AMiDST TOOLBOX

AMiDST TOOLBOX

Probabilistic Machine Learning

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

January, 2018 Geilo (Norway)

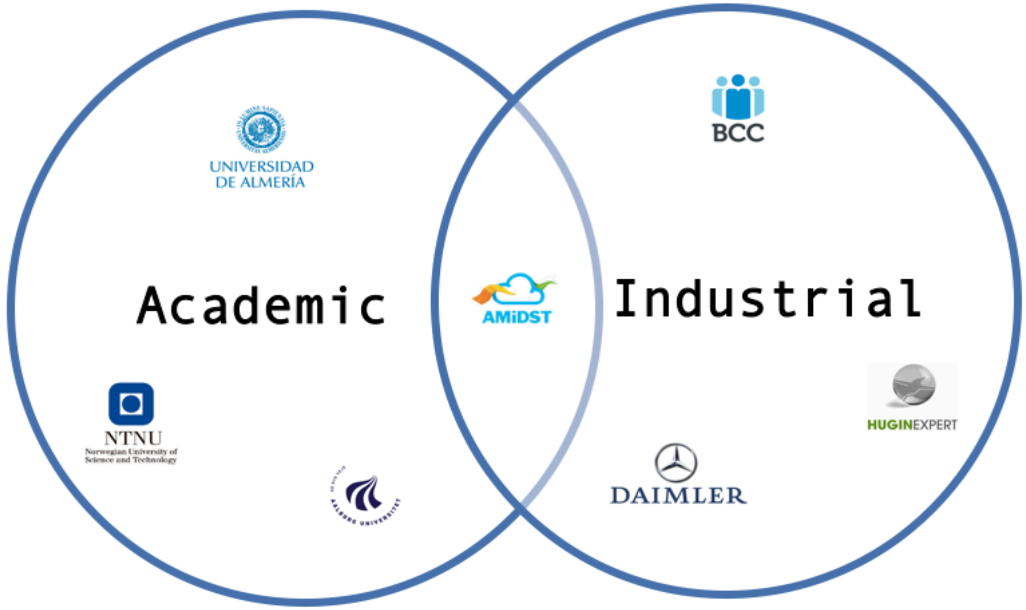
OUTLINE

- **Session 1**: Introduction to Probabilistic Machine Learning
  - Slides can be downloaded here.
- **Session 2**: Introduction to the AMIDST Toolbox.
  - Slides can be downloaded here.
  - Code exercises can be found here.
- **Session 3**: Coding an Intelligent Fire Detector System with the AMIDST Toolbox.
  - Slides can be downloaded here.
  - Code exercises can be found here.
- **Session 4**: Latent Variable Models.
  - Slides can be downloaded here.
- **Session 5**: Streaming data, Scalable Learning and Temporal Models with the AMIDST Toolbox.
  - Slides can be downloaded here.
  - Code exercises can be found here.
- **Session 6**: Future Trends in Probabilistic Machine Learning.
  - Slides can be downloaded here.

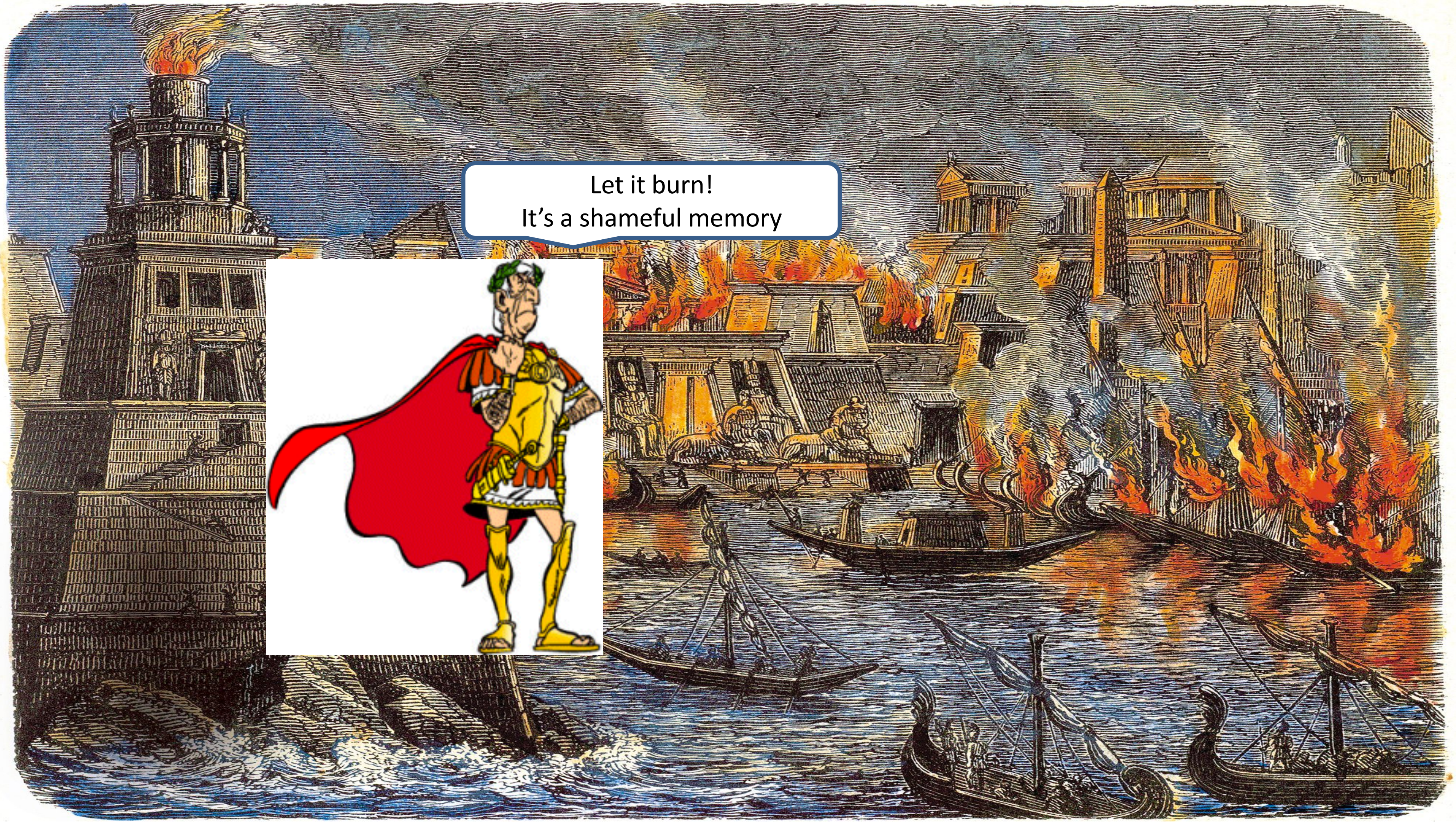
https://github.com/andresmasegosa/GeiloWinterSchool2018

The AMIDST consortium

AMiDST TOOLBOX



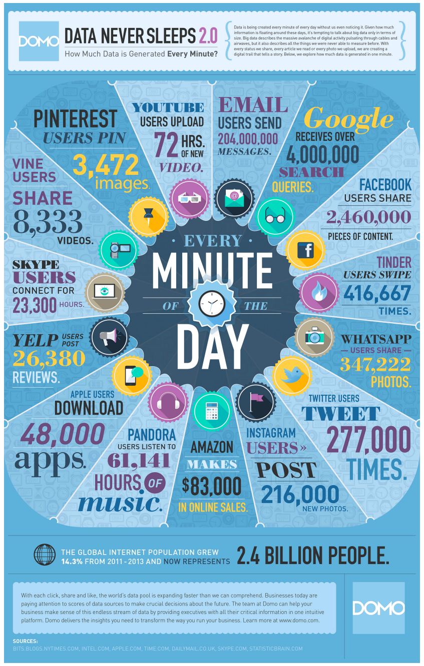
SEVENTH FRAMEWORK PROGRAMME


Let it burn!
It’s a shameful memory

3rd century BC

50 thousands papyrus = 20 thousands books

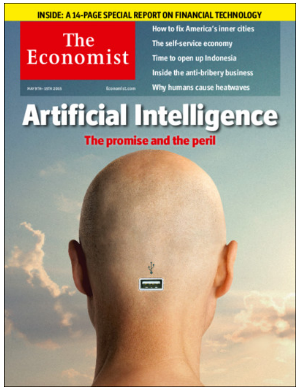
21st century DC



6 thousand books, 2 millions posts and news
daily


The Economist
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

Machine Learning



Text Modelling

1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says	Children	Garden

6	7	8	9	10
Bush	Building	Won	Yankees	Government
Campaign	Street	Team	Game	War
Clinton	Square	Second	Mets	Military
Republican	Housing	Race	Season	Officials
House	House	Round	Run	Iraq
Party	Buildings	Cup	League	Forces
Democratic	Development	Open	Baseball	Iraqi
Political	Space	Game	Team	Army
Democrats	Percent	Play	Games	Troops
Senator	Real	Win	Hit	Soldiers

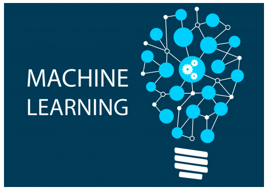
11	12	13	14	15
Children	Stock	Church	Art	Police
School	Percent	War	Museum	Yesterday
Women	Companies	Women	Show	Man
Family	Fund	Life	Gallery	Officer
Parents	Market	Black	Works	Officers
Child	Bank	Political	Artists	Case
Life	Investors	Catholic	Street	Found
Says	Funds	Government	Artist	Charged
Help	Financial	Jewish	Paintings	Street
Mother	Business	Pope	Exhibition	Shot

Topics found in 1.8M articles from the New York Times

[Hoffman, Blei, Wang, Paisley, JMLR 2013]

Knowledge Access (3rd century BC)

Knowledge Access (21rd century)



TensorFlow NumPy AMiDST TOOLBOX

What is Machine Learning?

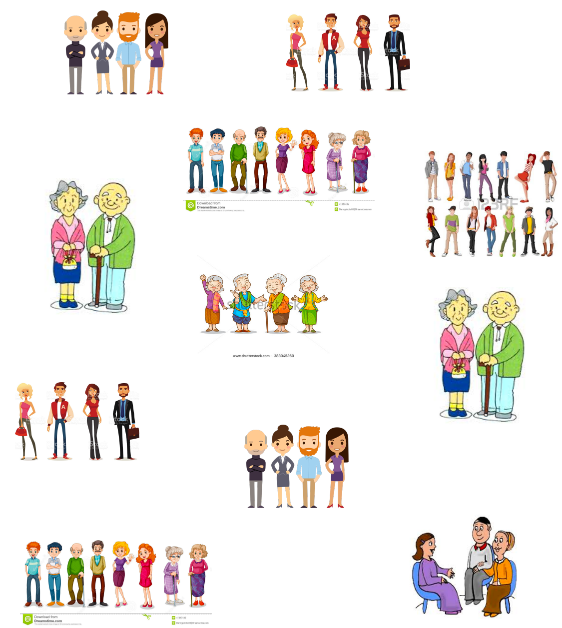
Manual Computer Programming

vs

What is Machine Learning?

