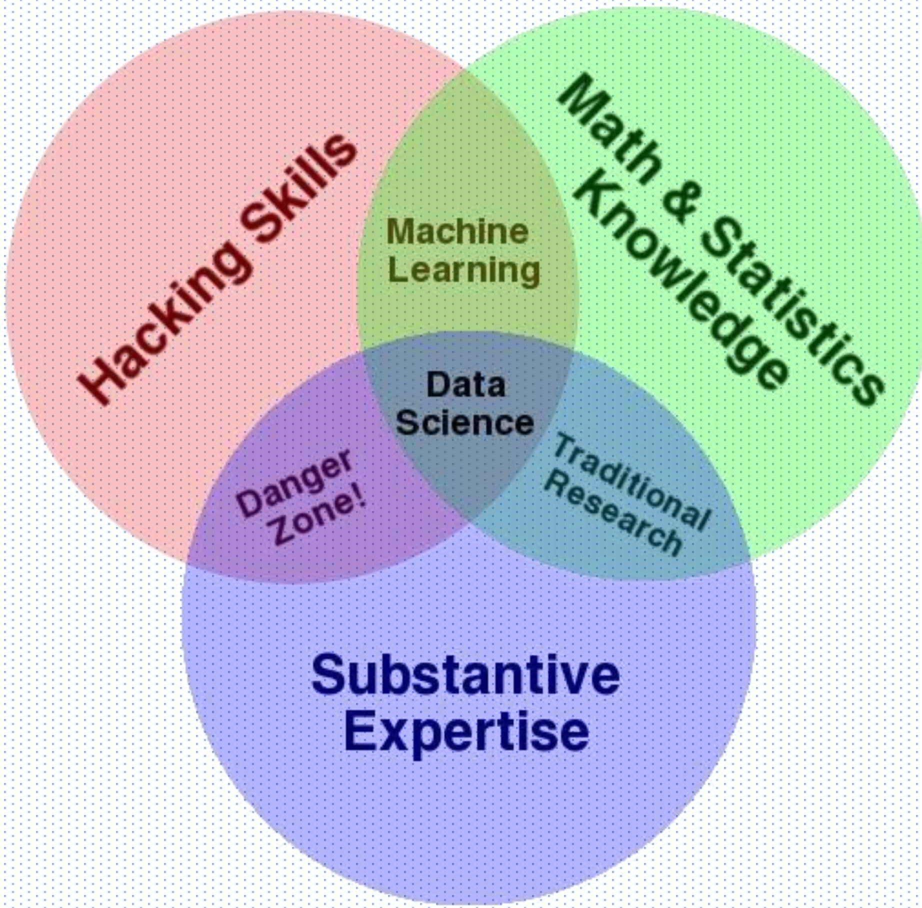
MSc in Data Analytics

Introduction to Machine Learning

Stavros A. Karkanis
Professor
University of Thessaly
sk@teiste.gr / karkanis@gmail.com

The Data Science Venn Diagram

Computer
Science



Statistics

Domain
knowledge

http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Data Science Overview

Why,

Where,

What,

How,

Who

- Where does data come from?
- What is Data Science?
- How to do Data Science?
- Who are Data Scientists?

Sponsored search

Google revenue around $50 bn/year from marketing, 97% of the companies revenue.

Sponsored search uses an auction – a pure competition for marketers trying to win access to consumers.

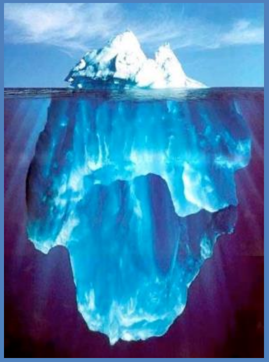
In other words, a competition for **models** of consumers – their likelihood of responding to the ad – and of determining the right bid for the item.

There are around 30 billion search requests a month. Perhaps a **trillion events** of history between search providers.

Google Adwords and Adsense

"Big Data" Sources

It's All Happening On-line



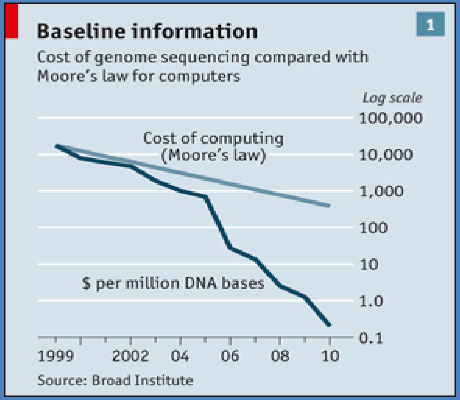
Every:
Click
Ad impression
Billing event
Fast Forward, pause,…
Server request
Transaction
Network message
Fault
…

User Generated (Web & Mobile)

