

AMiDST TOOLBOX

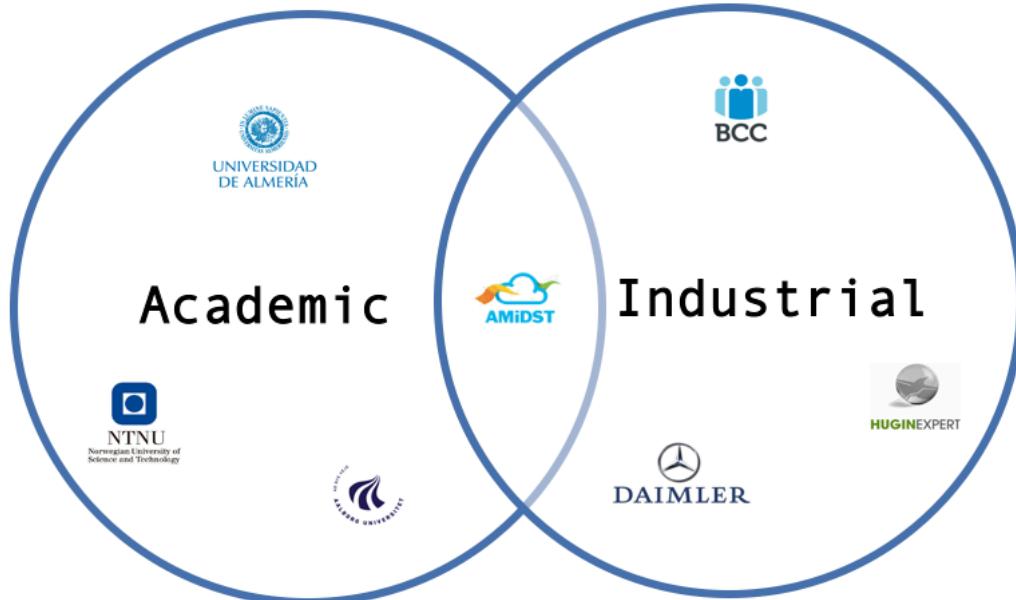
Session 1: Introduction to Probabilistic Machine Learning

