Machine Learning
for
Time Series and Sensor Data

Pablo Maldonado

Who am I?

- Pablo Maldonado

- PhD. “Applied” Mathematics, Game Theory and some RL, Univ. Paris VI.

- Freelance Trainer and Consultant since 2016, www.datastart.eu
  - Enterprise Training in Machine Learning & Data Science with ***R, Python and MATLAB***.
  - Building prototypes and data science teams.

- Back and forth from Academia (Czech Technical University, Ukrainian Catholic University).

Assumptions

- You know what the following words mean:
  - Supervised learning, unsupervised learning, training a model, feature selection, overfitting, underfitting.

- You can type code in Python (Jupyter notebooks). **Live coding.**

- You can clone/download the course repo
  - www.github.com/jpmaldonado/ml-for-iot

Why are time series / sensor special?

Some reasons

- Time series data is ordered
  - Meaningless to do random train/test split.
  - Two kinds of “supervised learning”:
    - Forecasting
    - Classification
- Continuum of features.
- Either too much or too little data.
- Normalization may not work the way you expect.

Continuum of features

- “Curse of dimensionality”: Intuition is meaningless in high dimensions.
  - Infinite dimensional oranges have all the pulp in the skin.
  - Much more points needed to sample points at a given tolerance level.
  - Almost everyone is my neighbour in high dimensions.
  - More features = better performance, Even more features = worst performance (Hughes effect).
- High correlation among features.

Too little data

Problem: Analyze nutrient content.

Nutritional content

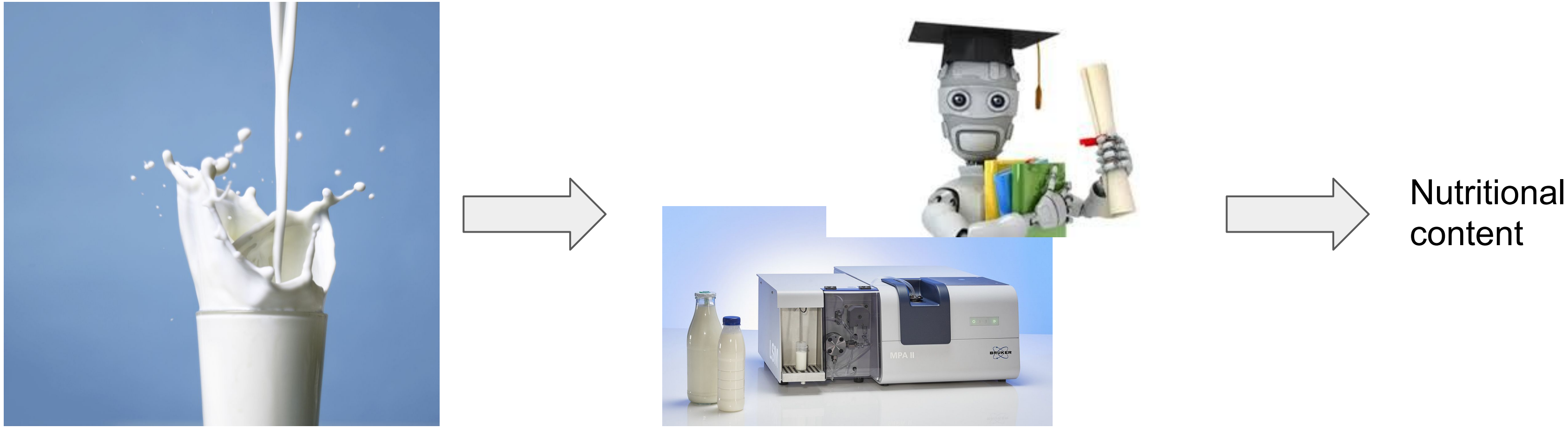
Too little data

Problem: Analyze nutrient content.

Nutritional content

Too little data

Problem: Analyze nutrient content.



Too little data

Problem: Analyze nutrient content.

You need annotated data here!