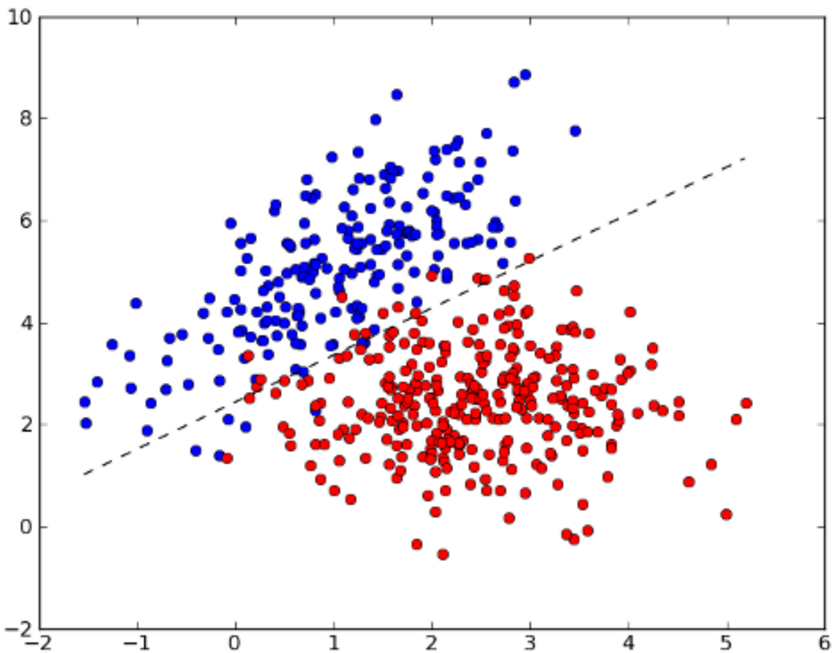
Automatic Computer Programming

vs



Supervised Learning

- Finding a functional mapping:



$$f_\theta : X \to Y$$

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

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

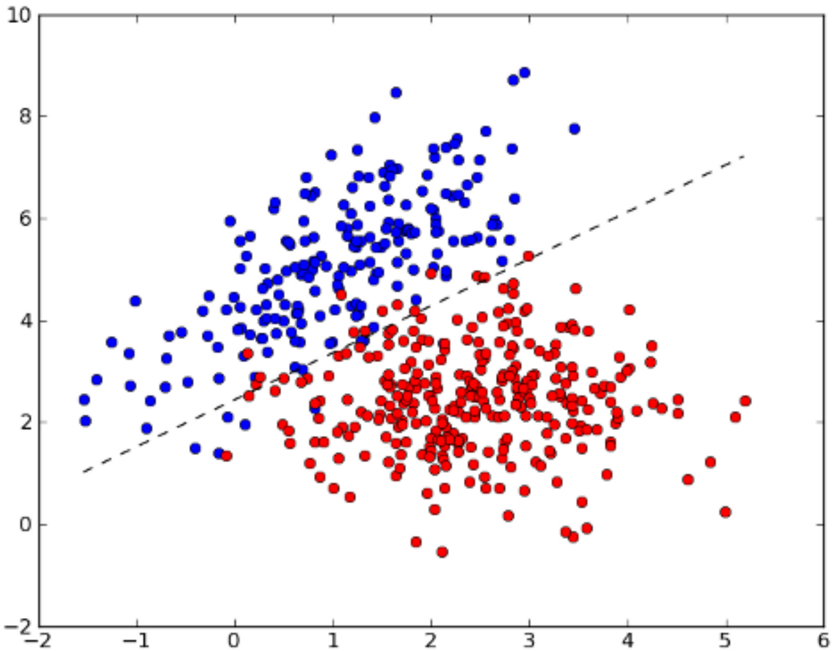
$$\theta \in \Re^2$$

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

AMiDST TOOLBOX

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

$$f_\theta : X \to Y$$



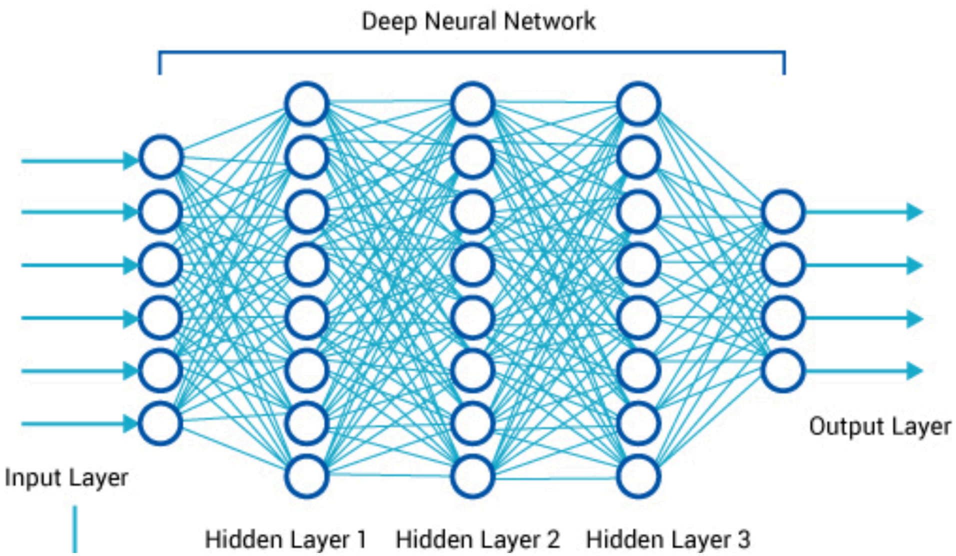
$$\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

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

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

Machine learning involves solving continuous optimization problems

SUPERVISED LEARNING



$$f_\theta : X \rightarrow Y$$

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

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

DNN are highly non-linear mappings

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

Supervised Learning

AMiDST TOOLBOX

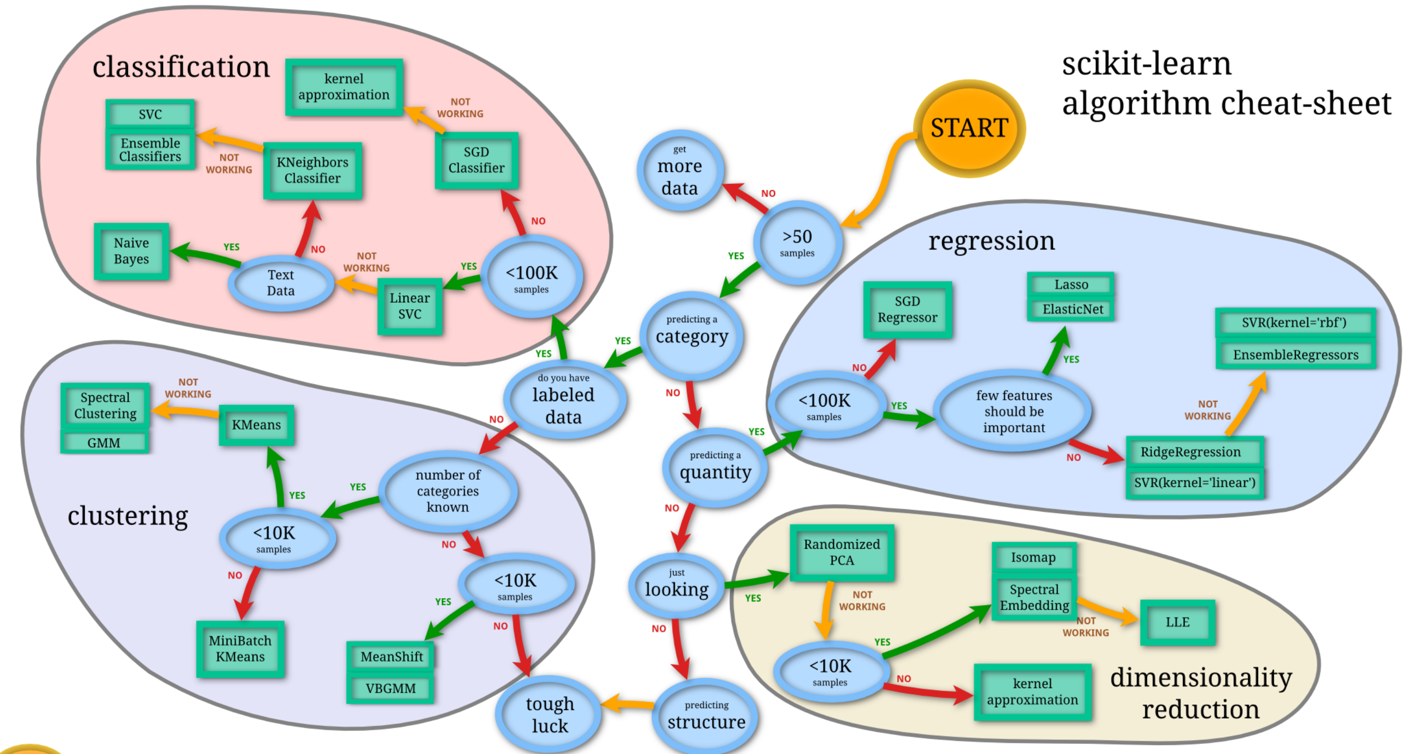
[Functional Mapping]

X (Input) → Y (Output)

Email →	Spam? (0/1)
Image →	Object (computer vision)
Audio →	Text (speech recognition)
English →	French (automatic translation)
(Corpus, Query) →	List (information retrieval)
(Ad, user) →	Click? (Ad Placing)
(Camera, Address) →	Steering Angle, Brake Force,…. (Autonomous Driving)