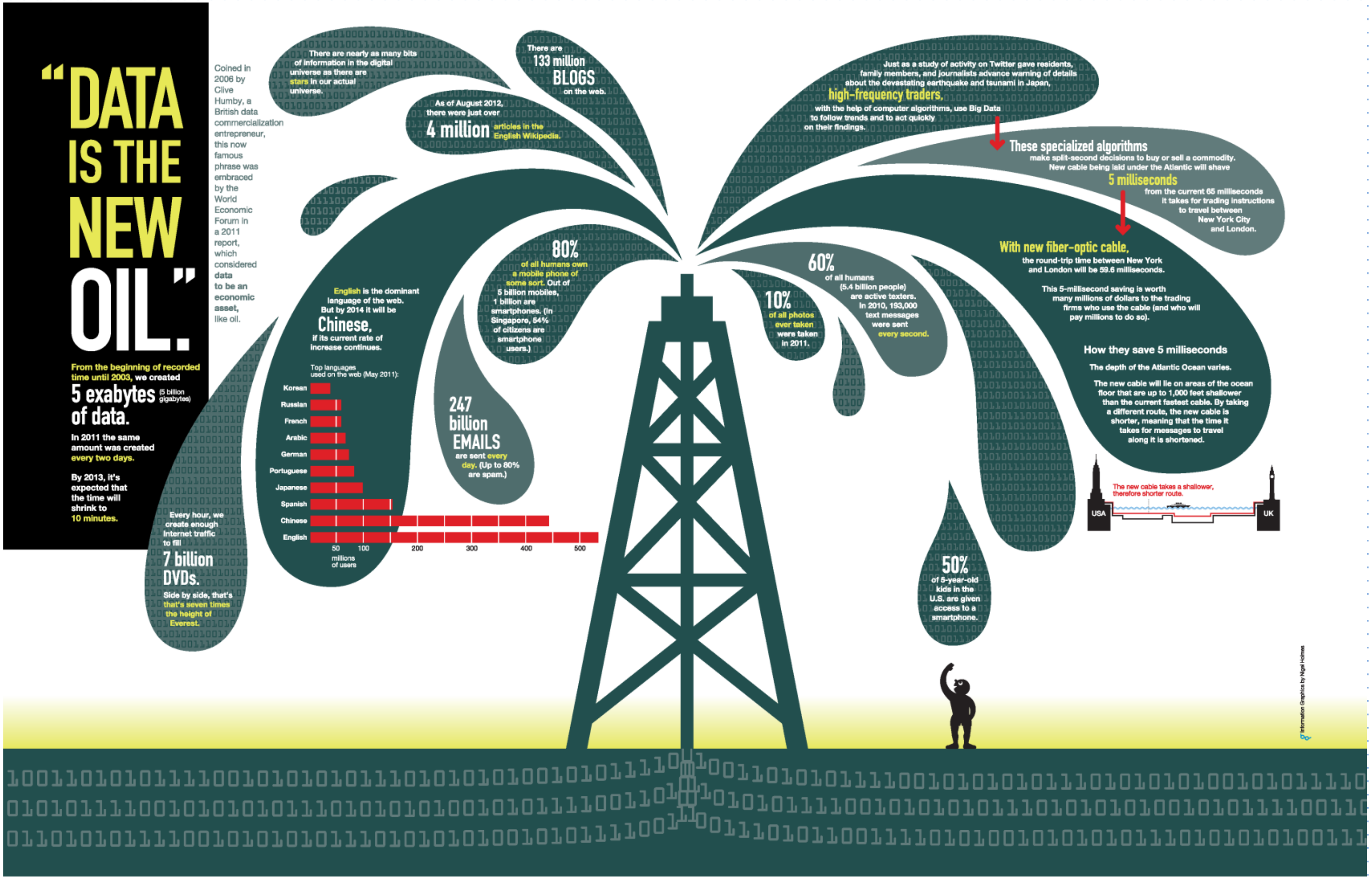
…
..

Internet of Things / M2M

Health/Scientific Computing



“Data is the New Oil” – World Economic Forum 2011



“Data is the New Oil” – World Economic Forum 2011

- “Data is the new oil." Coined in 2006 by Clive Huby, a British data commercialization entrepreneur, this now famous phrase was embraced by the World Economic Forum in a 2011 report,

- All human generated information up to 2003 is 5 exabytes. Same amount of data was generate every 2 days in 2011 and would be every 10 min NOW.

- Data is just like crude oil. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.

5 Vs of Big Data

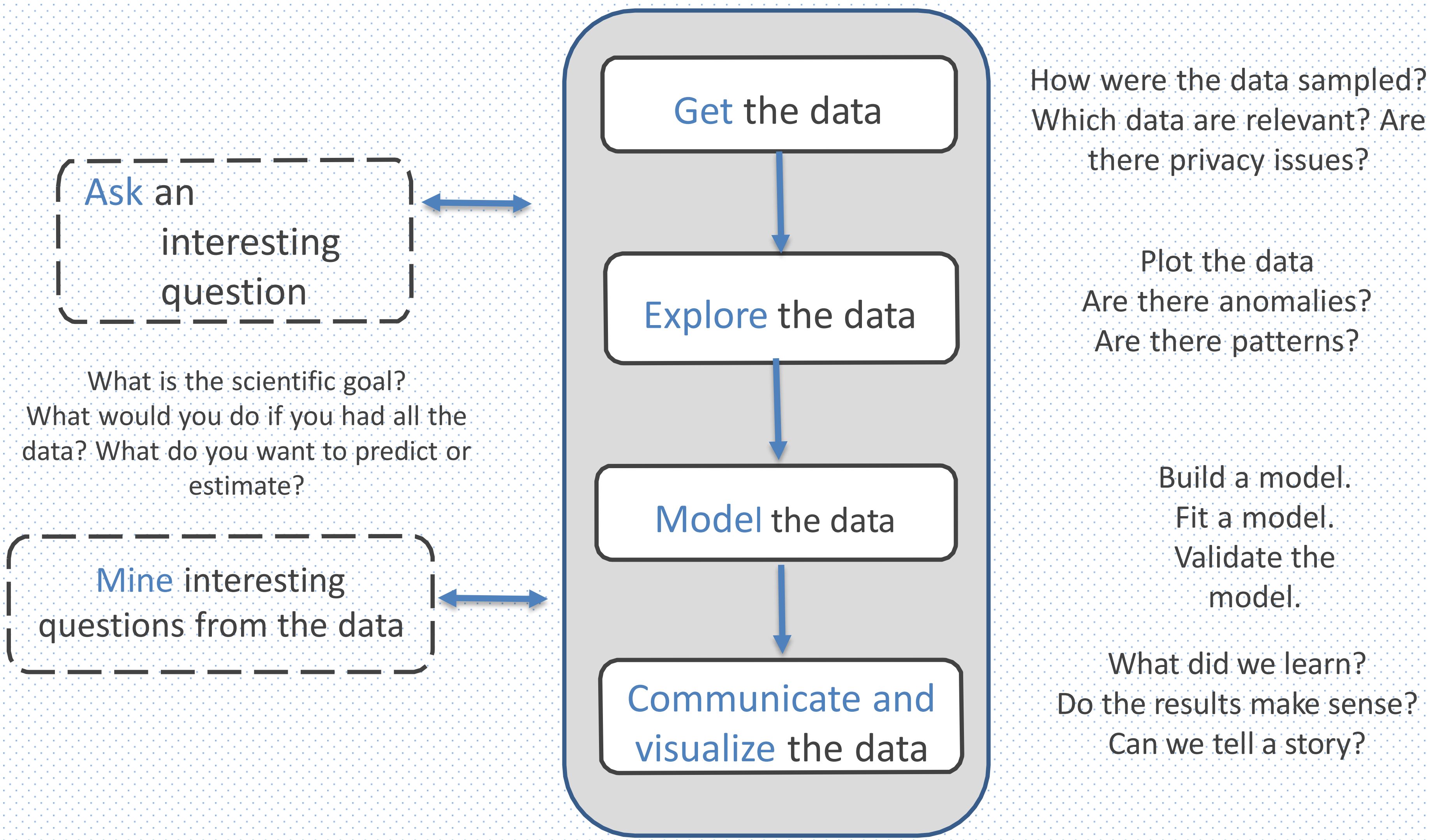
- Raw Data: **Volume**
- Change over time: **Velocity**
- Data types: **Variety**
- Data Quality: **Veracity**
- Information for Decision Making: **Value**

Data Science – A Definition

Data Science is the science which uses computer science, statistics and machine learning, visualization and human-computer interactions to collect, clean, integrate, analyze, visualize, interact with data to create data products.

