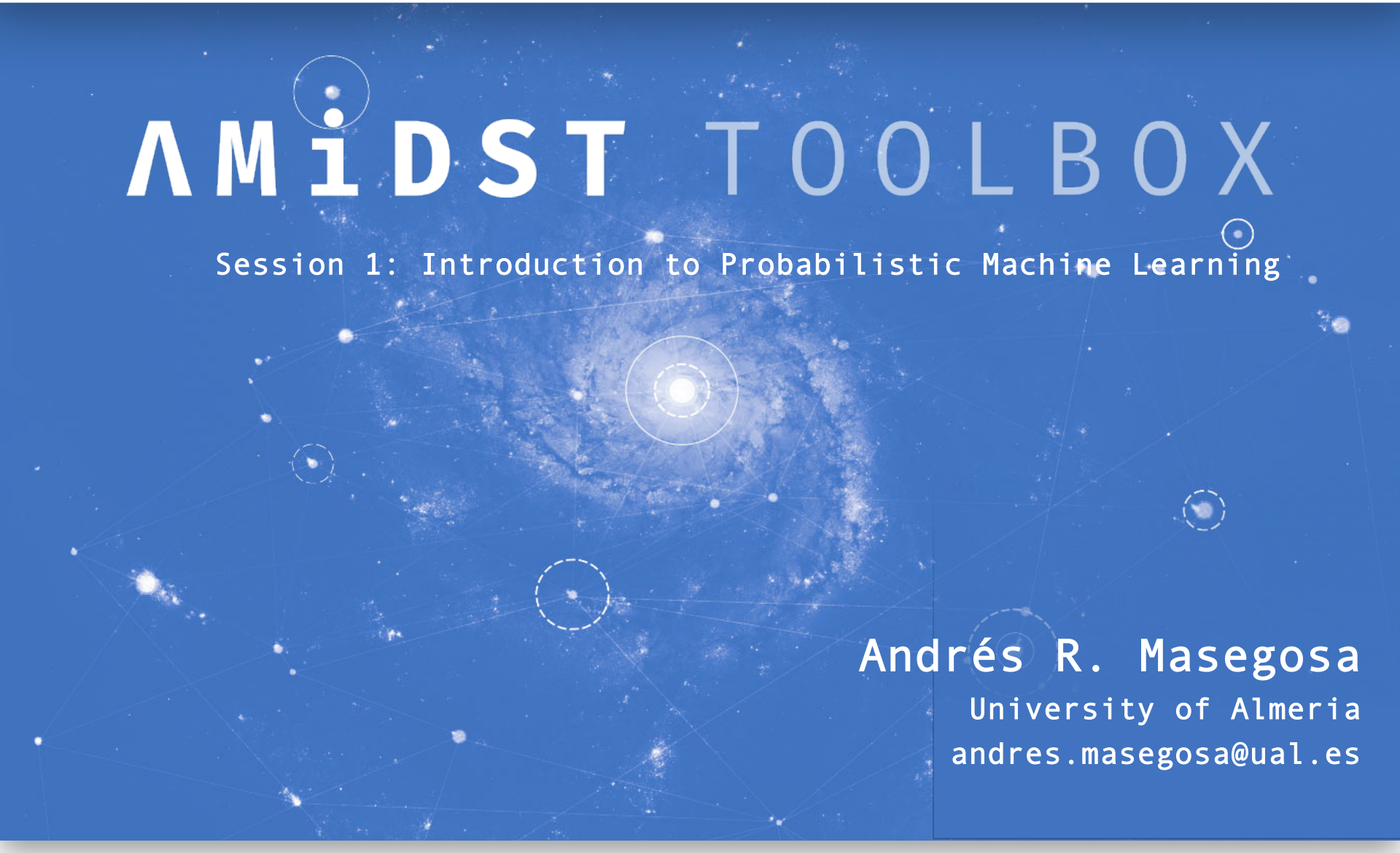


AMiDST TOOLBOX

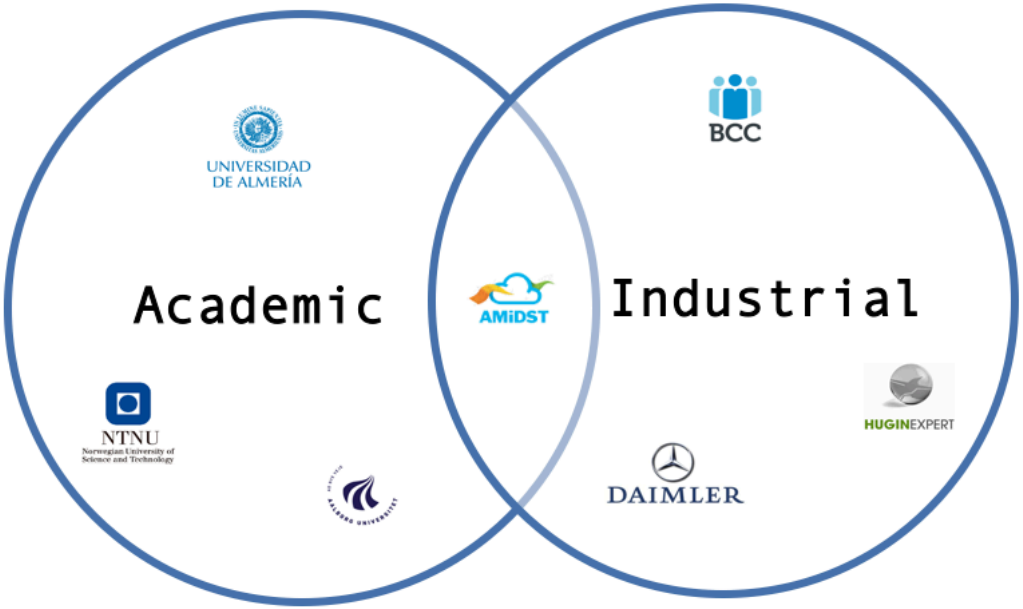
Session 1: Introduction to Probabilistic Machine Learning

Andrés R. Masegosa
University of Almeria
andres.masegosa@ual.es

January, 2018 Geilo (Norway)

