Class introduction

APSC 8280: Machine learning applied to plant science

03/19/2020

Outline

- Introductions

- Syllabus

- Introduction to Machine Learning

- Brief History of Machine Learning

- Lab 0: laptop setup

- Basic R programming I

Introductions

- Name

- Research

- Expectations from this class

- Something about you no one would have ever guessed

Syllabus: what to expect from this class

The basics

History of ML
Fundamentals of ML
Strengths and weaknesses of different ML models

Asking the right questions

Data cleaning and preprocessing
Feature engineering

Applied ML

Quantitative genetics
Plant disease epidemiology
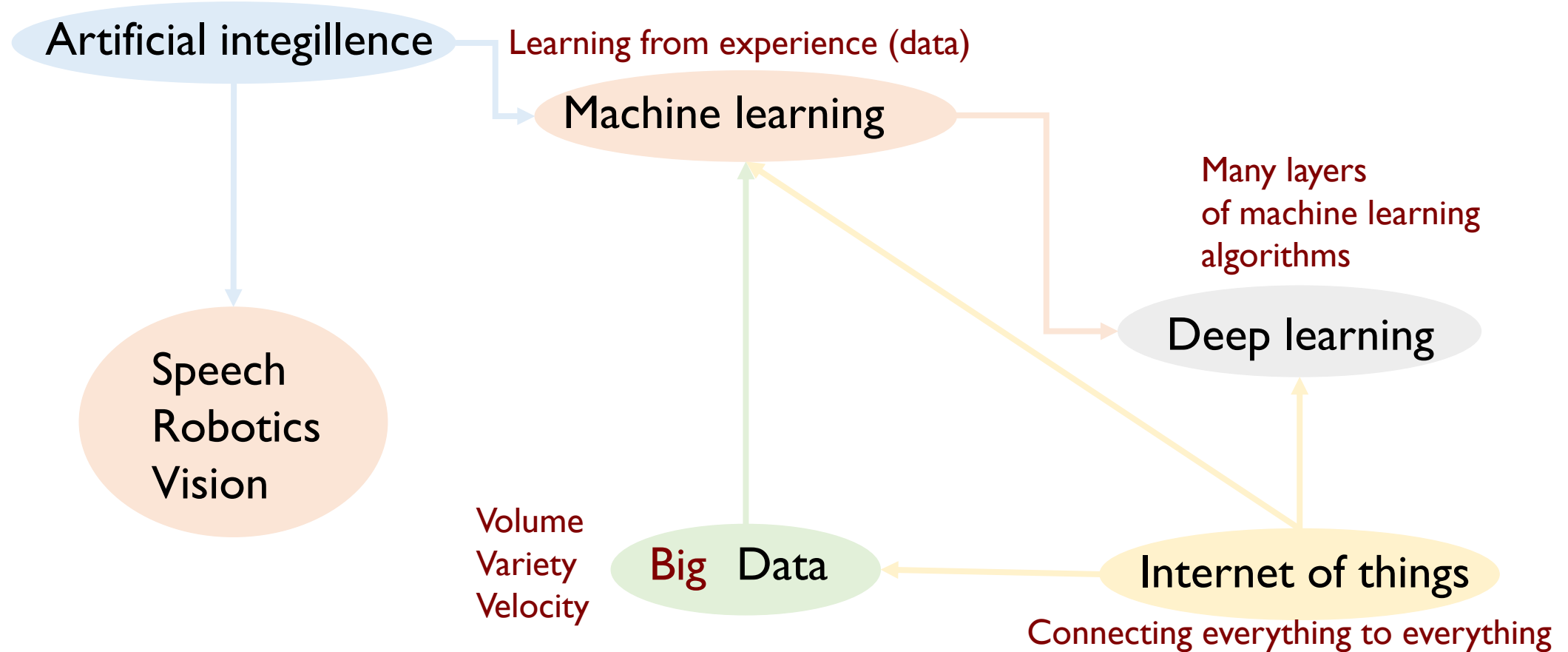Precision agriculture

Advanced topics

Deep learning
Computer vision
Reinforcement learning

What is not covered?