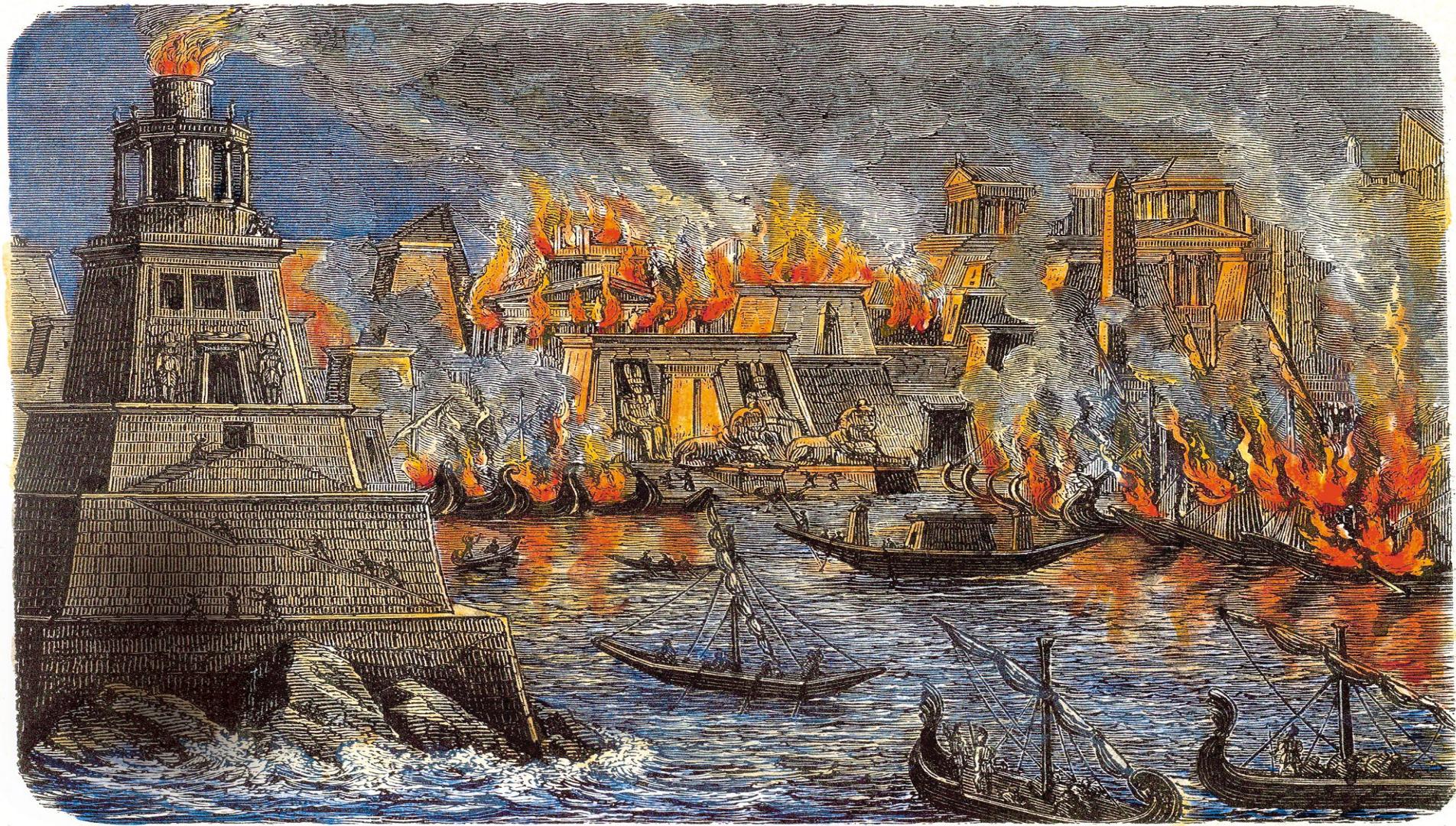
Andrés R. Masegosa

University of Almeria
andres.masegosa@ual.es

THE AMIDST CONSORTIUM

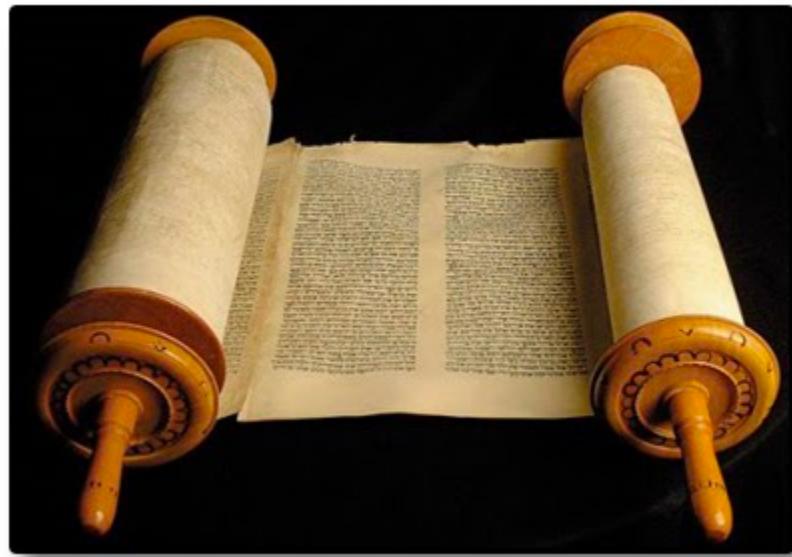
AMIDST
TOOLBOX







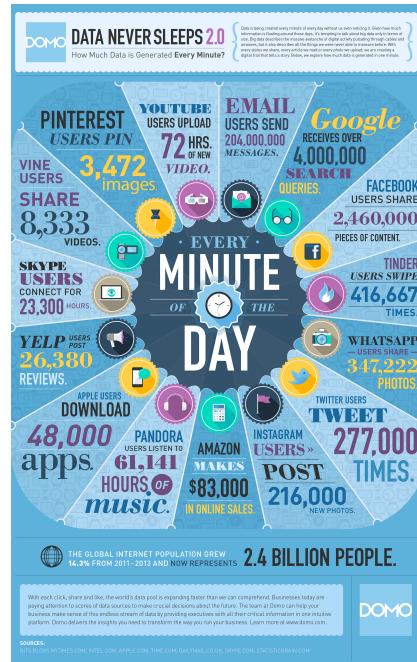
3rd century BC



50 thousands papyrus = 20 thousands books



21st century DC



6 thousand books, 2 millions posts and news daily



The Economist

FEBRUARY 27TH - MARCH 5TH 2010 Economist.com

Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

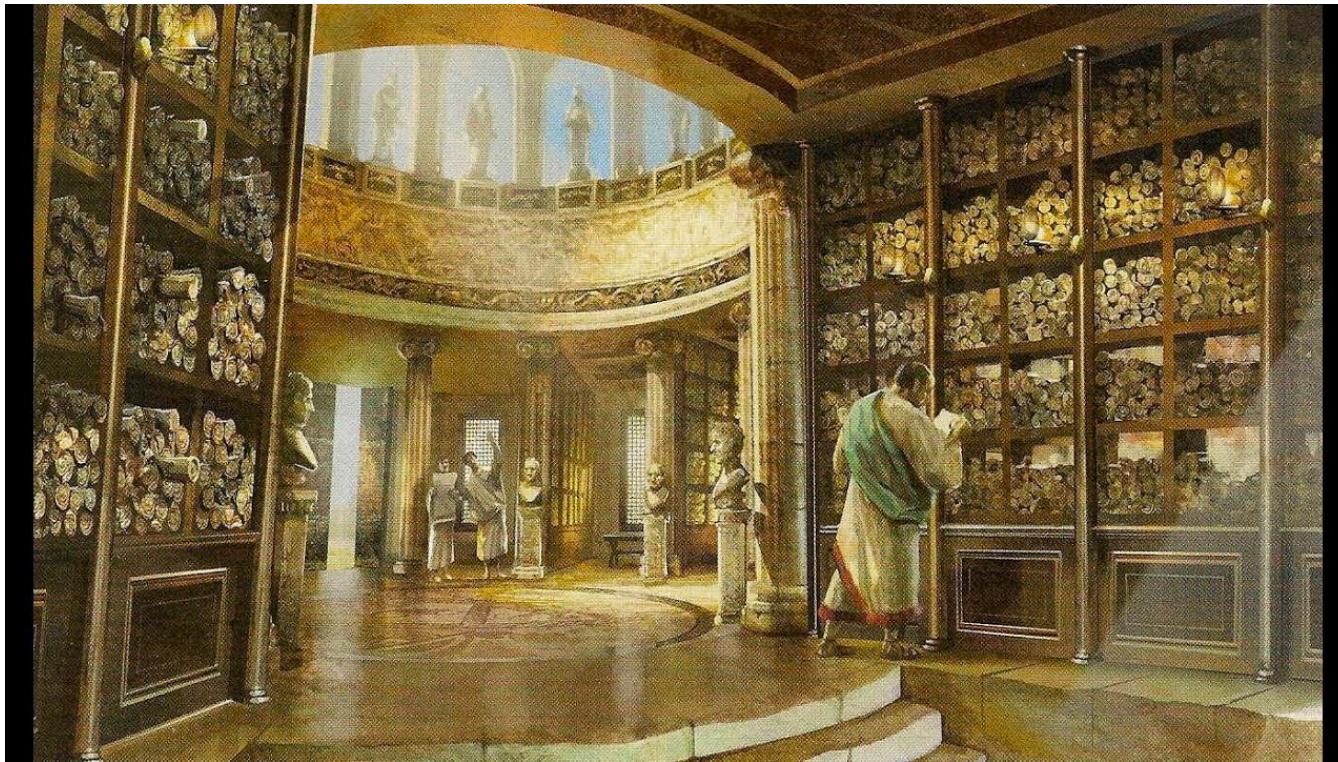
A black and white illustration of a man in a dark suit and tie standing under a large, multi-colored umbrella (green, yellow, red) that is tilted over a small, delicate flower. The man is holding a watering can and pouring water onto the flower. The background is a dense grid of binary code (0s and 1s).



Machine Learning



Acceso al conocimiento (siglo III a.c.)



Acceso al conocimiento (siglo XXI d.c.)

The central image is a dark teal square featuring a white brain-like network of nodes and connections, with the words "MACHINE LEARNING" written vertically to its left.

Below this is a photograph of a data center with several tall, black server racks standing in rows.

At the bottom of the box are three logos:

- TensorFlow**: A logo consisting of a stylized orange 'T' shape with a yellow gradient, next to the word "TensorFlow" in orange.
- NumPy**: A logo featuring a 3x3 grid of colored cubes (blue, yellow, and red) followed by the word "NumPy" in blue.
- ΛMiDST Toolbox**: A logo with the text "ΛMiDST" in white and "TOOLBOX" in grey, separated by a horizontal arrow pointing right, all contained within a blue rectangular background.