THE AMIDST CONSORTIUM

AMiDST TOOLBOX



SEVENTH FRAMEWORK PROGRAMME











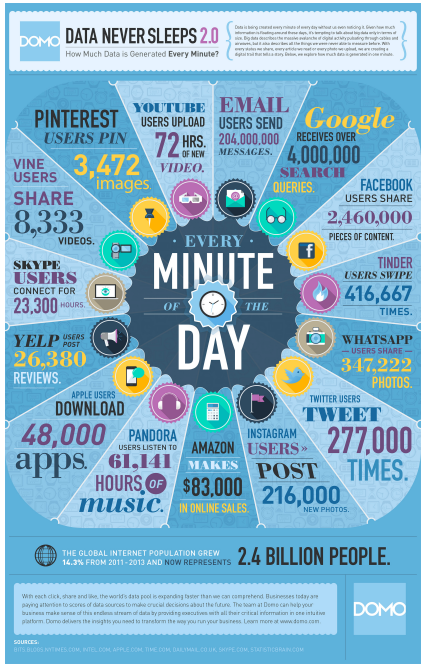



Let it burn!
It's a shameful memory

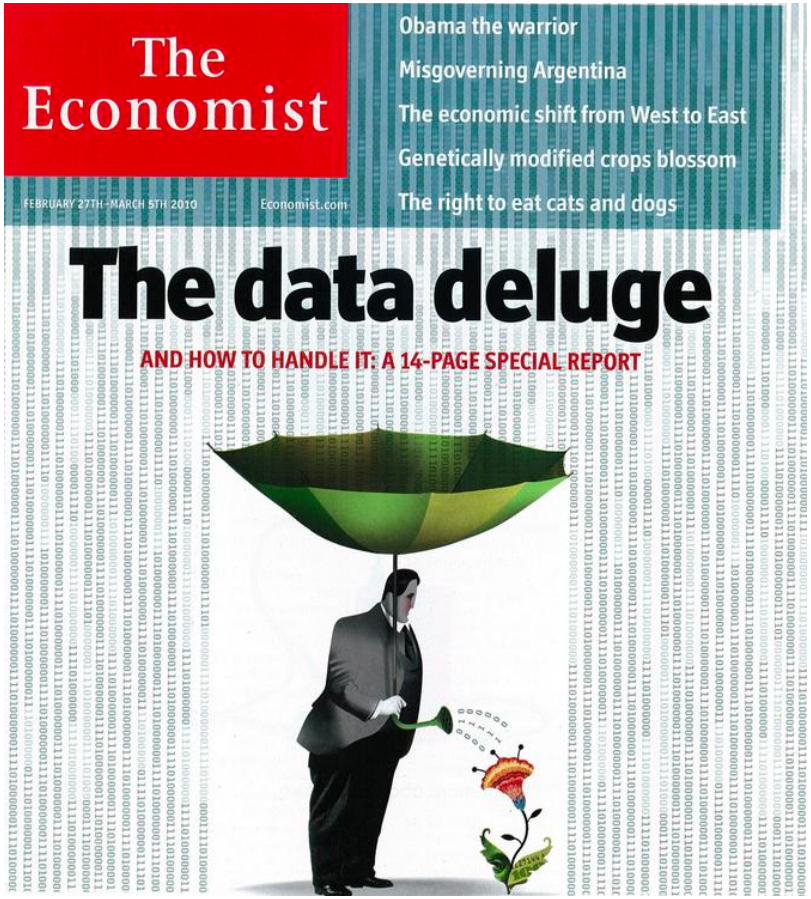
3rd century BC

50 thousands papyrus = 20 thousands books

21st century DC

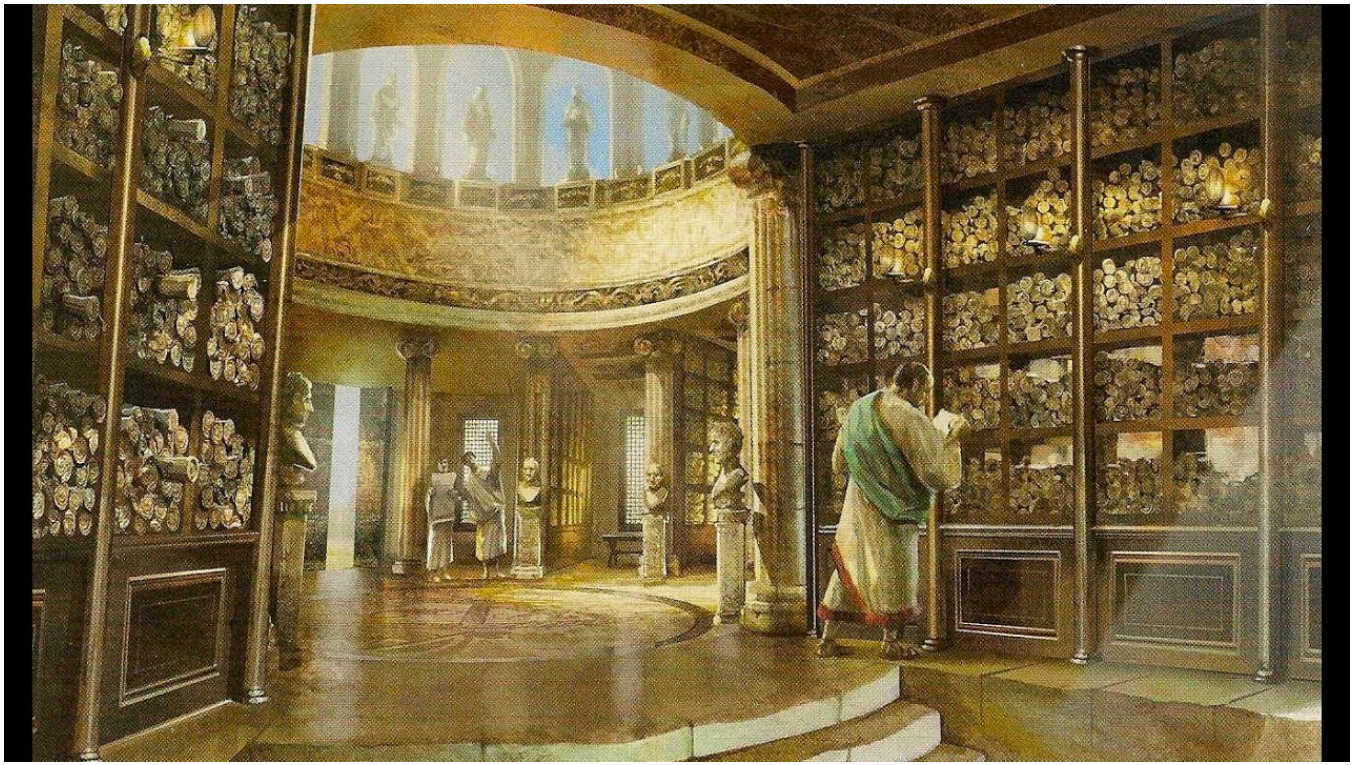


6 thousand books, 2 millions posts and news daily

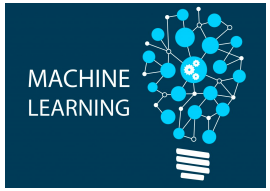

The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
FEBRUARY 27TH–MARCH 5TH 2010
Economist.com
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

Machine Learning

Acceso al conocimiento (siglo III a.c.)



Acceso al conocimiento (siglo XXI d.c.)



