Discussion of Session Titled, "Machine Learning in Agricultural Trade and Policy Settings"

Mateusz Filipski

Selected Paper prepared for presentation at the International Agricultural Trade Research Consortium's (IATRC's) 2019 Annual Meeting: Recent Advances in Applied General Equilibrium Modeling: Relevance and Application to Agricultural Trade Analysis, December 8-10, 2019, Washington, DC.

UNIVERSITY OF GEORGIA

Machine Learning in Agricultural Trade and Policy Settings:

Discussion on 3 presented papers

by

Mateusz Filipski

UGA, IFPRI

IATRC, Washington DC, Dec 9, 2019

Outline

- Overview of the session
- A couple of slides on each paper
- Tying it all together

Three papers

- Predicting trade flows with Decision Trees / Neural Networks (Gopi)
- Predicting famines with Decision Trees (Kathy)
- Predicting Yields with (Semiparametric) Neural Networks (Joe)

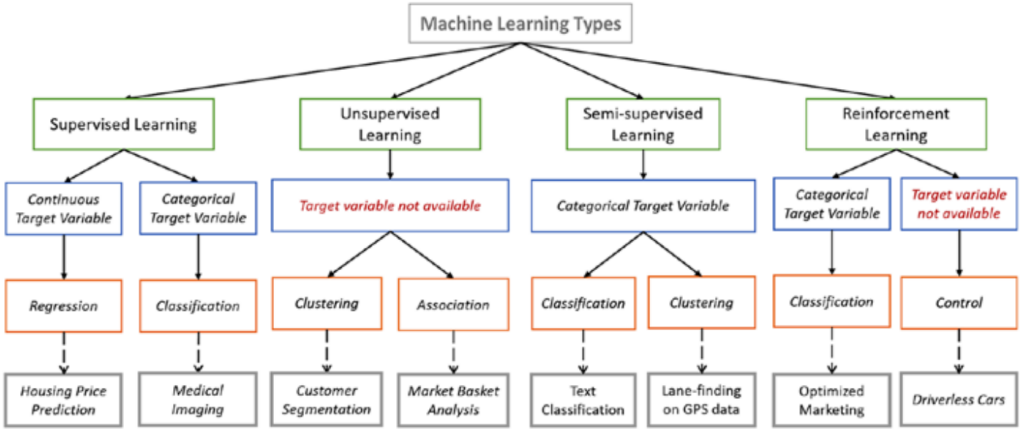
- ... Notice a pattern?

Three papers

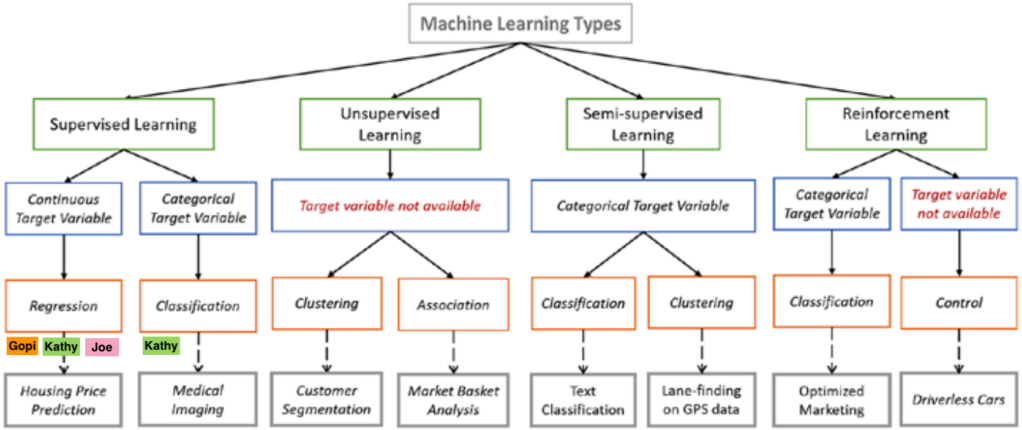
- Predicting trade flows with Decision Trees / Neural Networks (Gopi)
- Predicting famines with Decision Trees (Kathy)
- Predicting Yields with (Semiparametric) Neural Networks (Joe)

- ... Notice a pattern?