- Math behind machine learning algorithms and optimization

- Working with cloud and enterprise machine learning tools
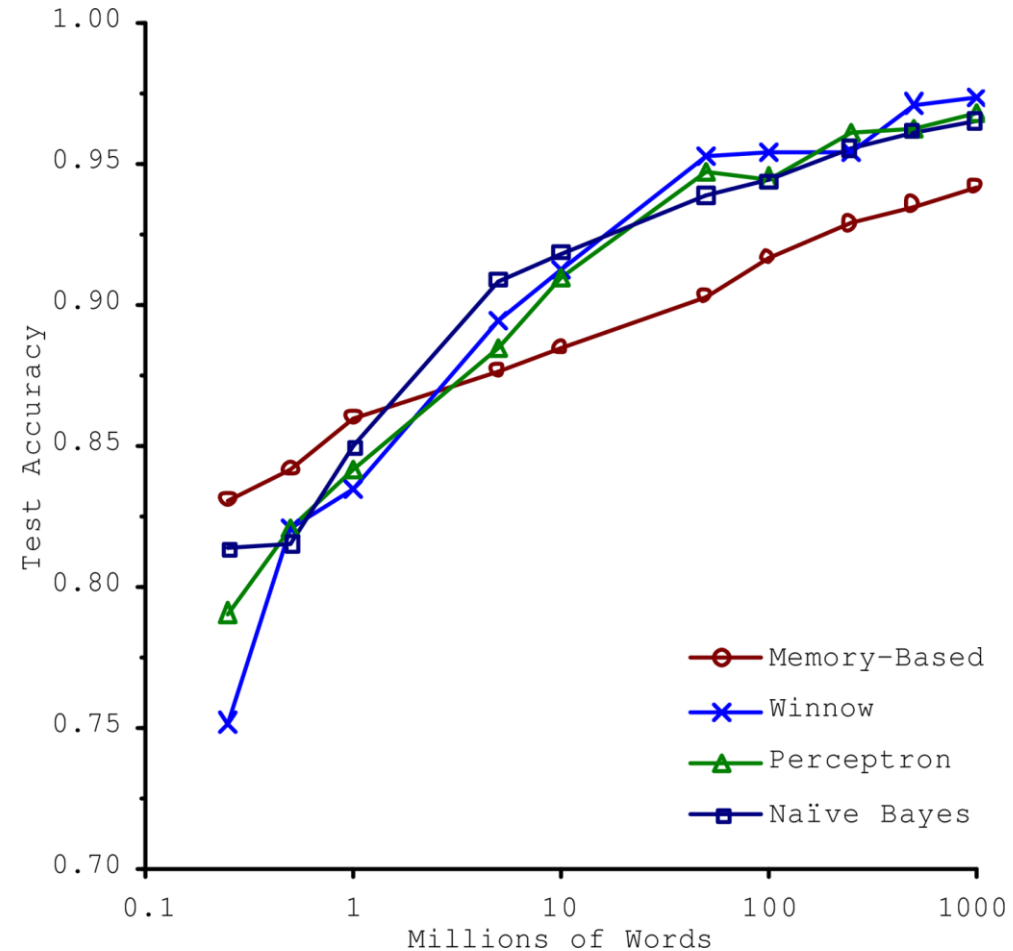
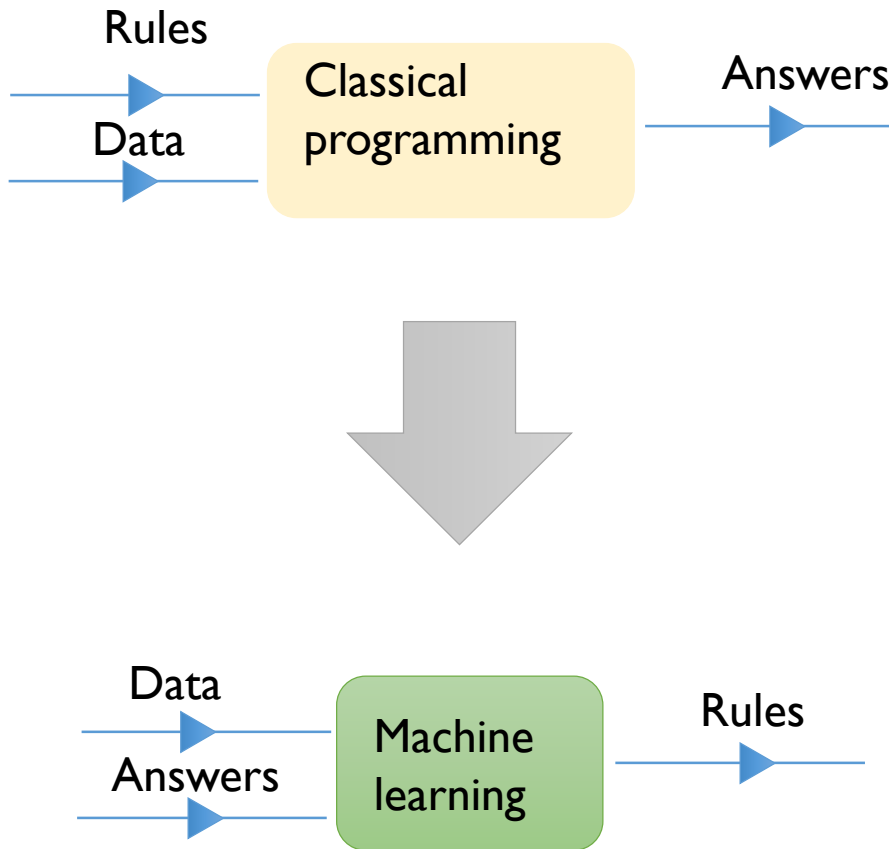- Deployment and maintenance of machine learning systems

What's what?

Getting computers to do things humans do

Artificial integillence

Learning from experience (data)

Machine learning

Many layers
of machine learning
algorithms

Speech
Robotics
Vision

Deep learning

Volume
Variety
Velocity

Big Data

Internet of things

Connecting everything to everything
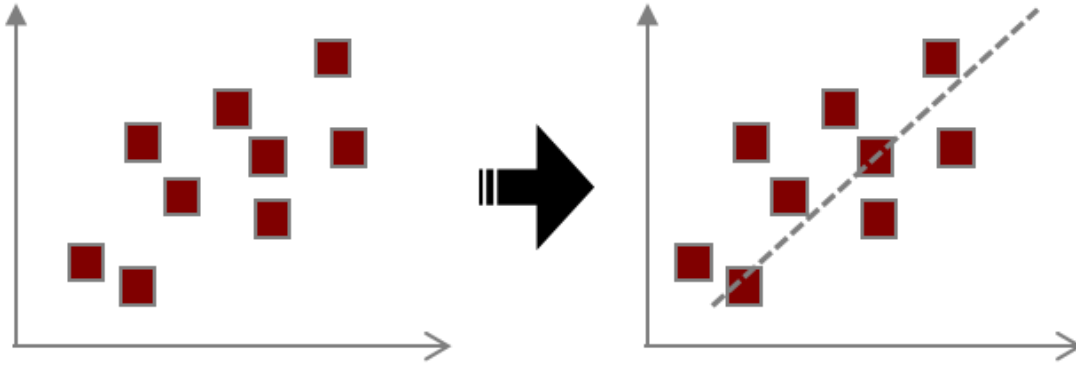
Machine learning: a paradigm shift

