ML in Economics and Finance: Where do We Go Now? - Part I

Raul Riva

FGV EPGE

December, 2025

INSPER - São Paulo

Intro

Who is this guy?

- I have just joined FGV EPGE as an Assistant Professor;
- I got my PhD in Finance at Northwestern University;
- Asset Pricing + Macro-Finance + Econometrics;

Who is this guy?

- I have just joined FGV EPGE as an Assistant Professor;
- I got my PhD in Finance at Northwestern University;
- Asset Pricing + Macro-Finance + Econometrics;

I am **not** an ML developer, but maybe a mildly sophisticated economist consumer

Where are we?

- Last 20-30 years: explosion of computation power and popularization of ML techniques;
- Last 15 years: we economists imported several techniques from CS and Stats;

Where are we?

- Last 20-30 years: explosion of computation power and popularization of ML techniques;
- Last 15 years: we economists imported several techniques from CS and Stats;
- Many challenges in this translation:
  - Causality vs pattern recognition;
  - Interpretability;
  - Sophisticated notions of equilibrium;
  - Time series dynamics;

Where are we?

- Last 20-30 years: explosion of computation power and popularization of ML techniques;
- Last 15 years: we economists imported several techniques from CS and Stats;
- Many challenges in this translation:
  - Causality vs pattern recognition;
  - Interpretability;
  - Sophisticated notions of equilibrium;
  - Time series dynamics;

Right now:

- Better understanding of the limitations of ”plug and play” ML;
- Great stuff: new hybrid methods designed by and for economists;
- Bad stuff: we are flooded with tutorials, books, videos, bootcamps...

Where do we go now?

- The Econ/Finance forecasting crowd was really fast in adopting ML...
- But what else? What is worth knowing about ML in Econ and Finance?

Where do we go now?

- The Econ/Finance forecasting crowd was really fast in adopting ML...
- But what else? What is worth knowing about ML in Econ and Finance?

What I will do:

- My own economist-crafted definition of ML methods and how to think about them;
- Three very cool agendas where ML can help economists
- Causality in HD, seriously heterogeneous treatment effects, and solving large models;

Where do we go now?

- The Econ/Finance forecasting crowd was really fast in adopting ML...
- But what else? What is worth knowing about ML in Econ and Finance?

What I will do:

- My own economist-crafted definition of ML methods and how to think about them;
- Three very cool agendas where ML can help economists
- Causality in HD, seriously heterogeneous treatment effects, and solving large models;

What I will not do:

- Teach you how to code;

Where do we go now?

- The Econ/Finance forecasting crowd was really fast in adopting ML...
- But what else? What is worth knowing about ML in Econ and Finance?

What I will do:

- My own economist-crafted definition of ML methods and how to think about them;
- Three very cool agendas where ML can help economists
- Causality in HD, seriously heterogeneous treatment effects, and solving large models;

What I will not do:

- Teach you how to code;
- Pretend I know how to prove the complicated theorems and walk you through proofs;

Where do we go now?

- The Econ/Finance forecasting crowd was really fast in adopting ML...
- But what else? What is worth knowing about ML in Econ and Finance?

What I will do:

- My own economist-crafted definition of ML methods and how to think about them;
- Three very cool agendas where ML can help economists
- Causality in HD, seriously heterogeneous treatment effects, and solving large models;

What I will not do:

- Teach you how to code;
- Pretend I know how to prove the complicated theorems and walk you through proofs;
- Lie to you and say you can easily perform any of this in Stata! 🙄

Who is this for?

- Students starting their empirical research agendas;
- Fellow empirical researchers trying to grasp what kind of ML tools might be useful;
- Someone coming from Stats or CS into Economics;

Who is this for?

- Students starting their empirical research agendas;
- Fellow empirical researchers trying to grasp what kind of ML tools might be useful;
- Someone coming from Stats or CS into Economics;

Who is this not for?

- Super sophisticated economists already deploying these techniques everywhere;
- Hardcore econometricians looking for open theoretical problems;

Who is this for?

- Students starting their empirical research agendas;
- Fellow empirical researchers trying to grasp what kind of ML tools might be useful;
- Someone coming from Stats or CS into Economics;

Who is this not for?

- Super sophisticated economists already deploying these techniques everywhere;
- Hardcore econometricians looking for open theoretical problems;

DISCLAIMER: These are **my** own views, based on **my** experience, and **my** own readings. Other people will disagree.