Turn data into data products.

Data Science – A Scheme



Data Science – Key Facets

- data sampling/cleaning in order to get an informative, manageable data set;

- data storage and management in order to be able to access data
  - especially big data - quickly and reliably during subsequent analysis;

- **exploratory data analysis**
  - **to generate hypotheses and intuition about the data;**

- **prediction**
  - **based on statistical tools such as regression, classification, and clustering;**

and

- communication of results through visualization, stories, and interpretable summaries.

Contrast: Databases

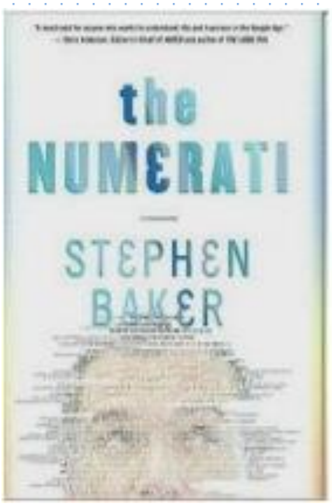
	Databases	Data Science
Data Value	“Precious”	“Cheap”
Data Volume	Modest	Massive
Examples	Bank records, Personnel records, Census, Medical records	Online clicks, GPS logs, Tweets, Building sensor readings
Priorities	Consistency, Error recovery, Auditability	Speed, Availability, Query richness
Structured	Strongly (Schema)	Weakly or none (Text)
Properties	Transactions, ACID*	CAP* theorem (2/3), eventual consistency
Realizations	SQL	NoSQL: MongoDB, CouchDB, Hbase, Cassandra, Riak, Memcached, Apache River, …

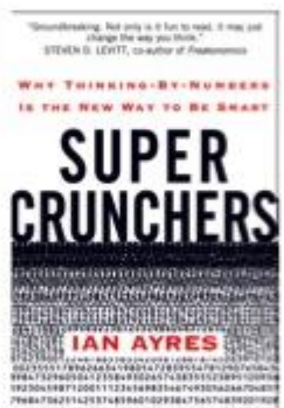
ACID = Atomicity, Consistency, Isolation and Durability

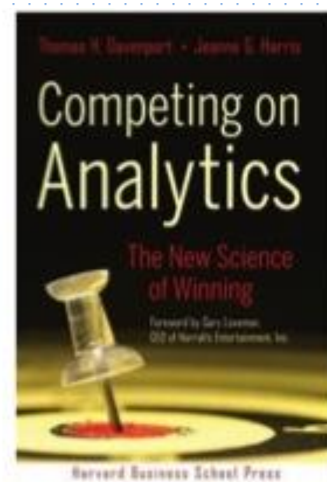
CAP = Consistency, Availability, Partition Tolerance

Contrast: Business Intelligence

Business Intelligence	Data Science
Querying the past	Querying the past present and future







Contrast: Machine Learning

Machine Learning
Develop new (individual) models
Prove mathematical properties of models
Improve/validate on a few, relatively clean, small datasets
Publish a paper

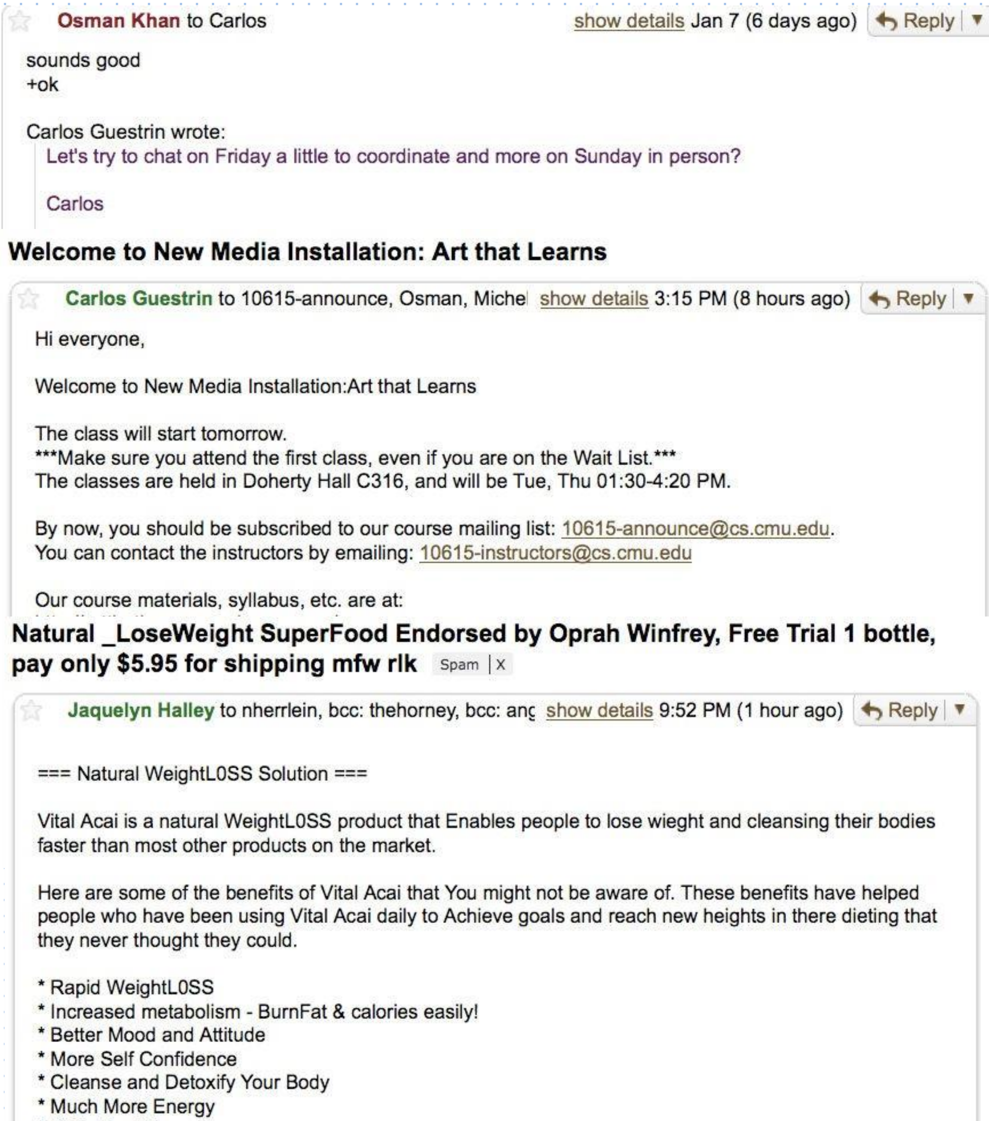
Data Science
Explore many models, build and tune hybrids
Understand empirical properties of models
Develop/use tools that can handle massive datasets
Take action!