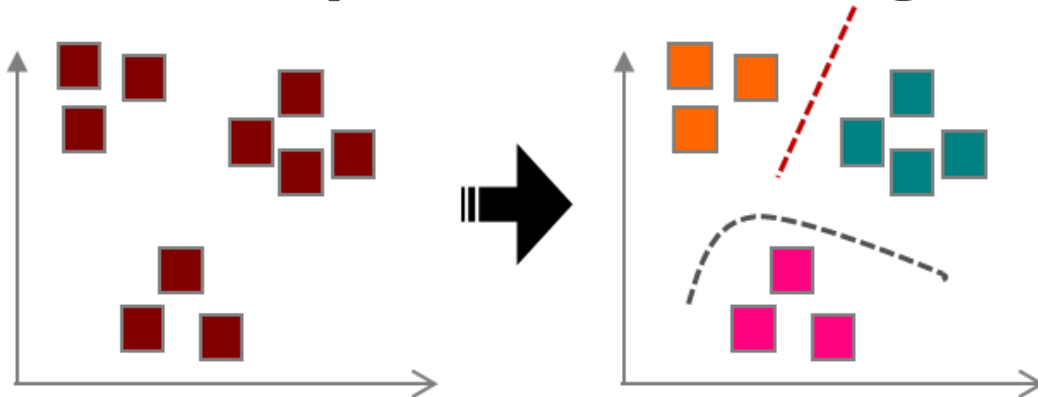## "The unreasonable effectiveness of data"

- Banko, Michele, and Eric Brill. "Scaling to very very large corpora for natural language disambiguation." Proceedings of the 39th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2001.
- Adapted from Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.
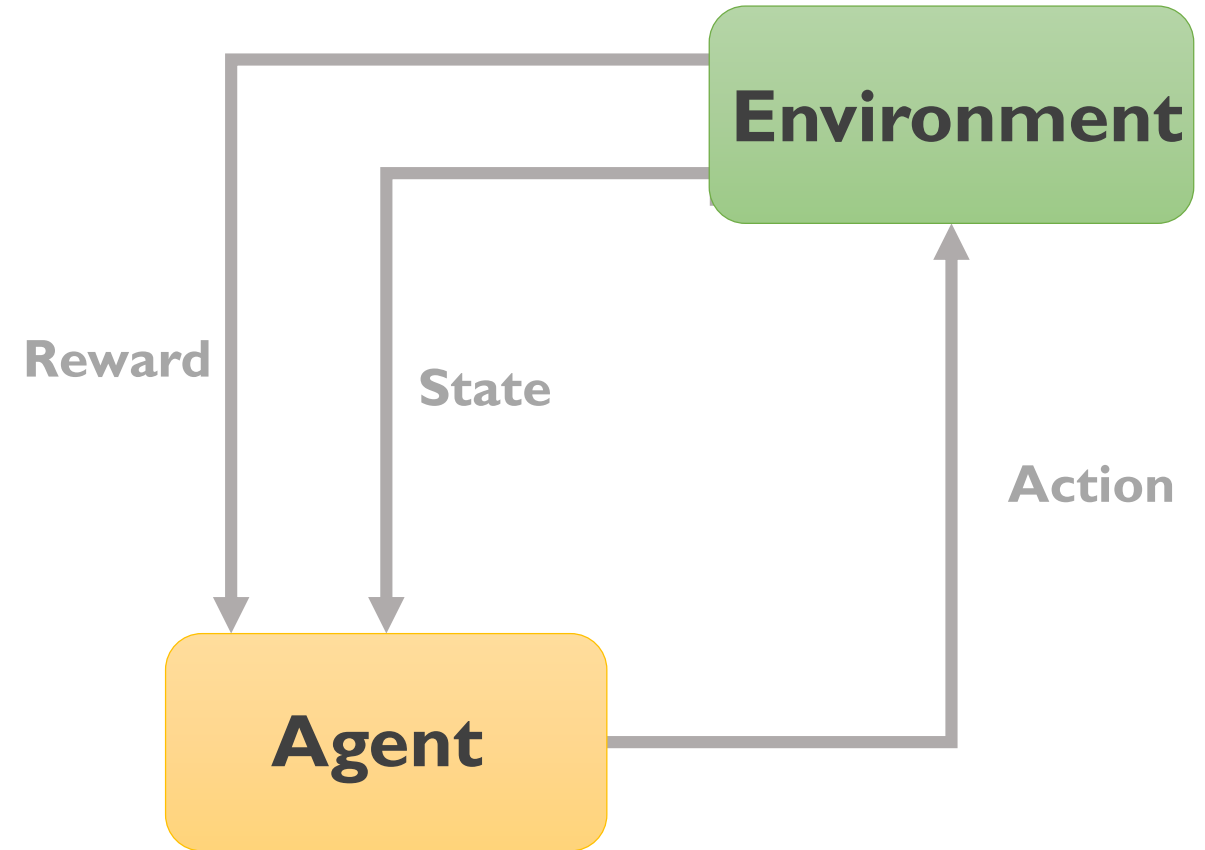
Types of machine learning

Supervised learning

Unsupervised learning

Reinforcement learning

Reward   State   Action

Environment

Agent

Adapted from Raschka, Sebastian, and Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing Ltd, 2019.