Flight Plan ✈️

1. What is ML, anyway?
2. Causality in High Dimensions

Today

3. (Seriously) Heterogeneous Treatment Effects
4. Solving Large-Scale General Equilibrium Models

Tomorrow

Flight Plan ✈️

1. What is ML, anyway?
2. Causality in High Dimensions

Today

3. (Seriously) Heterogeneous Treatment Effects
4. Solving Large-Scale General Equilibrium Models

Tomorrow

Please bring questions at any time!

Questions?

A General Framework

What is *Machine Learning*?

- Different fields = different definitions: CS, Stats, Operations Research, ...
- Many types: Supervised, Unsupervised, Reinforcement Learning, Deep Learning, ...
- More buzzwords = better consulting gigs! 🤑

What is *Machine Learning*?

- Different fields = different definitions: CS, Stats, Operations Research, ...
- Many types: Supervised, Unsupervised, Reinforcement Learning, Deep Learning, ...
- More buzzwords = better consulting gigs! 🤑
- Today and tomorrow: **Supervised Learning**;

What is *Machine Learning*?

- Different fields = different definitions: CS, Stats, Operations Research, ...
- Many types: Supervised, Unsupervised, Reinforcement Learning, Deep Learning, ...
- More buzzwords = better consulting gigs! 🤑
- Today and tomorrow: **Supervised Learning**;
- I will be brave enough and provide the one I think is really useful for Economists:

What is *Machine Learning*?

- Different fields = different definitions: CS, Stats, Operations Research, ...
- Many types: Supervised, Unsupervised, Reinforcement Learning, Deep Learning, ...
- More buzzwords = better consulting gigs! 🤑
- Today and tomorrow: **Supervised Learning**;
- I will be brave enough and provide the one I think is really useful for Economists:

(Supervised) **Machine Learning** is a set of tools that enable computationally-feasible data-driven search over high-dimensional functional spaces.

A General Framework

$$y = f(\boldsymbol{x}) + \varepsilon$$

- $y \in \mathbb{R}^k$ is some "target" or "outcome";
- $\boldsymbol{x} \in \mathbb{R}^p$ is a vector of "features", or "predictors", or "covariates";
- $f : \mathbb{R}^p \to \mathbb{R}^k$ is some unknown function;
- ε is some unobserved noise because the world is messy;

A General Framework

$$y = f(\boldsymbol{x}) + \varepsilon$$

- $y \in \mathbb{R}^k$ is some "target" or "outcome";
- $\boldsymbol{x} \in \mathbb{R}^p$ is a vector of "features", or "predictors", or "covariates";
- $f : \mathbb{R}^p \to \mathbb{R}^k$ is some unknown function;
- ε is some unobserved noise because the world is messy;

Question: given a function space $\mathcal{F}$, how to find $\hat{f} \in \mathcal{F}$ that approximates f well?

A General Framework

$$y = f(\boldsymbol{x}) + \varepsilon$$

- $y \in \mathbb{R}^k$ is some "target" or "outcome";
- $\boldsymbol{x} \in \mathbb{R}^p$ is a vector of "features", or "predictors", or "covariates";
- $f : \mathbb{R}^p \to \mathbb{R}^k$ is some unknown function;
- ε is some unobserved noise because the world is messy;

Question: given a function space $\mathcal{F}$, how to find $\hat{f} \in \mathcal{F}$ that approximates f well?

- Collect data $\{(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)\}$;
- Define some notion of "approximates well" $\implies$ (a loss function);
- Be explicit about $\mathcal{F}$;
- Be explicit about your optimization mechanism;

You are already doing ML!

Consider an outcome y_i, and a set of covariates $\boldsymbol{x}_i$ for $i = 1, \ldots, n$:

$$y_i = \alpha + \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- This is a linear regression model;
- The function space $\mathcal{F}$ is the set of all affine functions of the treatment and covariates;
- The loss function is the MSE: $\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$;
- OLS: minimize a convex loss function over the space of parameters;

You are already doing ML!

Consider an outcome y_i, and a set of covariates $\boldsymbol{x}_i$ for $i = 1, \ldots, n$:

$$y_i = \alpha + \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- This is a linear regression model;
- The function space $\mathcal{F}$ is the set of all affine functions of the treatment and covariates;
- The loss function is the MSE: $\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$;
- OLS: minimize a convex loss function over the space of parameters;

Conclusion: Linear regression is a (very simple) ML method! But there is so much more...