What is Machine learning?

Machine learning is everywhere!

And you are probably already using it

Spam filtering


Osman Khan to Carlos show details Jan 7 (6 days ago) Reply

sounds good
+ok

Carlos Guestrin wrote:
Let's try to chat on Friday a little to coordinate and more on Sunday in person?

Carlos

Welcome to New Media Installation: Art that Learns

Carlos Guestrin to 10615-announce, Osman, Michel show details 3:15 PM (8 hours ago) Reply

Hi everyone,

Welcome to New Media Installation:Art that Learns

The class will start tomorrow.
Make sure you attend the first class, even if you are on the Wait List.
The classes are held in Doherty Hall C316, and will be Tue, Thu 01:30-4:20 PM.

By now, you should be subscribed to our course mailing list: 10615-announce@cs.cmu.edu.
You can contact the instructors by emailing: 10615-instructors@cs.cmu.edu

Our course materials, syllabus, etc. are at:

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only $5.95 for shipping mfw rlk Spam | X

Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang show details 9:52 PM (1 hour ago) Reply

=== Natural WeightL0SS Solution ===

Vital Acai is a natural WeightL0SS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

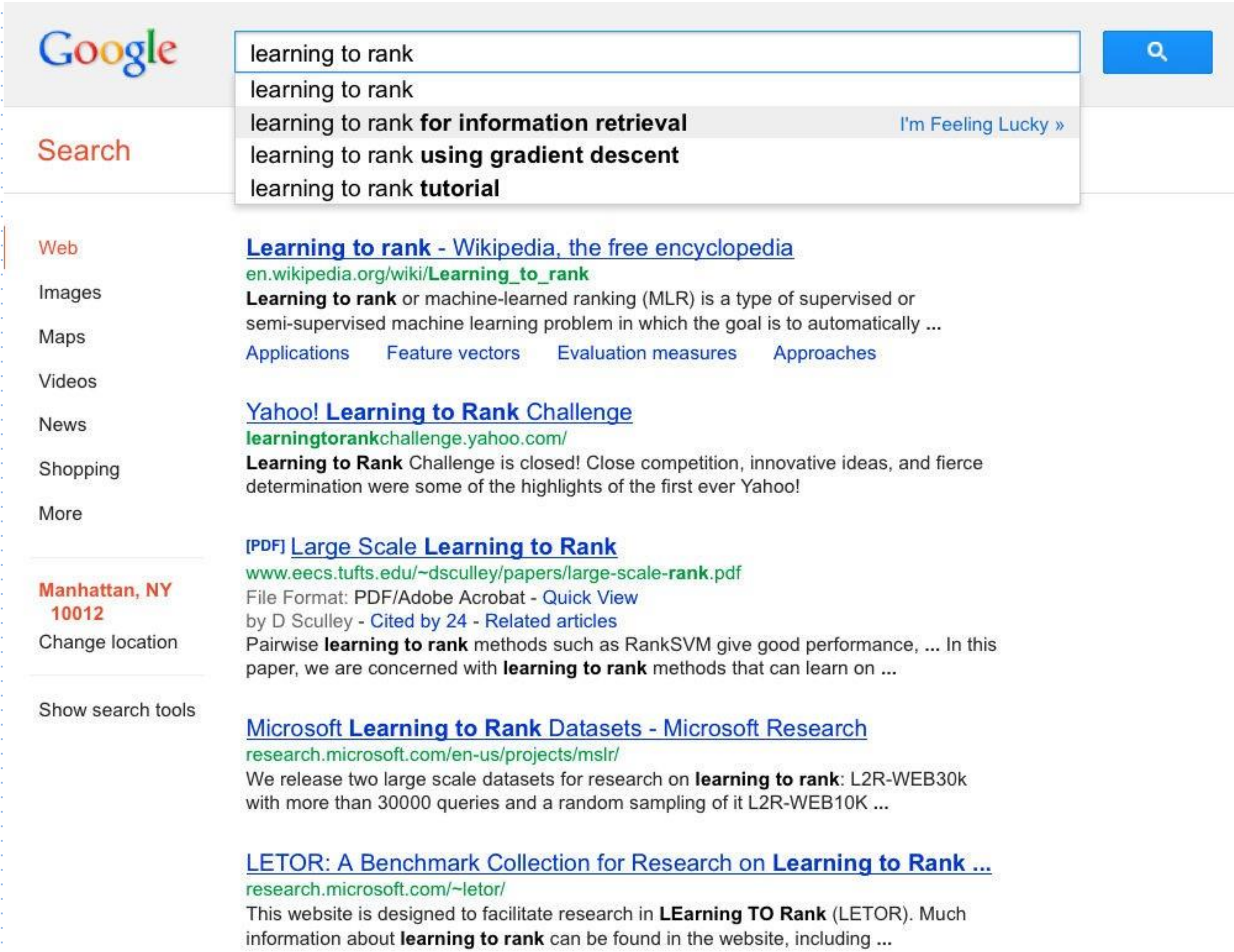
* Rapid WeightL0SS
* Increased metabolism - BurnFat & calories easily!
* Better Mood and Attitude
* More Self Confidence
* Cleanse and Detoxify Your Body
* Much More Energy

