Machine Learning for Economists: An Introduction

Sonan Memon
Pakistan Institute of Development Economics, Islamabad

30th June 2021

Introduction

i. This webinar aims to provide a brief overview of machine learning and its economic applications.

ii. For more technical slides: https://github.com/sonanmemon/Introduction-to-ML-For-Economists.

iii. My 4 lecture videos: https://www.youtube.com/watch?v=E9dLEAZW3L4.

iv. Hastie, T., Tibshirani, R., & Friedman, J. (2017). The elements of statistical learning.

v. Alpaydin, Ethem, Introduction to machine learning, MIT Press, 2020.

Outline

i. Introduction to ML and Its Relation with Econometrics
ii. Applications in industry and other sciences.
iii. Economics and Development Applications
iv. LASSO, Multi-Armed Bandit Problems and Latent Dirichlet Allocation Methods
v. AI and Ethical Dilemmas
vi. Conclusion

What is Machine Learning (ML)?

i. ML refers to the set of algorithms and computational methods which enable computers to learn patterns from training data without being explicitly programmed to do so.

ii. ML uses training data to learn patterns and makes predictions in out of sample based on new input data.

iii. Capacity to find complex, flexible and crucially *generalizable* structure in training data.

iv. ML algorithms can be thought of as complex function approximation techniques, outperforming traditional methods.

v. $ML \subset AI$. $AI \neq ML$.

ML and Econometrics

i. Identification of parameters usually not valid, *prediction* is the main goal of ML.
ii. Lesson: look for $\widehat{y}$ problems and not $\widehat{\beta}$ problems.
iii. ML can learn from econometrics and econometrics can learn from ML.
iv. See Mullainnathan and Speiss (2016) and Varian (2014) for more on this.

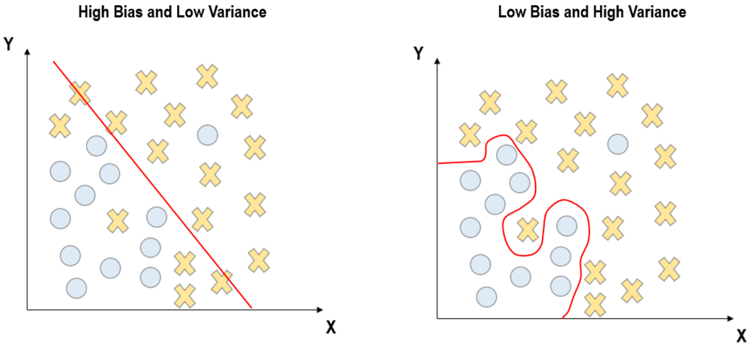
Typology of ML

i. ML can do *supervised learning* when the goal is to use data on X and Y to learn mapping $Y = f(X)$.
ii. Meanwhile, *unsupervised learning* digs out patterns and associations in input space X without data on output Y.
iii. Multi-armed bandits and Reinforcement Learning for Experimentation
iv. Causal Trees and Heterogeneous Treatment Effects

Variance Bias Trade off and Regularization

i. ML methods are powerful enough to fit nearly any input data very accurately which is why we get the *over fitting problem*.
ii. An overly precise fit of the data leads to low in sample bias but high out of sample variance and poor fit of the training data leads to low variance but high bias.
iii. We use *regularization* to tie the hands of the estimator so that it allows us to generalize beyond the training data.
iv. The extent of regularization is determined by *tuning parameter*.

Variance Bias Trade off



Cross Validation for Tuning Parameter

i. Split the data into K folds of roughly equal size. The hold out method is repeated K times by using each subset only once for testing only and all others for training only.

ii. Each time k is chosen to be test set, train and estimate model on remaining subsets.

iii. We choose tuning parameter so that the average *cross validation* error across the K folds is minimized.