TensorFlow

NumPy

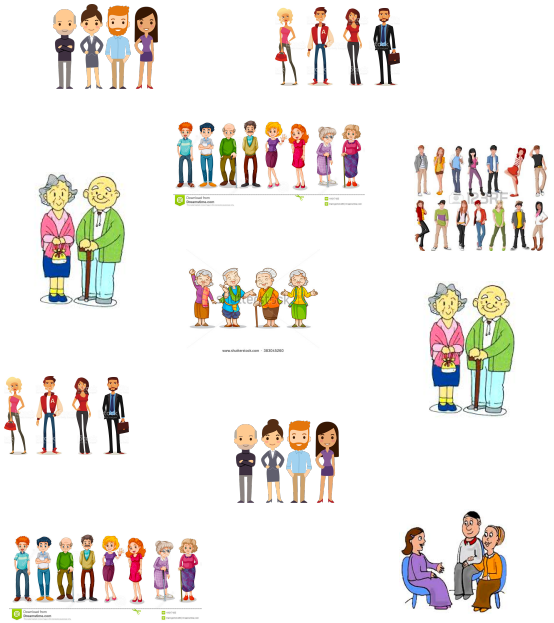
AMiDST TOOLBOX

Manual Computer Programming

vs

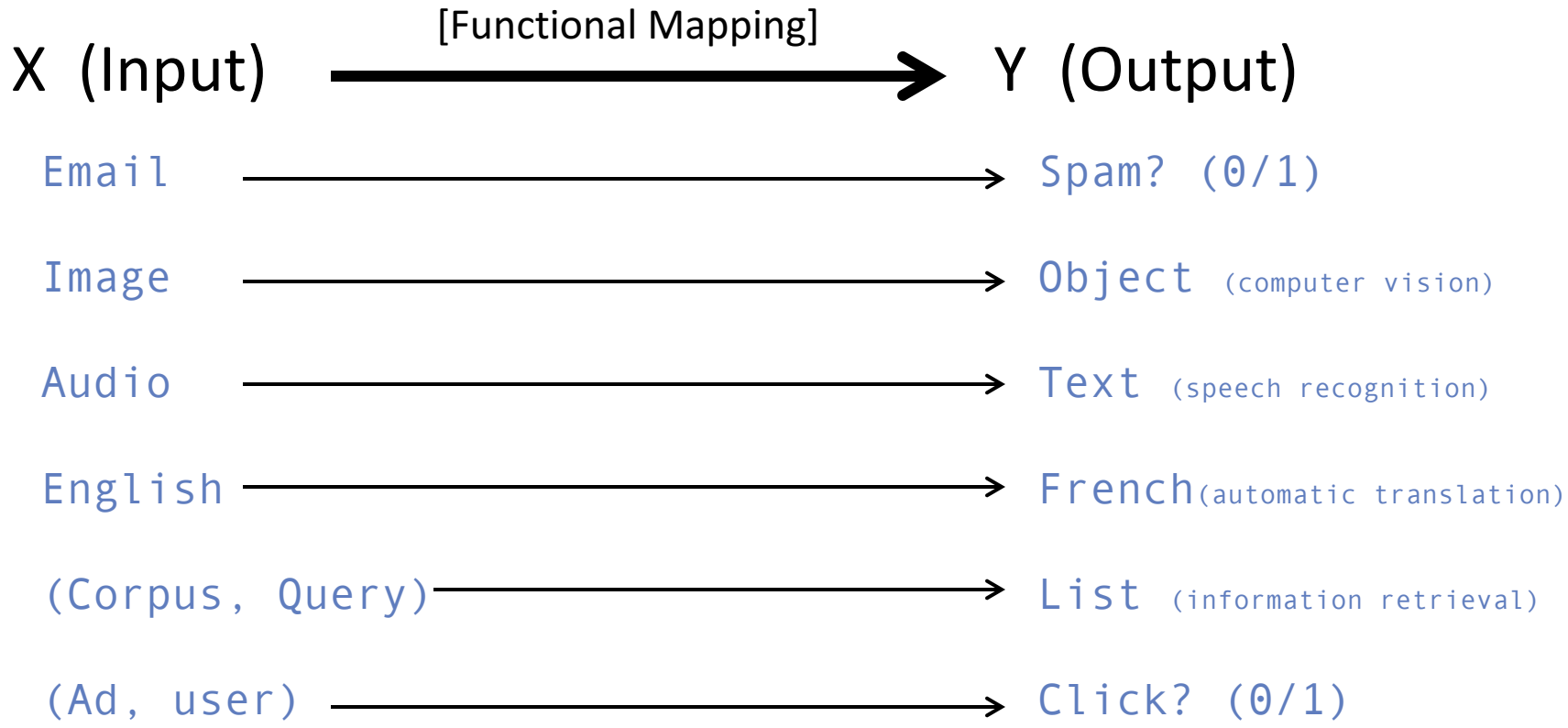
Automatic Computer Programming

vs



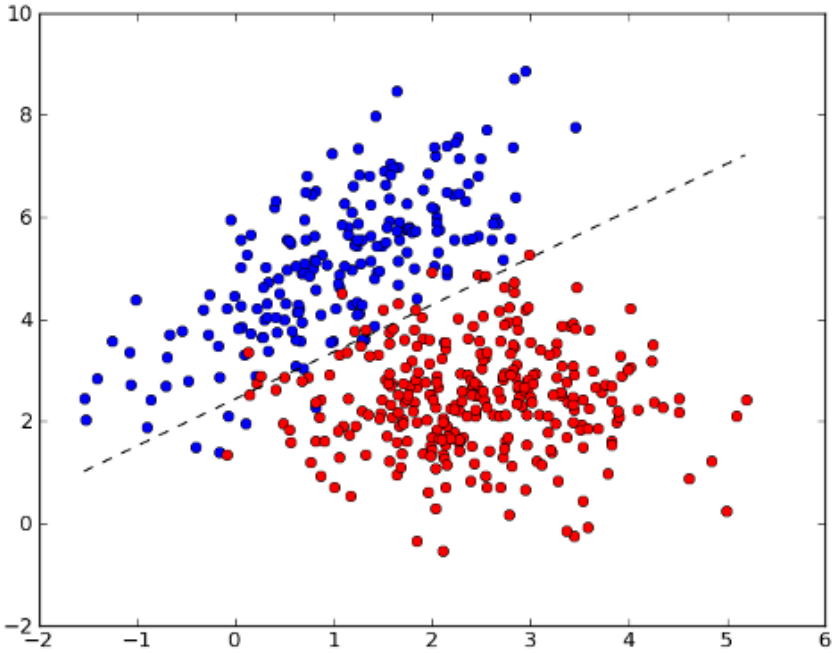
Machine Learning

- 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:



$$f : X \rightarrow Y$$

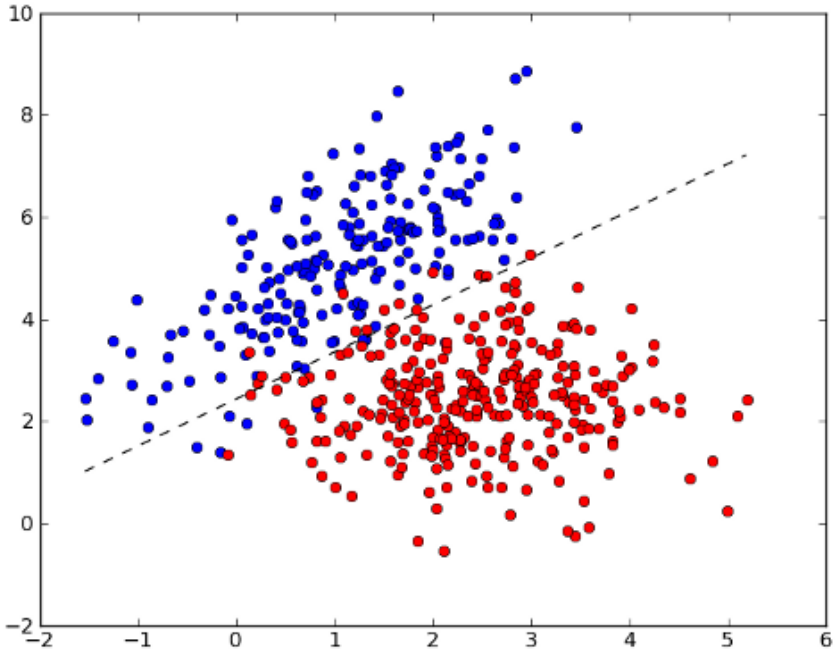
$$x \in \Re^2 \qquad y \in \{Red, Blue\}$$

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

$$\theta \in \Re^2$$

The mapping problem reduces to find the suitable $\theta^\star$.

- How do we find $\theta^\star$?
  - We learn it from data!