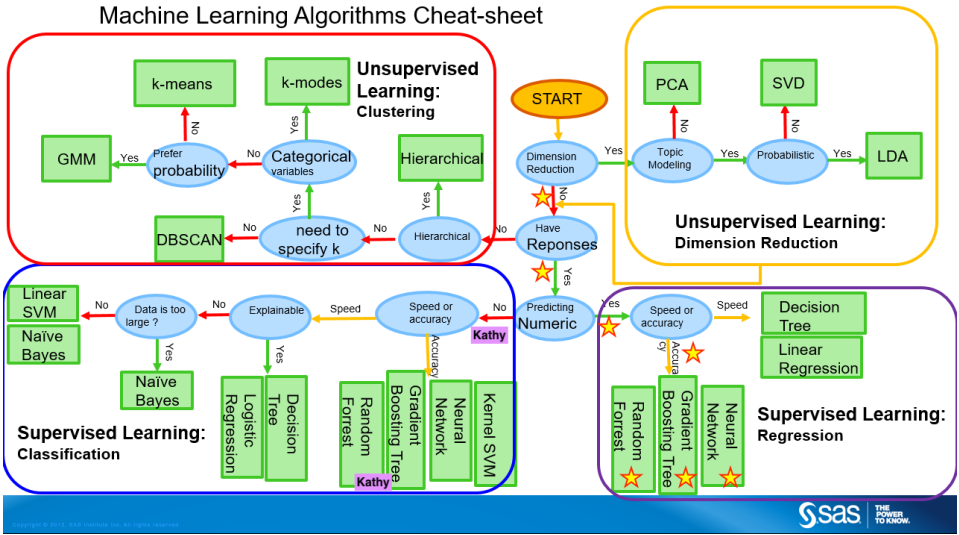
The world of ML



The world of ML



Navigating the world of ML



Trade flows paper: What I love about it

- Completely new look at a (very) old question
- Leverages the volumes of data we have
- Uses known theory (gravity model) but frees us from structure.

Trade flows paper: What I'm missing

- Plot the data and draw a tree
- Are you really using the right model as comparison? (the PPML)
- You can defy gravity!

Trade flows paper: What I'm missing

- Plot the data and draw a tree
- Are you really using the right model as comparison? (the PPML)
- You can defy gravity!

Trade flows paper: Let's Defy Gravity!

Table 3: Ranking Variables by Informatio

Variables	Beef	Corn	Rice
Population_Destination	2	1	1
GDP Per Capita_Destination	1	2	2
Population_Origin	3	3	4
GDP Per Capita_Origin	4	5	7
Distance	5	6	6
Time	6	4	3
Longitude_Destination	7	7	5
Latitude_Destination	8	8	8
Political Stability	9	9	9

- Still a bilateral framework i/j
- Maybe country k matters?
- Eg: USA $\Rightarrow$ Brazil/China trade

Famines paper: What I love about it

- Major issue, huge potential.
- Can address the "Where" question with high precision (+ who?)
- Leverages existing HH data + satellite data $\Rightarrow$ cheap.

Famines paper: What I'm missing

- Are you really using the right decision tool for comparison? (a logit)
- LSMS as ground truth is small data (esp. at cluster-level)
- HH variables limit your ability to expand (ex: flooring material?)
- Very few high-frequency variables matter in the results

Famine paper: High-frequency variables

Variable	Importance
roof_natural	0.11
cell_phone	0.09
floor_dirt_sand_dung	0.08
number_celphones	0.05
roof_iron	0.04
day1rain	0.04
clust_beans_price	0.04
lhz_maxdaysnorain	0.03
lhz_nuts_mktthin	0.03
asset_index	0.03
Household_head_age	0.03
clust_maize_price	0.03
dist_road	0.03
clust_maize_mktthin	0.02

Malawi

- Most important "high-frequency variable" is sixth.
- Should show how results vary over time (say, 5 maps at different times, with different clusters identified as insecure)

Yield paper: What I love about it

- (*focusing just on the ML part of the paper*)
- Pioneers the upcoming changes: ML for inputs to chained models (here: crop + cost models).
- All input data is remote-sensed: easily scalable and re-usable
- Fairly mature methodology

Yield paper: What I'm missing

- ML not always superior to simple OLS.
- Superiority of ML is very regional. This suggests there is a major missing variable.
- What difference does the ML really make in terms of the final projections?

It's the end of the world as we know it...

- With ML, Prediction becoming central: $\beta \Rightarrow \widehat{y}$
- Incremental changes: better inputs for existing models
- Seismic changes: completely new ways of dealing with data / models
- Exciting time to think about new ways of doing things

- ... And I feel fine! ♪♫

It's the end of the world as we know it...

- With ML, Prediction becoming central: $\beta \Rightarrow \widehat{y}$
- Incremental changes: better inputs for existing models
- Seismic changes: completely new ways of dealing with data / models
- Exciting time to think about new ways of doing things

- ... And I feel fine! ♪♫

ML is not without pitfalls

Source: https://xkcd.com/1838/