Spam

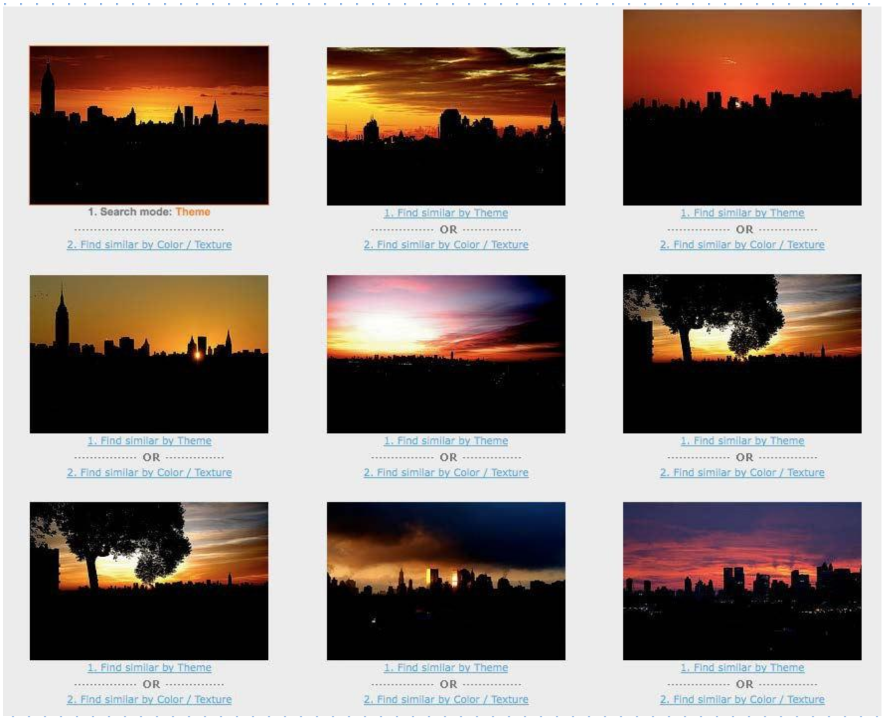
vs.

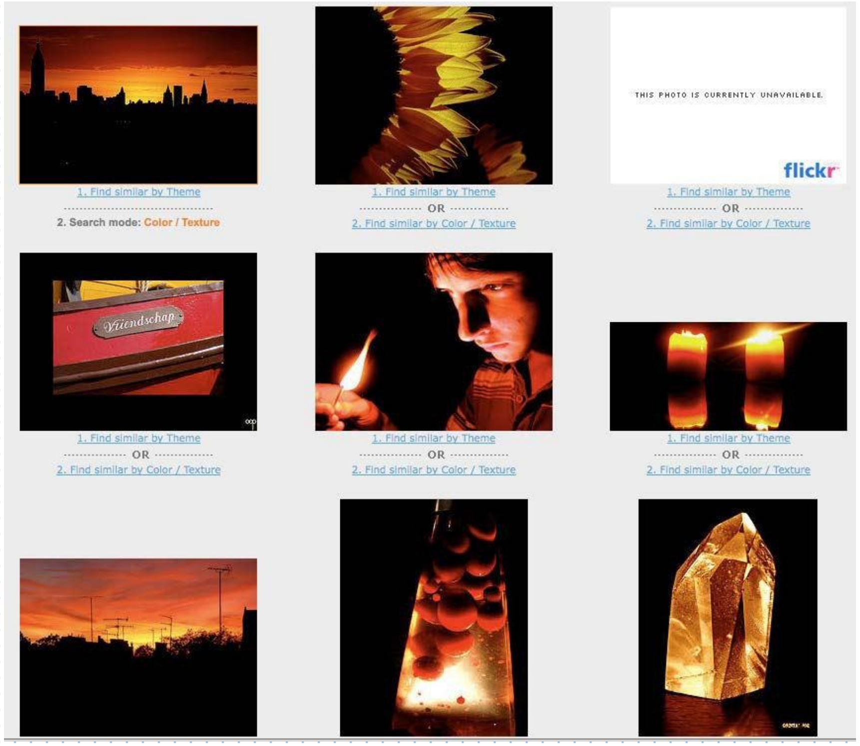
Not spam

Web search


Google
learning to rank
learning to rank
learning to rank for information retrieval
I'm Feeling Lucky »
learning to rank using gradient descent
learning to rank tutorial
Search
Web
Images
Maps
Videos
News
Shopping
More
Manhattan, NY 10012
Change location
Show search tools
Learning to rank - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Learning_to_rank
Learning to rank or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ...
Applications
Feature vectors
Evaluation measures
Approaches
Yahoo! Learning to Rank Challenge
learningtorankchallenge.yahoo.com/
Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!
[PDF] Large Scale Learning to Rank
www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf
File Format: PDF/Adobe Acrobat - Quick View
by D Sculley - Cited by 24 - Related articles
Pairwise learning to rank methods such as RankSVM give good performance, ... In this paper, we are concerned with learning to rank methods that can learn on ...
Microsoft Learning to Rank Datasets - Microsoft Research
research.microsoft.com/en-us/projects/mslr/
We release two large scale datasets for research on learning to rank: L2R-WEB30k with more than 30000 queries and a random sampling of it L2R-WEB10K ...
LETOR: A Benchmark Collection for Research on Learning to Rank ...
research.microsoft.com/~letor/
This website is designed to facilitate research in LEarning TO Rank (LETOR). Much information about learning to rank can be found in the website, including ...

