

Introduction to Building Machine Learning Models with a Simple Manufacturing Application

Andy Henderson, Ph.D.
Chief Technology Officer
Praemo Inc.

Matthew Krugh, Ph.D.
Postdoctoral Research Fellow
Clemson University



- Why ML and where do you start?
- Dataset overview
- Importing and cleaning
- Analysis
- Modeling



Why Deploy ML?

- Volume of Data (existing and being newly created) is too LARGE!
 - New products and processes creating variation in data
 - Time savings (saving people time)
 - Find hidden or unexpected phenomena
-
- ML learns automatically from historical data to identify unexpected behaviors and patterns and predict future trends



Inspiration

Human Intelligence

“...cognitive capacity and stock of knowledge available at a given time.” (Maier)

“Ability to think abstractly and rationally and to derive purposeful knowledge and action from it.” (Duden)

→ Knowledge and Adaptation on the basis of information obtained from previously acquired data acquisition.

Machine

A machine which perceives its condition and environment and learns from and adapts to it.

Artificial General Intelligence (Strong AI)

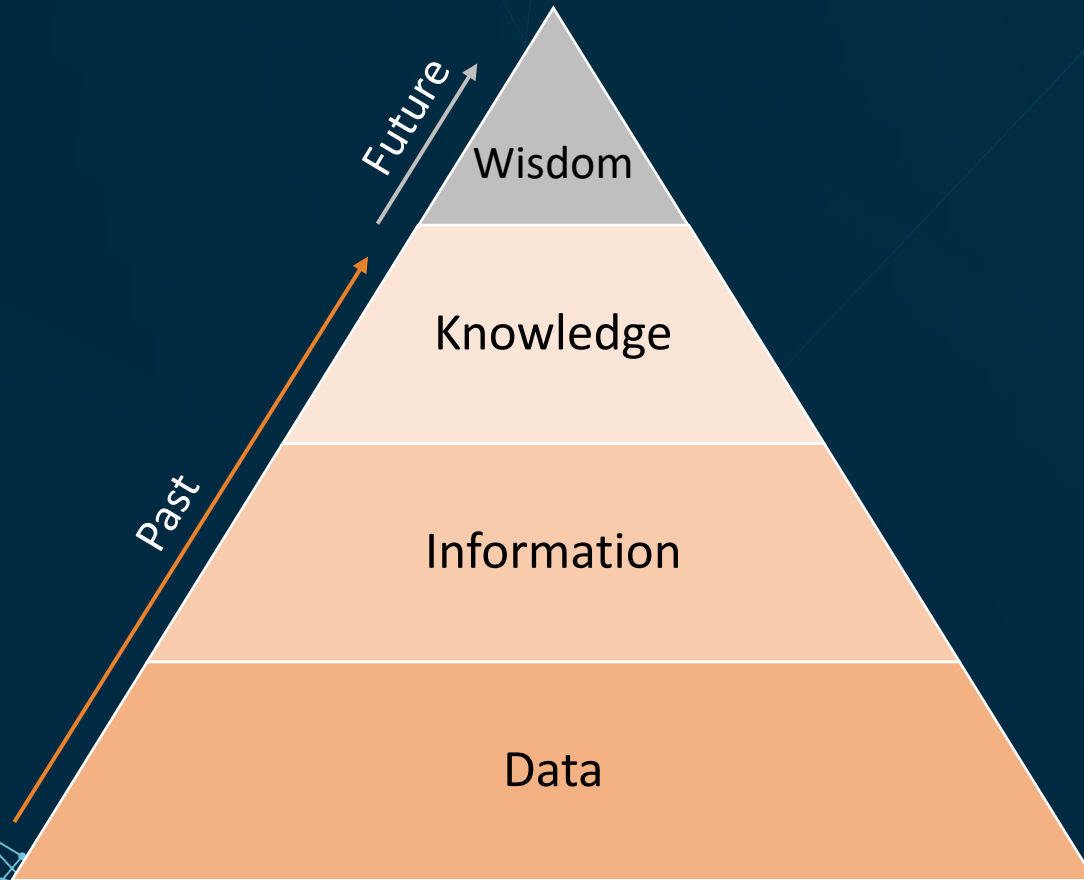
The capability of a machine to learn anything that a human being can.

vs.

Artificial Narrow Intelligence (Weak AI)

The capability of a machine to learn a specific problem at least as good as but rather better than a human being can.

What is AI?



Without analytical processing, data are just a series of bits. Data is important, but don't stop there.