Manual Computer Programming



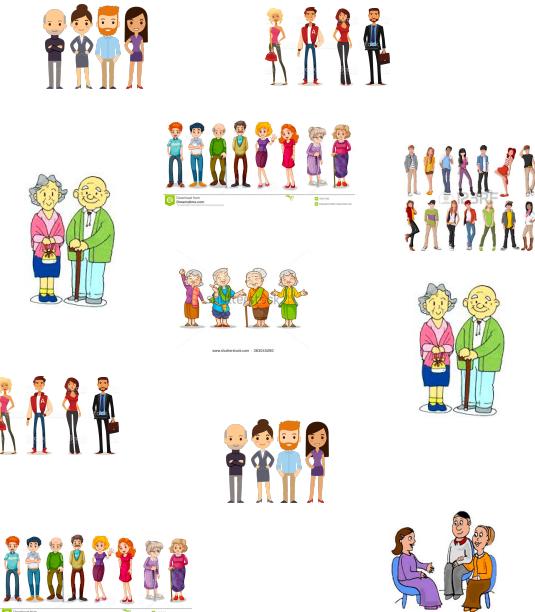
vs



Automatic Computer Programming



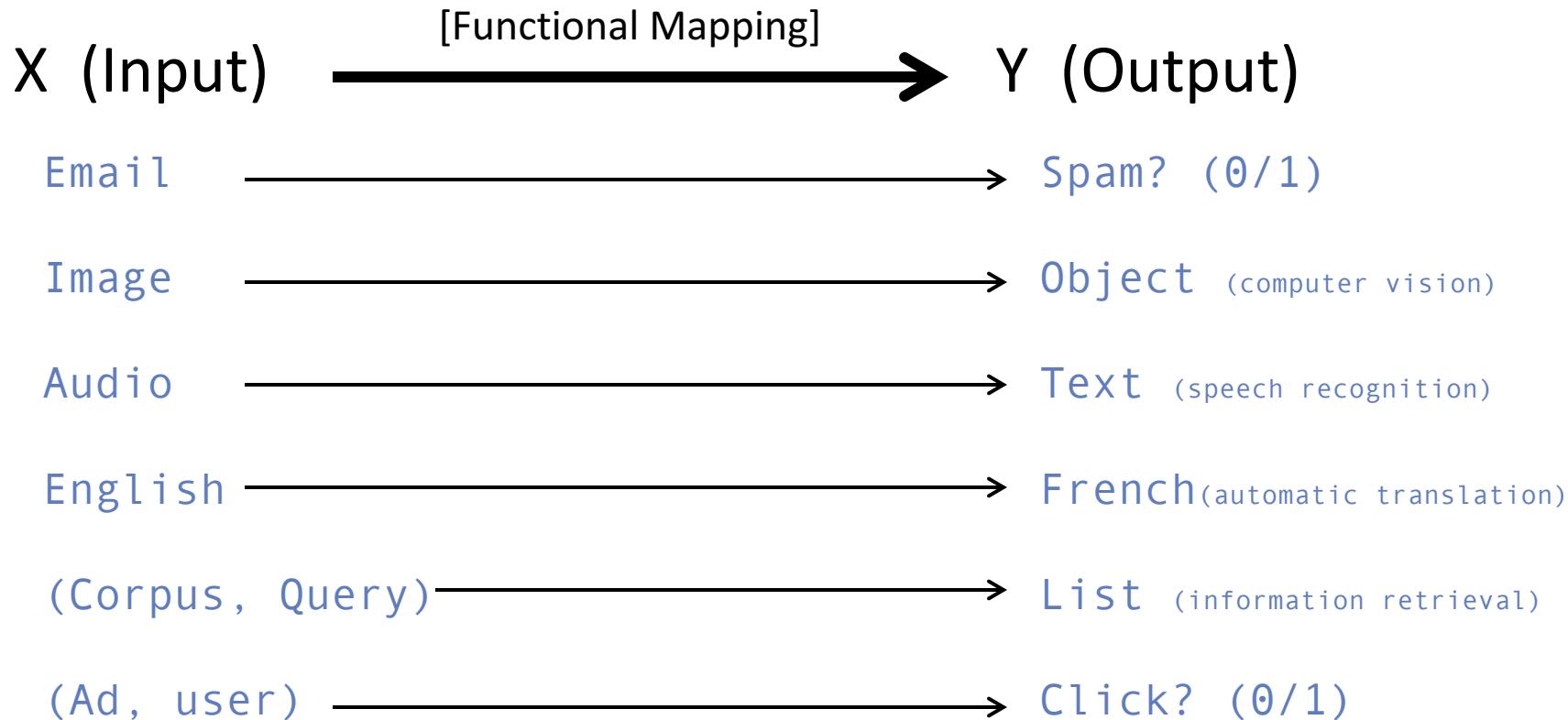
VS



Machine Learning

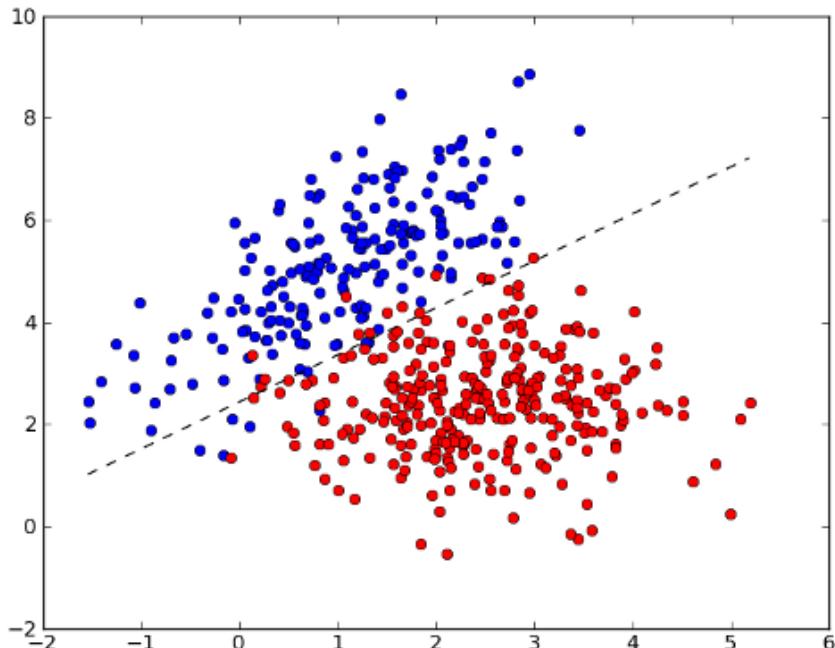


- Supervised Learning:



Andrew Ng: Artificial Intelligence is the New Electricity.
<https://www.youtube.com/watch?v=21EiKfQYZXc&t=1206s>

- Mappings as a function-finding problem:



$$f : X \rightarrow Y$$

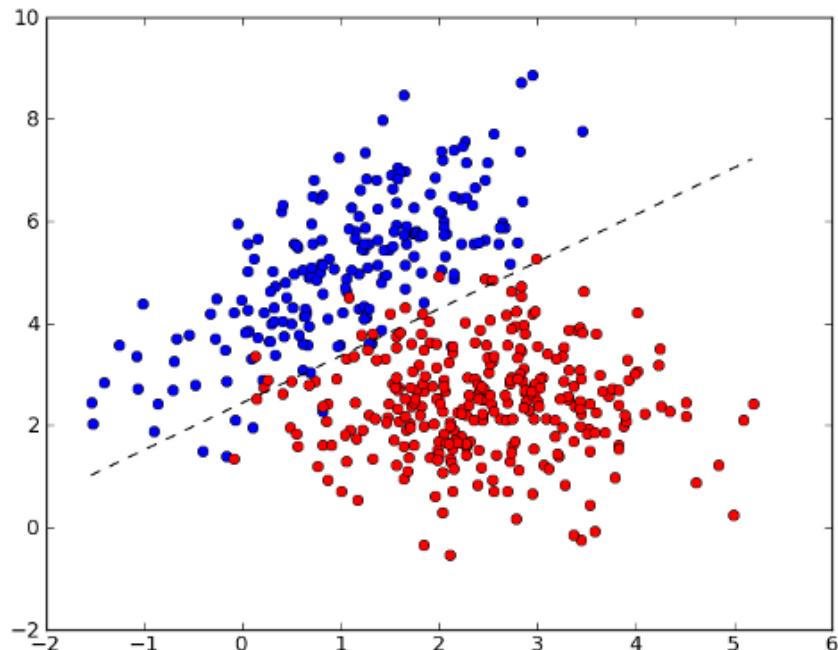
$$x \in \mathbb{R}^2 \quad y \in \{\text{Red}, \text{Blue}\}$$

$$f(x; \theta) = \begin{cases} \text{Blue} & \theta^T x \geq 0 \\ \text{Red} & \theta^T x < 0 \end{cases}$$

$$\theta \in \mathbb{R}^2$$

The mapping problem reduces to find the suitable θ^* .

- How do we find θ^* ?
 - We learn it from data!