Given image, find similar images

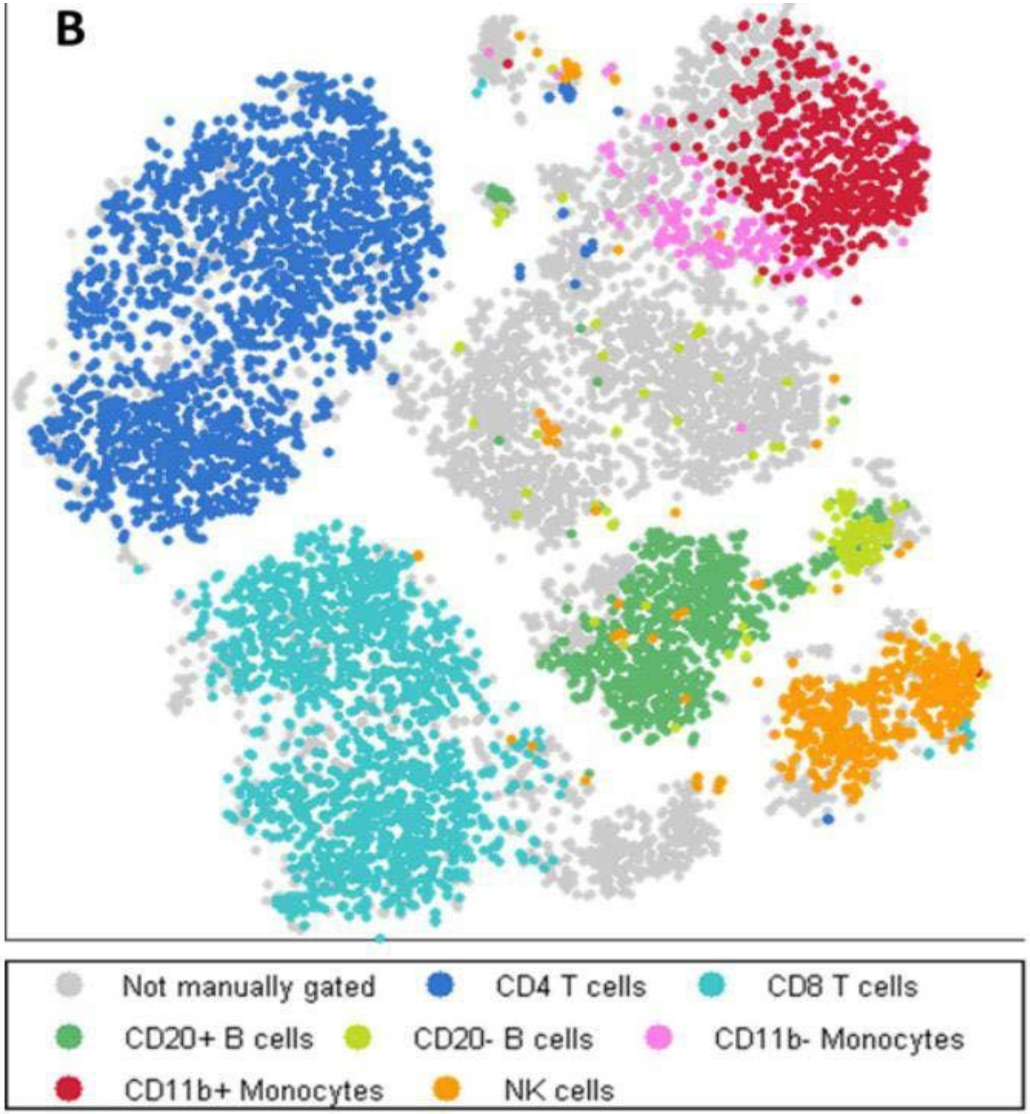




Recommendation systems

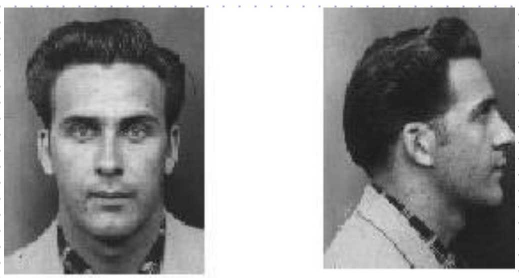
Stock market

Discovering new cancer subtypes



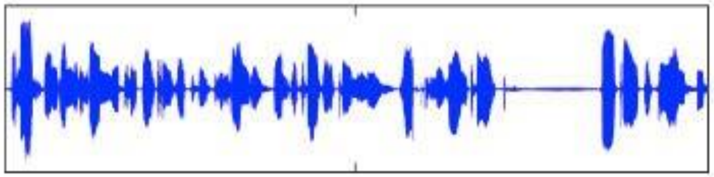
Beating human Go masters

Face recognition



Example training images
for each orientation

Speech recognition



Machine learning

- Grew out of work in AI
- New capability for computers Examples:
  - **Database mining**
    - Large datasets from growth of automation/web
    - E.g., Web click data, medical records, biology, engineering
  - **Applications can’t program by hand**
    - eg, Automonous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision
  - **Self-customizing programs**
    - e.g., Amazon, Netflix product recommendations
  - **Understanding human learning**
    - e.g. brain, real AI

Growth of Machine Learning

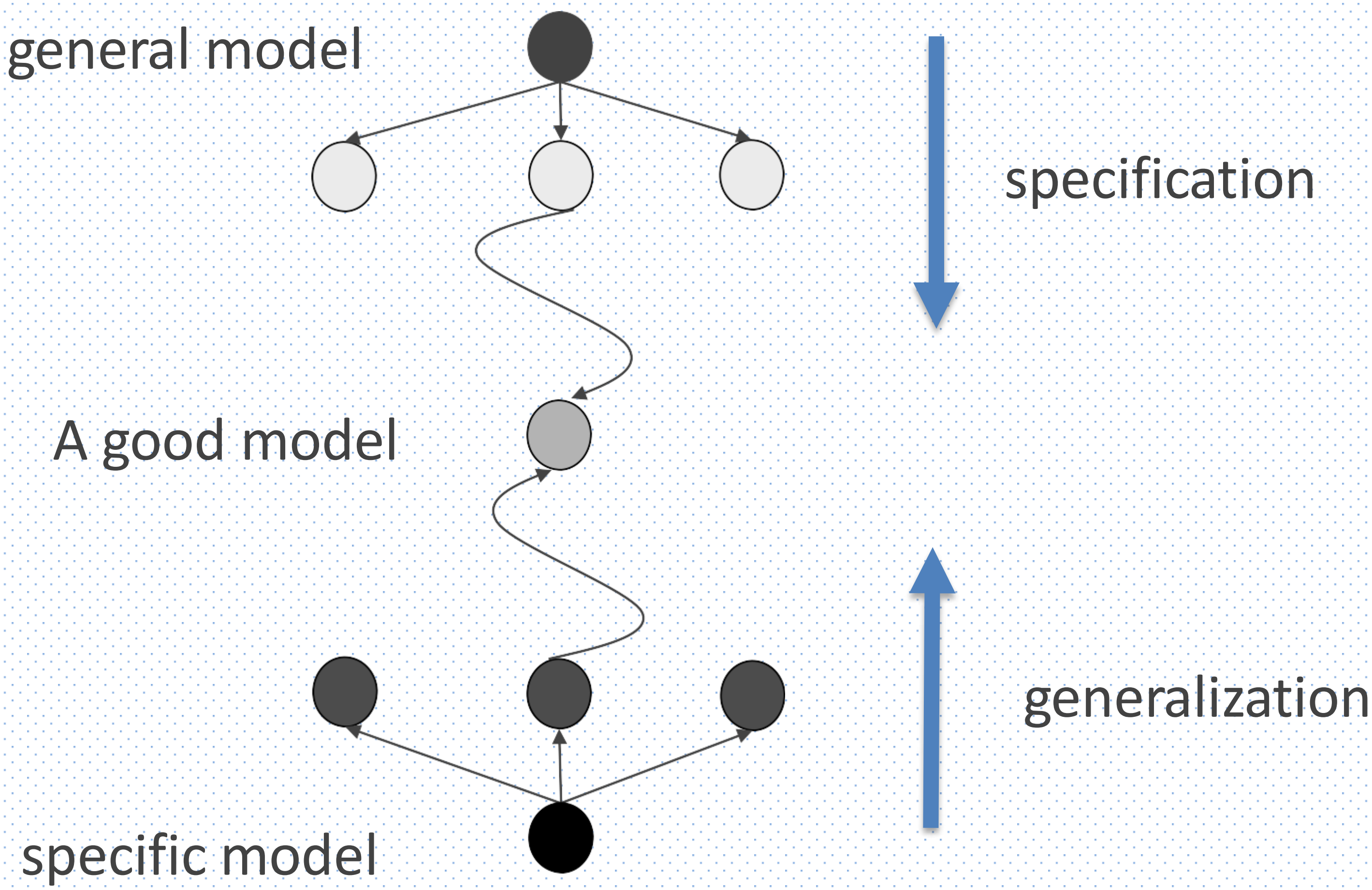
- **Machine learning is preferred approach to**
  - Speech recognition, Natural language processing
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
  - Sensor networks
  - …
- **This trend is accelerating**
  - Big data
  - Improved machine learning algorithms
  - Faster computers
  - Good open-source software