Nutritional content

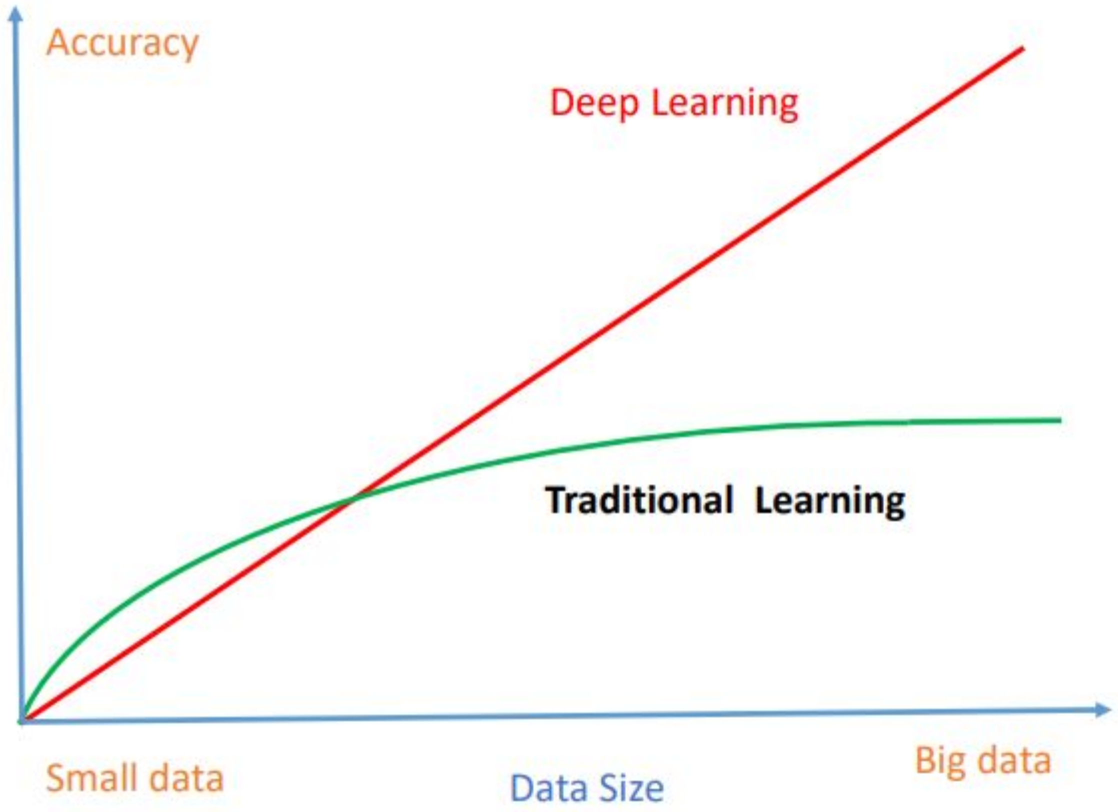
Too much data

- The most profitable industry that uses machine learning. **Which one it is?**

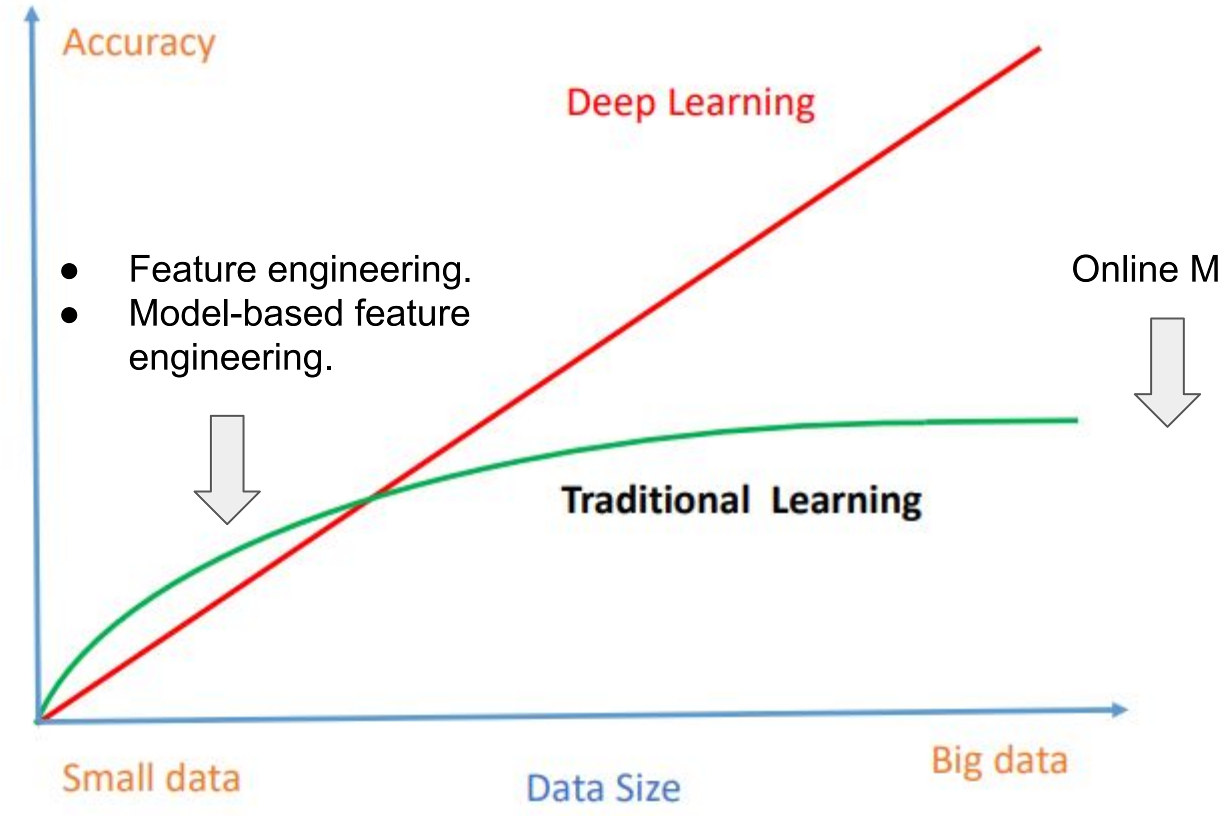
- Data does not fit in memory.

- Even worse, there is so much of it that storing it on disk is becomes impossible and pointless.

Ng Curve


Accuracy
Deep Learning
Traditional Learning
Small data
Data Size
Big data

Our Course



- Feature engineering.
- Model-based feature engineering.

Online ML

Why should we care?

- 5G might bring much more devices into common use.

- Wide range of applications:
  - Agriculture
  - Food industry
  - Industrial machines

- Few or systematic, manageable errors.

Questions?