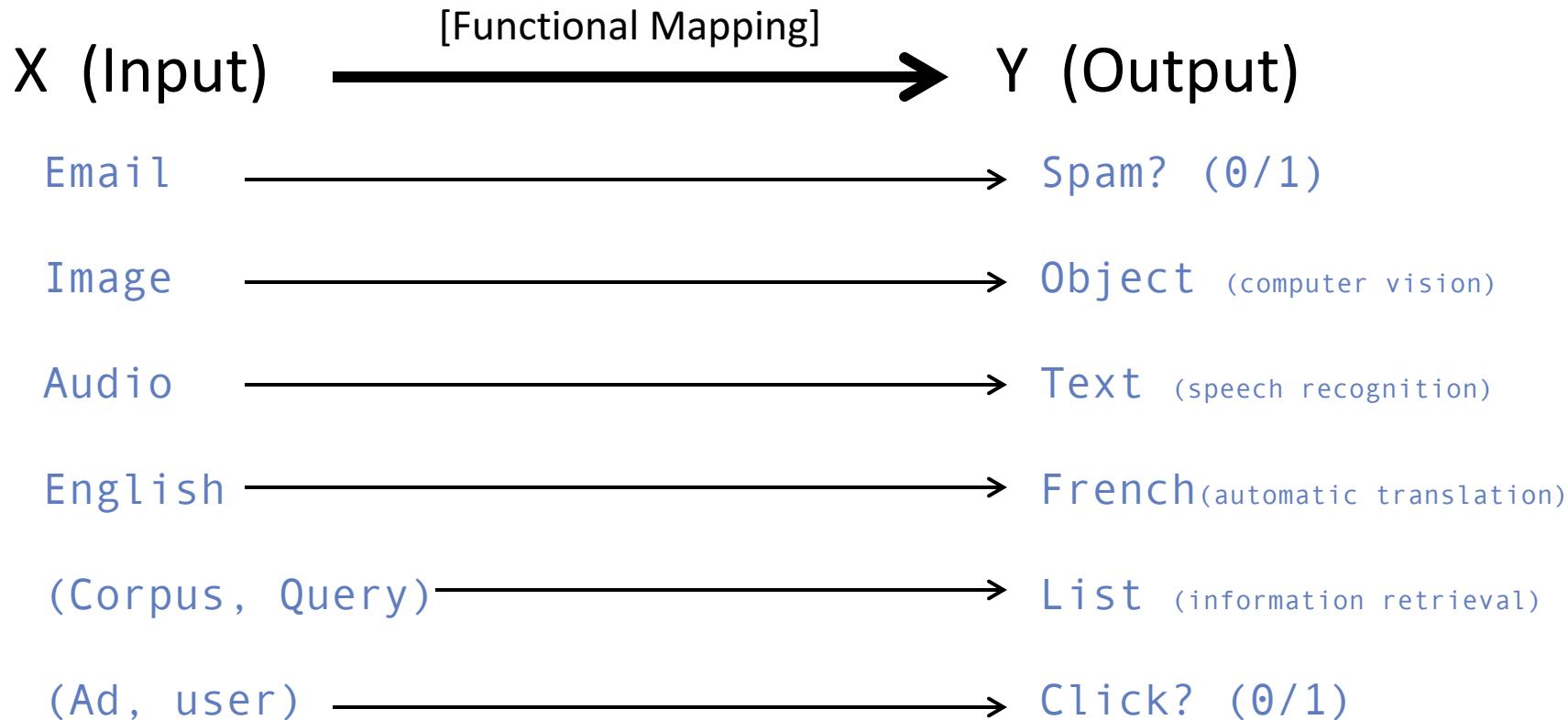
$$\mathcal{D} = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

$$\ell((x, y); \theta) = \begin{cases} 0 & f(x) = y \\ 1 & f(x) \neq y \end{cases}$$

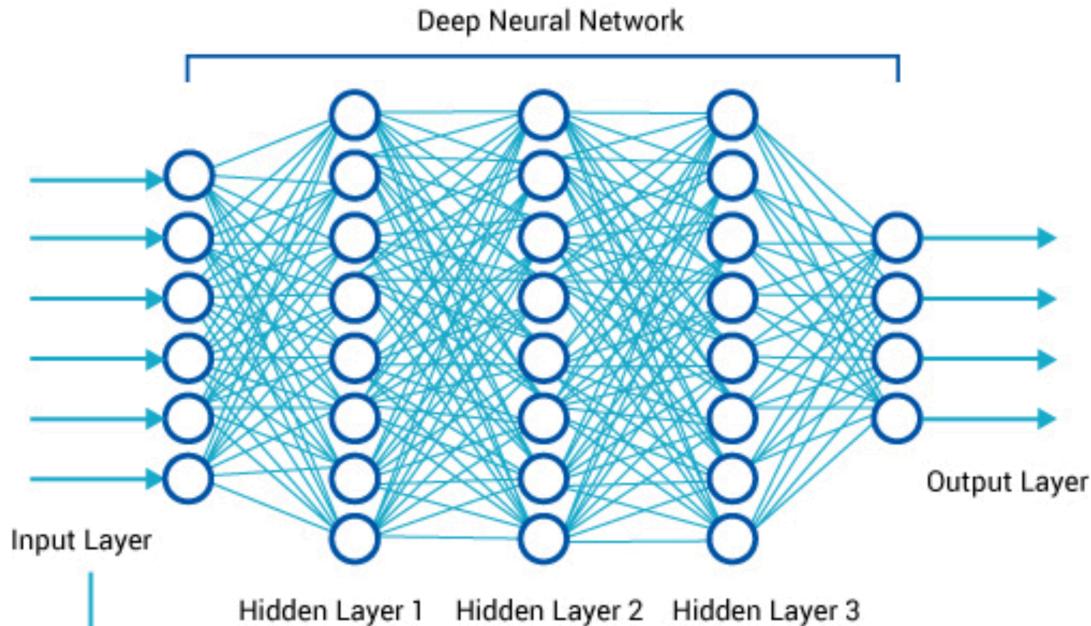
$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n \ell((x_i, y_i); \theta)$$

Machine learning involves solving continuous optimization problems

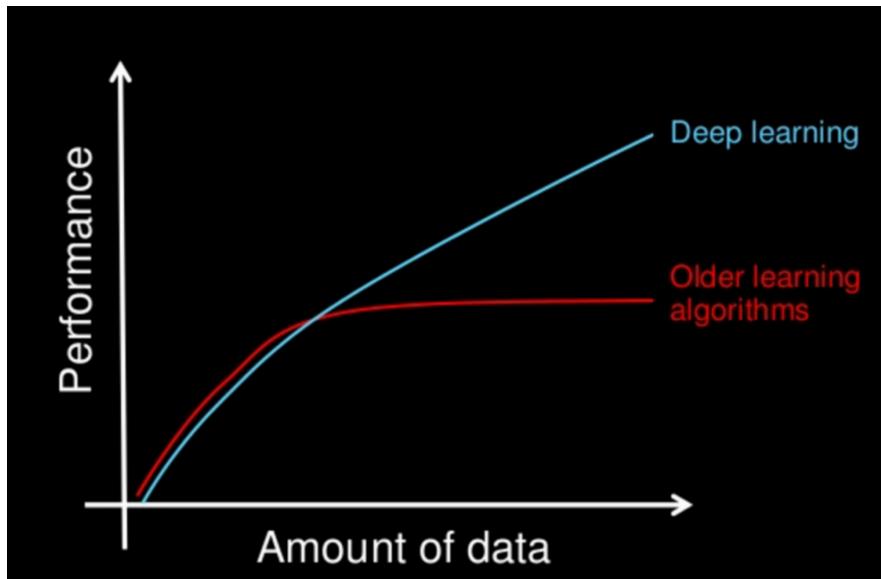
- Supervised Learning:



Andrew Ng: Artificial Intelligence is the New Electricity.
<https://www.youtube.com/watch?v=21EiKfQYZXc&t=1206s>



DNN are highly non-linear mappings



$$f : X \rightarrow Y$$

$$\mathcal{D} = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n \ell((x_i, y_i); \theta)$$

Andrew Ng: Artificial Intelligence is the New Electricity.
<https://www.youtube.com/watch?v=21EiKfQYZXc&t=1206s>