Hold on... Isn't Machine Learning just Non-Parametric Estimation?

- The general framework I used could be used in a non-parametric estimation class...
- Why do we need new tools? We already have the good and old kernel regression!

Hold on... Isn't Machine Learning just Non-Parametric Estimation?

- The general framework I used could be used in a non-parametric estimation class...
- Why do we need new tools? We already have the good and old kernel regression!
- Well... there is the curse of dimensionality! If $p \approx 6$, you are already in trouble!

Hold on... Isn't Machine Learning just Non-Parametric Estimation?

- The general framework I used could be used in a non-parametric estimation class...
- Why do we need new tools? We already have the good and old kernel regression!
- Well... there is the curse of dimensionality! If $p \approx 6$, you are already in trouble!

OLS

- Leverages linearity (strong!);
- Easy to compute and interpret;

Fully Non-Parametric Methods

- Extreme flexibility;
- Super data hungry!

Machine Learning = a *compromise*: richer parametrizations while still computationally feasible in high dimensions.

Questions?

Causality in High Dimensions

Preliminaries

- No estimator will lead to causality by itself – only careful design will;

Preliminaries

- No estimator will lead to causality by itself – only careful design will;
- ML methods were *not* created to tackle causality problems;
- See Varian (2014), Mullainathan and Spiess (2017), Athey and Imbens (2019), and Masini, Medeiros, and Mendes (2023);

Preliminaries

- No estimator will lead to causality by itself – only careful design will;
- ML methods were *not* created to tackle causality problems;
- See Varian (2014), Mullainathan and Spiess (2017), Athey and Imbens (2019), and Masini, Medeiros, and Mendes (2023);
- Kleinberg et al. (2015): many policy-relevant questions are prediction problems!

Preliminaries

- No estimator will lead to causality by itself – only careful design will;
- ML methods were *not* created to tackle causality problems;
- See Varian (2014), Mullainathan and Spiess (2017), Athey and Imbens (2019), and Masini, Medeiros, and Mendes (2023);
- Kleinberg et al. (2015): many policy-relevant questions are prediction problems!
- Belloni 🇧🇷, Chernozhukov, Hansen and co-authors took it even further:
  - Computing the propensity score *is* forecasting!
  - The first-stage regression in an IV context *is* forecasting!

Treatment Effects in High Dimensions

Suppose you're interested in the treatment effect $\theta_0 \in \mathbb{R}$:

$$y_i = d_i \theta_0 + \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- $y_i \in \mathbb{R}$ is an outcome;
- $d_i \in \mathbb{R}$ is a treatment;
- $\boldsymbol{x}_i \in \mathbb{R}^p$ is a vector of available covariates;
- ε_i is some unobserved noise with $\mathbb{E}[\varepsilon_i \mid d_i, \boldsymbol{x}_i] = 0$;

Treatment Effects in High Dimensions

Suppose you're interested in the treatment effect $\theta_0 \in \mathbb{R}$:

$$y_i = d_i\theta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- $y_i \in \mathbb{R}$ is an outcome;
- $d_i \in \mathbb{R}$ is a treatment;
- $\boldsymbol{x}_i \in \mathbb{R}^p$ is a vector of available covariates;
- ε_i is some unobserved noise with $\mathbb{E}[\varepsilon_i \mid d_i, \boldsymbol{x}_i] = 0$;
- You have an i.i.d. sample $\{y_i, d_i, \boldsymbol{x}_i\}_{i=1}^n$ and we allow for $p \gg n$;

Treatment Effects in High Dimensions

Suppose you're interested in the treatment effect $\theta_0 \in \mathbb{R}$:

$$y_i = d_i\theta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- $y_i \in \mathbb{R}$ is an outcome;
- $d_i \in \mathbb{R}$ is a treatment;
- $\boldsymbol{x}_i \in \mathbb{R}^p$ is a vector of available covariates;
- ε_i is some unobserved noise with $\mathbb{E}[\varepsilon_i \mid d_i, \boldsymbol{x}_i] = 0$;
- You have an i.i.d. sample $\{y_i, d_i, \boldsymbol{x}_i\}_{i=1}^n$ and we allow for $p \gg n$;
- **Goal**: estimate θ_0 and get a confidence interval;