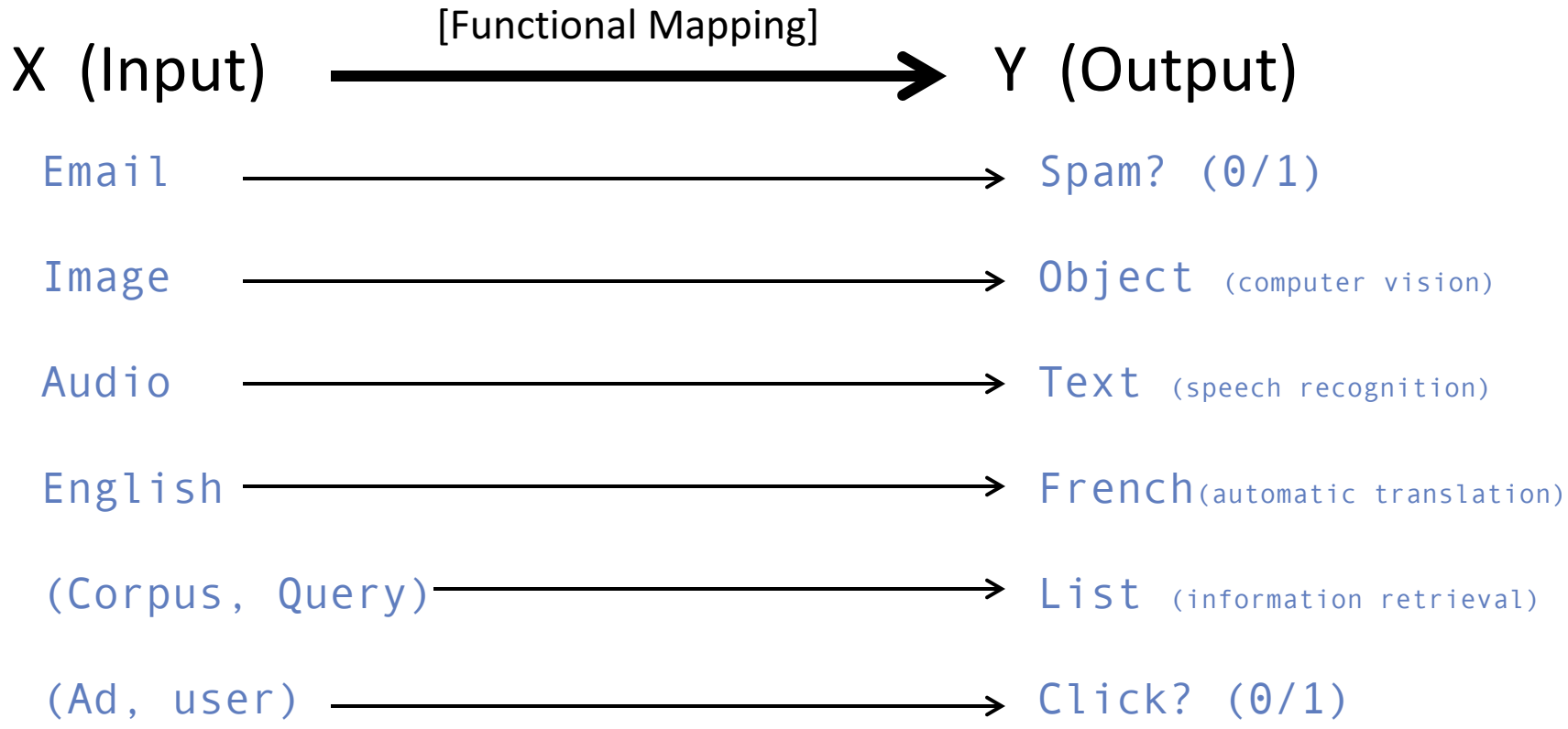
$$\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

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

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

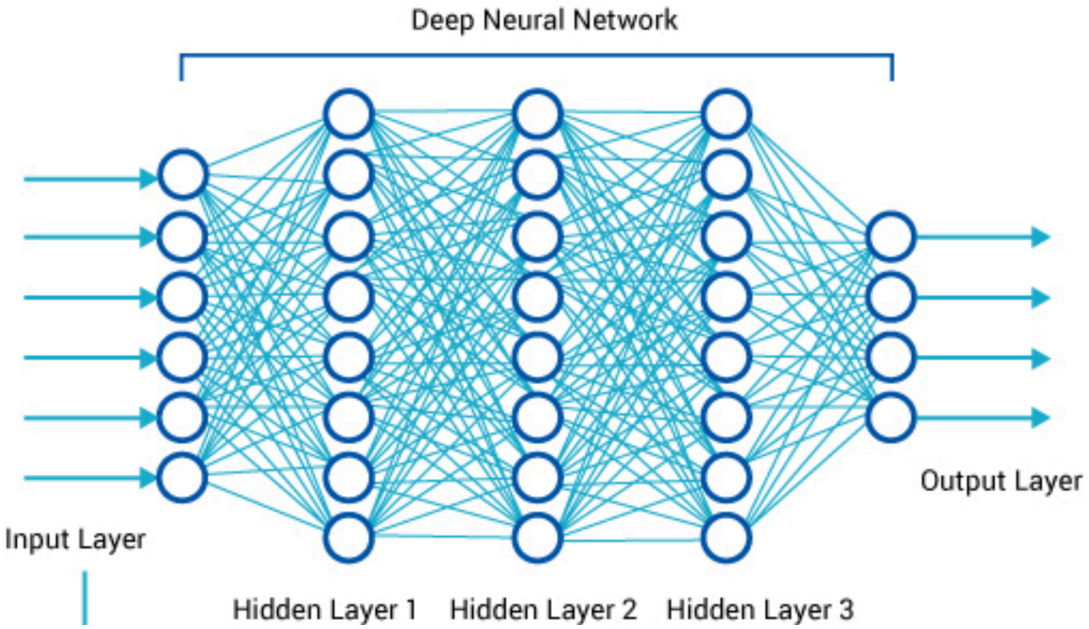
Machine learning involves solving continuous optimization problems

- Supervised Learning:



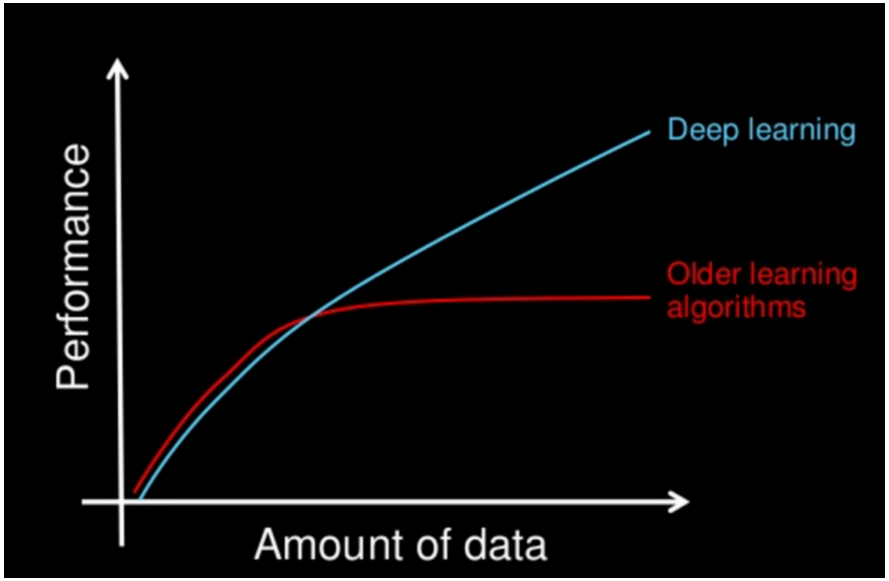
Andrew Ng: Artificial Intelligence is the New Electricity.
https://www.youtube.com/watch?v=21EiKfQYZXc&t=1206s

Deep Neural Networks



DNN are highly non-linear mappings

Deep Learning



$$f : X \to Y$$

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

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

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:

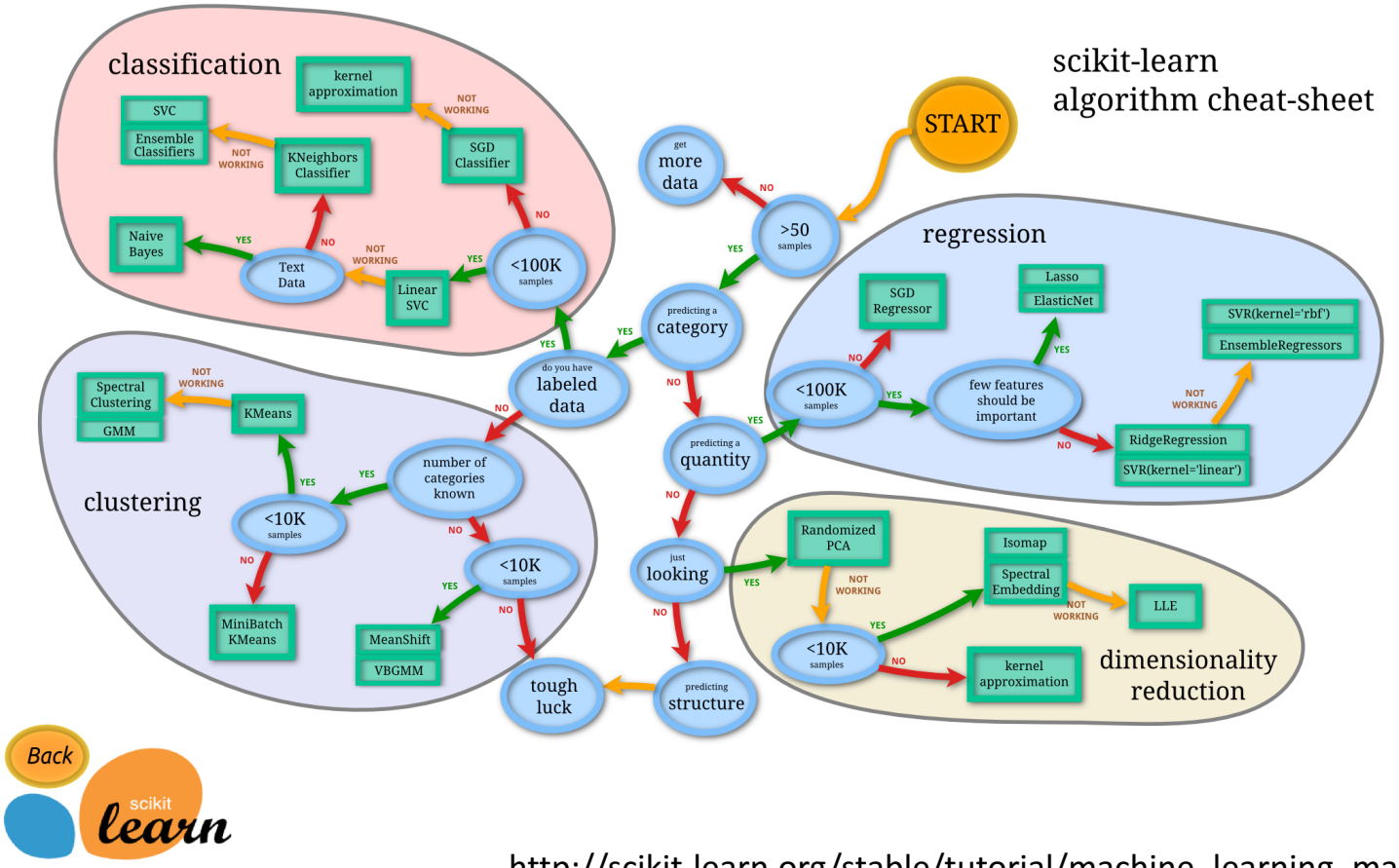
$$\sum_{k=1}^{n}\left\|\left(\mathbf{\mu} + a_k\mathbf{e}\right) - \mathbf{x}_k\right\|^2$$

  - **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

Theoretical Issues

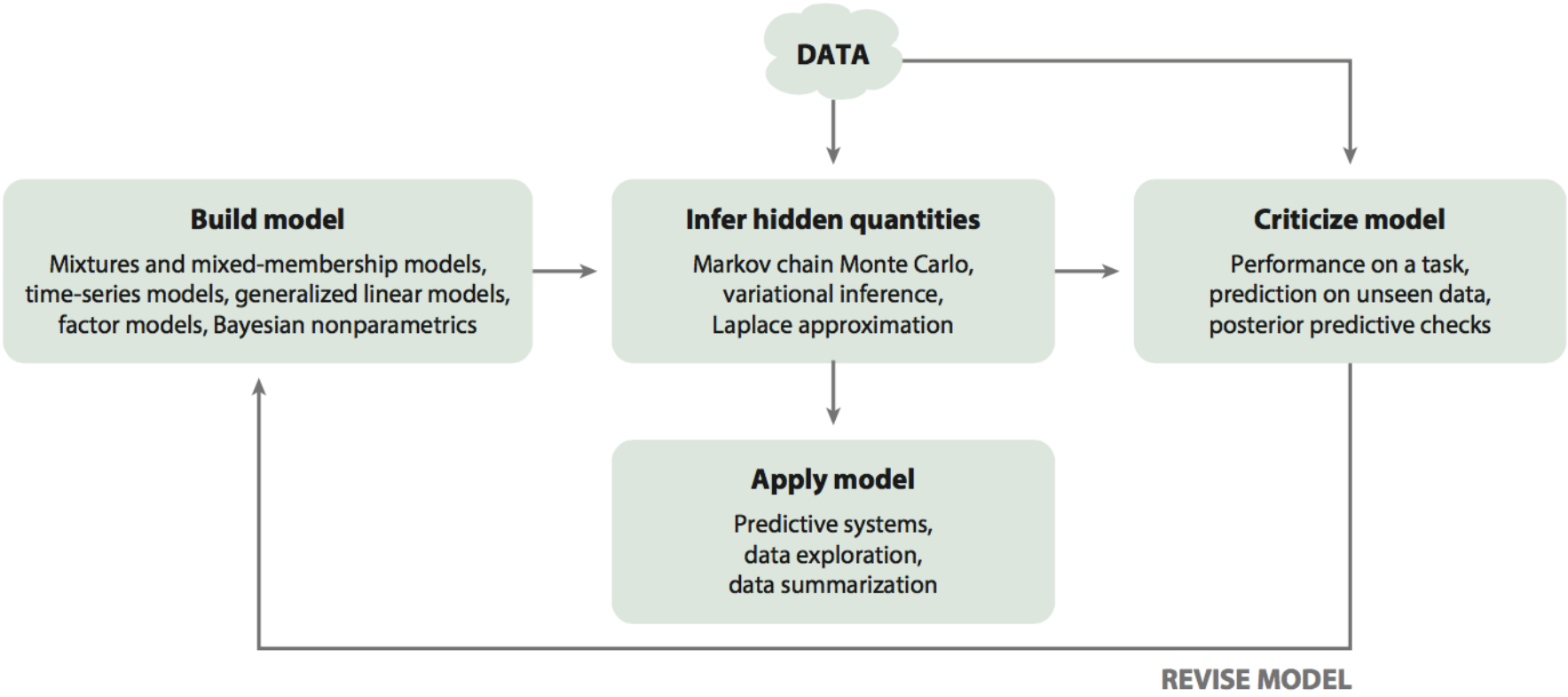
- Black Box Approaches
  - No Model Interpretability
  - No Knowledge Extraction
  - No understanding in how decisions are made
- Uncertainty Quantification
  - No Model Uncertainty
  - No Predictions Uncertainty

- 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.