Some tasks best solved by learning

Prediction

➢ Future stock prices or currency exchange rates

➢ Which movies will a person like

➢ The performance of crossing any two inbred and tester parents

Recognizing patterns

➢ Objects in real scenes

➢ Facial identities or facial expressions

➢ Spoken words

Recognizing anomalies

➢ Unusual sequences of credit card transactions

➢ Unusual patterns of sensor readings in a nuclear power plant

Adapted from Geoffrey Hinton slides

What has machine learning achieved so far?

Human-like skills

✓ Image classification

✓ Speech transcription

✓ Handwriting transcription

✓ Autonomous driving

Commerce

✓ Ad targeting

✓ Search results

✓ Customer service

✓ Digital assistants (e.g. Alexa)

✓ Reduction in fraudulent credit card transactions

✓ Stock market and investing

Healthcare

✓ Personalized/precision medicine

✓ Drug discovery

✓ Discovery of genetic sequences linked to diseases

✓ Medical imaging diagnosis

✓ Identifying disease and diagnosis

Government/Politics

✓ Police work (crime prediction)

✓ Election results prediction

✓ Traffic management
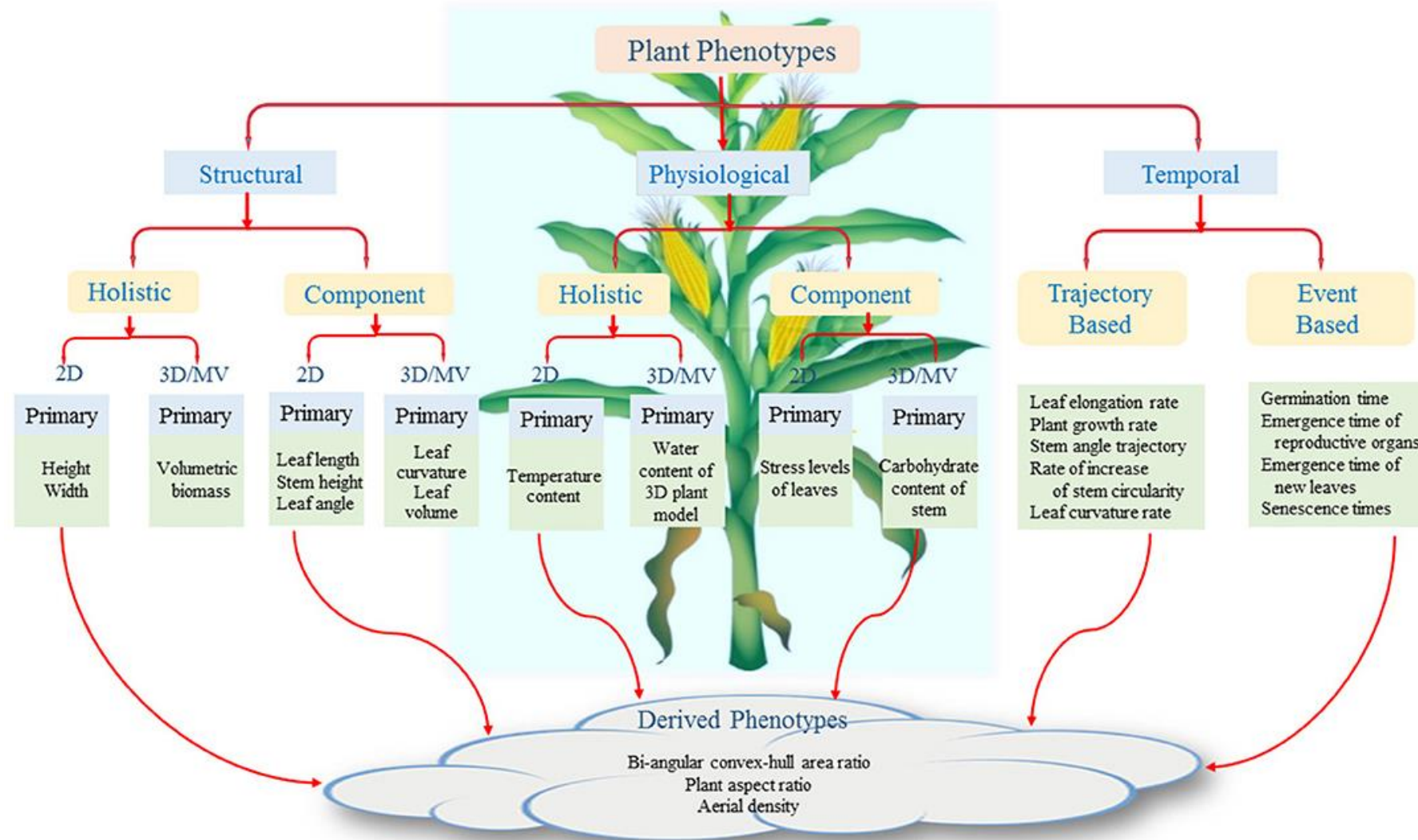
Manufacturing/BioTech