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

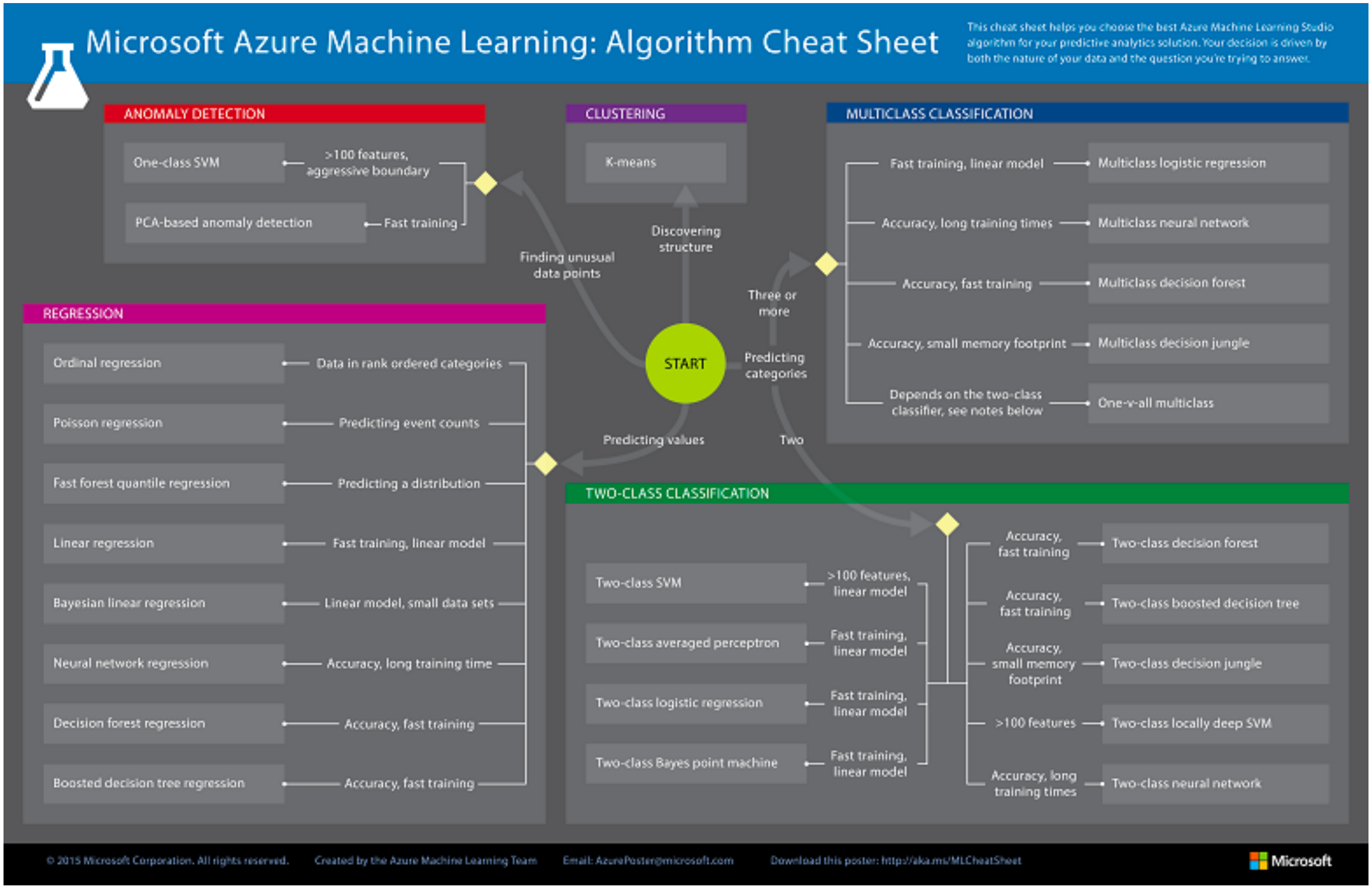
Function-approximation Machine Learning



Back

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

Machine Learning Landscape



Practical Issues

- 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.

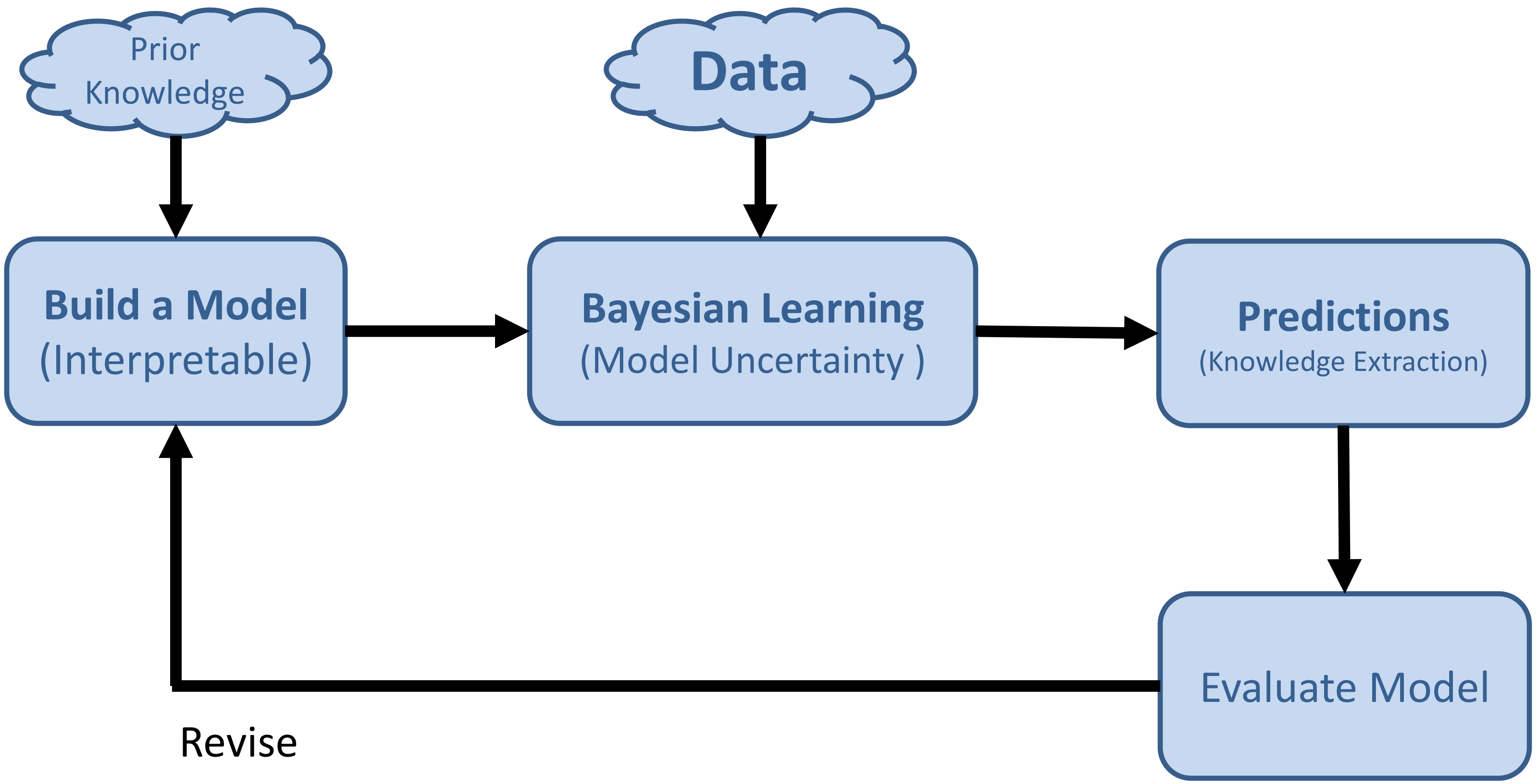
Theoretical Issues

- Black Box Approaches
  - No Model Interpretability
  - No understanding in how decisions are made

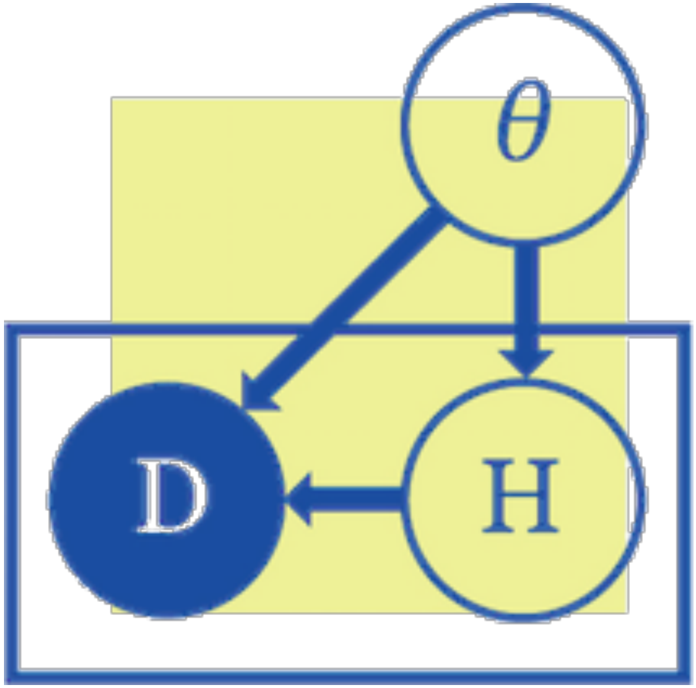
- Uncertainty Quantification
  - No Predictions Uncertainty
  - No Model Uncertainty

Probabilistic (Bayesian) Machine Learning

Model-Based Machine Learning

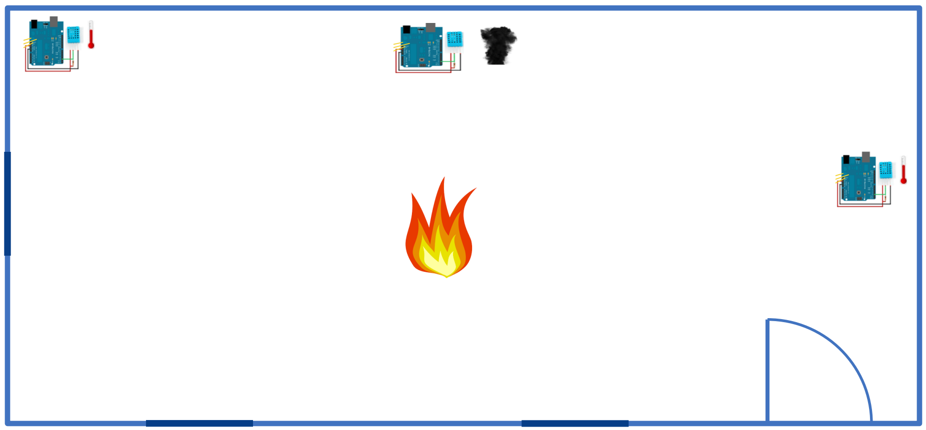


Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.