Treatment Effects in High Dimensions

Suppose you're interested in the treatment effect $\theta_0 \in \mathbb{R}$:

$$y_i = d_i\theta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- $y_i \in \mathbb{R}$ is an outcome;
- $d_i \in \mathbb{R}$ is a treatment;
- $\boldsymbol{x}_i \in \mathbb{R}^p$ is a vector of available covariates;
- ε_i is some unobserved noise with $\mathbb{E}[\varepsilon_i \mid d_i, \boldsymbol{x}_i] = 0$;
- You have an i.i.d. sample $\{y_i, d_i, \boldsymbol{x}_i\}_{i=1}^n$ and we <u>allow</u> for $p \gg n$;
- **Goal**: estimate θ_0 and get a confidence interval;

Question: what will happen if you try OLS here?

- Let's say you believe only a few β_j's are $\neq 0 \implies$ "*sparsity*" in $\boldsymbol{\beta}$;
- But you do not know which ones!

Treatment Effects in High Dimensions

- Let's say you believe only a few β_j's are $\neq 0 \implies$ "*sparsity*" in $\boldsymbol{\beta}$;
- But you do not know which ones!
- What about using your economic intuition to select a subset of controls?

Treatment Effects in High Dimensions

- Let's say you believe only a few β_j's are $\neq 0 \implies$ *"sparsity"* in $\boldsymbol{\beta}$;
- But you do not know which ones!
- What about using your economic intuition to select a subset of controls?
- Applying Econ theory is always a good idea, but:
    - You might not get a meaningful reduction with theory alone;
    - Your referee might not agree with your choices;
    - You might get lost in a sea of robustness checks...

- Let's say you believe only a few β_j's are $\neq 0 \implies$ "*sparsity*" in $\boldsymbol{\beta}$;
- But you do not know which ones!
- What about using your economic intuition to select a subset of controls?
- Applying Econ theory is always a good idea, but:
  - You might not get a meaningful reduction with theory alone;
  - Your referee might not agree with your choices;
  - You might get lost in a sea of robustness checks...
- Good news: ML researchers devoted a lot of attention to *sparse regressions*!

Welcome to SBE, Mr. LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) estimator solves:

$$\hat{\boldsymbol{\delta}} \equiv \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \lambda \sum_{j=1}^{p} |\delta_j| \right\}$$

- $\lambda \geq 0$ is a tuning parameter that controls the amount of penalization ("*regularization*");
- $\boldsymbol{w}_i$ is a general vector of regressors of size p;

Welcome to SBE, Mr. LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) estimator solves:

$$\hat{\boldsymbol{\delta}} \equiv \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \lambda \sum_{j=1}^{p} |\delta_j| \right\}$$

- $\lambda \geq 0$ is a tuning parameter that controls the amount of penalization ("*regularization*");
- $\boldsymbol{w}_i$ is a general vector of regressors of size p;
- The ℓ_1 penalty $\sum_{j=1}^{p} |\delta_j|$ induces sparsity in $\hat{\boldsymbol{\delta}}$;

Welcome to SBE, Mr. LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) estimator solves:

$$\hat{\boldsymbol{\delta}} \equiv \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \lambda\sum_{j=1}^{p}|\delta_j| \right\}$$

- $\lambda \geq 0$ is a tuning parameter that controls the amount of penalization ("*regularization*");
- $\boldsymbol{w}_i$ is a general vector of regressors of size p;
- The ℓ_1 penalty $\sum_{j=1}^{p}|\delta_j|$ induces sparsity in $\hat{\boldsymbol{\delta}}$;
- If $\lambda = 0$, we get OLS; if $\lambda \to \infty$, we get $\hat{\boldsymbol{\delta}} = \mathbf{0}$;

Welcome to SBE, Mr. LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) estimator solves:

$$\hat{\boldsymbol{\delta}} \equiv \arg\min_{\boldsymbol{\delta} \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - \boldsymbol{w}_i' \boldsymbol{\delta})^2 + \lambda \sum_{j=1}^{p} |\delta_j| \right\}$$

- $\lambda \geq 0$ is a tuning parameter that controls the amount of penalization ("*regularization*");
- $\boldsymbol{w}_i$ is a general vector of regressors of size p;
- The ℓ_1 penalty $\sum_{j=1}^{p} |\delta_j|$ induces sparsity in $\hat{\boldsymbol{\delta}}$;
- If $\lambda = 0$, we get OLS; if $\lambda \to \infty$, we get $\hat{\boldsymbol{\delta}} = \mathbf{0}$;
- For intermediate values of λ, some $\hat{\delta}_j$'s will be exactly zero!

