Machine Learning and Production of Public Data

Jeff Chen
Chief Innovation Officer
APDU 2019

The views expressed here are those of the authors and do not represent those of the U.S. Bureau of Economic Analysis or the U.S. Department of Commerce.
What’s Machine Learning?

Got SkyNet?

I’m sorry, Dave.
This Person Does Not Exist

Source: https://thispersondoesnotexist.com
Harry Potter as written by AI

Part 1

“The Malfoys!” said Hermione.

Harry was watching him. He looked like Madame Maxime. When she strode up the wrong staircase to visit himself.

“I’m afraid I’ve definitely been suspended from power, no chance—indeed?” said Snape. He put his head back behind him and read groups as they crossed a corner and fluttered down onto their ink lamp, and picked up his spoon. The doorbell rang. It was a lot cleaner down in London.

Hermione yelled. The party must be thrown by Krum, of course.

Harry collected fingers once more, with Malfoy. “Why, didn’t she never tell me. …” She vanished. And then, Ron, Harry noticed, was nearly right.

“Now, be off,” said Sirius, “I can’t trace a new voice.”

He punished Uncle Vernon, so loudly that she could barely use it with Viktor Krum, these faces in one side of their previous year. Mr. Dursley again was this champion. He was about to give the Remembrall there…. every time he
What’s Machine Learning?

\[ y_i = f(X) \]
What’s Machine Learning?

\[ y_i = f(X) \]

Function (or “algorithm”)

Input Variables

Target (outcome)
A simple example
Fitting a line

House Price

Size
Economics generally is interested in parameter estimation.
ML is interested in *prediction*.
ML is interested in *prediction*.
Pursuit of the auto-magical algorithm!
Machine learning focuses on building scalable data pipelines to allow prediction can be central to any business process.

Phase 1: Train and deploy an algorithm
Machine learning focuses on building scalable data pipelines to allow prediction can be central to any business process.

Phase 2: Score new data
Methods in Economics
• Linear Regression
• ARIMA and other time series
• Quantile Regression

Methods for Prediction
• Linear Regression
• ARIMA
• Quantile Regression
• Ensemble Models
• Gradient Boosting
• Regularized Regression
• Adaptive Boosting
• Regression Trees
• Random Forests
• Support Vector Machines
• Multi-Adaptive Regression Splines
• Convolutional Neural Networks
• Recurrent Neural Networks
• Long Short Term Memory Networks
• Generative Adversarial Networks
• Latent Dirichlet Allocation
• + many more
What is BEA doing in this space?

- Improved Imputation
- Short-range prediction
- Anomaly detection
- Seasonality Detection
Service Sector Source Data Timing

End of Quarter

Advance Estimate

When we’d like it to be available

Second Estimate

When source data are available
What is BEA doing in this space?

Short-term prediction using machine learning (for services sector estimates)

Traditional Data

Alternative Data
Computation-Driven Prediction

- Project working title named “MacGyvertron” as Angus MacGyver is able to turn a win when the odds are overwhelmingly against him.
- Trains millions of models on BEA’s cloud computing infrastructure, then constructs an ensemble.
- Assessed over 1000 economic and social variables ranging from aggregate credit card transactions to employment estimates.
- Methodology helps spot parts of PCE that are improvable.
(1) Physician Offices

(2) Software Publishers

(3) Motor Vehicle Repair and Maintenance

(4) Medical Labs
Simulated impacts on PCE Estimates

• It is possible to improve all advance PCE estimates using ML models, but only approximately **one-third** of PCE service series are can yield ‘sure-fire’ improvements.

• Average **12%** reduction in revisions to PCE, but can be as much as **20%** as the model improves.

• There are larger revision reductions to subcomponents of PCE per quarter.
  - *Health*: **+11%**
  - *Transportation*: **+25%**
  - *Personal Care and Clothing*: **+27%**
Beyond

- Short-range prediction
- Improved Imputation
- Anomaly detection
- Seasonality Classification