Most important requirement for most AI application is data. You need to understand the types of data to understand how to handle it

- Qualitative: *Categorical data in ordered (nominal) or unordered (ordinal) way*
- Quantitative: *Numerical data in discrete or indiscrete way*
- Single values: *Time-independent data points*
- Time-series: *Time-dependent series of data points*
- Structured: *Data which can be represented in a table*
- Unstructured: *Images, signals, sounds, ...*
- Historical: *Stored data from the past*
- Present: *Data from current processes*

Data
Science

Machine
Learning

Deep
Learning

Artificial
Intelligence

Data Science

- Statistical methods to analyze hypotheses about data

Machine Learning

- Statistical learning methods to build models

Deep learning

- Artificial neural networks that try to mimic the human brain as we understand it

Artificial Intelligence

- Systems that resemble human intelligence (learn, reason, convey information)

Machine Learning and AI Overview

How do we change traditional programming around for machine learning?



Machine Learning and AI Overview

How do we change traditional programming around for machine learning?

Traditional Programming

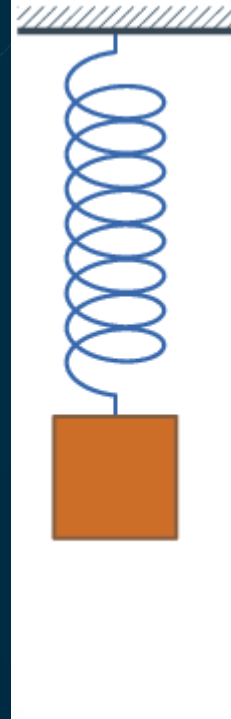


Machine Learning



How are things learned?

- Observe a set of examples – Training Data
 - Displacement measurements relative to mass changes
- Infer something about the process that generated the data
 - Fit a polynomial curve using linear regression
- Use inference to make predictions about previously unseen data – Test Data
 - Predict displacement for other weights



