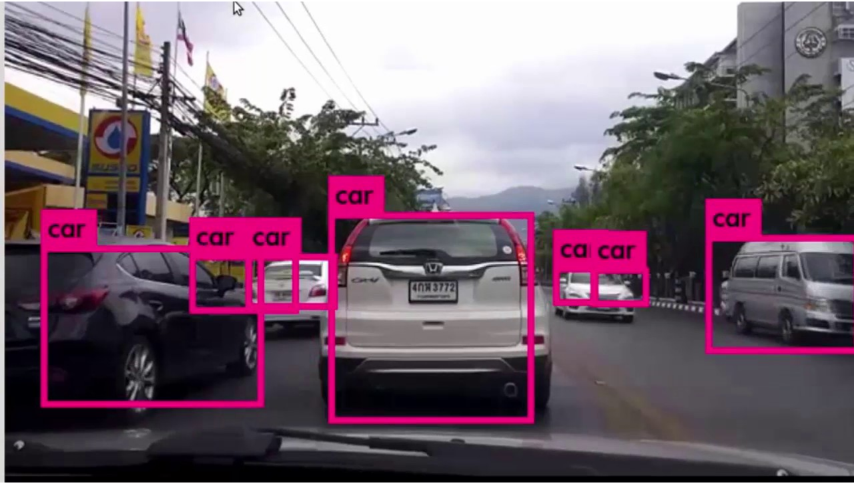
Different Types of Data = Different Types of Learning

Supervised Learning
Goal: Learn a mapping from inputs x to outputs y given $\mathrm{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$
$\mathbf{x}_i \in \mathbb{R}^D$, $y_i \in \mathbb{R}$
$(\mathbf{x}_i, y_i)$ is called an **example**

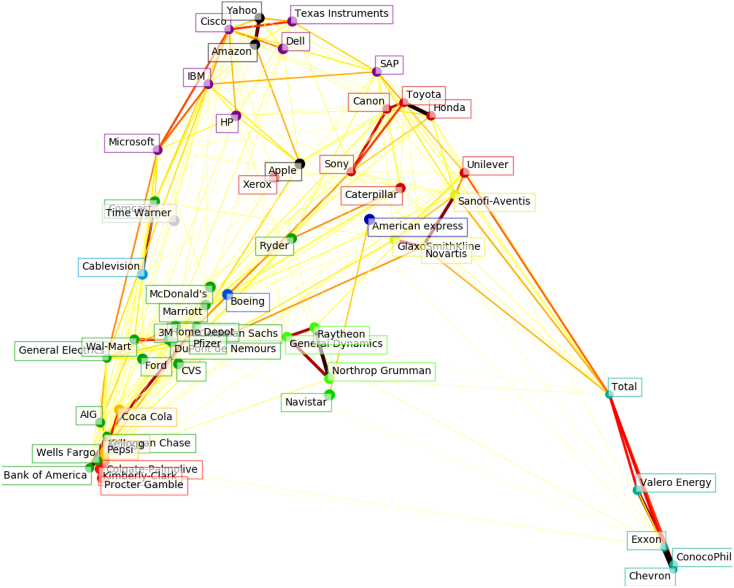
Unsupervised Learning
Goal: Find "interesting" patterns in $\mathrm{D} = \{\mathbf{x}_i\}_{i=1}^N$
$\mathbf{x}_i \in \mathbb{R}^D$

- $\mathbf{x}$ represents the set of **features** or **covariates**
- y represents the **target** variable or **label**

Supervised Learning: An Example



Unsupervised Learning: An Example



Source: https://sklearn.org/auto_examples/applications/plot_stock_market.html

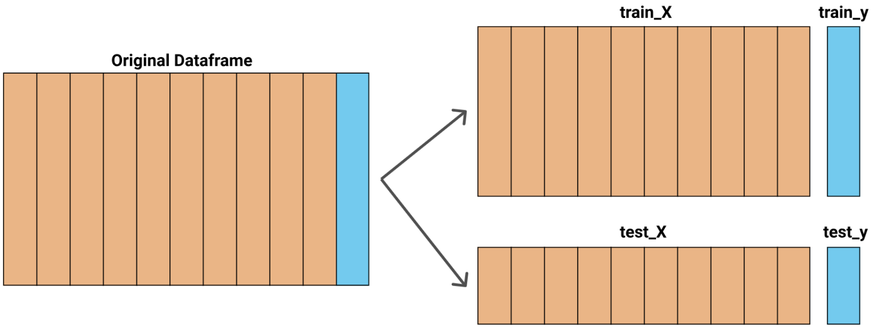
Capacity, Overfitting and Underfitting

- The objective of ML algorithms is to perform well on previously unseen examples, i.e., **generalize**
- However, performance on unseen inputs is not observable
- Question 1: How do you estimate performance on unseen examples?
- Question 2: How do you choose among different models?