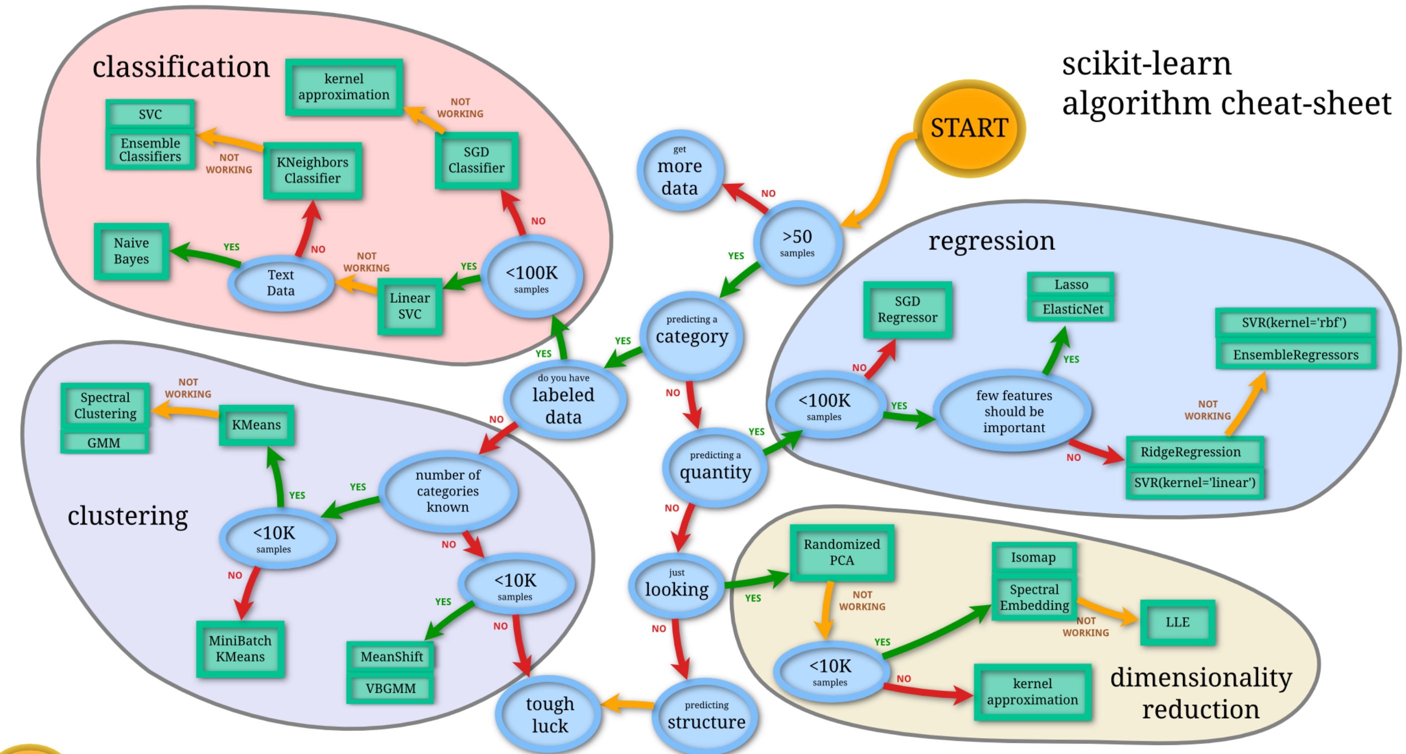
Probabilistic Graphical Models

Fire Detection from smoke and temperature sensors

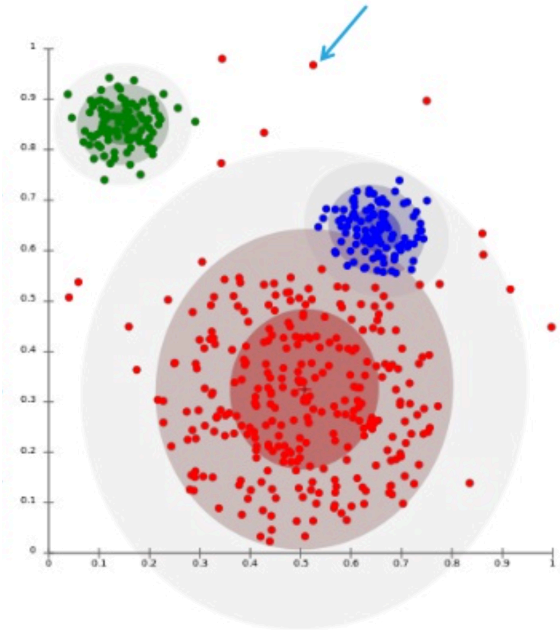


- Data Collected
  - Tons of observations in normal settings (no fire).
  - No observations in the presence of fire.

Function-approximation Machine Learning

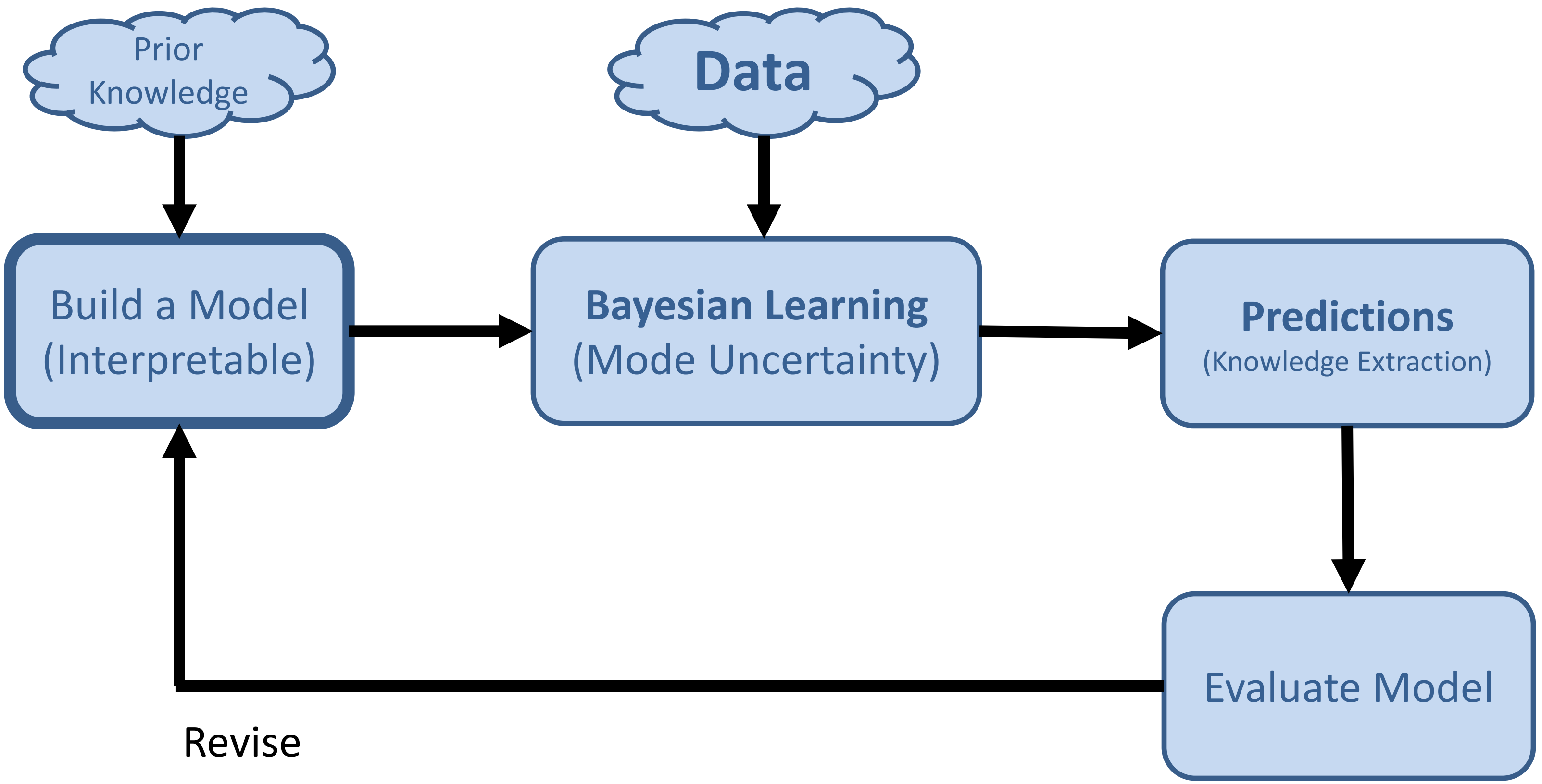


http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

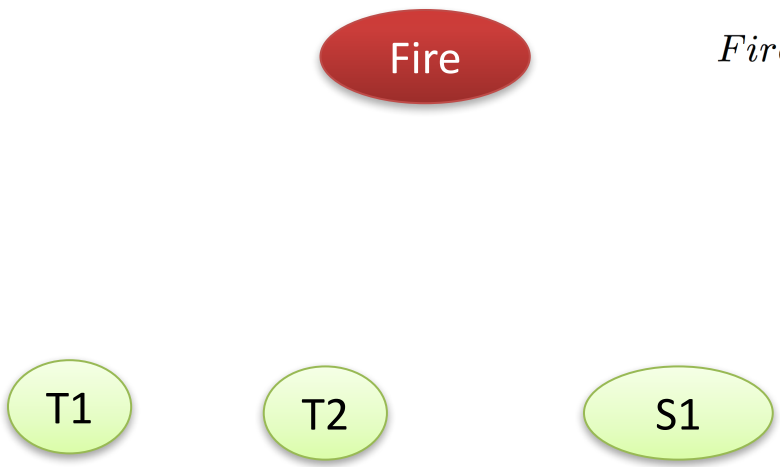


Black Box Approach:

Anomaly Detection with (streaming) K-means



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



$$Fire \sim Binomial(\rho)$$

$$T1, T2 \sim Normal(\mu, \sigma)$$

$$S_1 \sim Binomial(\rho)$$

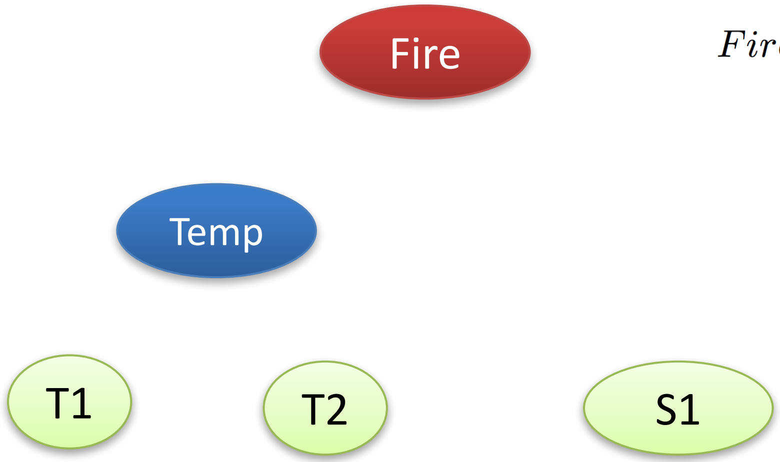
Probabilistic Modeling

Every relevant object is a random variable.