- Beyond supervised classification

- **K-means clustering's loss function:**

$$\sum_{i=1}^n \sum_{k=1}^K z_{ik} \|x_i - \mu_k\|^2$$

- **Dimensionality Reduction's loss function:**

$$\sum_{k=1}^n \|(\boldsymbol{\mu} + a_k \mathbf{e}) - \mathbf{x}_k\|^2$$

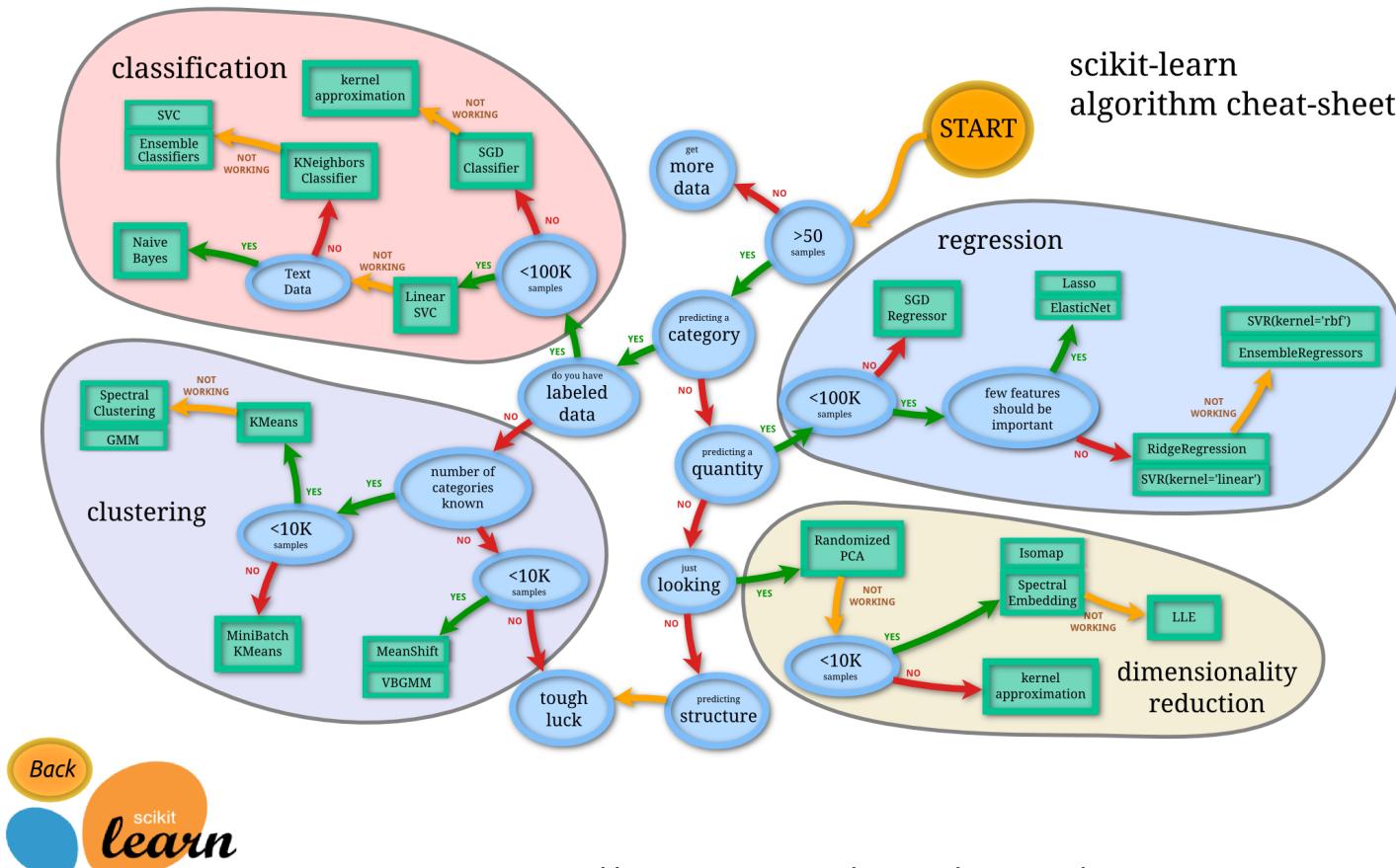
- **Collaborative Filtering's loss function:**

$$\sum_{(i,j):r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2$$

Andrew Ng. Coursera. Machine Learning.
<https://en.coursera.org/learn/machine-learning>



Function-approximation Machine Learning



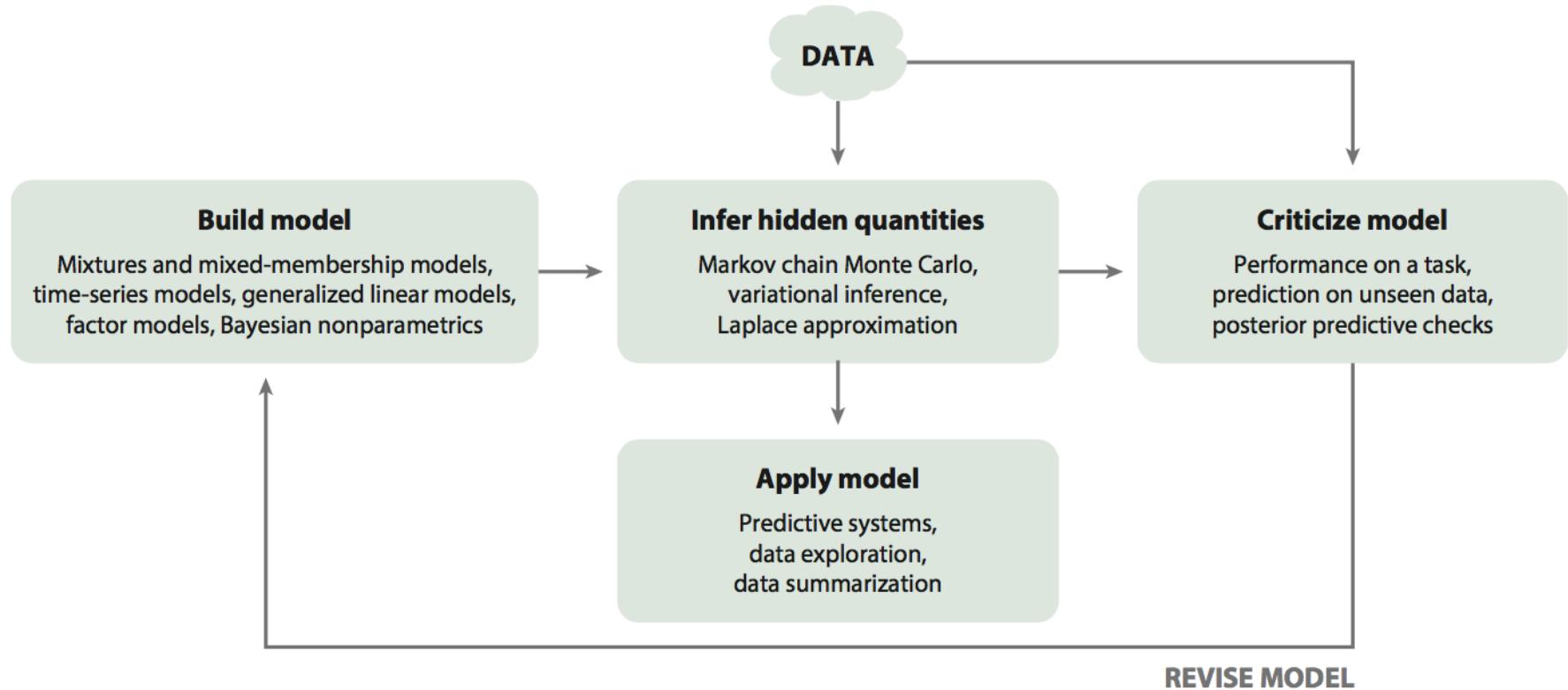
http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

- Black Box Approaches
 - No Model Interpretability
 - No Knowledge Extraction
 - No understanding in how decisions are made
- Uncertainty Quantification
 - No Model Uncertainty
 - No Predictions Uncertainty