Cross Validation

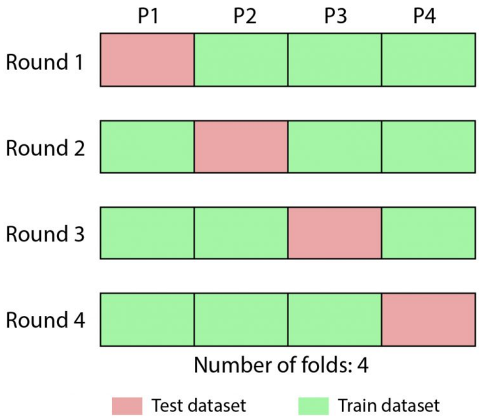
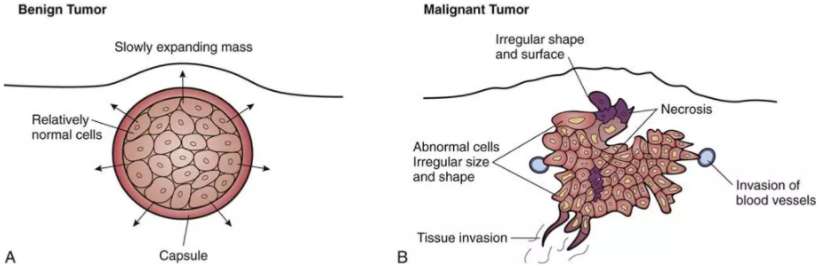


Figure: Cross Validation

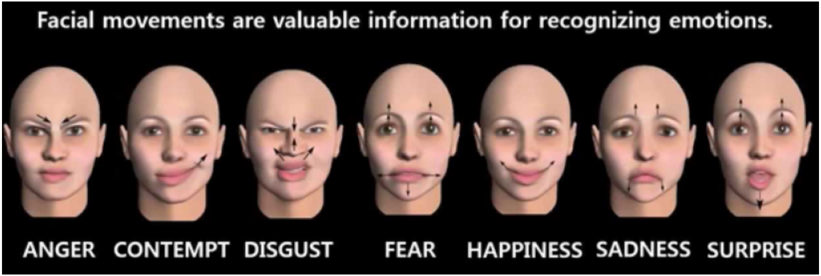
Applications Outside Economics

i. DeepFace for Face Recognition and Emotion Recognition.
ii. Deep Mind's AlphaGo program.
iii. Spam Filtering and Optimal Character Recognition.
iv. Medical Diagnosis and Genomics.
v. Speech Recognition, chat bots, automated translation and natural language processing.
vi. Self Driving Cars.

ML in Medical Diagnosis



Emotion Recognition



Applications in Economics

i. LASSO for macroeconomic forecasting, big data in neuroeconomics and big scanner data from supermarkets.
ii. Using multi-armed bandits to allocate optimal labor market interventions to improve job finding rate for Syrian refugees.
iii. Using LDA to analyze central bank communication McMahon and Hansen (2016) and understand the impact on news reporting on household inflation expectations Larsen et al (2021).

Applications in Development

i. Using satellite, mobile CDR data or a combination of the two to predict poverty for high spatial granularity and at high frequency and low cost Steel et al (2017); Aiken et al (2020).
ii. Improving tax compliance in India by using ML to identify "suspicious" firms which are less likely to file tax returns Mittal et al (2018).
iii. Using nighttime data and day time satellite data to measure the extent of urbanization, which can help in urban planning Goldbatt et al (2018).

Applications in Development

i. Parthasarathy et al (2017) use topical modeling and other text as data methods on an original corpus of village assembly transcripts from rural Tamil Nadu, India.

ii. They find that women are at a disadvantage relative to men; women are less likely to speak, set the agenda, and receive a relevant response from state officials.

iii. Although quotas for women on village councils have little impact on the likelihood that they speak, they improve the likelihood that female citizens are heard.