Fire Detector Model

$$Fire \sim Binomial(\rho)$$

$$Temp \sim Normal(\mu, \sigma)$$



$$T1, T2 \sim Normal(\mu, \sigma)$$

$$S_1 \sim Binomial(\rho)$$

Latent Variables

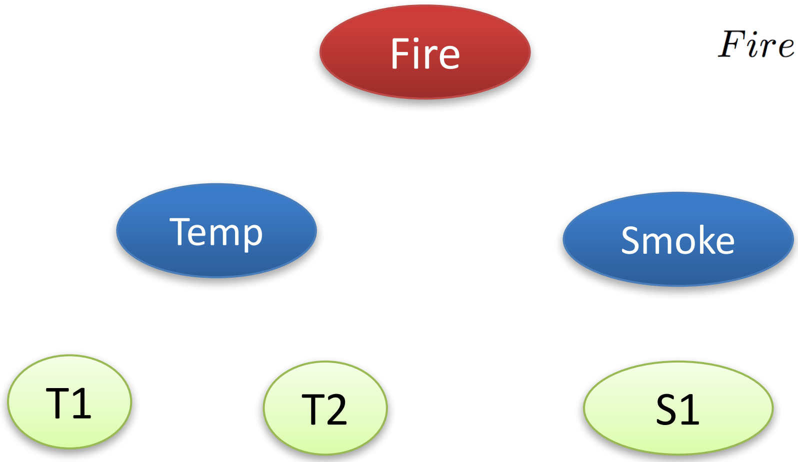
Non-observable relevant mechanisms

FIRE DETECTOR MODEL

$Fire \sim Binomial(\rho)$

$Temp \sim Normal(\mu, \sigma)$

$Smoke \sim Binomial(\rho)$



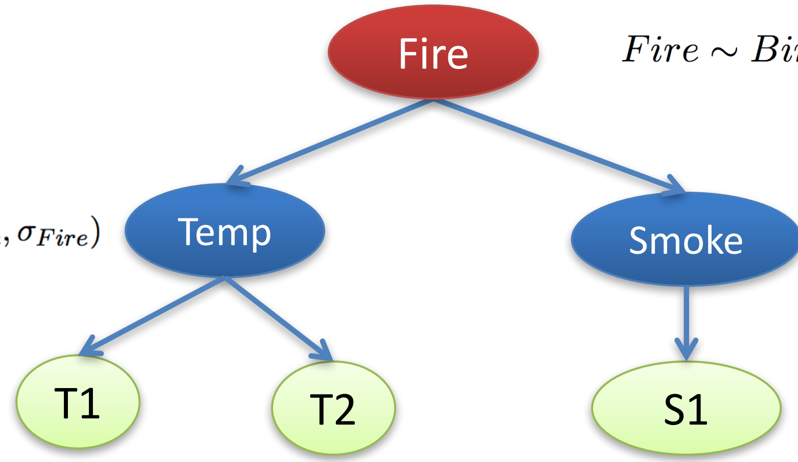
$T1, T2 \sim Normal(\mu, \sigma)$

$S_1 \sim Binomial(\rho)$

Latent Variables

Non-observable relevant mechanisms

Fire Detector Model



$Fire \sim Binomial(\rho)$

$Temp|Fire \sim Normal(\mu_{Fire}, \sigma_{Fire})$

$Smoke|Fire \sim Binomial(\rho_{Fire})$

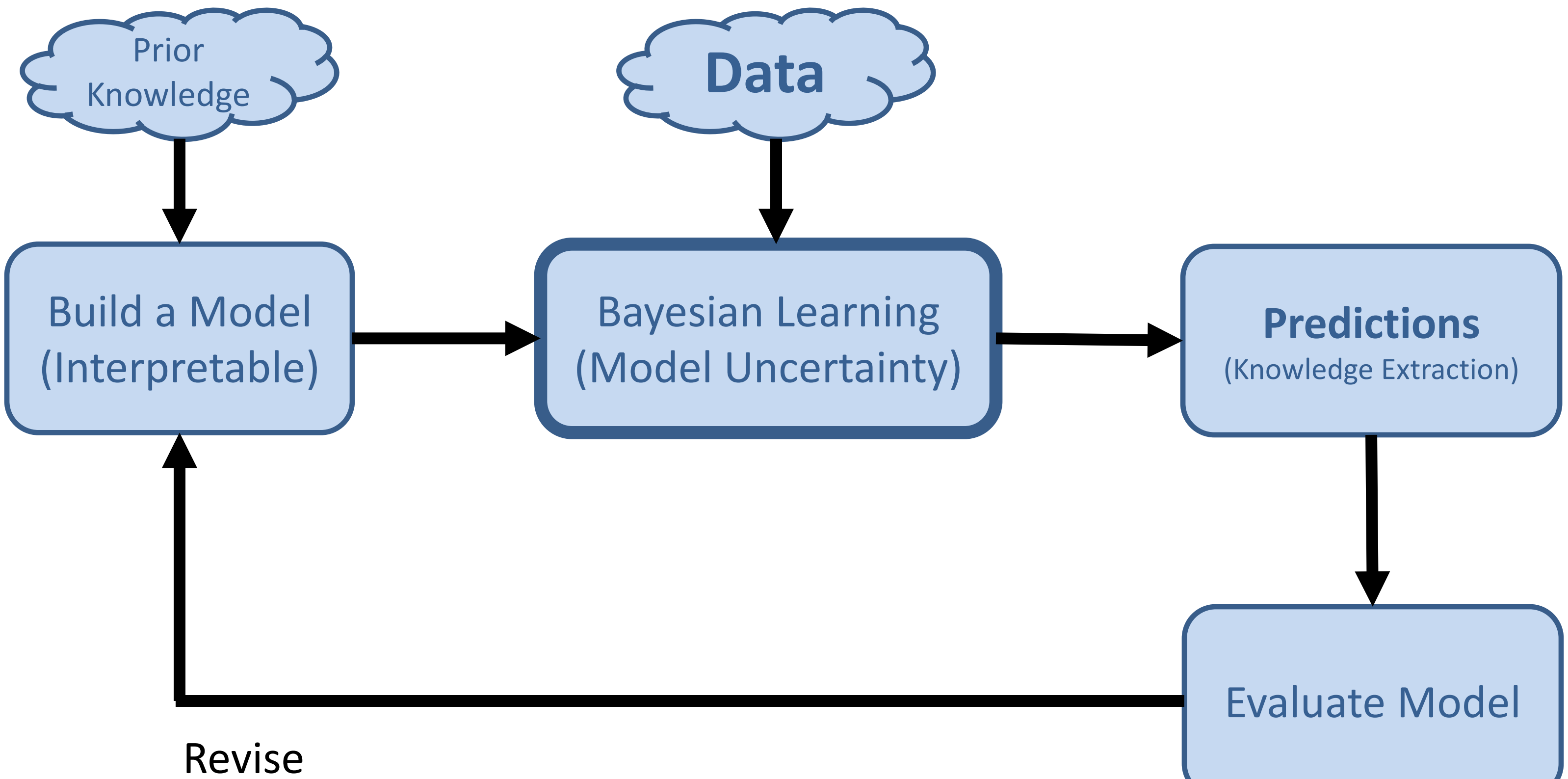
$T1, T2|Temp \sim Normal(\mu_{Temp}, \sigma)$

$S_1|Smoke \sim Binomial(\rho_{Smoke})$

Causal Relationships

They can be extracted for the mechanism itself

Code: Session3

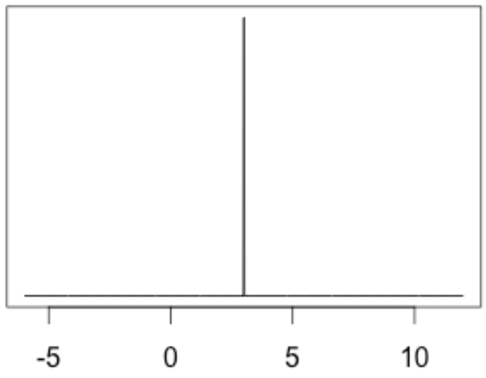


Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.

$$P(\theta|\mathbf{D})$$

Bayesian Learning

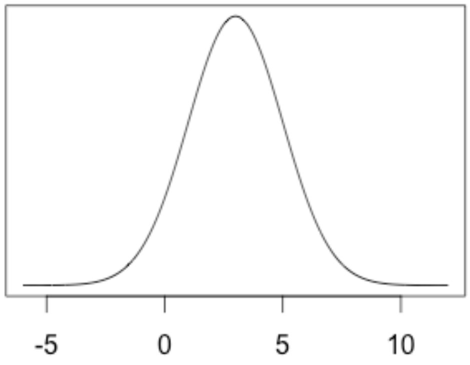
Model Uncertainty



$$\theta^{\star}$$

[Point Estimate]