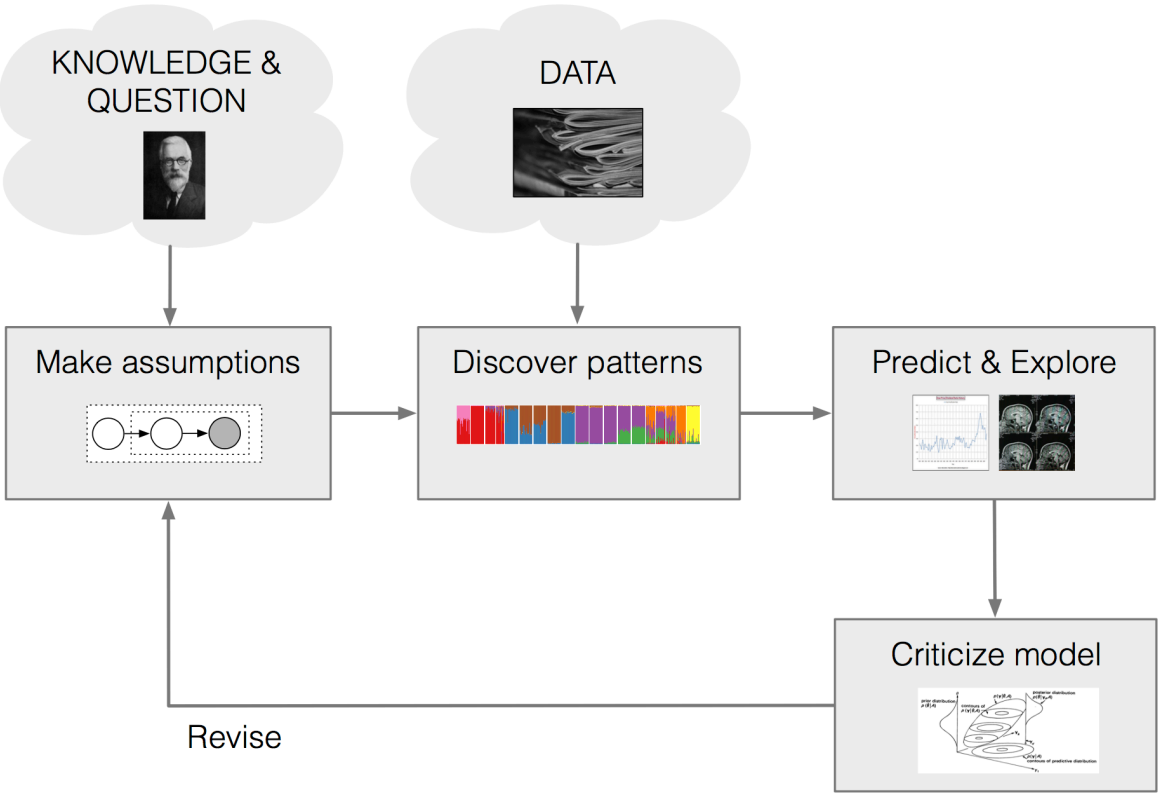
Probabilistic (Bayesian) Machine Learning

BOX's Loop

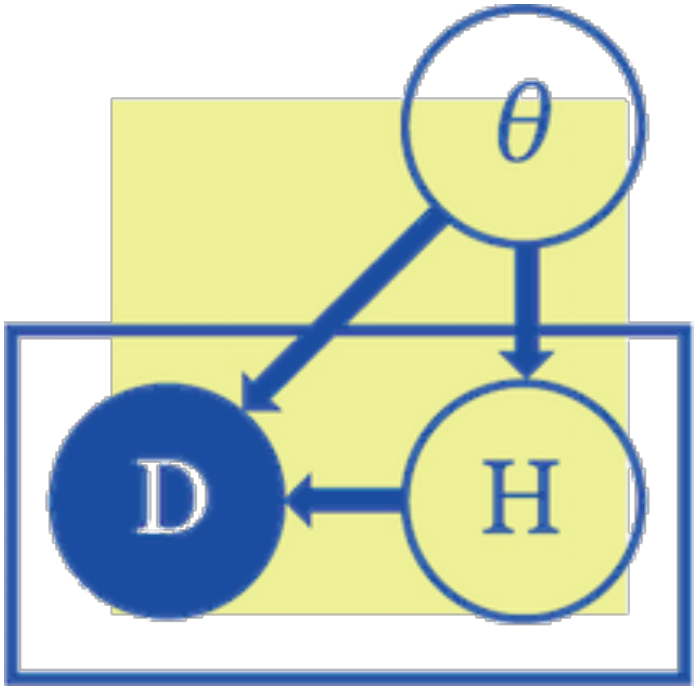


Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.

BOX's Loop



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

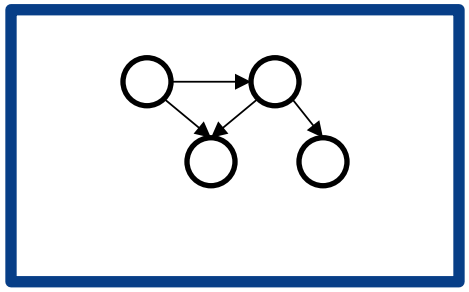
(Stochastic Gradient Descent)

Openbox Models

[Probabilistic Graphical Models]

Data

[+Prior Information]



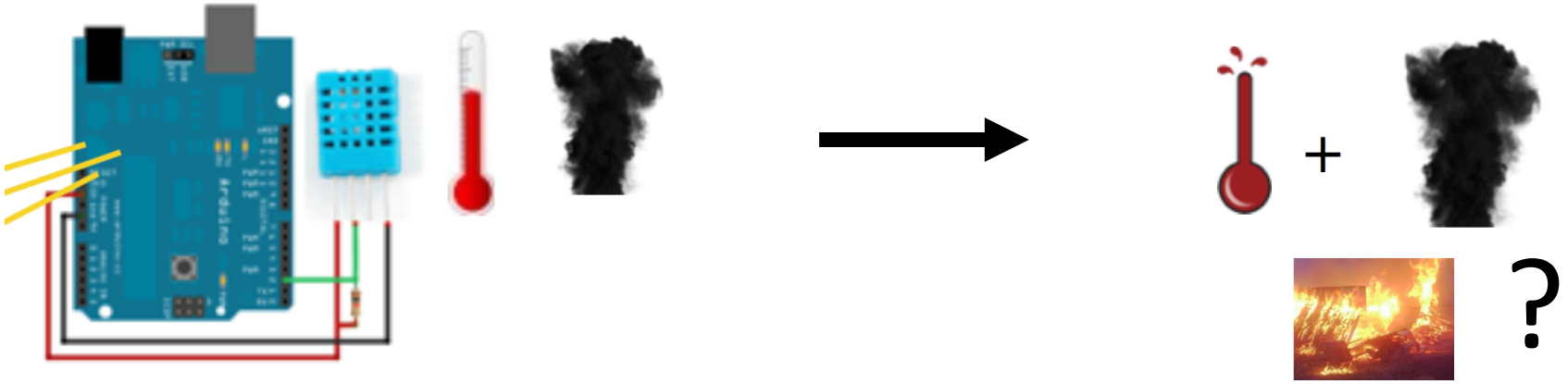
Knowledge

[+ Predictions]