Tested Mass vs Distance

Mass 1 = Distance 1
Mass 2 = Distance 2
Mass 3 = Distance 3
Mass n = Distance n



Untested Mass vs Distance

Mass 1.5 = Predicted Distance 1.5

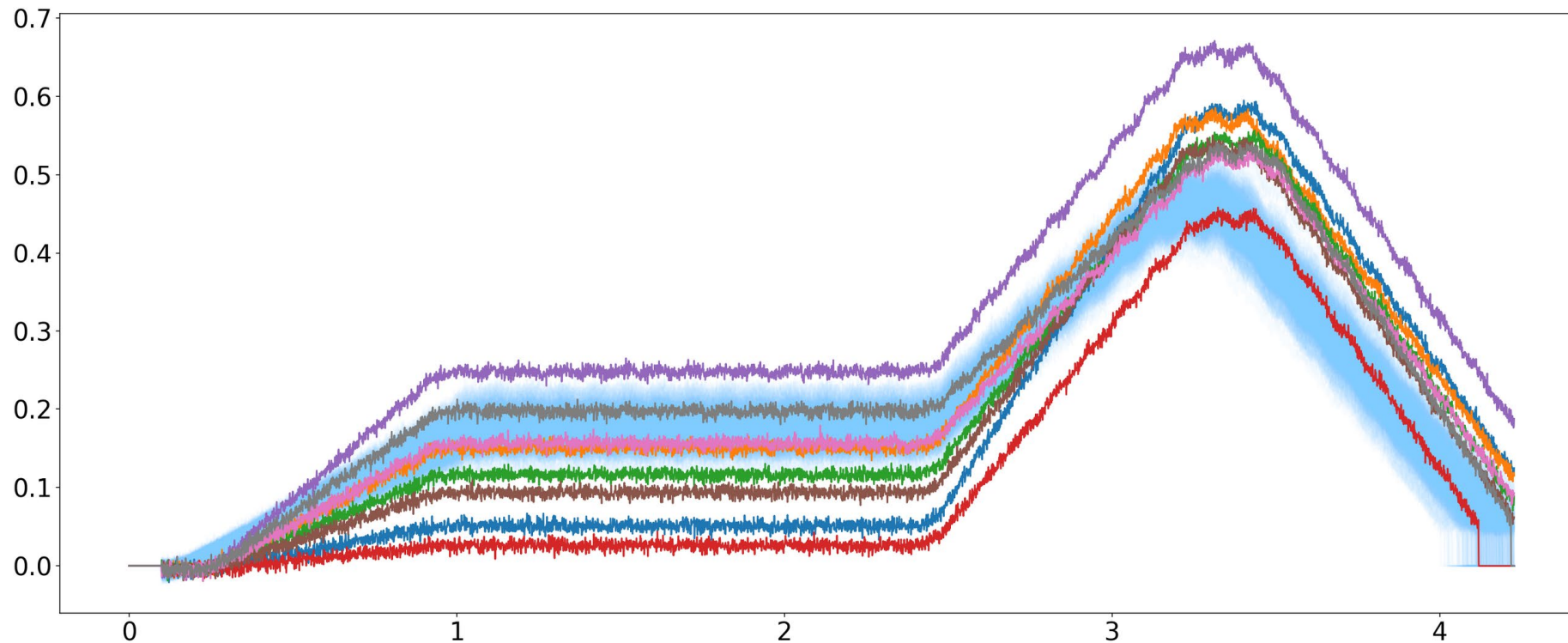


- Question being asked (Objectives)
- Knowledge and/or experience
- Company culture
- Timeline
- Being willing to take that first step...



Dataset - Synthetic Process Dataset

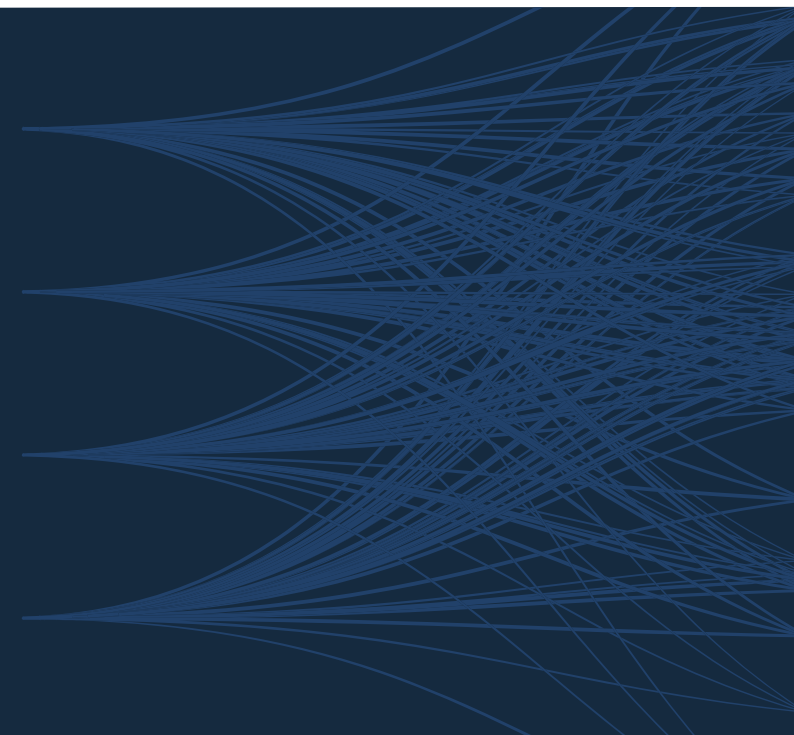
Training Data
Dark lines represent 'bad' data
Light blue represents 'good' cycles