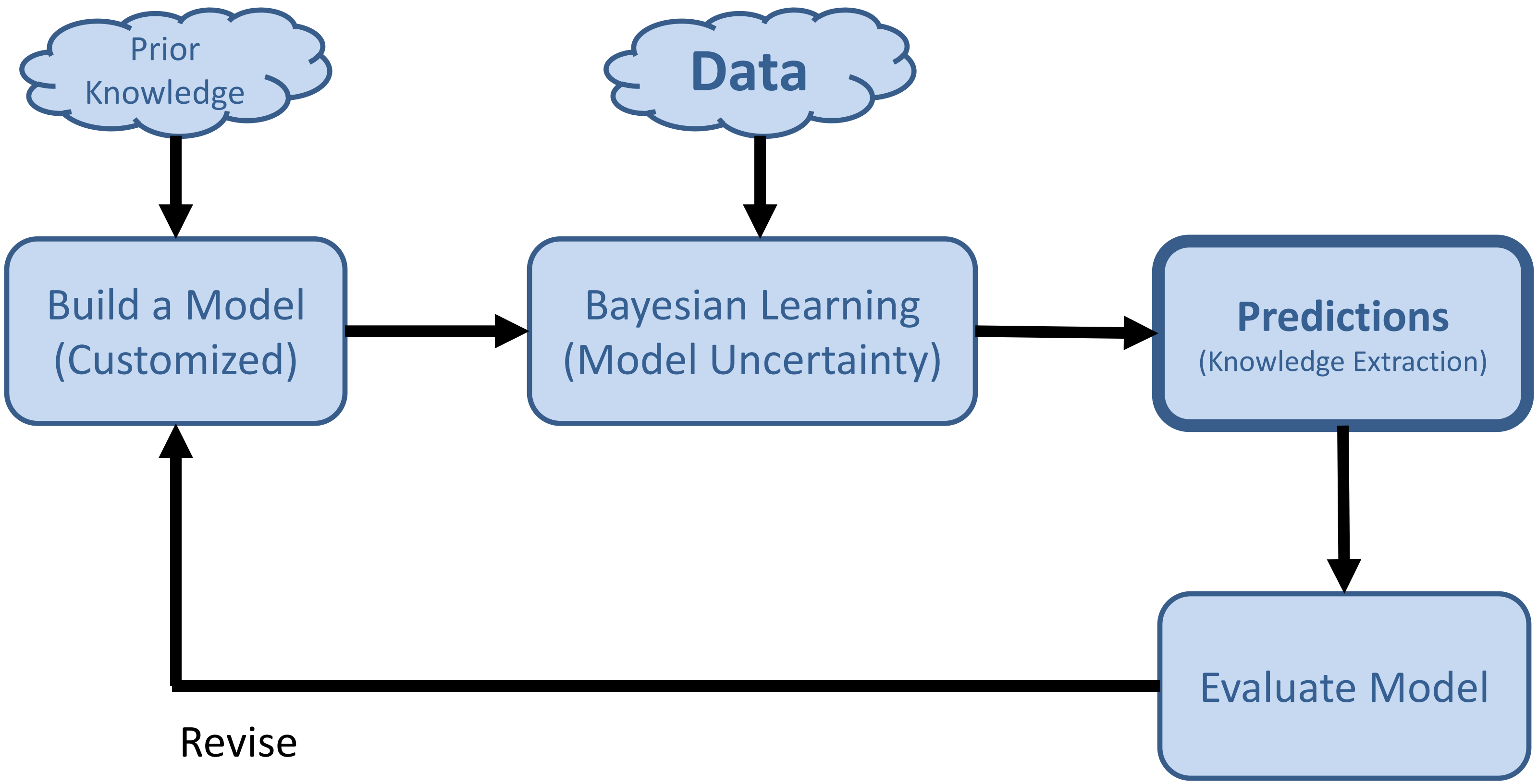
vs



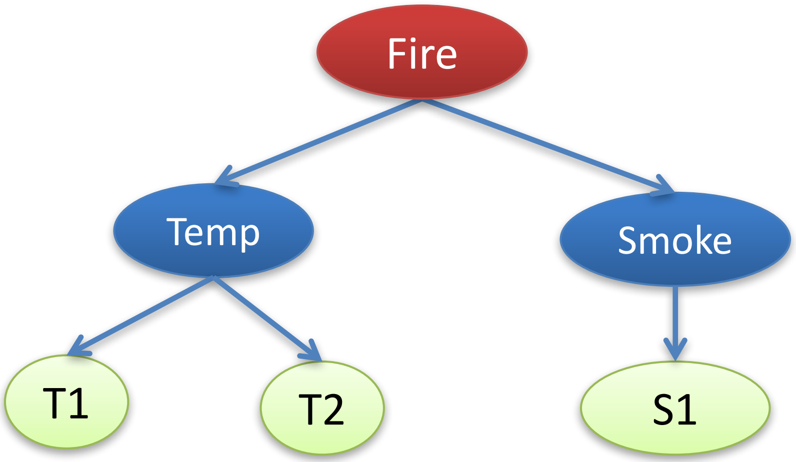
$$p(\theta|D)$$

[Bayesian Estimate]

Example: $y = \theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_k \cdot x_k$



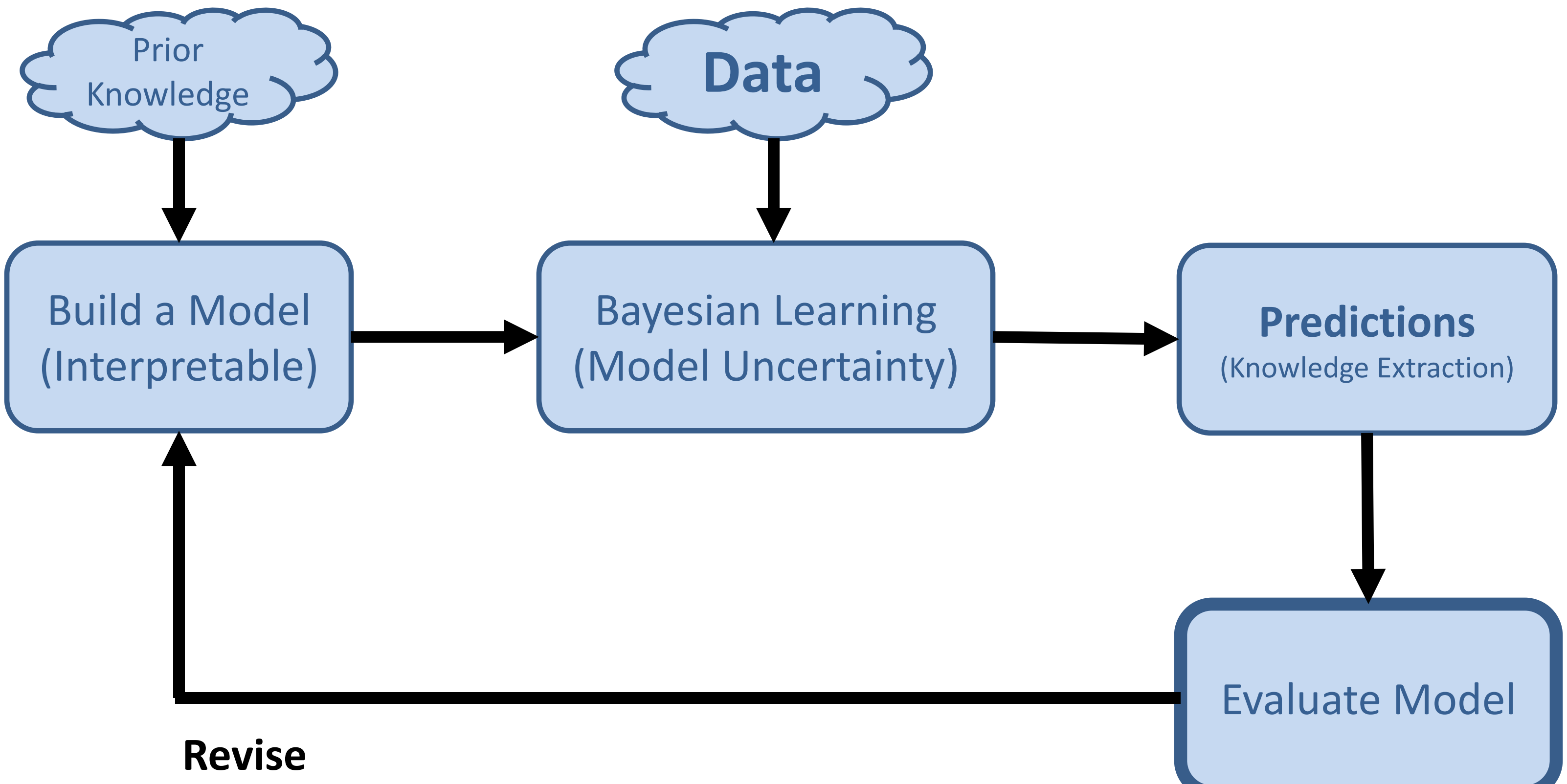
Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.