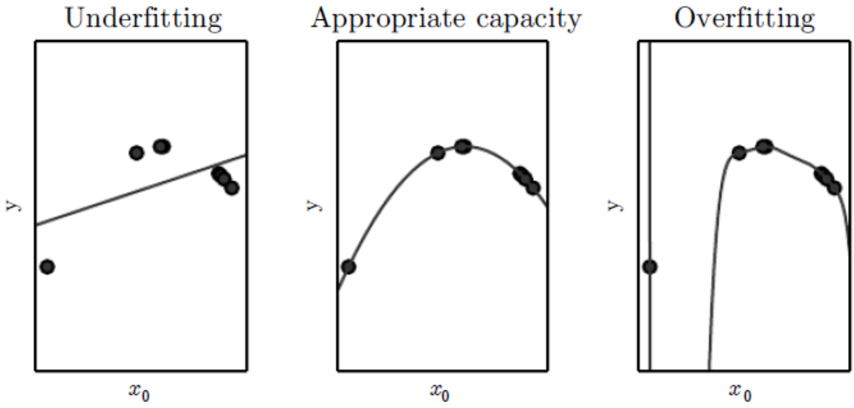
Estimating Generalization Performance

- Split your dataset into a **training** set and a **test** set
- Estimate the performance of your model on the test set to get an unbiased estimate of the performance
- About 30% of your data for small datasets
- About 1% of your data for large datasets

Estimating Generalization Performance



Capacity, Overfitting and Underfitting

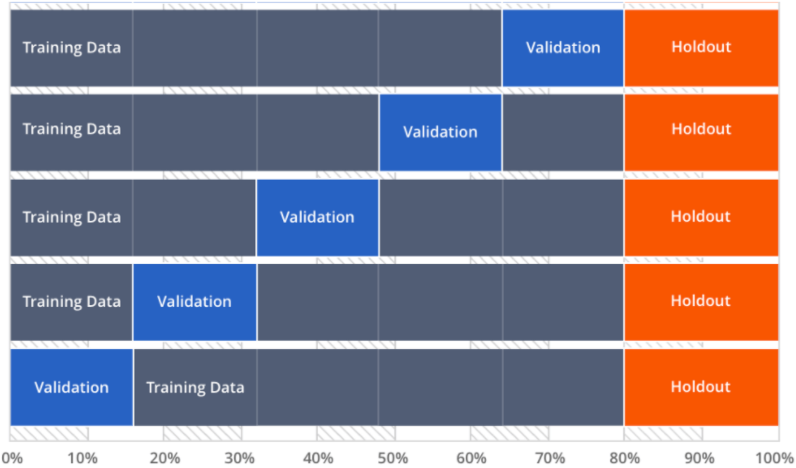


Choosing Among Models

Hold-out validation scheme

- Split your **training** set into a **training (!)** set and a **validation** set
- Choose the model with the best performance on the validation set

Cross-Validation



Regularization

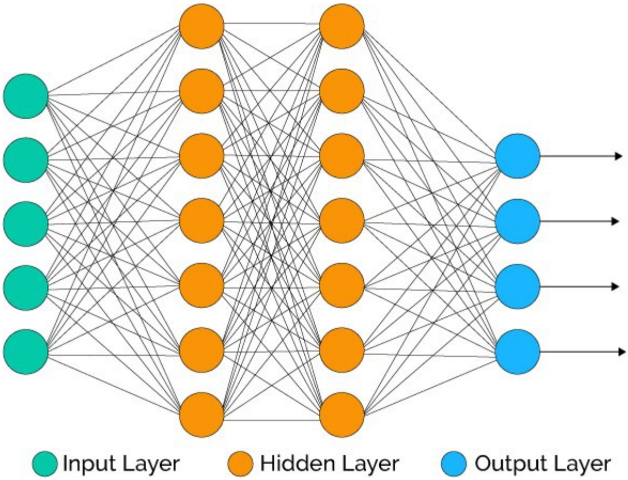
- The idea of regularization is to penalize complex models

Example: Linear Regression

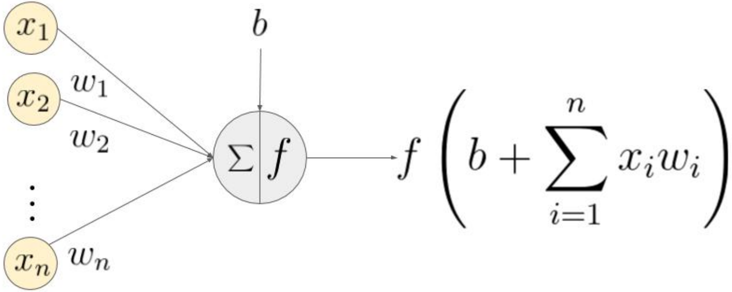
$$\min_\beta \left\{ \frac{1}{\mathrm{n}} \|\mathrm{y} - \mathrm{X}\beta\|_2^2 + \lambda \mathrm{R}(\beta) \right\}$$

- LASSO: $R(\beta) = \|\beta\|_1 = \sum_{i=1}^{n} |\beta_i|$
- Ridge: $R(\beta) = \|\beta\|_1 = \sum_{i=1}^{n} \beta_i^2$
- Choose λ using the validation scheme

Example: Neural Networks



Example: Neural Networks



An example of a neuron showing the input (x_1 - x_n), their corresponding weights (w_1 - w_n), a bias (b) and the activation function f applied to the weighted sum of the inputs.

Example: Neural Networks

https://playground.tensorflow.org/

References

Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
Deep Learning.
MIT Press, 2016.
http://www.deeplearningbook.org.

Kevin P. Murphy.
Machine Learning: A Probabilistic Perspective.
The MIT Press, 2012.