Welcome to SBE, Mr. LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) estimator solves:

$$\hat{\boldsymbol{\delta}} \equiv \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{w}_i'\boldsymbol{\delta})^2 + \lambda\sum_{j=1}^{p}|\delta_j| \right\}$$

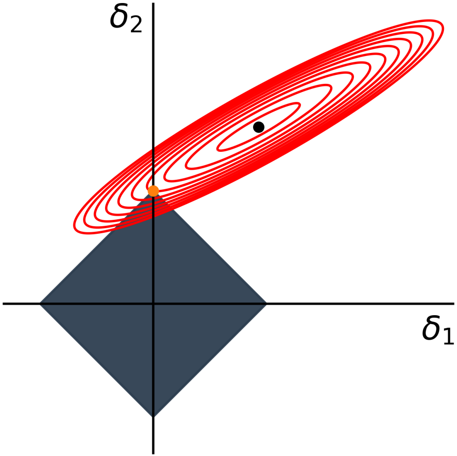
- $\lambda \geq 0$ is a tuning parameter that controls the amount of penalization ("*regularization*");
- $\boldsymbol{w}_i$ is a general vector of regressors of size p;
- The ℓ_1 penalty $\sum_{j=1}^{p}|\delta_j|$ induces sparsity in $\hat{\boldsymbol{\delta}}$;
- If $\lambda = 0$, we get OLS; if $\lambda \to \infty$, we get $\hat{\boldsymbol{\delta}} = \mathbf{0}$;
- For intermediate values of λ, some $\hat{\delta}_j$'s will be exactly zero!
- $\hat{\boldsymbol{\delta}}$ is gives up unbiasedness for much lower variance;
- This problem is still feasible if $p \gg n$ and it is convex $\implies$ fast computation;

The Geometry of LASSO

For $c > 0$, consider the following:

$$\tilde{\boldsymbol{\delta}} \equiv \arg\min_{\boldsymbol{\delta} \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - \boldsymbol{w}_i' \boldsymbol{\delta})^2 \right\}$$

$$\text{subject to } \sum_{j=1}^{p} |\delta_j| \leq c$$



- Think about the Lagrangian of this problem!
- For every λ, there is a c such that $\hat{\boldsymbol{\delta}} = \tilde{\boldsymbol{\delta}}$;

A Naive Approach

Generalizations

Appendix and References

References i

Athey, Susan and Guido W. Imbens (Aug. 2019). "Machine Learning Methods That Economists Should Know About". en. In: *Annual Review of Economics* 11.1, pp. 685–725. ISSN: 1941-1383, 1941-1391. DOI: 10.1146/annurev-economics-080217-053433. URL: https://www.annualreviews.org/doi/10.1146/annurev-economics-080217-053433 (visited on 12/03/2025).

Kleinberg, Jon et al. (May 2015). "Prediction Policy Problems". en. In: *American Economic Review* 105.5, pp. 491–495. ISSN: 0002-8282. DOI: 10.1257/aer.p20151023. URL: https://pubs.aeaweb.org/doi/10.1257/aer.p20151023 (visited on 12/02/2025).

Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes (Feb. 2023). "Machine learning advances for time series forecasting". en. In: *Journal of Economic Surveys* 37.1, pp. 76–111. ISSN: 0950-0804, 1467-6419. DOI: 10.1111/joes.12429. URL: https://onlinelibrary.wiley.com/doi/10.1111/joes.12429 (visited on 12/03/2025).

Mullainathan, Sendhil and Jann Spiess (May 2017). "Machine Learning: An Applied Econometric Approach". en. In: *Journal of Economic Perspectives* 31.2, pp. 87–106. ISSN: 0895-3309. DOI: 10.1257/jep.31.2.87. URL: https://pubs.aeaweb.org/doi/10.1257/jep.31.2.87 (visited on 12/02/2025).

Varian, Hal R. (May 2014). "Big Data: New Tricks for Econometrics". en. In: *Journal of Economic Perspectives* 28.2, pp. 3–28. ISSN: 0895-3309. DOI: 10.1257/jep.28.2.3. URL: https://pubs.aeaweb.org/doi/10.1257/jep.28.2.3 (visited on 12/02/2025).