$$p(Fire = true|t1, t2, s1)$$

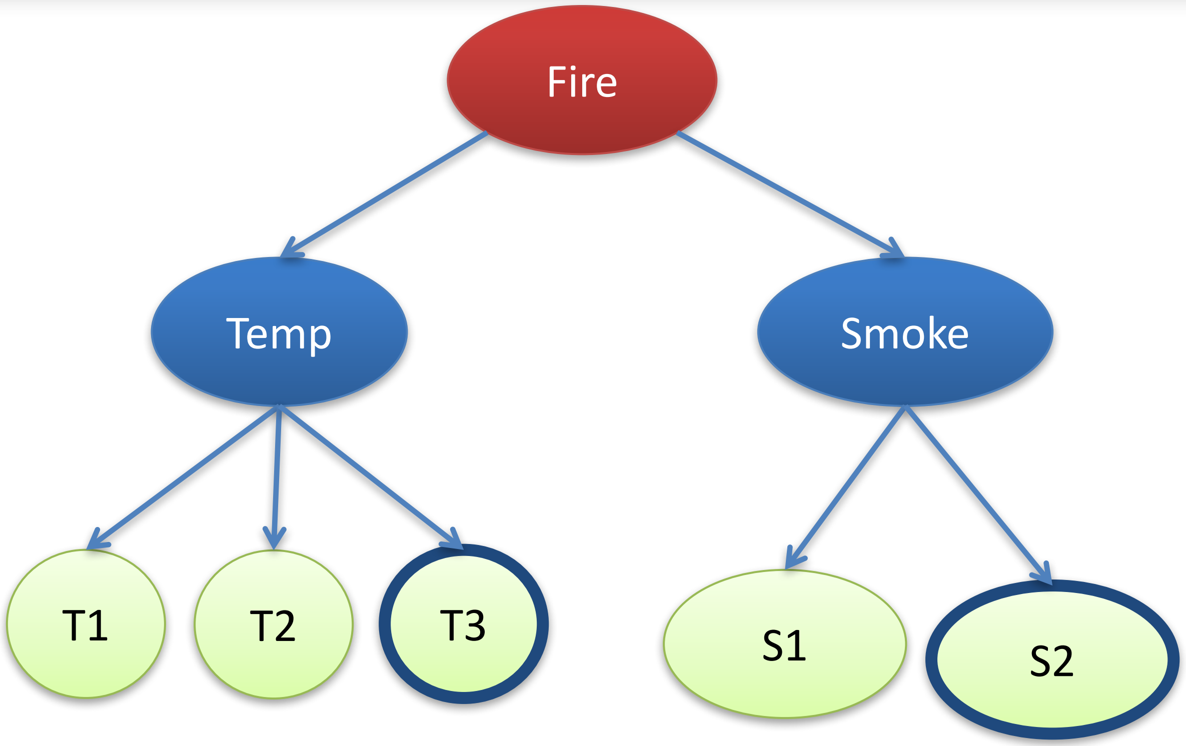
Query the Model

Computing posterior probabilities given evidence



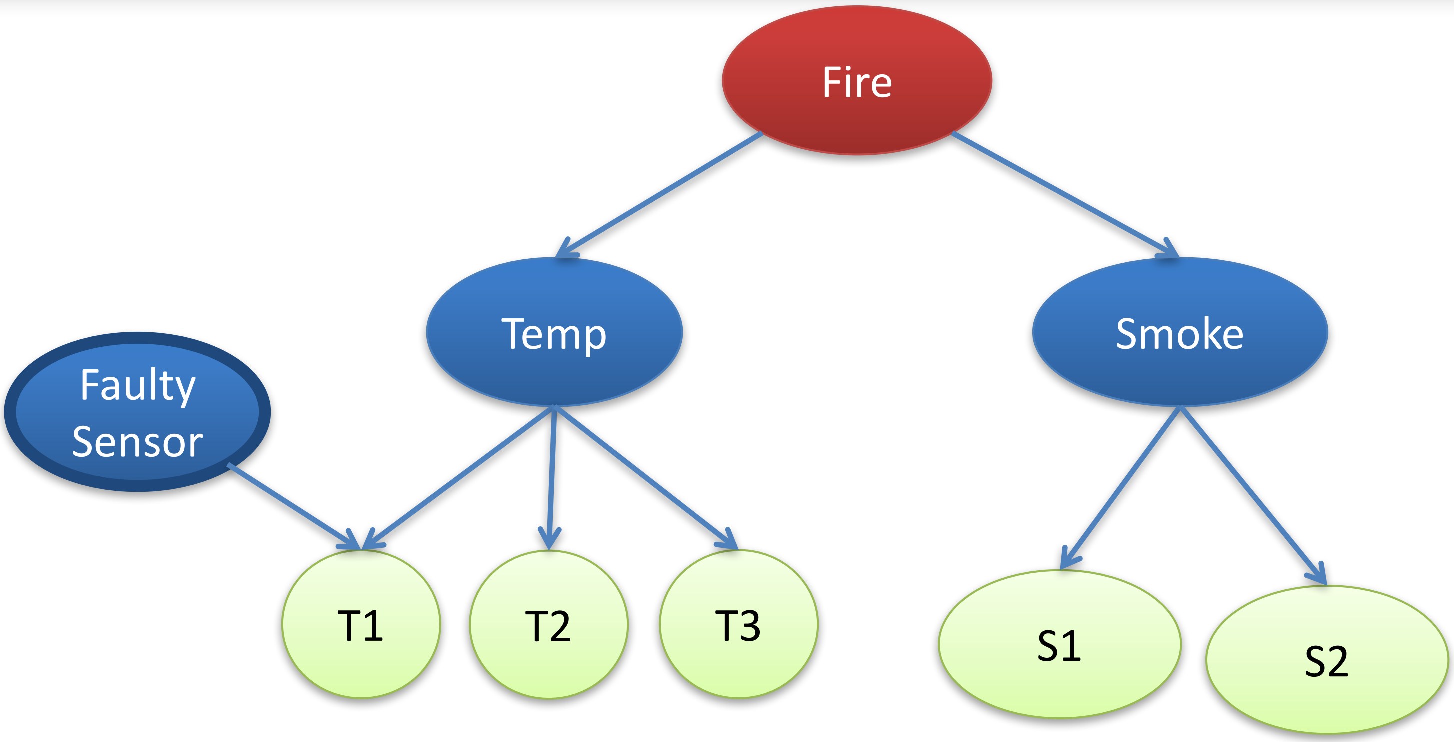
Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.

Example of PGMs



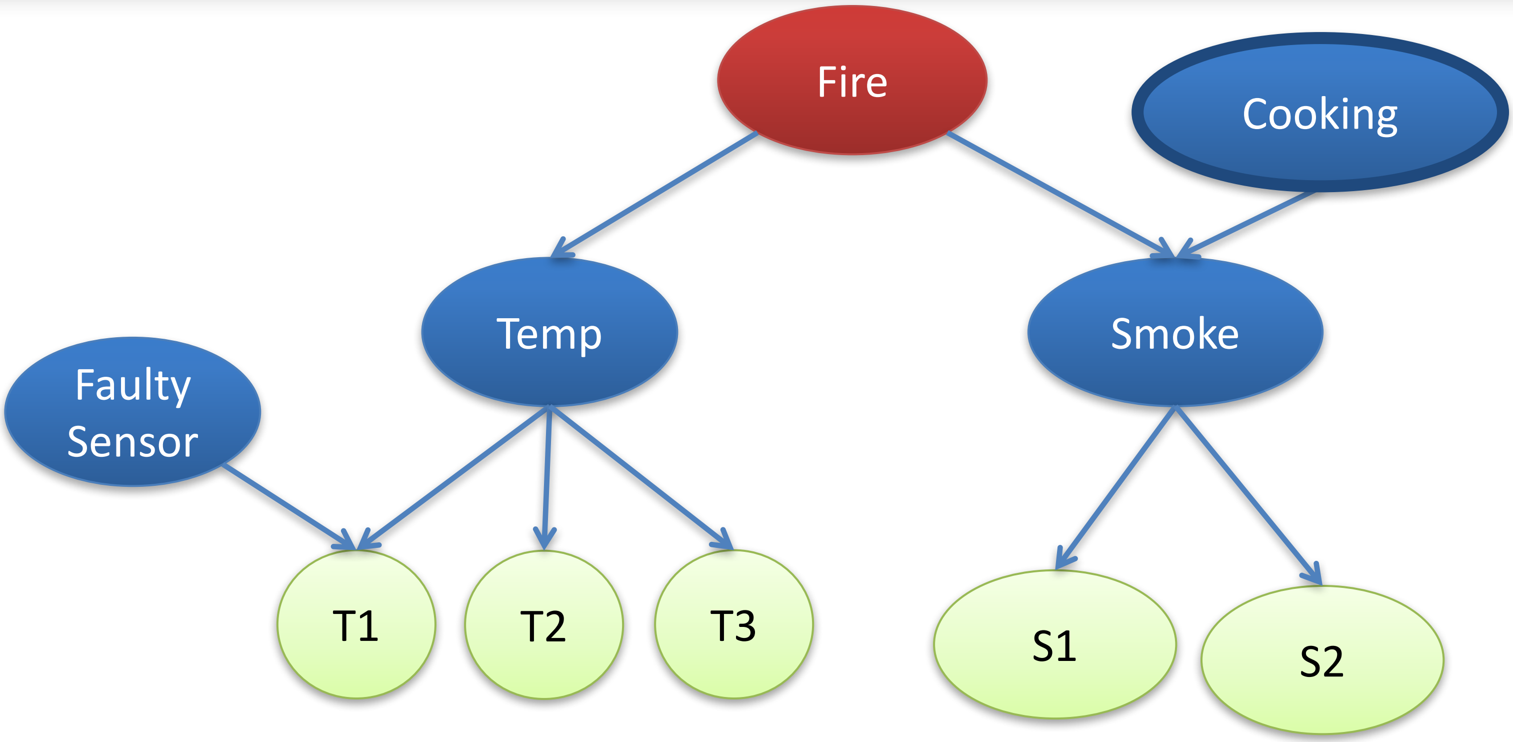
What if I have more sensors?

Example of PGMs



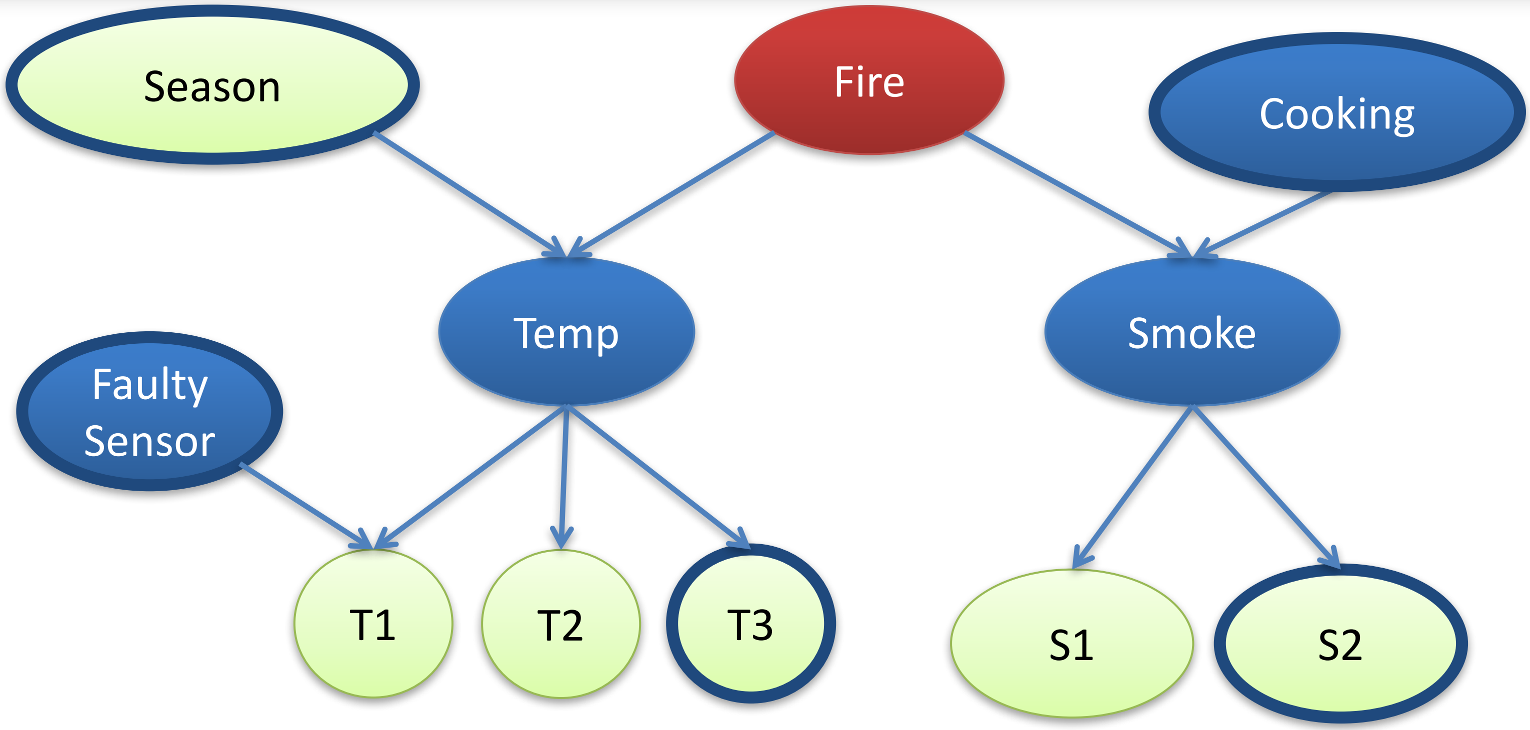
What if a sensor fails?

Example of PGMs



What if the system is placed in a kitchen?

Example of PGMs



What is normal indoor temperature changes through the season?