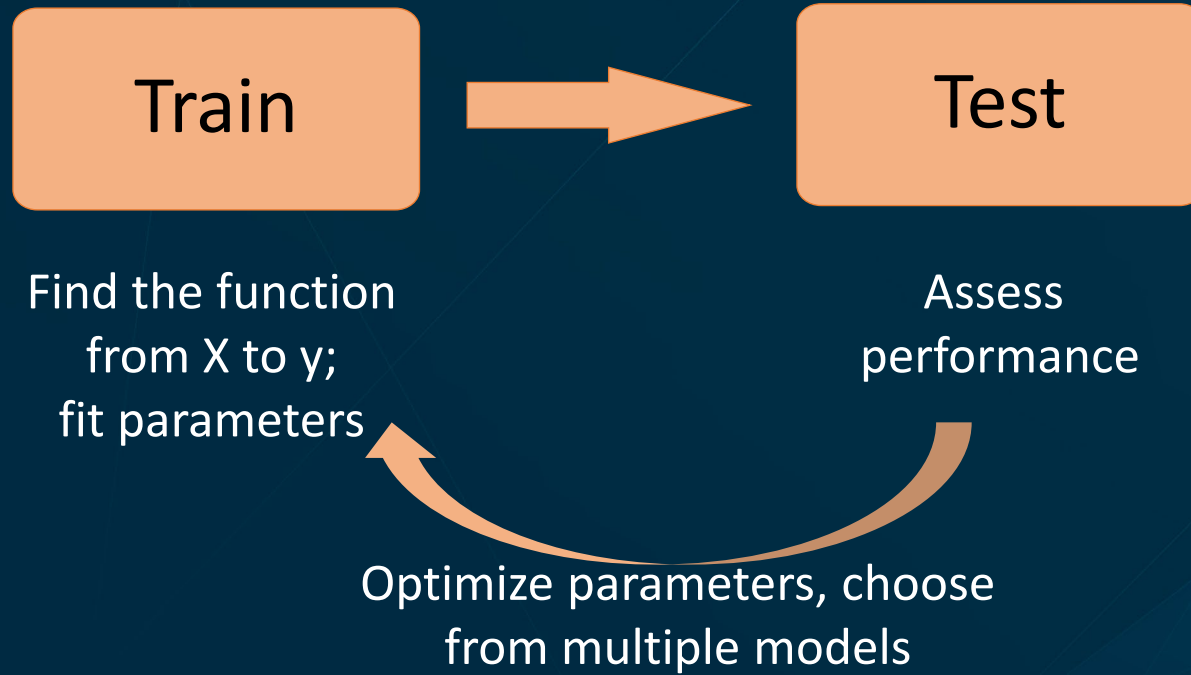
Importing and Cleaning Data



Data Analysis



Modeling Process Overview

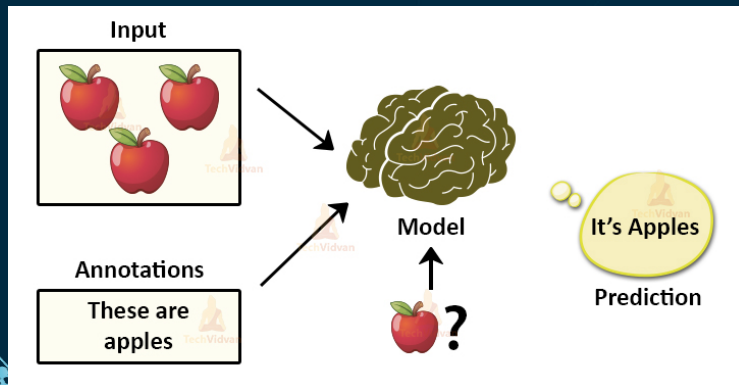


The Data used at each stage should be independent from the other stages



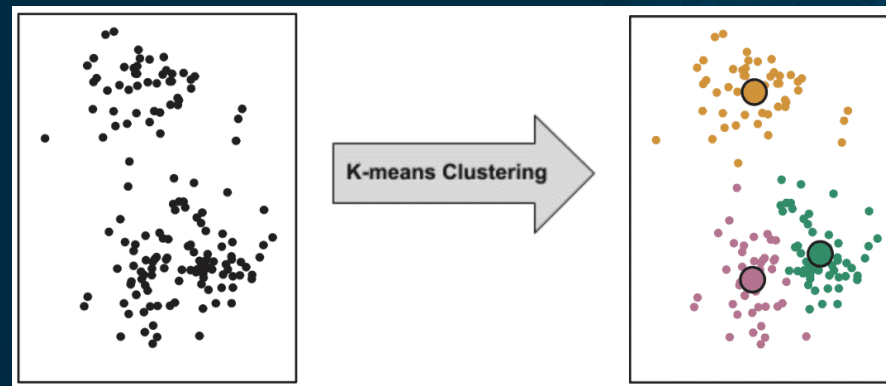
Supervised Learning

- The model compares its predictions with a real result and updates its parameters according to that.
- Required: Training data with known results for the predicted output (labeled data)