What is Machine Learning

- Arthur Samuel (1959):
  - *"the field of study that gives computers the ability to learn without being explicitly programmed."*

- Tom Mitchell (1998). Well-posed learning problem:
  - *"A computer program is said* ***to learn*** *from experience E with respect to* ***some class*** *of tasks T and performance* ***measure*** *P, if its performance at tasks in T, as measured by P,* ***improves with experience*** *E."*

What is learning?



Some Challenges

- massive data (500k users, 20k movies, 100m ratings)
- curse of dimensionality (very high-dimensional problem)
- missing data (90% of data missing)
- extremely complicated set of factors (e.g. that affect people’s ratings of movies: actors, directors, genre, ...)
- need to avoid overfitting (test data vs. training data)
- Unbalanced data (95% vs. 5% class distribution)

Turing Award

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing

3

Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By James Vincent | Mar 27, 2019, 6:02am EDT

f SHARE

From left to right: Yann LeCun | Photo: Facebook; Geoffrey Hinton | Photo: Google; Yoshua Bengio | Photo: Botler AI

Supervised learning

- Definition
- Formalizing supervised learning
  - Instance space and features
  - Label space
  - Hypothesis space

Supervised learning

We are ***given a data set*** and already know what our ***correct output*** should look like, having the idea that there is a ***relationship*** between the input and the output.

Regression & Classification

- Supervised learning problems are categorized into "regression" and "classification" problems.
  - In a regression problem, we are trying to predict results within a **continuous output**, meaning that we are trying to map input variables to some continuous function.
  - In a classification problem, we are instead trying to predict results in a **discrete output**. In other words, we are trying to map input variables into discrete categories.

Example

- *Given data about the size of houses on the real estate market, try to predict their price*. Price as a function of size is a ***continuous output***, so this is ***a regression problem***.

- How to ***turn*** this example into a ***classification*** problem?

  – Make our output about whether the house "sells for more or less than the asking price."

  – Here we are classifying the houses based on price into two discrete categories.

The badges game

- Attendees of the 1994 conference on *Computational Learning Theory* received conference badges labeled + or –

- Only one person (Haym Hirsh) knew the function that generated the labels

- Depended *only* on the attendee's name

- The task for the attendees: Look at as many examples as you want in the conference and find the unknown function

Let’s play

Name	Label
Claire Cardie	-
Peter Bartlett	+
Eric Baum	?
Haym Hirsh	?
Shai Ben-David	?
Michael I. Jordan	?

How were the labels generated?

Let’s play

Name	Label
Claire Cardie	-
Peter Bartlett	+
Eric Baum	
Haym Hirsh	+
Shai Ben-David	
Michael I. Jordan	-

What is the label for “Peyton Manning”?

What about “Eli Manning”?

Let’s play

Name	Label
Claire Cardie	-
Peter Bartlett	+
Eric Baum	
Haym Hirsh	+
Shai Ben-David	
Michael I. Jordan	-

How were the labels generated?

$$If\ length\ of\ first\ name\ \leq 5\ than + else\ -$$

Questions

1. Are you sure you got the correct function?

2. How did you arrive at it?

3. Learning issues:
   - Is this prediction or just modeling data?
   - How did you know that you should look at the letters?
   - All words have a length. Background knowledge.
   - What "learning algorithm" did you use?

Instances and Labels

Running example: Automatically tag news articles

China Open: Maria Sharapova beats Ekaterina Makarova in three sets

Share

Maria Sharapova will next play Romanian world two Simona Halep

Maria Sharapova beat fellow Russian Ekaterina Makarova in three sets to reach the China Open third round.

The five-time Grand Slam champion, ranked 104th in the world, beat world number 33 Makarova 6-4 4-6 6-1.

Wildcard entry Sharapova will next play Romanian world two Simona Halep, who was leading 6-2 2-1 when Slovak opponent Magdalena Rybarikova retired with illness.

Halep **lost** to Sharapova in the US Open first round in August.

Two-time Wimbledon champion Petra Kvitova beat America's Varvara Lepchenko 6-4 6-4 in Beijing.

The Czech 12th seed will play Denmark's world number six Caroline Wozniacki in the next round.

Share

Top Stories

Bale out of Wales' crucial World Cup qualifiers

Chelsea lose striker Morata to hamstring injury

Starc warns England to expect a pace barrage

Related to this story

Sharapova wins epic opener in Beijing

Unwell Muguruza retires at China Open

Robson safe after Vegas shooting

From Around the Web

Tennis Icon John McEnroe Sells Southampton Home...

Pete Sampras Lists His Los Angeles home at $8.9 Million

Athens Taxi Driver Got Paid €50 In Bitcoins In 2010, Now He's A...

