Machine Learning Tools useful for the Applied Economist

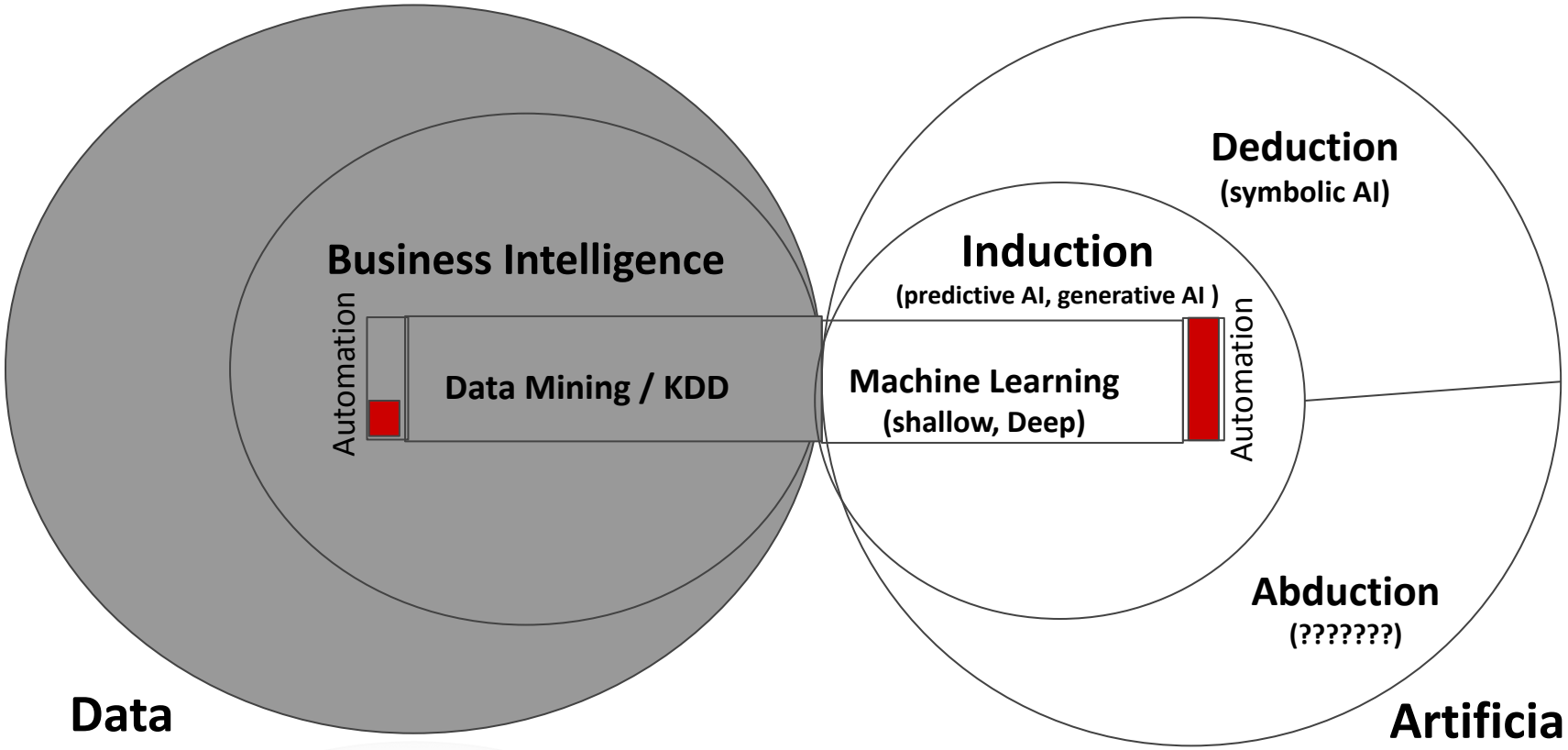
R.J Méndez-Marcano

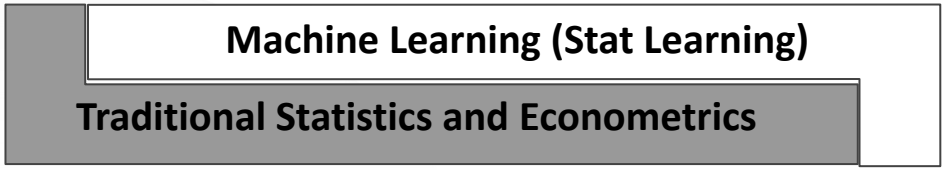
Devtech, Fall 2023 (Sessions 1, 2)

Machine Learning Tools useful for the Applied Economist

SESSION 1 (09/13/2023)