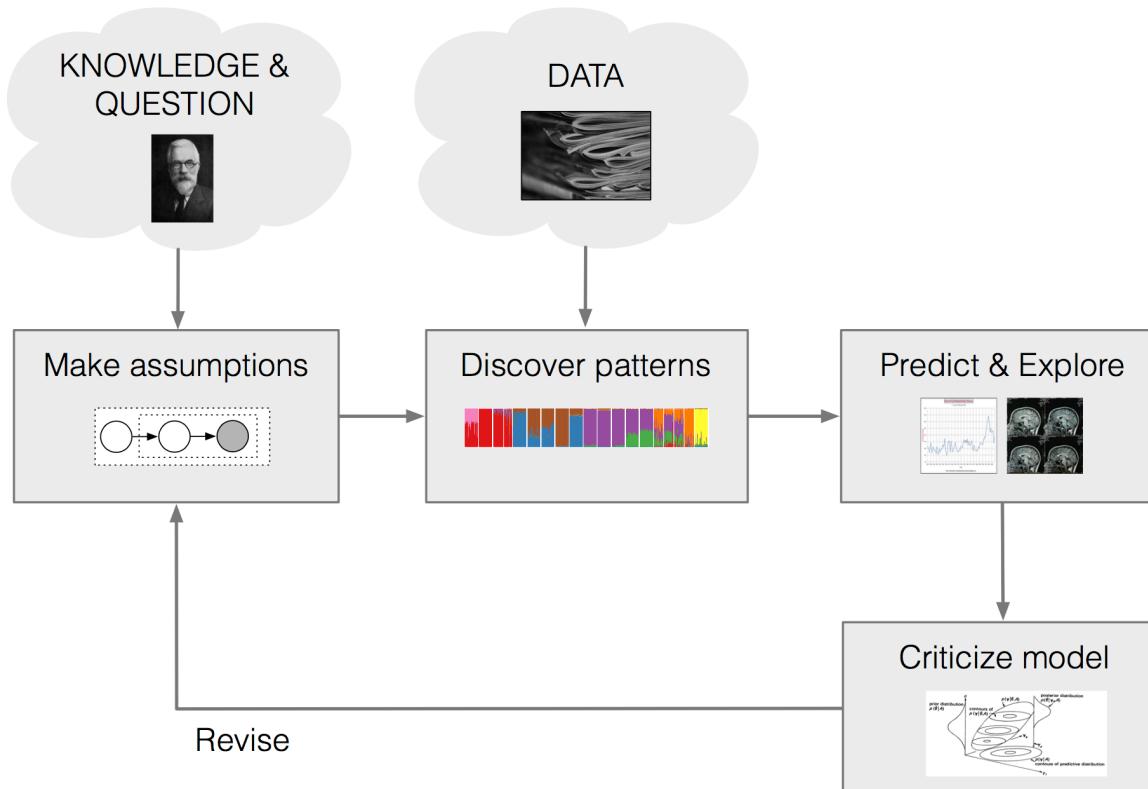
- High Cognitive Burden
 - Daunting number of algorithms and models.
 - Hard to master most of them.
- Algorithms can not be easily customized.
 - Real A.I. apps require ad-hoc adaptations.
 - Even Harder to adapt/modify existing algorithms.

Probabilistic (Bayesian) Machine Learning

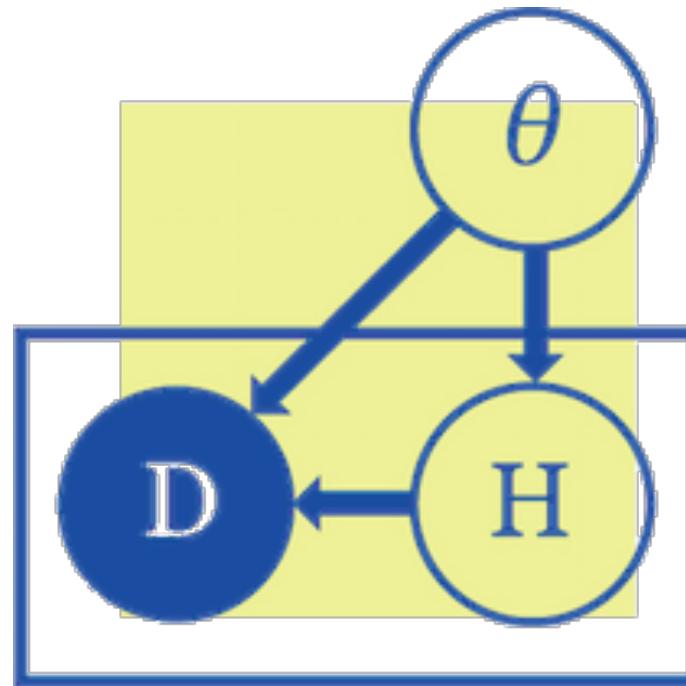
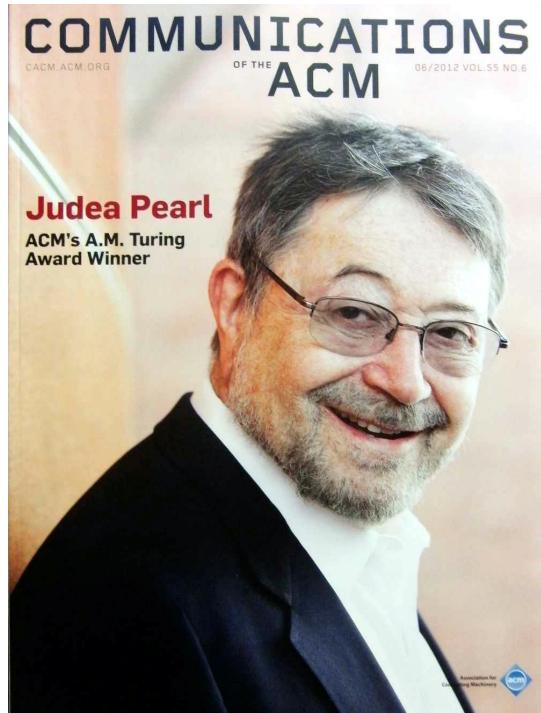




Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.



Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.



Probabilistic Graphical Models



Blackbox Models

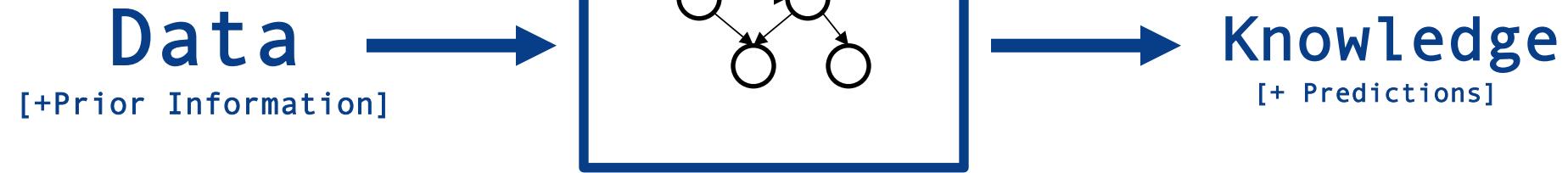
(kernel methods, deep learning, ensembles...)



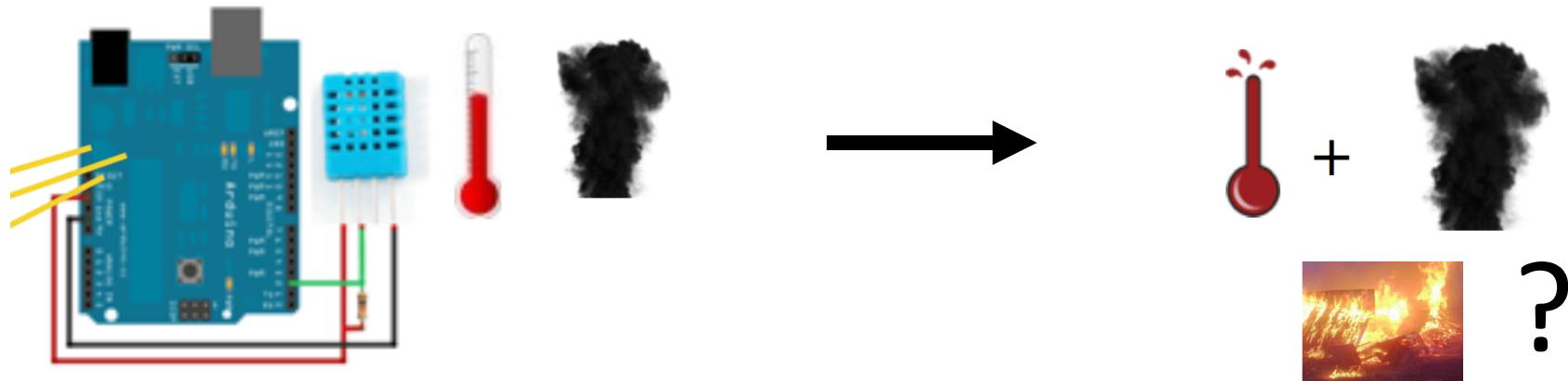
Loss Minimization
(Stochastic Gradient Descent)