Voronoi Polygons for Dhaka and Bangladesh

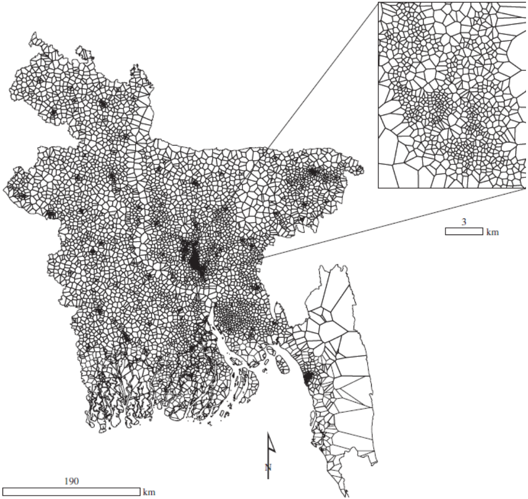


Figure: Source is Steel et al (2017)

Poverty Map for Bangladesh

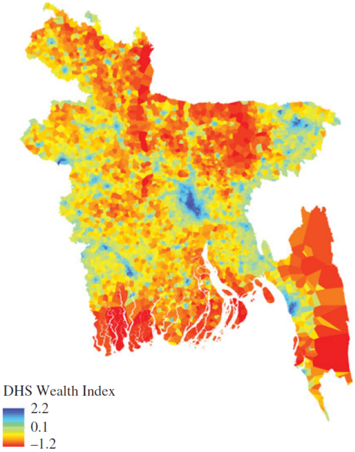


Figure: Source is Steel et al (2017)

Bogus Firms and Tax Evasion

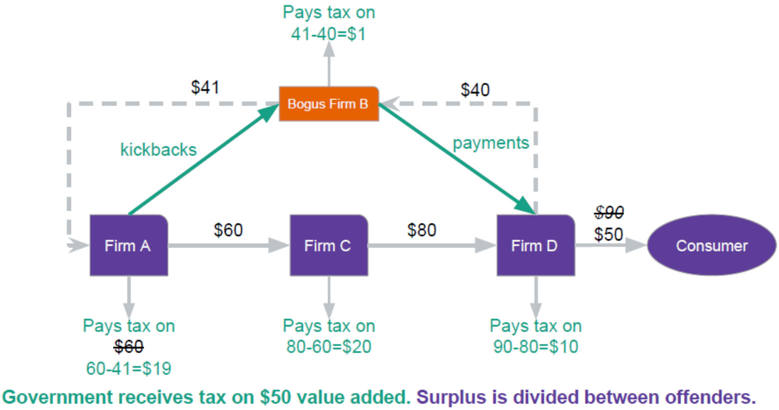


Figure: Source is Mittal et al (2018)

Using ML to Improve Tax Compliance

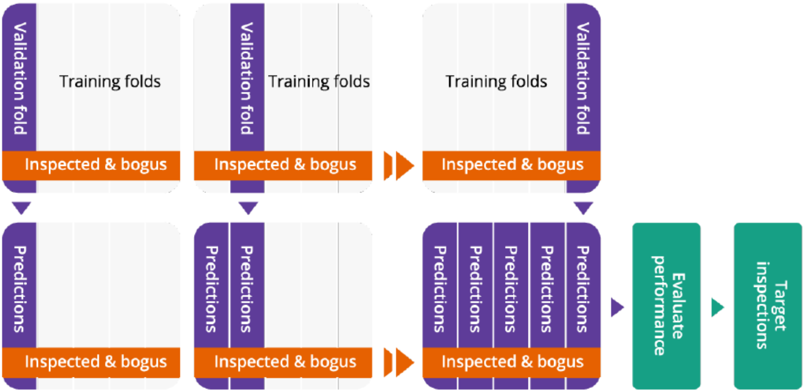


Figure: Source is Mittal et al (2018)

Classification of Lit Up Areas

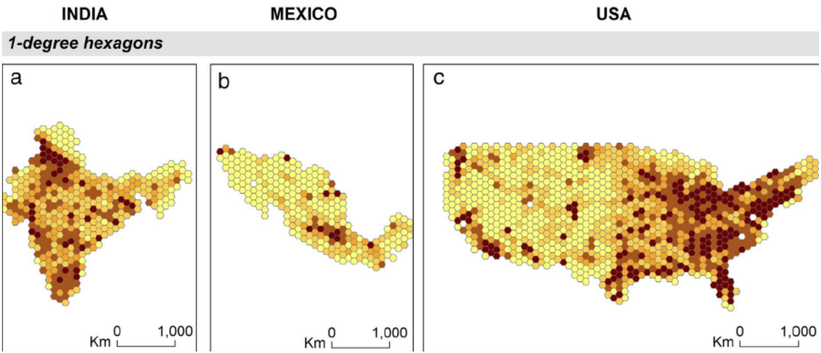


Figure: Source is Goldbatt et al (2018)