Big **Data** + **Computational Power** ⇒





Statistical Tools

(models, estadísticos, tests,, algorithms)

Typical Economic Data

- Observational
- Numerical/Categorical
- **Highly Structured**: Table form

} Limited complexity

  - **Small**: sheet of paper ⟶ Regression Oveparameterization
    - high dimensionality
    - non-linearity

Typical Economic Database

Obs	Individual	Date	Predictor1	Predictor2	⇒	Response
1	Argentina	1990	4120	SI		2.3
2	Argentina	2022	4230	NO		2.8
3	Brazil	1990	5180	SI		45
4	Brazil	2022	5024	SI		4.2
⇓						
389	Zimbabwe	1990	120	NO		5.0
390	Zimbabwe	2022	230	NO		9.0

Observations	ML: Instances
Explanatory or independent variables, predictors	ML: attributes, features, inputs
Dependent variable, response	ML: class attribute, target

Standard Economic Data Analysis/Modelling

- Descriptive Statistical Analysis
- Multivariate Statistical Analysis → ML: Unsupervised Learning
- **Regression Analysis** (Modelling) → ML: **Supervised Learning**
  - Numerical dependent variable → ML: Regression Learning
  - Categorical dependent variable → ML: Classification Learning

Regression model complexity

(Relation between mean response and parameters)

- **Linear/Generalized Linear** → Pen and pencil algorithms

  → ML teach to take advantage of automatization
  → ML teach how to deal with Overparameterization (regularization)

- Non-linear/Complex → Computationally intensive algorithms
  → ML teach how to model any degree of complexity
  → ML teach how to deal with Overfitting (cross-validation, testing)

Regression model purpose

- Explanatory or Structural (Academic) → **Teach to ML**

  ⇒ Individual direct causality identification and quantification involved

- Forecasting / Predictiction (Business) → **Learn from ML**

  ⇒ Just observable correlation involved

⇒ Economist traditional practice:

- Use of standard significance tests/goodness-of-fit
- Neglect Bias/Variance Trade-Off
- No out-of-sample forecasting evaluation

Estimation & Data Partitioning

Estimation	ML: Training
Estimation Sample	ML: Training Sample
Out-of-Sample	ML: Testing Sample
	ML: Validation Sample
Out-of-Sample forecasting accuracy	Testing

Typical Economic Database

Obs	Individual	Date	Predictor1	Predictor2	⇒	Response
1	Argentina	1990	4120	SI		2.3
2	Argentina	2022	4230	NO		2.8
3	Brazil	1990	5180	SI		45
4	Brazil	2022	5024	SI		4.2
⇓						
389	Zimbabwe	1990	120	NO		5.0
390	Zimbabwe	2022	230	NO		9.0

Observations	ML: Instances
Explanatory or independent variables, predictors	ML: attributes, features, inputs
Dependent variable, response	ML: class attribute, target