Fire Detection

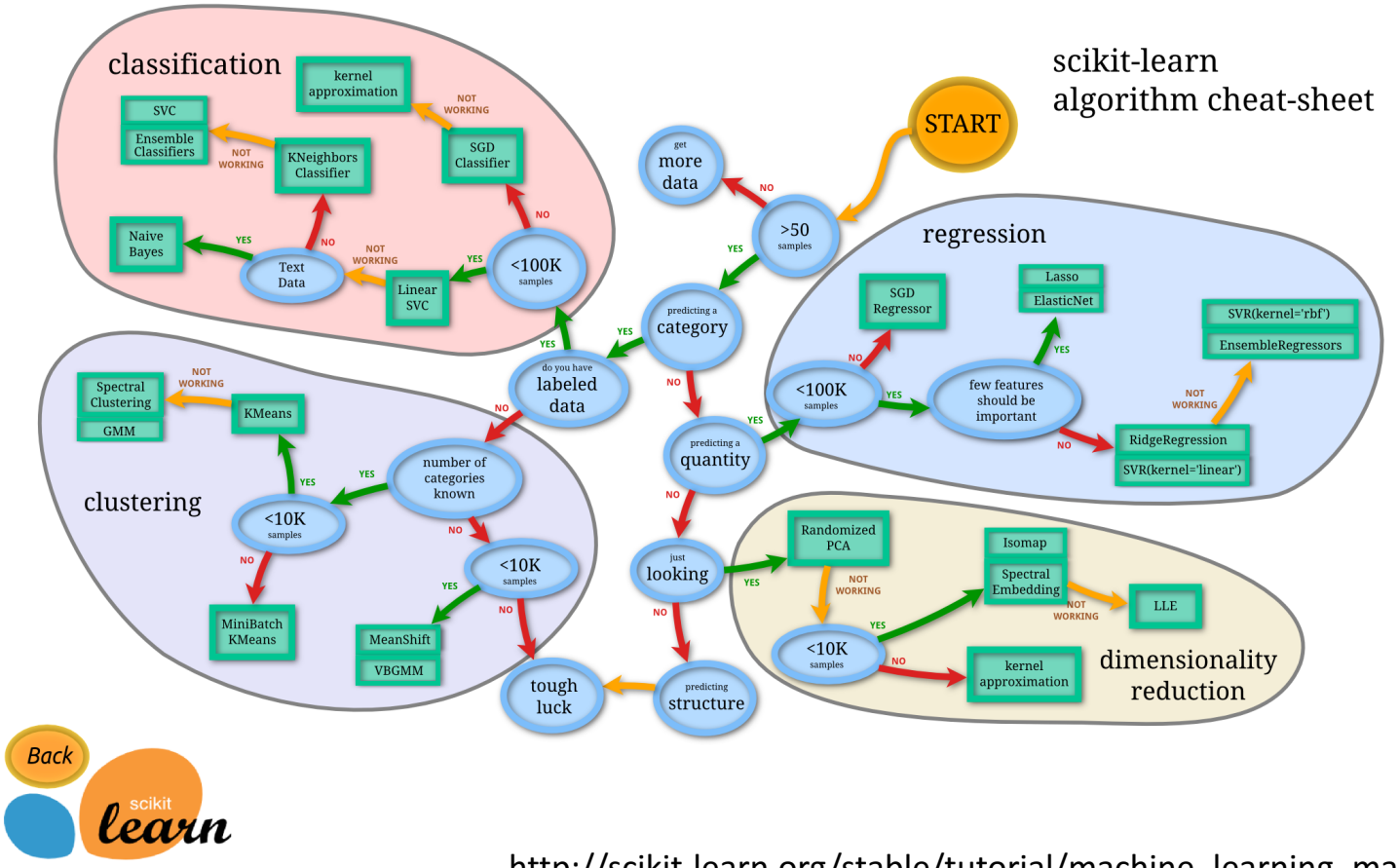
AMiDST TOOLBOX

Fire Detection from smoke and temperature sensors

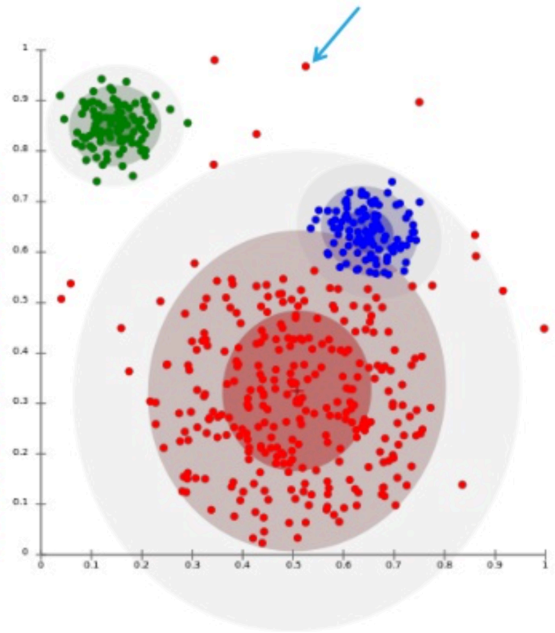


- Data Collected
  - Tons of observations in normal settings (no fire).
  - No observations in the presence of fire.

Function-approximation Machine Learning

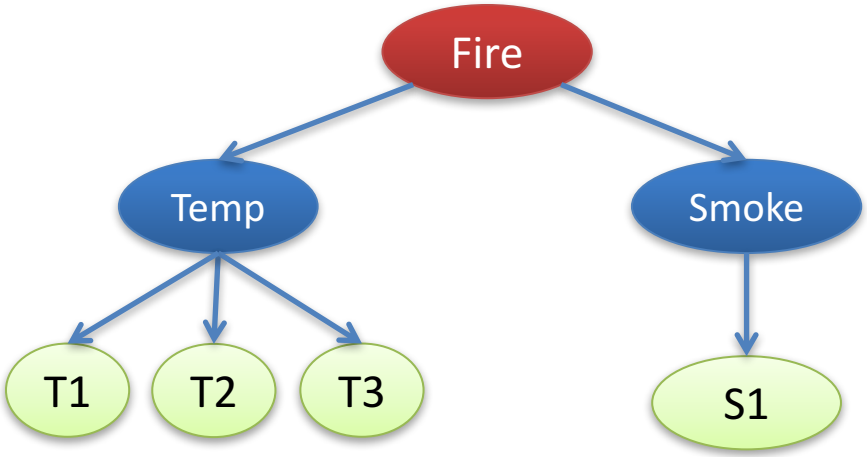


http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html



Black Box Approach:

Anomaly Detection with (streaming) K-means

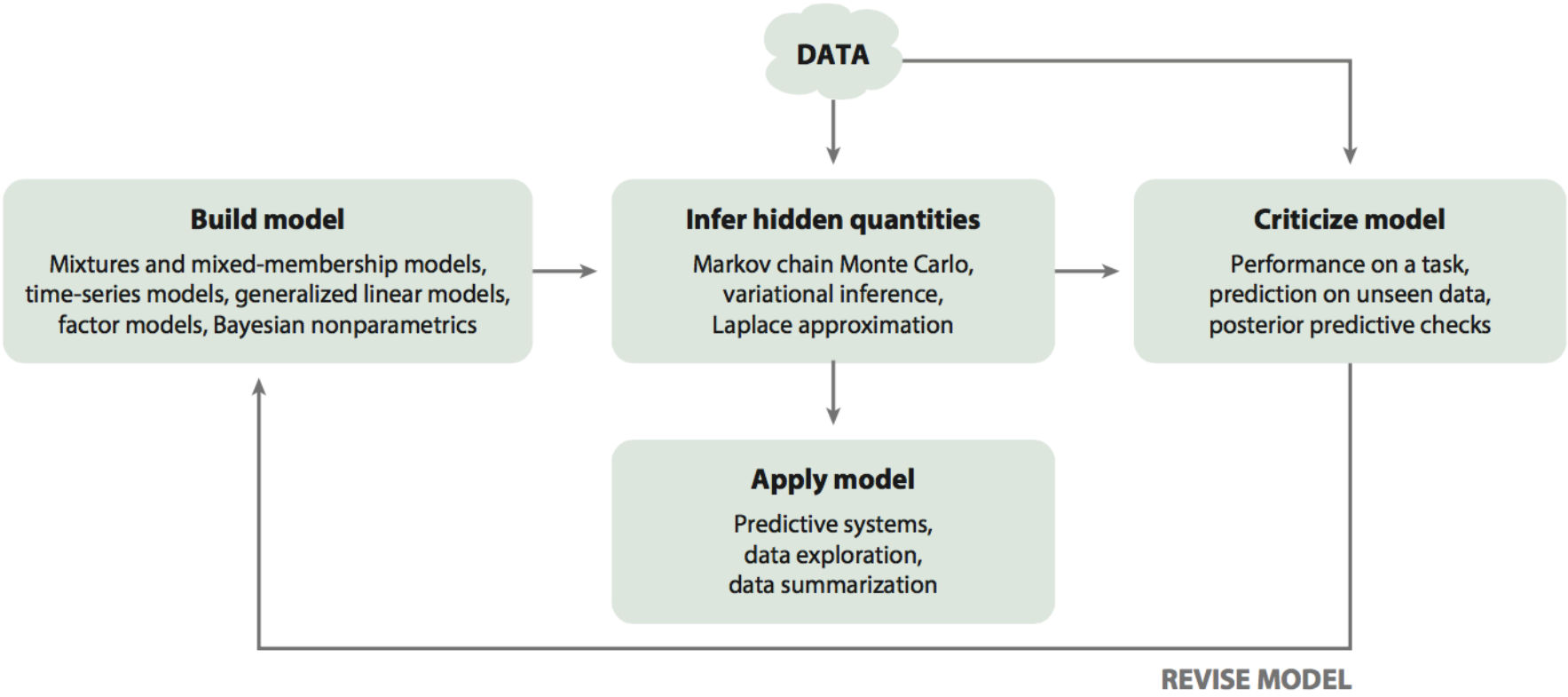


$$p(Fire = True|t_1, t_2, t_3, s_1)$$

Causal Modeling

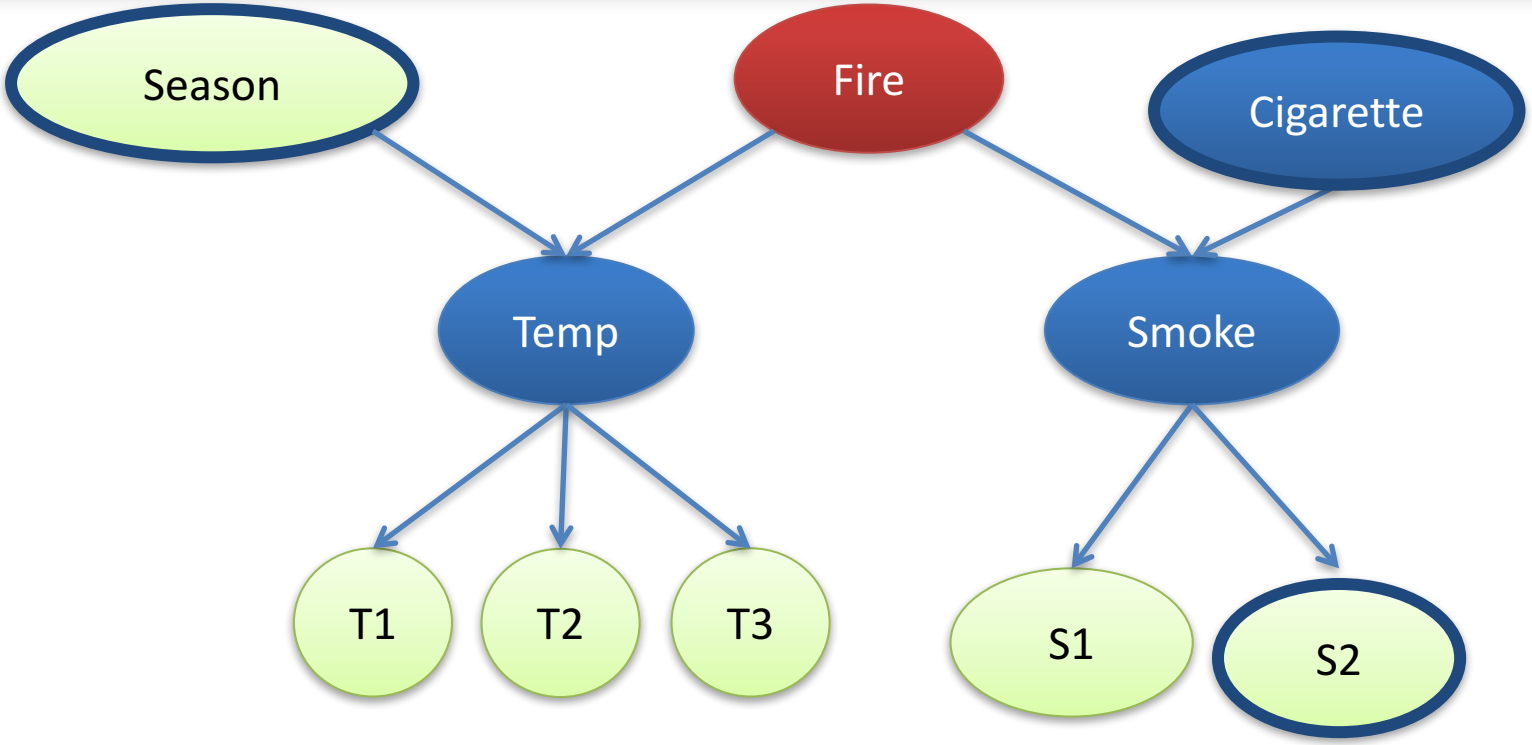
Fire, Temp and Smoke are latent variables.

BOX's Loop



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

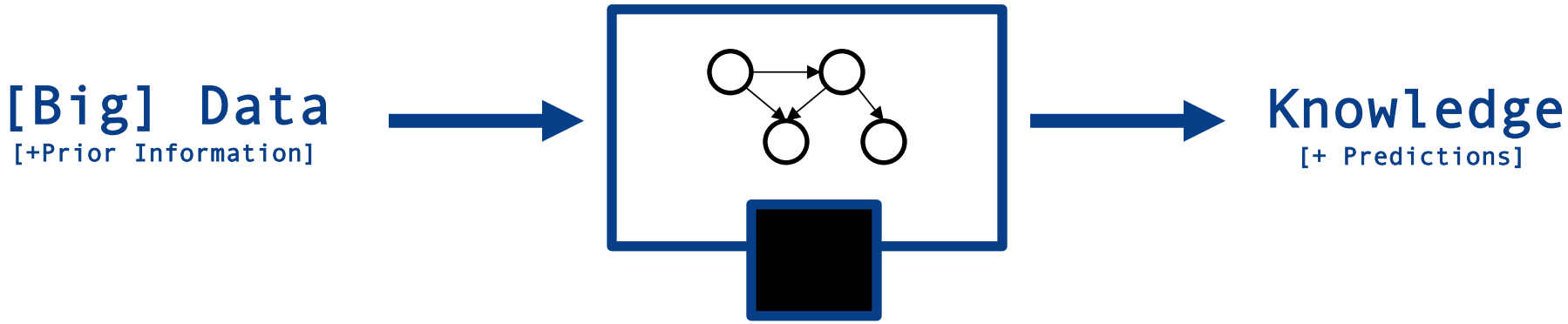


$$p(Fire = True|t_1, t_2, t_3, s_1, s_2, season)$$

Easy to Modify

Openbox Models

[Probabilistic Graphical Models]



[Scalable] Bayesian Inference Engine

(Powered by Variational Methods)

Thanks for your attention

www.amidsttoolbox.com

contact@amidsttoolbox.com

@AmidstToolbox