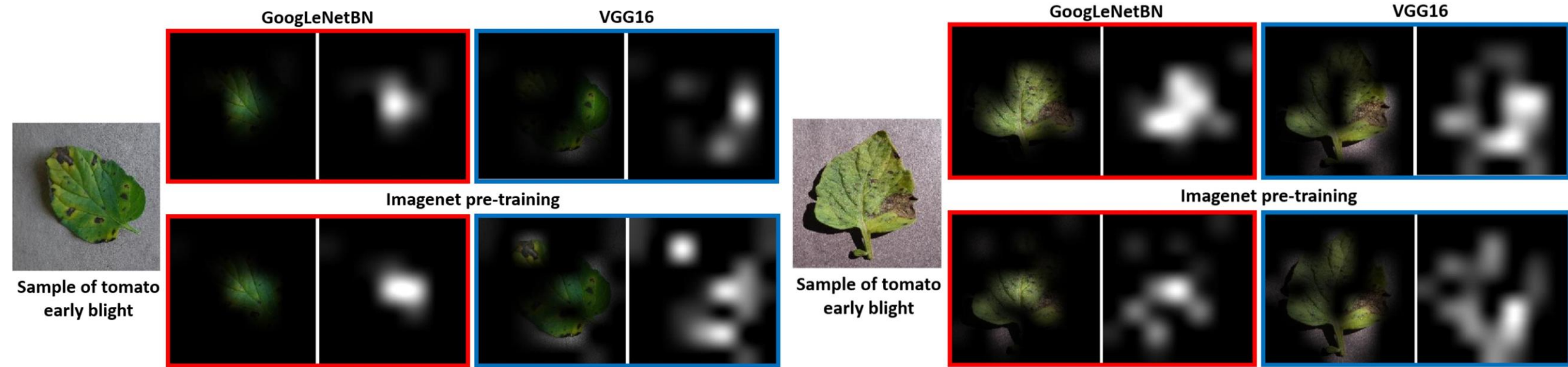✓ Smart manufacturing

✓ Product design

And lots more!!!

Examples of ML in plant science
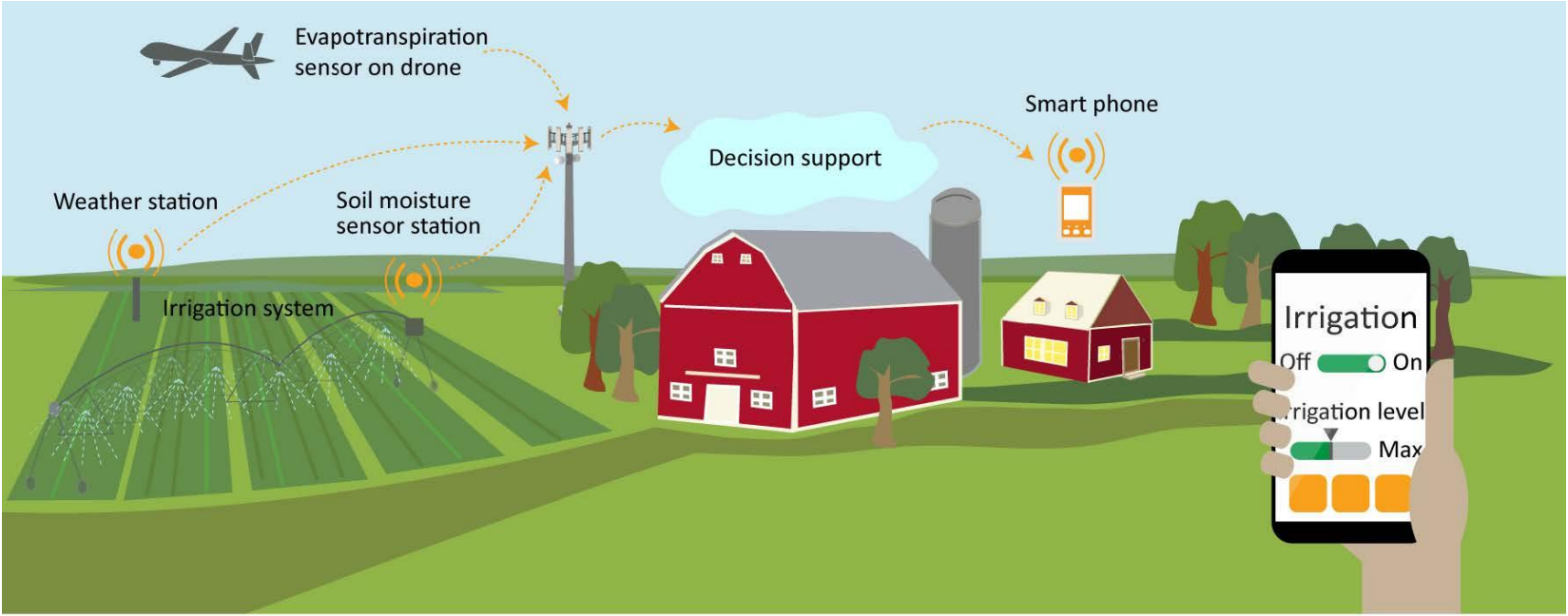
High throughput phenomics

Choudhury, Sruti Das, Ashok Samal, and Tala Awada. "Leveraging image analysis for High-Throughput plant phenotyping." *Frontiers in plant science* 10 (2019).

Examples of ML in plant science

Plant disease recognition and characterization

Lee, Sue Han, et al. "New perspectives on plant disease characterization based on deep learning." Computers and Electronics in Agriculture 170 (2020): 105220.

Examples of ML in plant science

Precision agriculture

Crop yield prediction

	Tester 1	Tester 2	Tester 3
Inbred 1	X	X	X
Inbred 2	X	?	?
Inbred 3	?	X	X
Inbred 4	?	?	X
Inbred 5	X	X	?