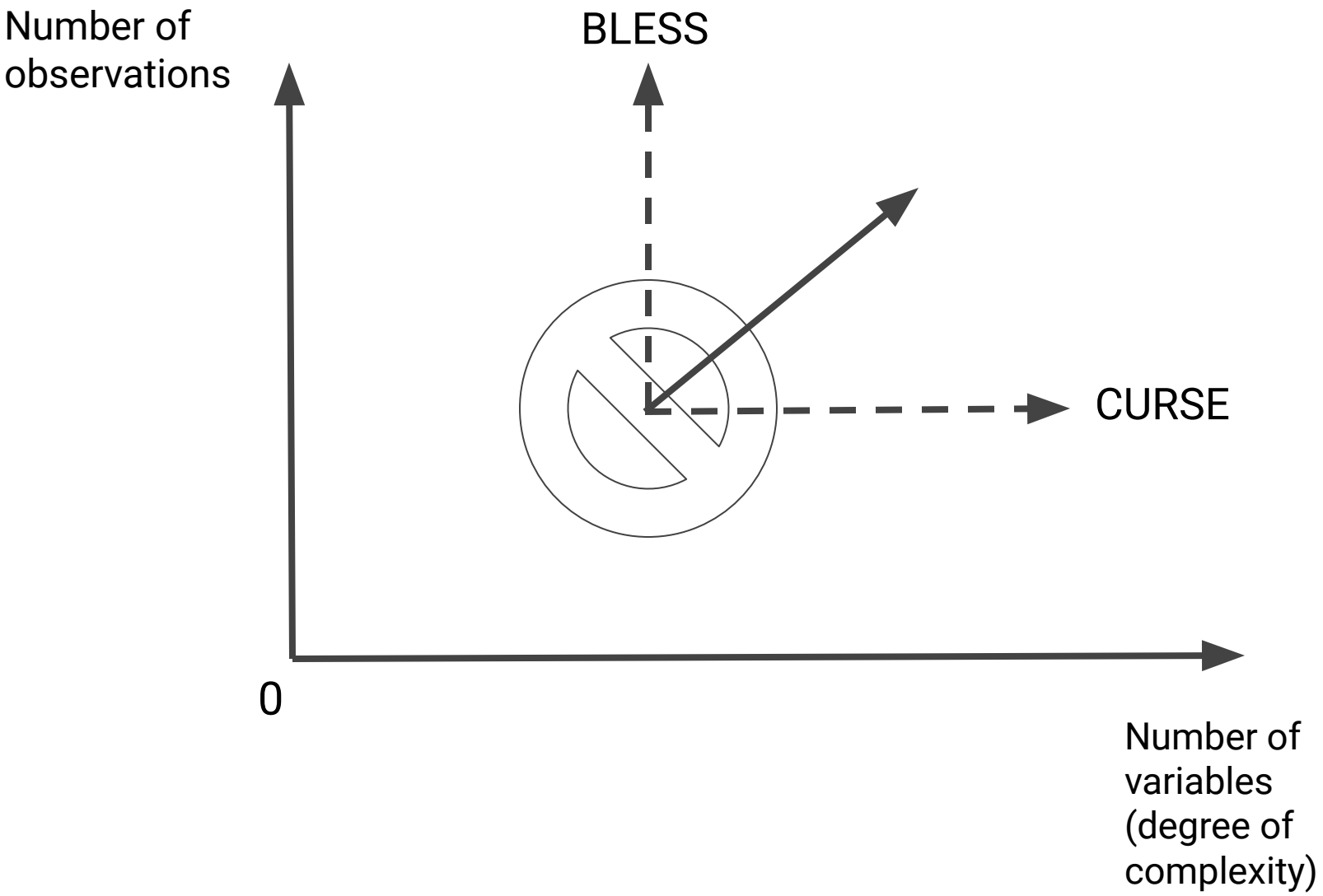
Economic Data is growing much larger but…

- BIG DATA: GPT-3 → 570000 MB
- Economic data:
  - Macroeconomic → PWT-10 = 6 MB
  - Microeconomic → NLSY97 = 14000 MB

Size dimensions: height vs. weight

- Observations → Increase complexity → Overfitting risk
  - Individuals
  - Dates per individuals
    - Frequency
    - Span
- Explanatory Variables or Predictors → Overparam risk

Standard OLS Estimation of Linear Model



What is useful to learn from ML

- Incorporating in estimation and evaluation even of standard linear regression models (forecasting purpose):
  - Semi-automatization
  - Cross-validation and testing
  - Regularization or shrinkage
  - Ensembling

- For supervised learning, try a bit more complex models appropriate for the volume and nature of economic nada
  - Decision trees: YES
  - Random Forest: YES
  - Support Vector Machine: ??
  - Multi-Layers Neural Networks: NO (DL)
  - Simpler neural networks: ??
- For unsupervised learning, a bit of the same
  - Autoecoders: NO (DL)
  - Principal Component Anaysis: YES
  - K-means: YES
  - Hierarchical Clustering: YES

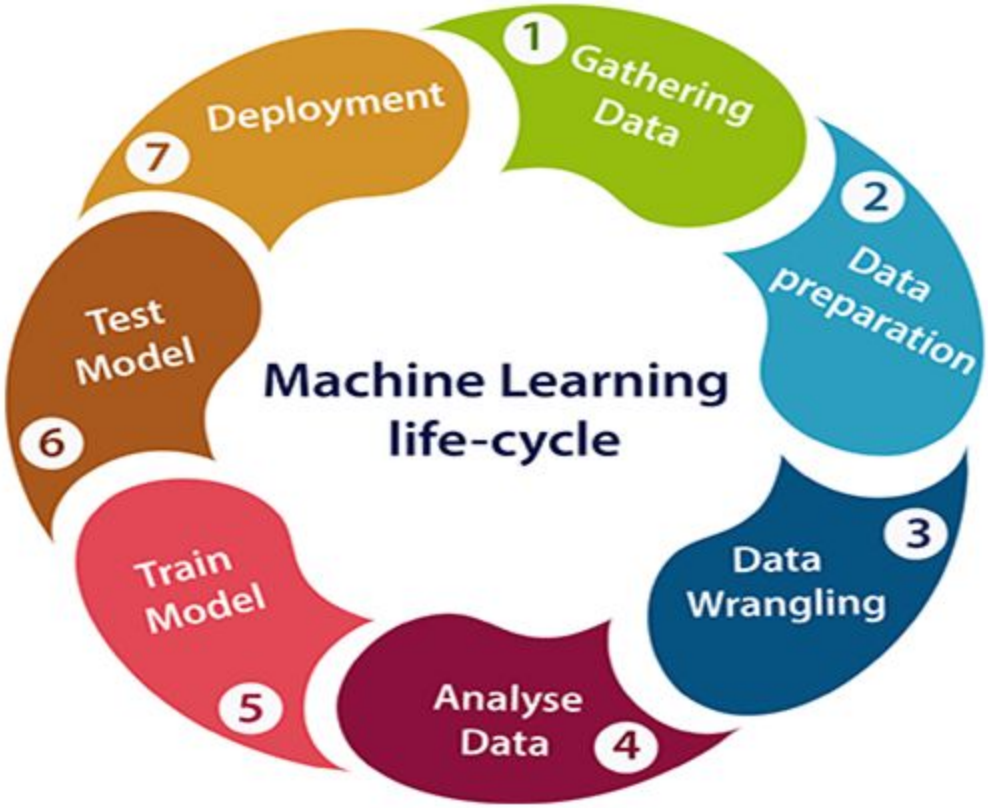
Machine Learning Tools useful for the Applied Economist