Openbox Models

[Probabilistic Graphical Models]

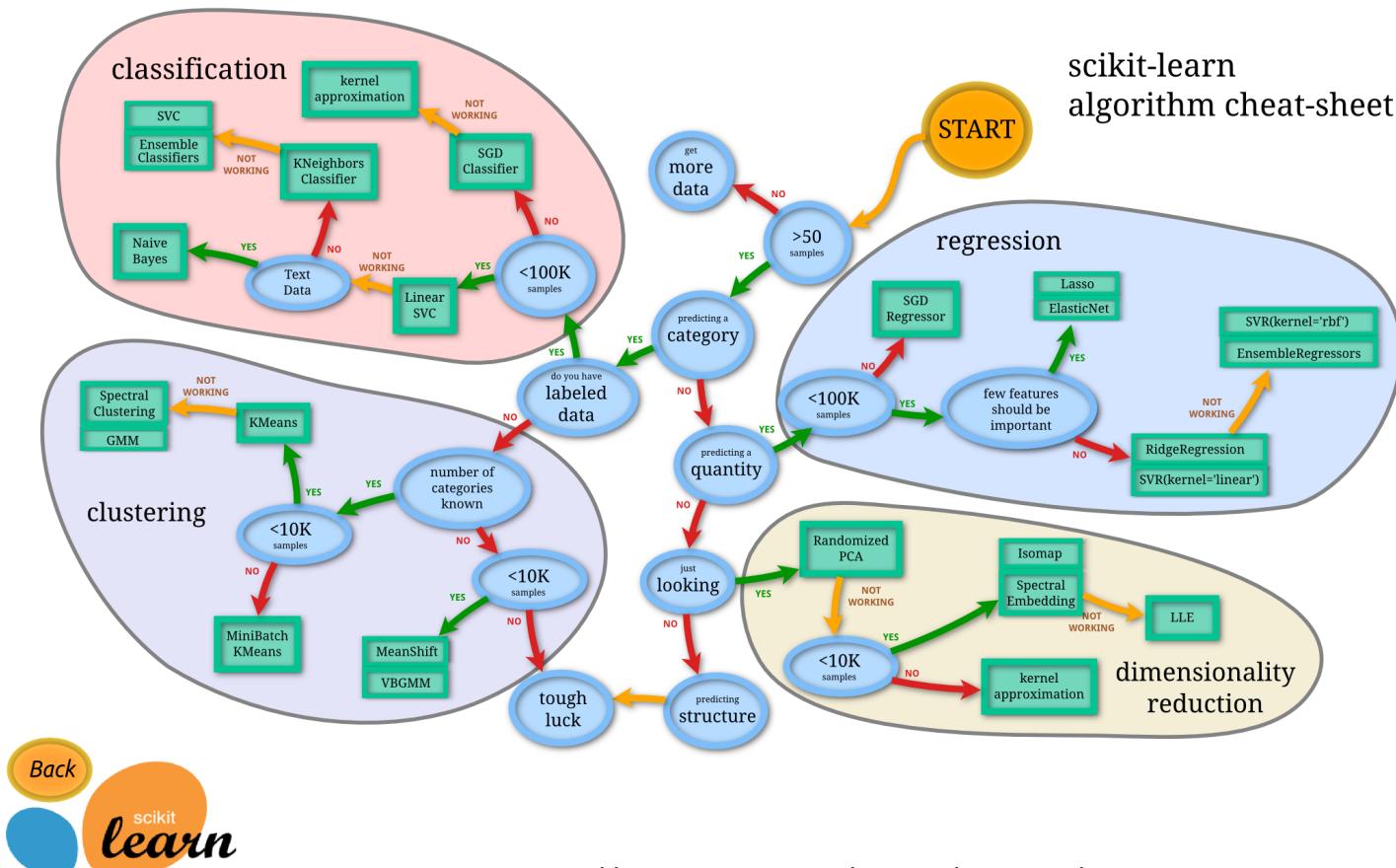


Fire Detection from smoke and temperature sensors

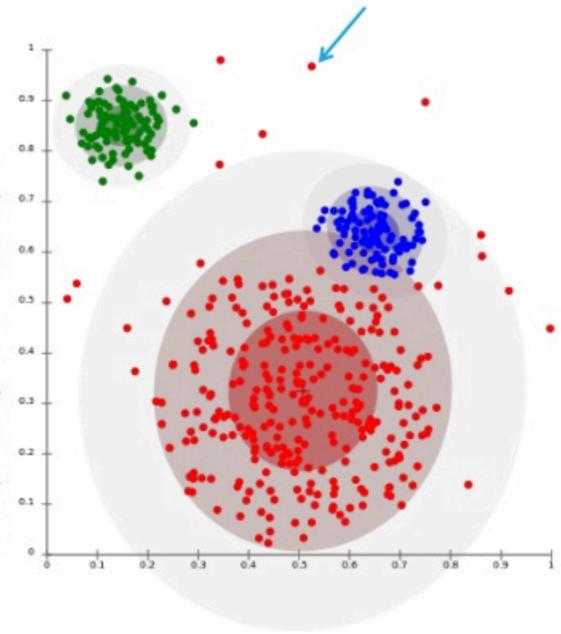


- Data Collected
 - Tons of observations in normal settings (no fire).
 - No observations in the presence of fire.