Illustration of Built (Bottom) Versus Non-Built Areas

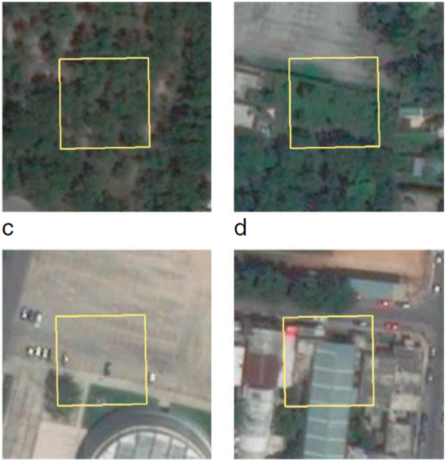


Figure: Source is Goldblatt et al (2018)

Classification of Built Up Areas in India

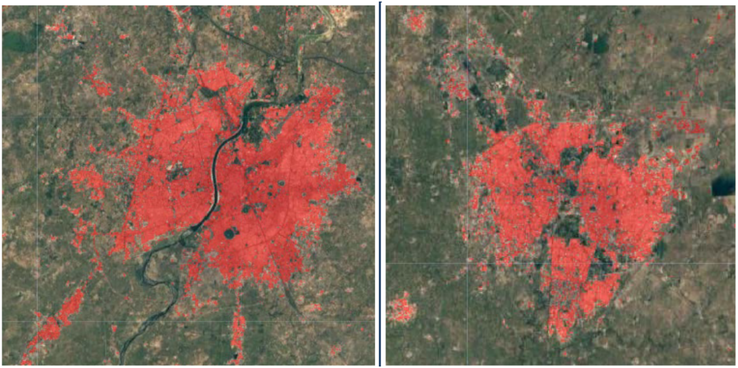


Figure: Source is Goldblatt et al (2018)

Classification of Built Areas in India

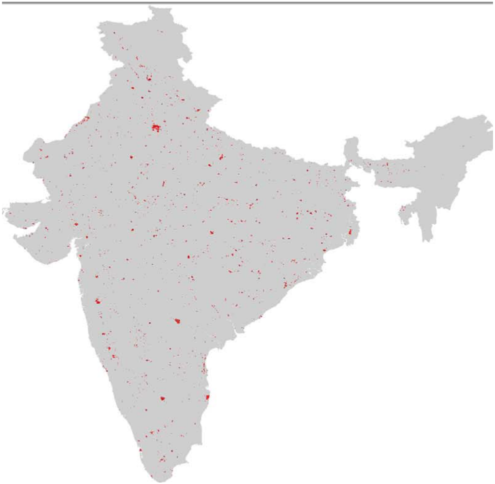


Figure: Source is Goldblatt et al (2018)

Women are Not Heard

Female 1: In Pattupalli village, so far, **there is no fair price shop.** They are keeping it in the Women's Health Building. Women are quarrelling. The village people want it built new. There is fight in the panchayat. So people are going to the neighboring village. But the patta [titled] land owners are preventing them from using their land for going to the next village, so resolution should be passed for construction of a ration shop here.

Male 1: For so many years, **there is no ration shop here**. Only rental shops are here. So long, it was in rented place and now it is kept in Women's Health Association. Now, women ask for the building and want a fair price shop built. So there is a lot of problem. Please establish for us a ration shop.

Male (President): **Regarding this ration shop, we should talk with MLA** [Member of the Legislative Assembly] and BDO [Block Development Officer]. The request will be made...