SESSION 2 (09/20/2023)

SAS Viya for Learners

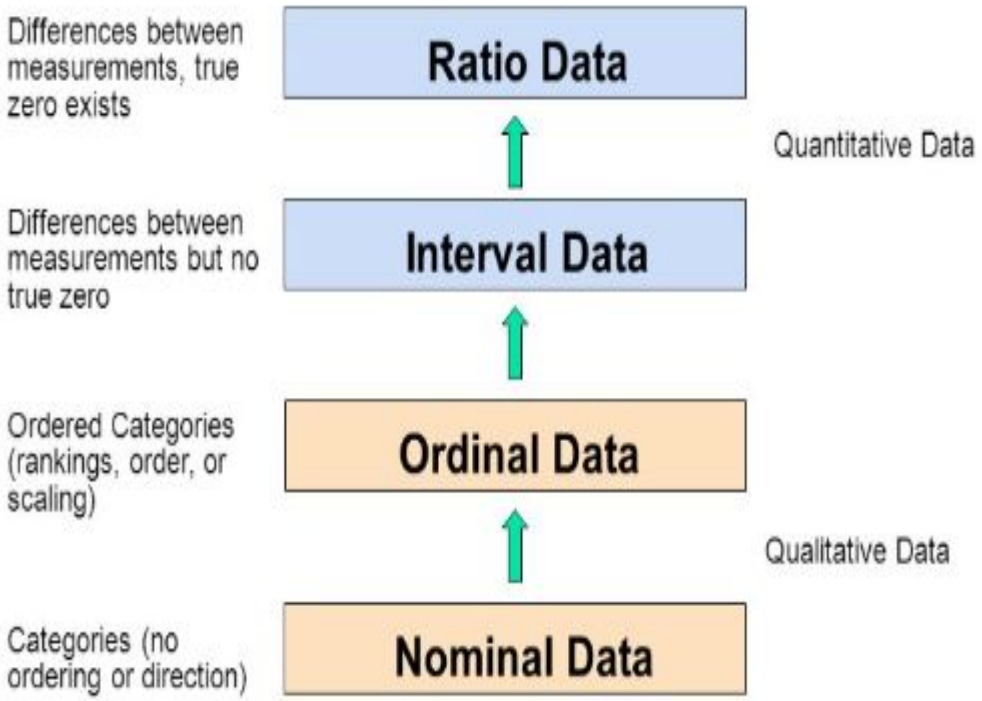
- Link: https://www.sas.com/en_us/software/viya-for-learners.html
- Link: http://documentation.sas.com/doc/en/vdmmlcdc/8.2/vdmmlug/p0psvjbnx1y57mn1taay7h8d4ve0.htm

Machine Learning Project Life Cycle



"The most important thing in the complete process is to understand the problem and to know the purpose of the problem. Therefore, before starting the life cycle, we need to understand the problem because the good result depends on the better understanding of the problem."

Reference: https://www.javatpoint.com/machine-learning-life-cycle



Scales of Measurement

"Knowing the scale of measurement for a variable is an important aspect in choosing the right statistical analysis."

Reference: https://www.graphpad.com/support/faq/what-is-the-difference-between-ordinal-interval-and-ratio-variables-why-should-i-care/

Machine Learning Tools useful for the Applied Economist

R.J Méndez-Marcano

Devtech, Fall 2023