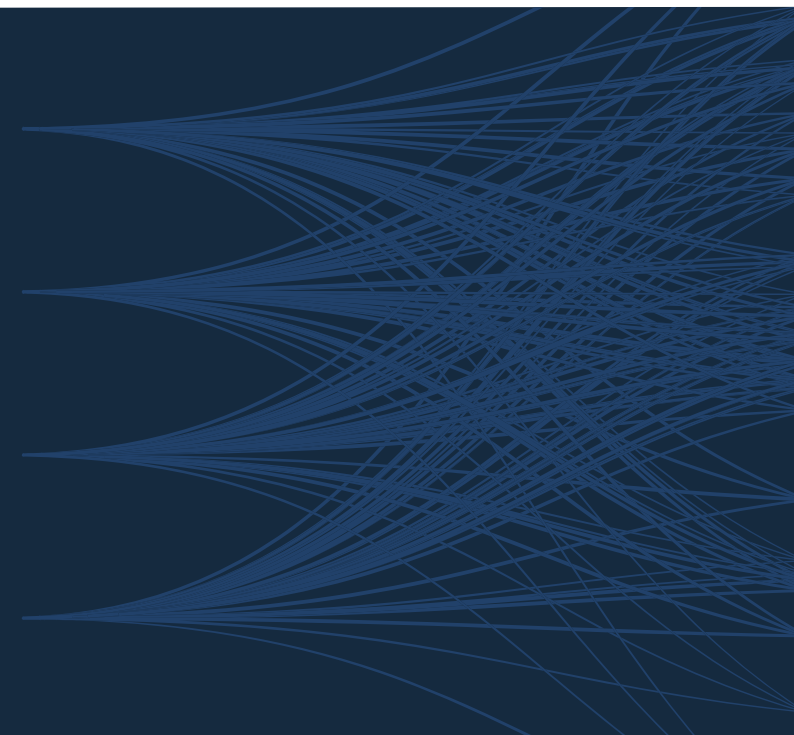


Unsupervised Learning

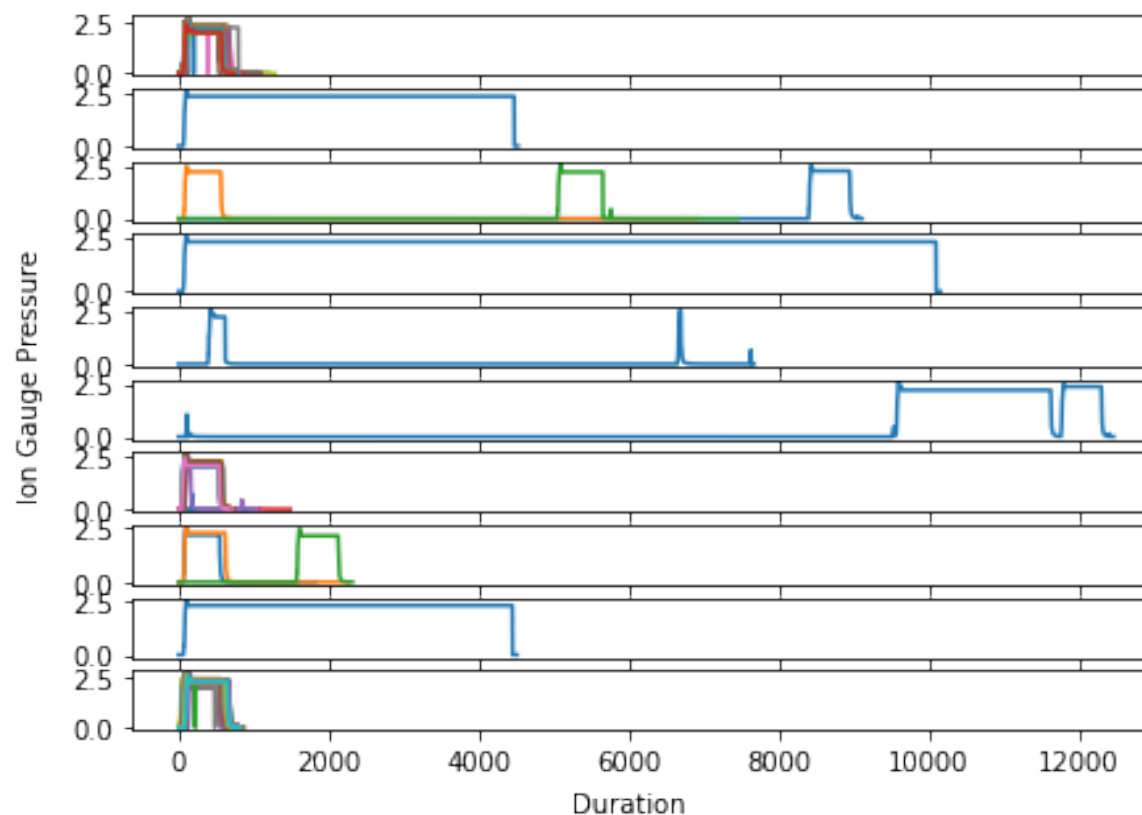
- The model clusters data or
- The model reduces the data to make it more interpretable
- Required: Usually a large data set to recognize patterns



Model Building



- Explore data to determine how to parse
- Unsupervised Learning to cluster based on similarity
- Next... label clusters to run supervised learning



Andy Henderson, Ph.D.

Email:

andy.henderson@praemo.com

Phone: (864) 230-9015

LinkedIn:

<https://www.linkedin.com/in/andyhend/>

Website:

<https://praemo.com>

Matthew Krugh, Ph.D.

Email:

mkrugh@clemson.edu

Phone: (412) 401-7764

LinkedIn:

<https://www.linkedin.com/in/matthew-krugh>

Website:

<https://clemsonvac.com/>