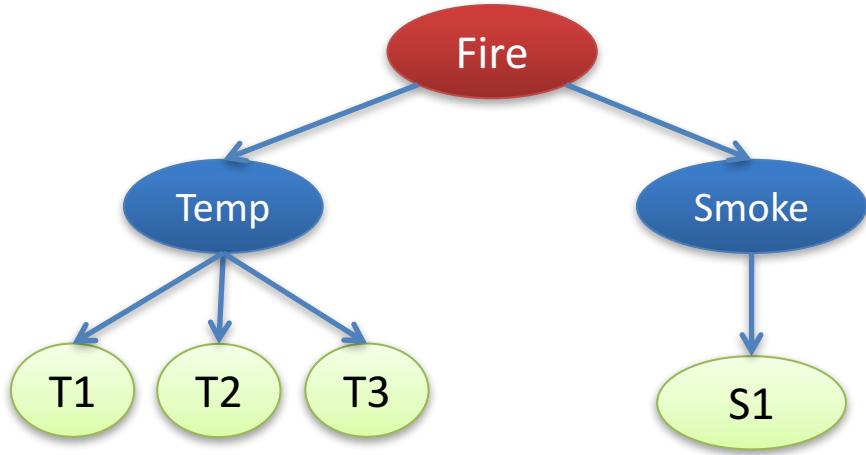
Function-approximation Machine Learning



http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html



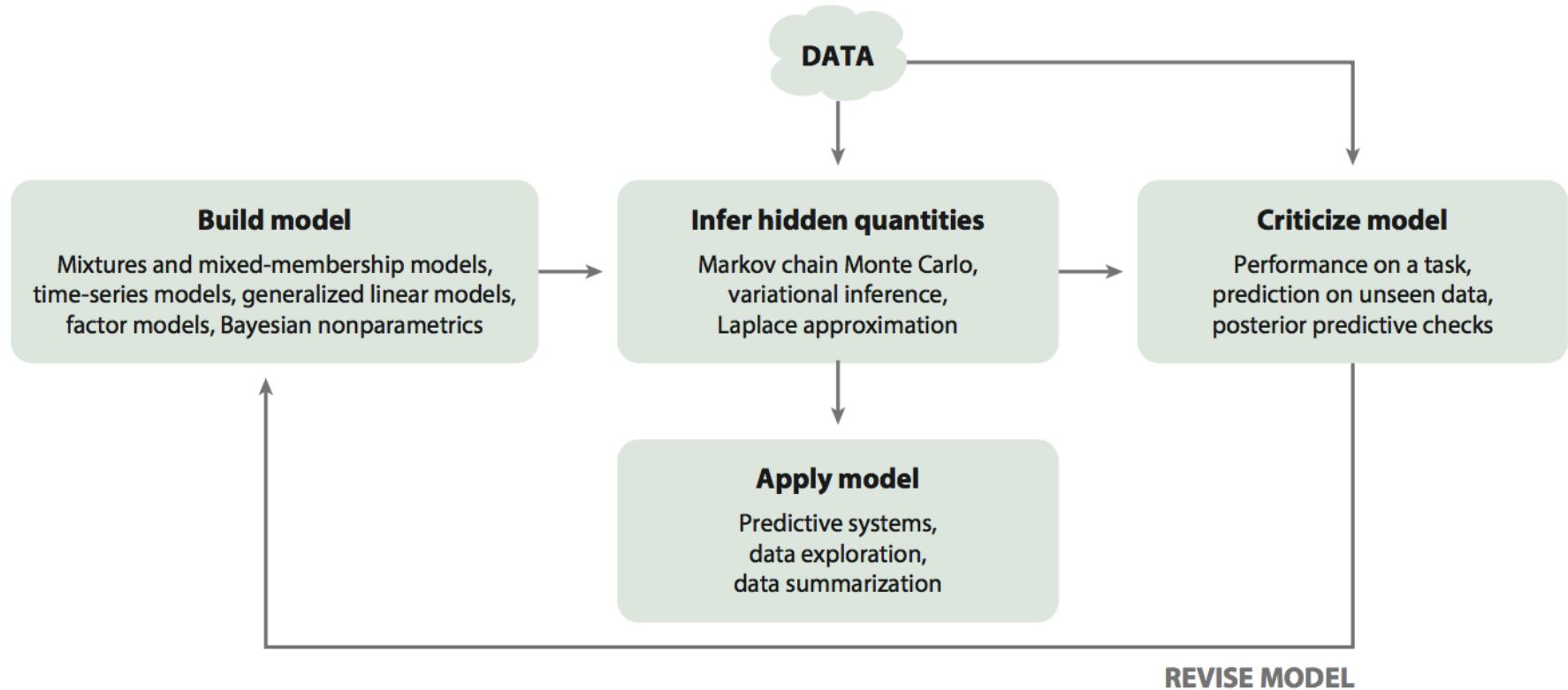
Black Box Approach:
Anomaly Detection with (streaming) K-means



$$p(Fire = True | t_1, t_2, t_3, s_1)$$

Causal Modeling

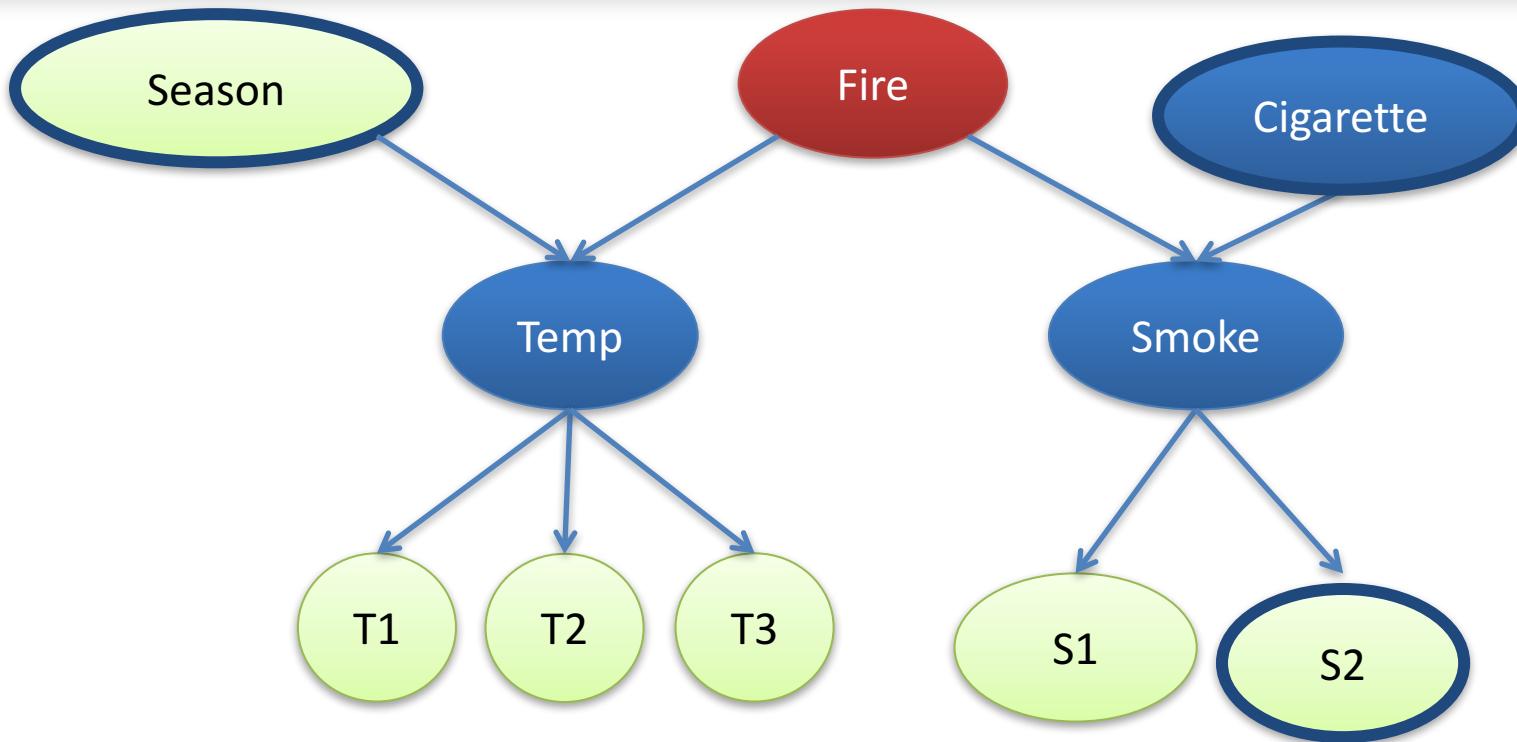
Fire, Temp and Smoke are latent variables.



Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.

EXAMPLE OF PGMS

AMIDST
TOOLBOX



$$p(Fire = True | t_1, t_2, t_3, s_1, s_2, season)$$

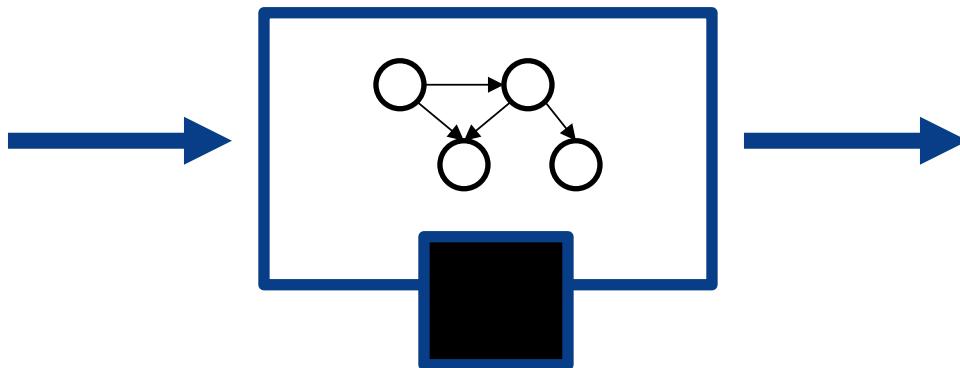
Easy to Modify



Openbox Models

[Probabilistic Graphical Models]

[Big] Data
[+Prior Information]



[Scalable] Bayesian Inference Engine
(Powered by Variational Methods)

Thanks for your attention

www

www.amidsttoolbox.com

@

contact@amidsttoolbox.com



[@AmidstToolbox](https://twitter.com/AmidstToolbox)

AMIDST
→ TOOLBOX