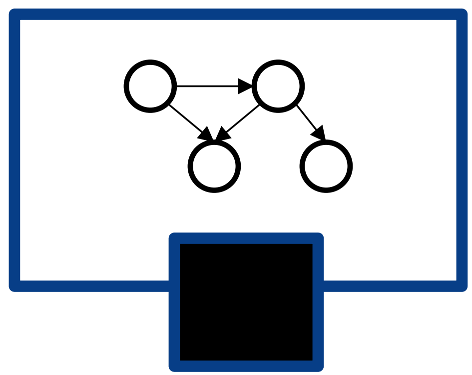
AMIDST APPROACH

Openbox Models

[Probabilistic Graphical Models]

[Big] Data

[+Prior Information]

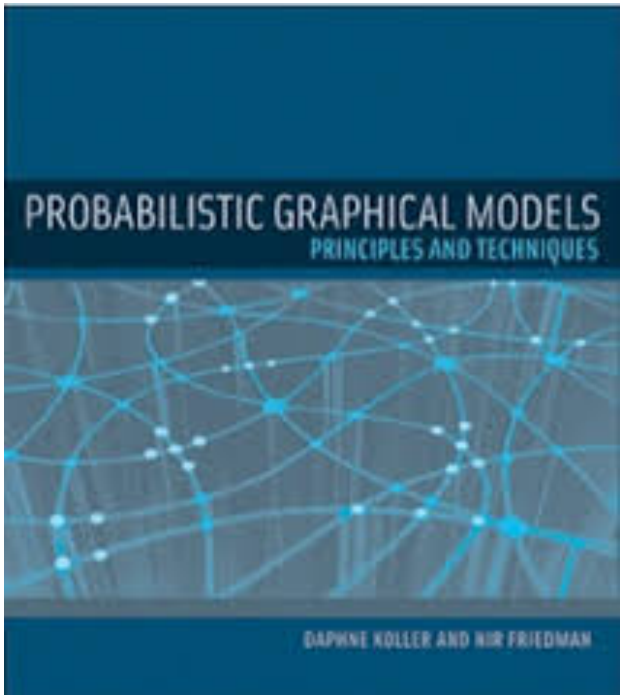


Knowledge

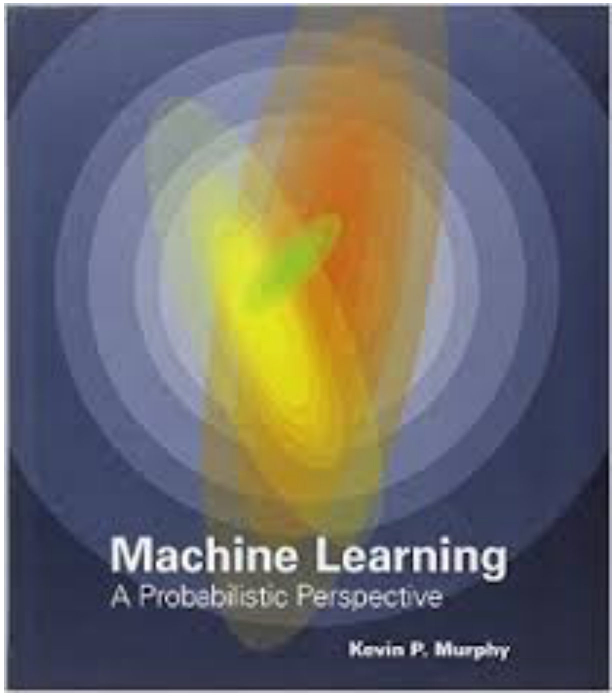
[+ Predictions]

[Scalable] Bayesian Inference Engine

(Powered by Variational Methods)



+



Probabilistic **Graphical Models**

Probabilistic Machine Learning

Thanks for your attention

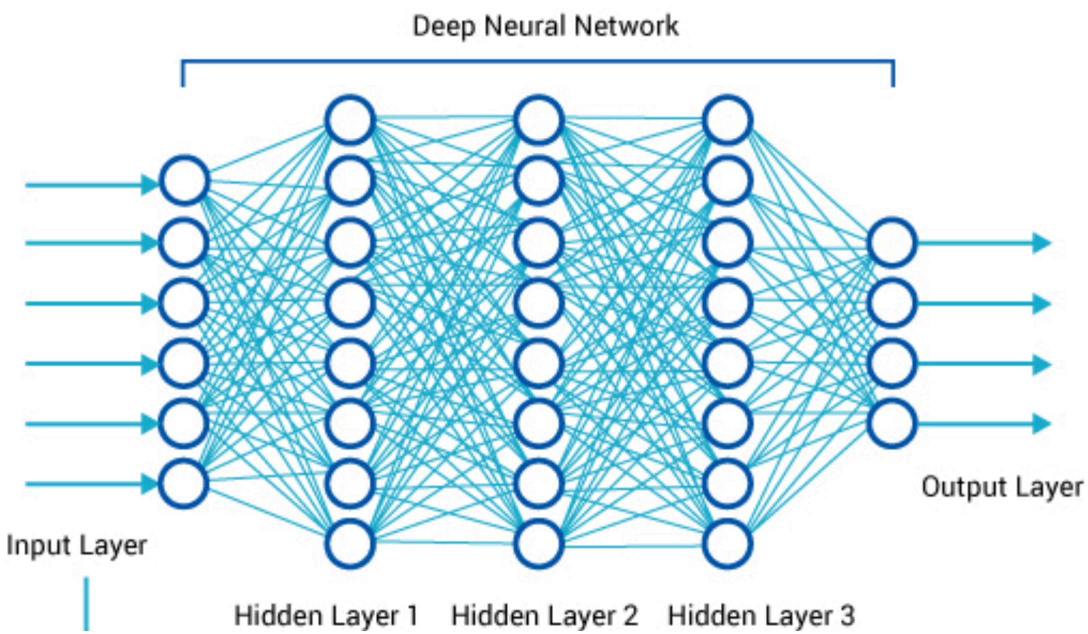
www www.amidsttoolbox.com

@ contact@amidsttoolbox.com

@AmidstToolbox

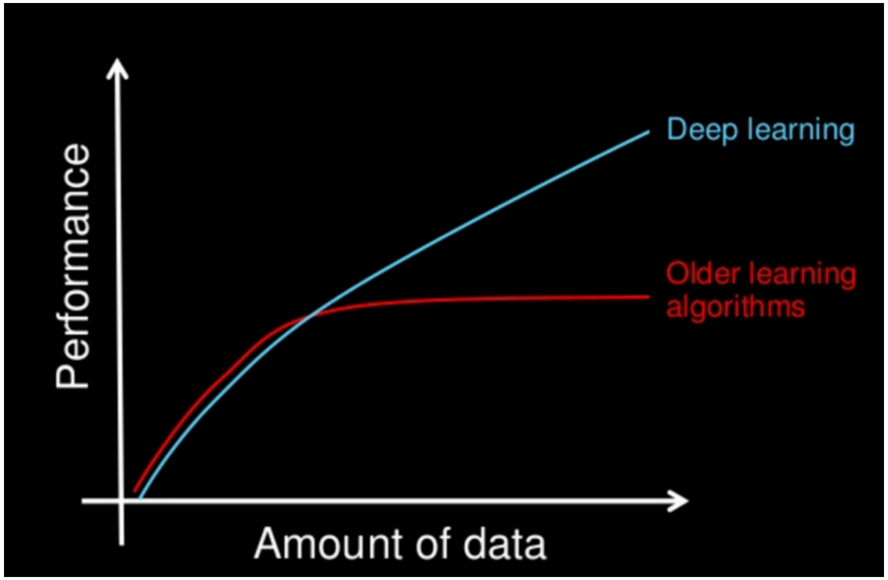
AMiDST TOOLBOX

What about Deep Learning?



DNN are highly non-linear mappings

DEEP LEARNING



$$f_\theta : X \rightarrow Y$$

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

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

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:

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

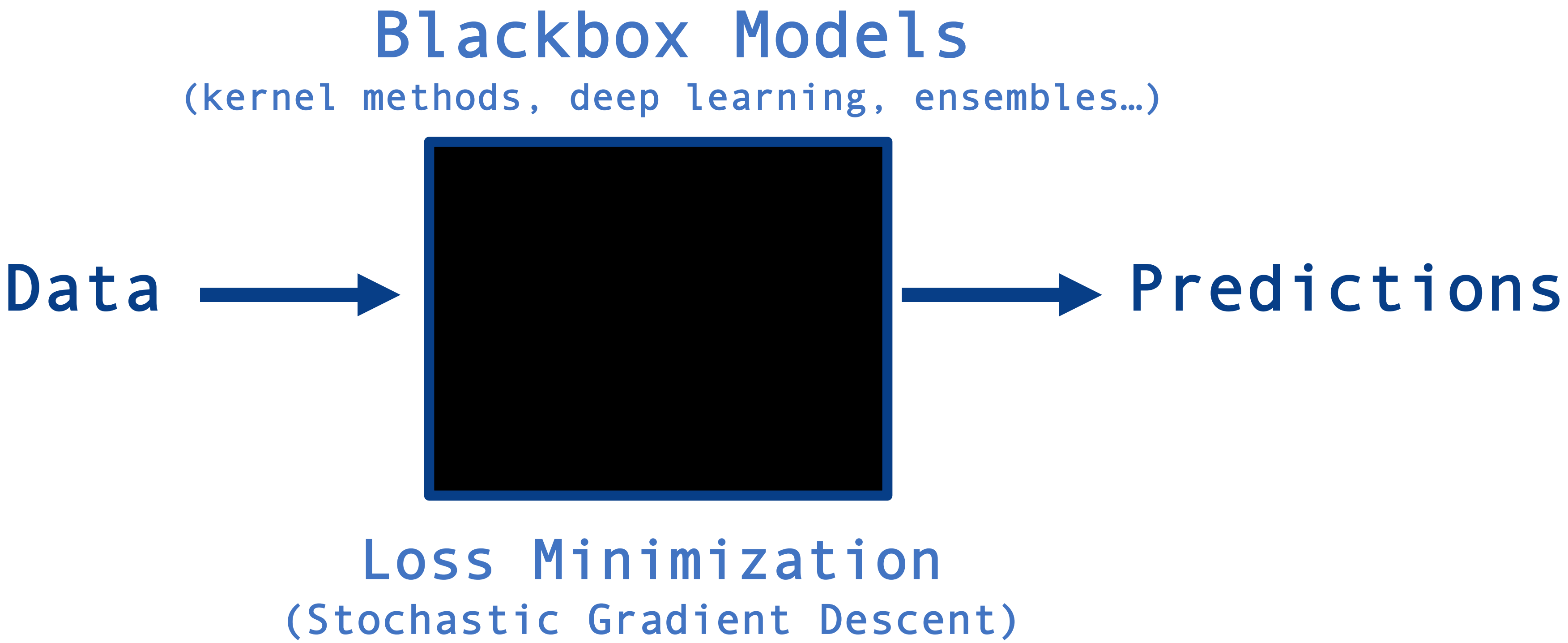
  - **Dimensionality Reduction**'s loss function:

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

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

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

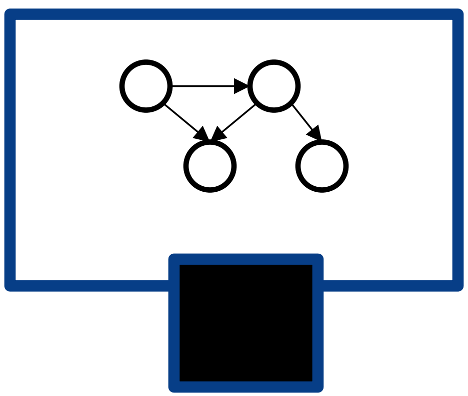
Andrew Ng. Coursera. Machine Learning.
https://en.coursera.org/learn/machine-learning


Blackbox Models
(kernel methods, deep learning, ensembles…)
Data
Predictions
Loss Minimization
(Stochastic Gradient Descent)

Probabilistic Machine Learning

Openbox Models
[Probabilistic Graphical Models]

Data
[+Prior Information]



Knowledge
[+ Predictions]

Black-Box Learning Engine
(Bayesian inference)