Figure: Source: Parthasarathy et al (2017)

Overview of Methods

i. Supervised Learning: LASSO, deep neural nets, regression trees, random forests and classification trees.
ii. Unsupervised Learning: LDA, clustering algorithms.
iii. Experimental Settings: Multi-Armed Bandit Problems.
iv. I will briefly talk about LASSO, multi-armed bandit problems and LDA.

LASSO

i. LASSO solves:
$\widehat{\beta_{lasso}} = \underset{\beta}{argmin} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\}$ s.t $\sum_{j=1}^{p} |\beta_j| \leq c$.

ii. Making c sufficiently small with cause some of the coefficients to be exactly zero $\implies$ *selection*.

iii. The retained set of coefficients will also be shrunk toward zero since LASSO favors *sparsity* $\implies$ *shrinkage*.

iv. Used for optimal variable selection when number of covariates is large.

Multi-Armed Bandits (MAB)

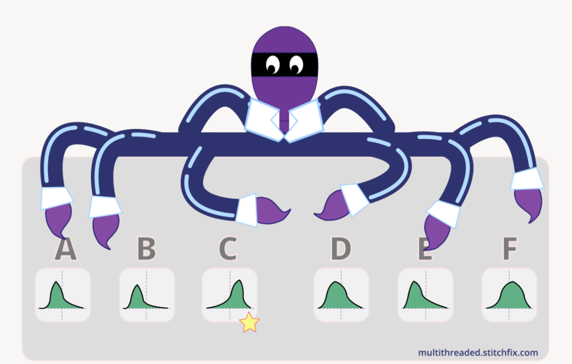
i. A/B testing is inefficient since we allocate fixed number of units to each treatment, some of which may be sub-optimal.
ii. With MAB's, we have prior treatment assignment probabilities for every arm, which are updated over time as we assign incoming units to various treatments and observe outcomes.
iii. MAB helps us learn the pay off distribution of various treatments more optimally.
iv. MAB solves the *exploration* versus *exploitation* trade off.

Multi-Armed Bandits (MAB)

i. In a Bernoulli context, where each arm can either return success or failure, we use the Beta distribution for each arm since it has support over $[0, 1]$.
ii. The octopus diagram on upcoming slide illustrates the updating of Beta distributions over time for each of the arms.
iii. With *contextual bandits*, the payoffs of various treatments can vary by individual characteristics e.g genetics, sex and age can determine the success distribution for medical drugs.

Two Armed Bandit in Casinos

Multi-Armed Bandit



Application Based on Cario et al (2021)

i. Cario et al (2021) use contextual bandits for adaptive, targeted treatment assignment in a field experiment, focused on improving job finding rate for Syrian refugees in Jordan.

ii. The algorithm balances the goal of maximizing participant welfare and precision of treatment effect estimates.

iii. After four months, cash has a sizable effect on employment and earnings of Syrians but information provision and psychological nudge do not.