Quiz: supervised or unsupervised?

Spam detection   **supervised**

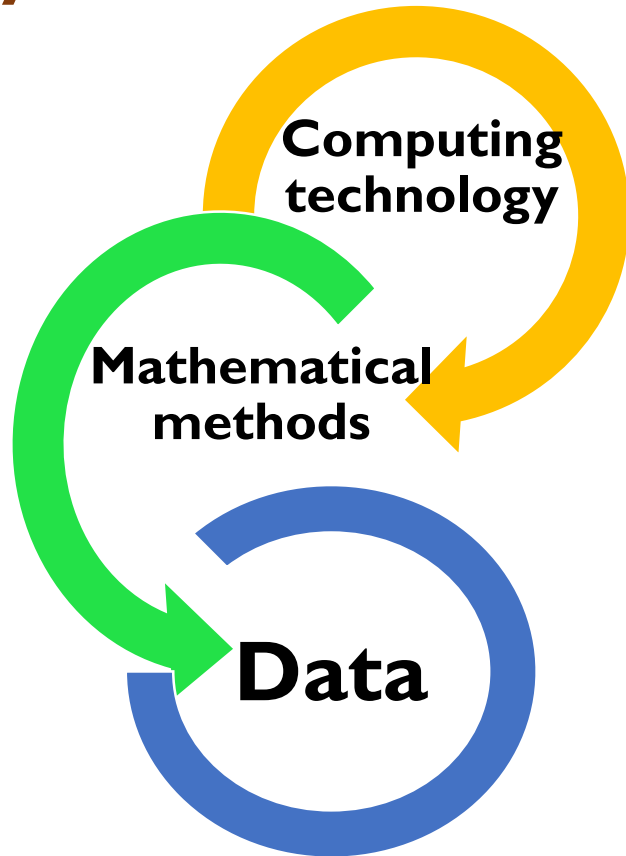Anomaly detection   **unsupervised**

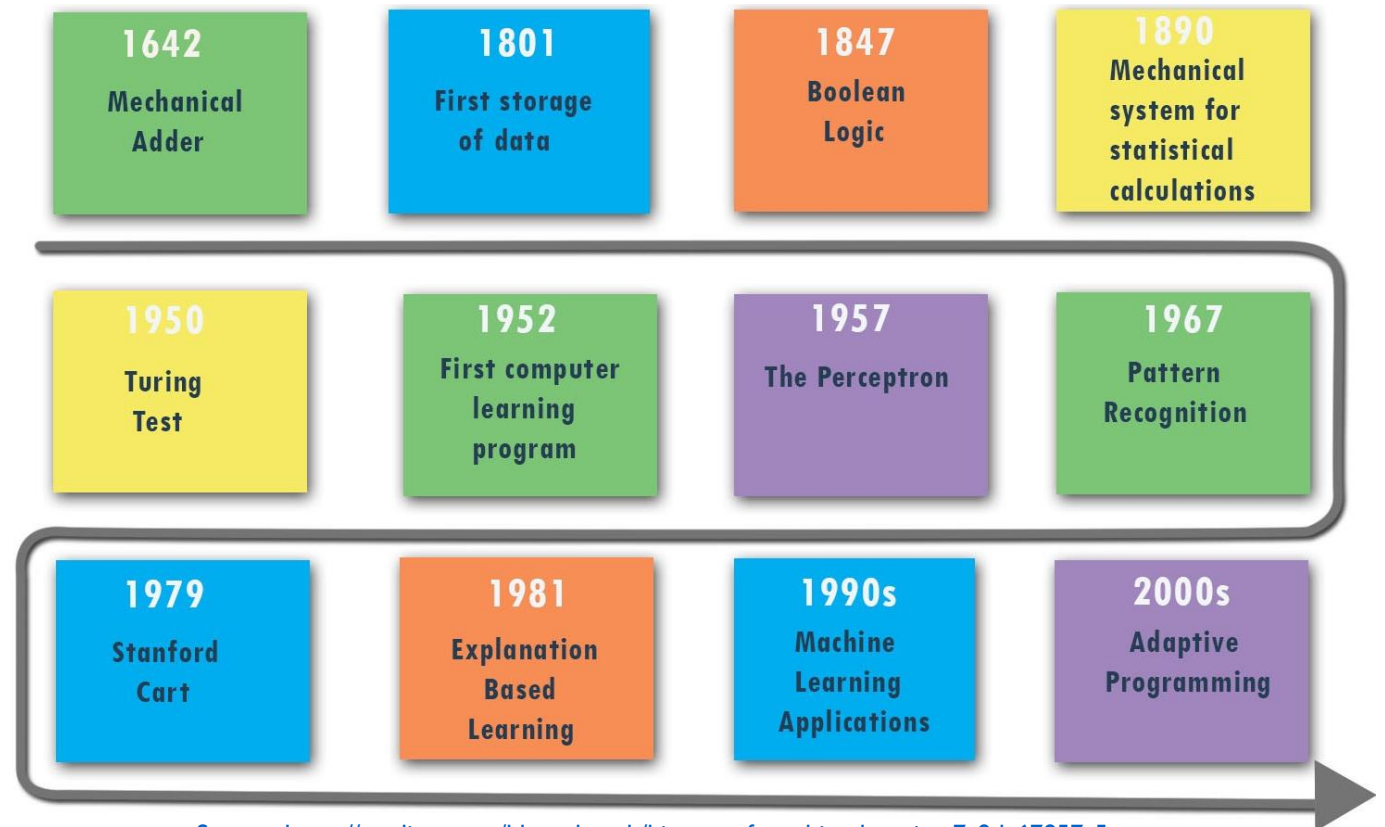Handwriting recognition   **supervised**

Object recognition   **Unsupervised/supervised**
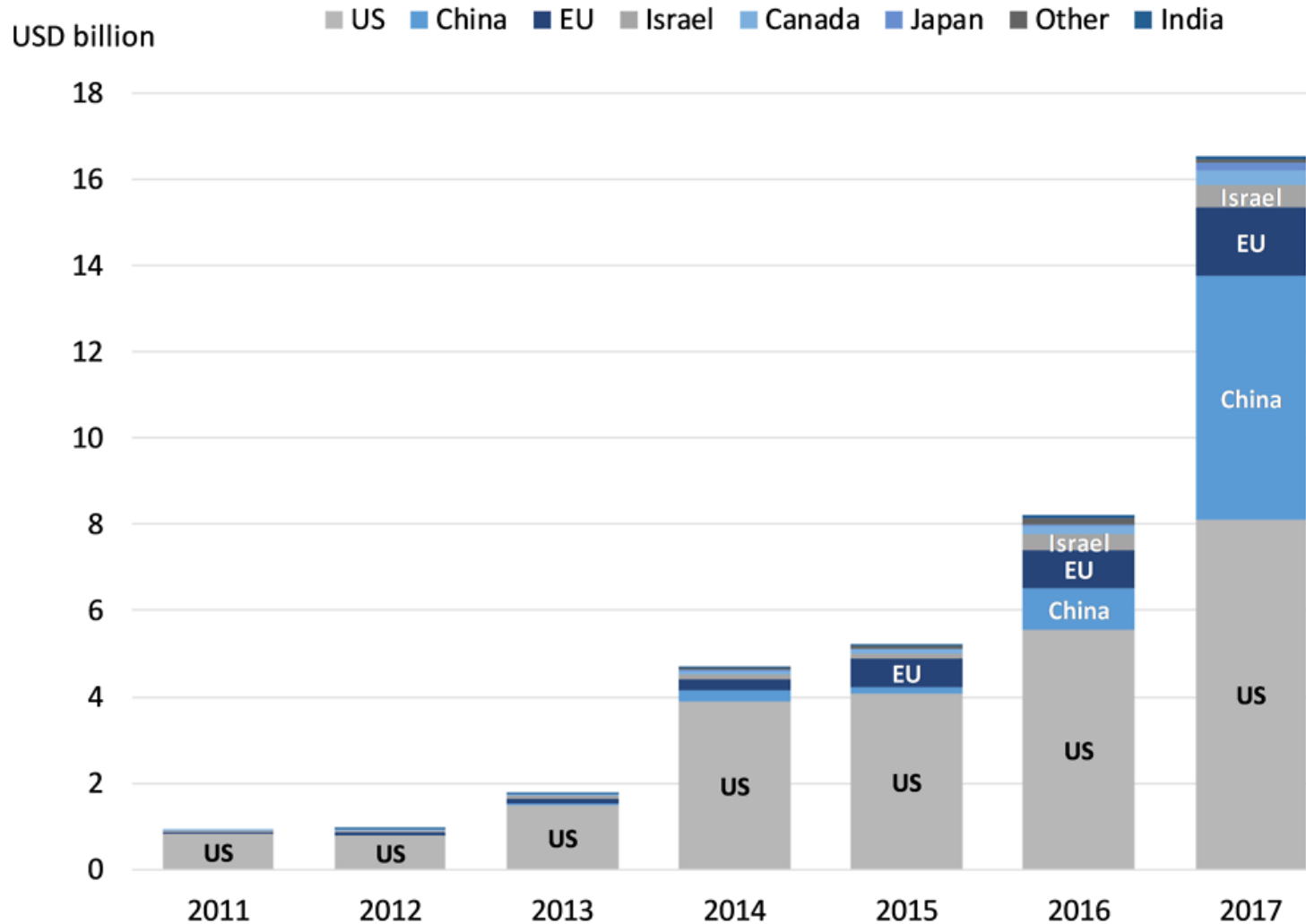
History of machine learning

Cycle of advancement

- Computing technology
- Mathematical methods
- Data

Brief history

1642 Mechanical Adder	1801 First storage of data	1847 Boolean Logic	1890 Mechanical system for statistical calculations
1950 Turing Test	1952 First computer learning program	1957 The Perceptron	1967 Pattern Recognition
1979 Stanford Cart	1981 Explanation Based Learning	1990s Machine Learning Applications	2000s Adaptive Programming