Adaptive Treatment Assignment

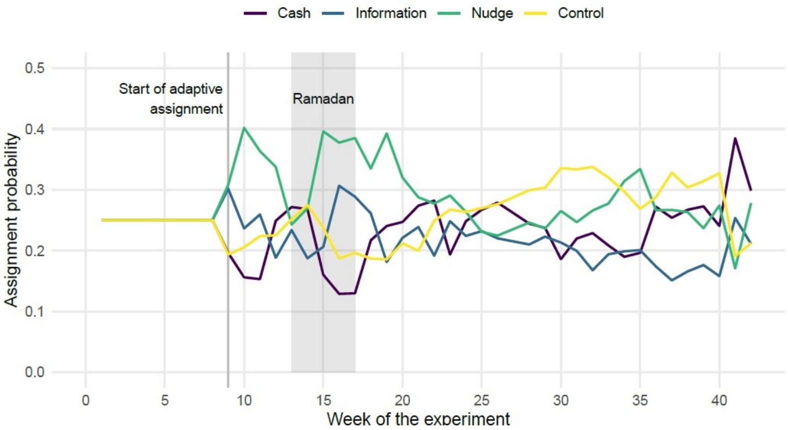
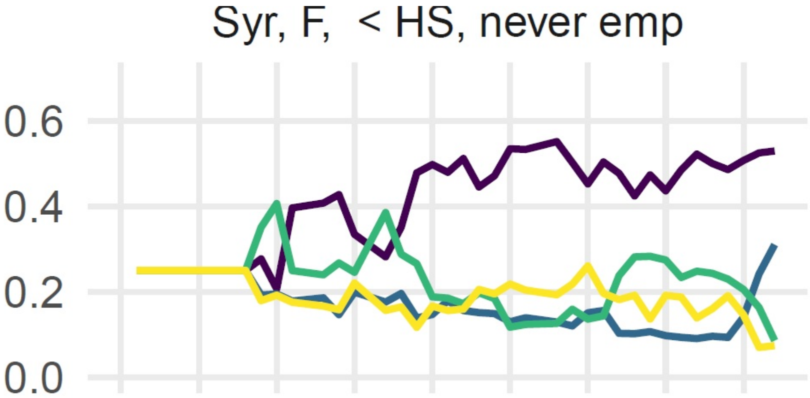


Figure: Source: Cario et al (2021)

Treatment Assignment for Specific Strata



Latent Dirichlet Allocation

i. LDA is a hierarchical Bayesian model, developed by Blei et al (2003) for topical modeling of text corpora.

ii. Estimates predetermined number of K topics based on high dimensional text data on documents.

iii. A topic is a probability distribution over the underlying topic probabilities and a word is a probability distribution over topics.

iv. LDA allows for multiple topics to exist within a particular document and also allows topics to vary across documents.

v. Extracts a sparse and meaningful representation from textual data.

Words and Topics Simplex

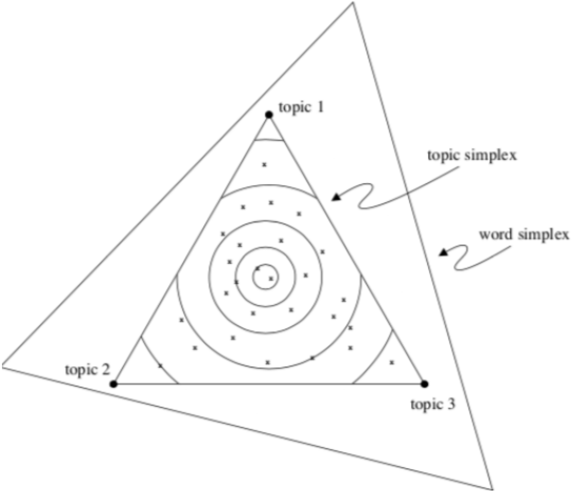


Figure: Source: Blei et al (2003).

Word Clouds for Estimated Topics

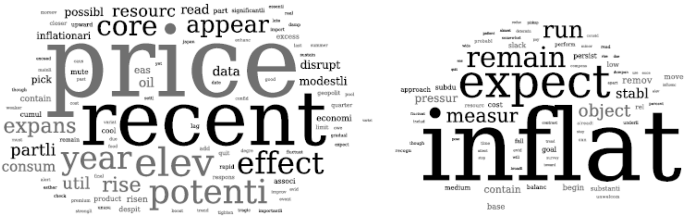


Figure: Source: Hansen and McMahon (2017).

AI and Ethics

i. When used for policy decisions, the lack of transparency of ML, a *black box* raises ethical quandaries.
ii. Amazon scrapped its AI tool for hiring because it showed significant bias against female job applicants.
iii. Algorithms designed to assess recidivism probability to make bail approval decision in US biased against African Americans.
iv. The goals of AI in industry are mainly speed, accuracy and efficiency which can clash with legislation, oversight and auditing that ethical considerations require.

Conclusion

i. ML has opened up a plethora of new opportunities for research in economics and development.
ii. It is about time we leverage the power of ML for academic and policy relevant research in Pakistan.
iii. Limited availability of big and thick data will continue to be a challenge.

Thank you

Q and A