Source: https://medium.com/bloombench/history-of-machine-learning-7c9dc67857a5

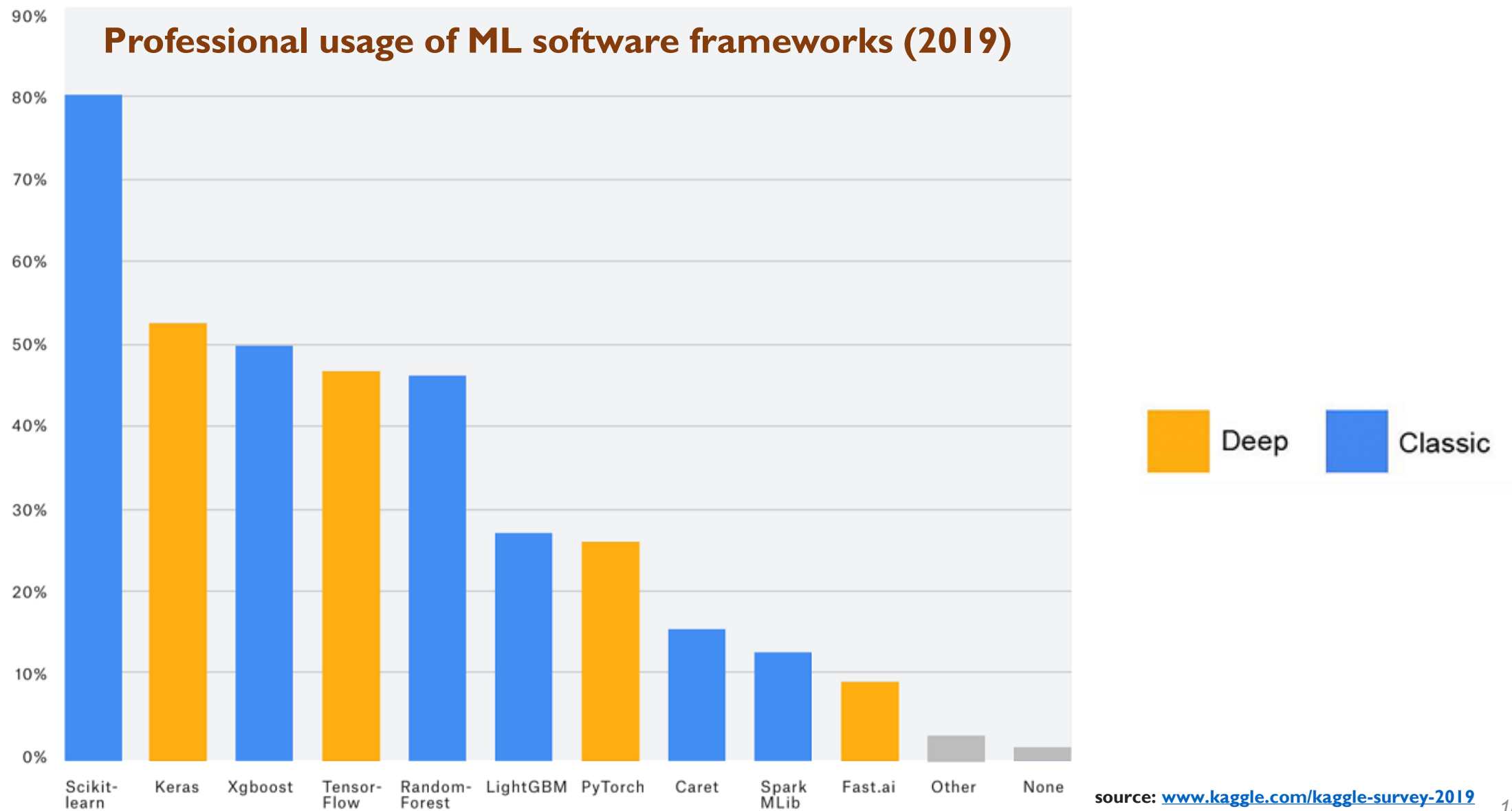**Explore the history of machine learning with Google:**

https://cloud.withgoogle.com/build/data-analytics/explore-history-machine-learning/

Current landscape

Total estimated investments in AI startups (2011-2017)

Current landscape

Professional usage of ML software frameworks (2019)

Lab 0: Laptop setup

Reproducible science

Question: Is reproducibility the same as replicability?

- **Replication:** the confirmation of results and conclusions from one study obtained independently in another

- **Reproducibility:** minimum standard for assessing the value of scientific claims, particularly when full independent replication of a study is not feasible

We want to move from this ...

Does this look familiar?

☐	Name	Date modified	Type
Ⓡ	Rscript_4_21_2016.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_4_22_2016a.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_4_22_2016b.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_4_24_2016.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_final.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_final_final.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_really_final.R	5/1/2016 3:03 PM	R File
Ⓡ	Rscript_really_really_final_final.R	5/1/2016 3:03 PM	R File

to more like this …

```
▼ 📁 letters
   ▶ 📁 cache          Store processed data
   ▶ 📁 config
   ▶ 📁 data           Store raw data
   ▶ 📁 diagnostics
   ▶ 📁 doc
   ▶ 📁 graphs         Store figures
   ▶ 📁 lib
   ▶ 📁 logs
   ▶ 📁 munge          Scripts to preprocess data
   ▶ 📁 profiling
     📄 README
   ▶ 📁 reports        Notebooks/markdowns: exploratory and final results
   ▶ 📁 src            All scripts for analyses
   ▶ 📁 tests          Tests to make sure code works as it should
     📄 TODO
```

Setting up your workstation

❑ R/R studio
- https://www.r-project.org/
- https://rstudio.com/products/rstudio/#Desktop

❑ Git / GitHub

- https://help.github.com/en/github/getting-started-with-githubhttps://desktop.github.com/
- https://towardsdatascience.com/getting-started-with-git-and-github-6fcd0f2d4ac6
- https://happygitwithr.com/index.html

❑ Jupyter lab/ notebooks
- https://www.storybench.org/install-r-jupyter-notebook/
- https://bence.ferdinandy.com/2018/11/22/fastest-way-to-set-up-r-with-jupyter-on-windows/

❑ Data science cookie cutters
- R packages (http://r-pkgs.had.co.nz/)
- R Project template (http://projecttemplate.net/index.html)
- Tools for writing reproducible research in R (https://github.com/benmarwick/rrtools)

❑ Package management (packrat: https://rstudio.github.io/packrat/)