Play this for 1 minute and see why everyone is addicted

An instance of a news article that needs to be classified

Sports

A label

Instances and Labels

Running example: Automatically tag news articles

Mapped by the classifier to

Sports

Business

Politics

Entertainment

***Instance Space**: All possible news articles*

***Label Space:** All possible labels*

Instances and Labels

X : Instance Space

The set of examples that need to be classified

$$y = f(x)$$

The goal of learning: Find this target function

Y: Label Space

The set of all possible labels

e.g: The set of all possible names, documents, sentences, images, emails etc.

e.g: {Spam, Not-Spam}, {+,-}, etc.

Instances and Labels

X : Instance Space

The set of examples that need to be classified

Target function
$y = f(x)$

Y: Label Space

The set of all possible labels

Learning algorithm only sees examples of the function f in action

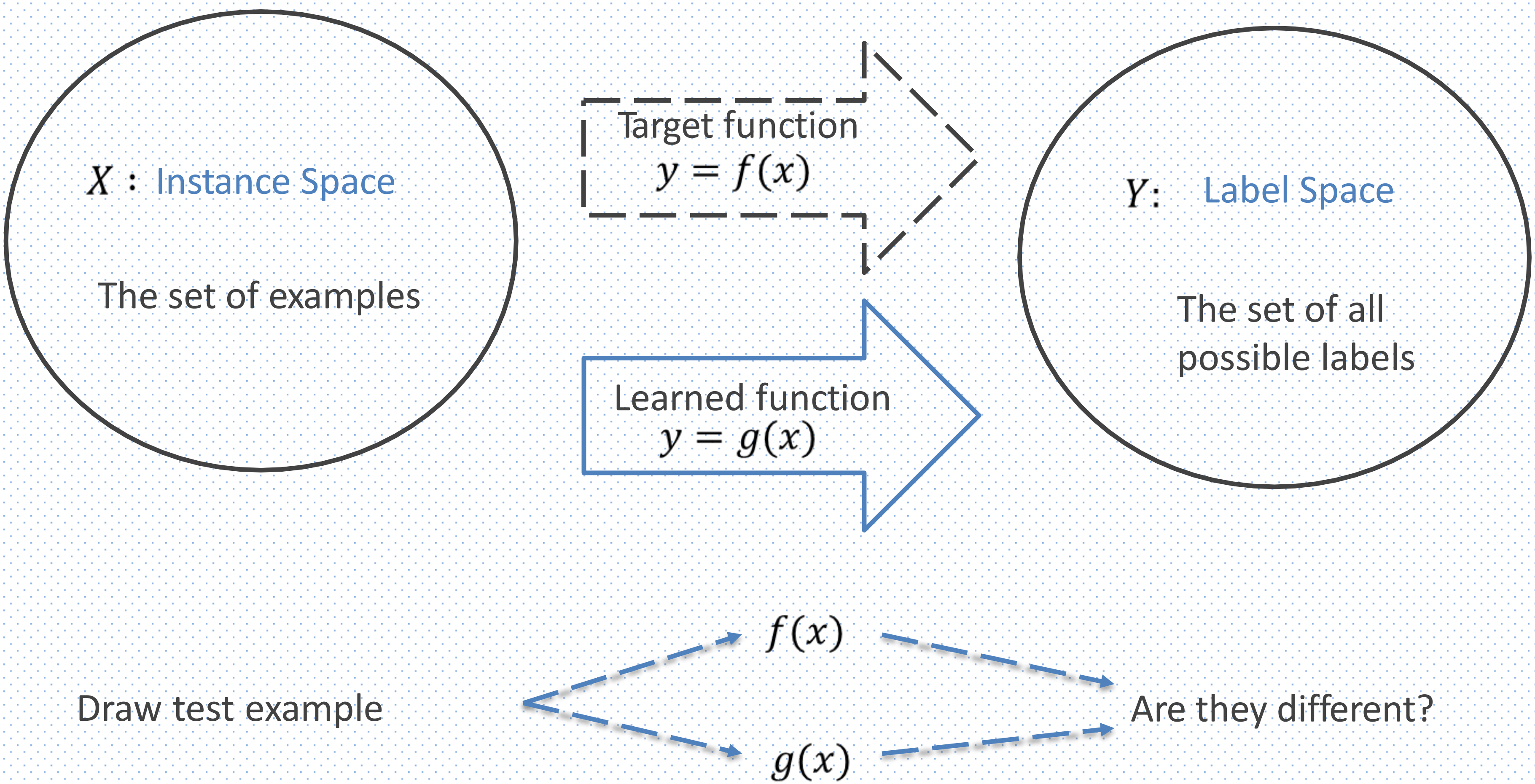
$(x_1, f(x_1))$
$(x_2, f(x_2))$
$(x_3, f(x_3))$
…
$(x_n, f(x_n))$

Labeled **training data**

Learning algorithm

A learned function $g: X \rightarrow Y$

Supervised learning: Training



Apply the model to many test examples and compare to the target’s prediction

Supervised learning: General setting

- Given: Training examples of the form $< x_1, f(x_1)>$
- The function f is an unknown function

- Typically the input x is represented in a feature space
  - Example: $x \in \{0,1\}$ or $x \in R^n$
  - A deterministic mapping from instances in your problem (eg: news articles) to features
- For a training example x, the value of $f(x)$ is called its label
- Goal: Find a good approximation of f
- The label determines the kind of problem we have

  - Binary classification: $f(x) \in \{0,1\}$
  - Multiclass classification: $f(x) \in \{1,2,3,\ldots,K\}$
  - Regression: $f(x) \in R$

Binary classification

Where the label space consists of two elements

- Spam filtering
  - Is an email spam or not?
- Recommendation systems
  - Given user's movie preferences, will she like a new movie?
- Malware detection
  - Is an Android app malicious?
- Time series prediction
  - Will the future value of a stock increase or decrease with respect to its current value?

On using supervised learning

1. What is our instance space?

   What are the inputs to the problem? What are the **features**?

2. What is our label space?

   What is the **prediction task**?

3. What is our hypothesis space?

   What function should the **learning algorithm** search over?

4. What is our learning algorithm?

   How do we learn from **labeled data**?

5. What is our loss function or evaluation metric?

   What is **success**?

Instances and Labels

X : Instance Space

The set of examples that need to be classified

The goal c
Find this t

lea

Eg: The set of all possible names, documents, sentences, images, emails etc.

Designing an appropriate feature representation of the instance space is crucial

Instances $x \in X$ are defined by features/attributes

Example: Boolean features

- Does the email contain the word “free”?

Example: Real valued features

- What is the height of the person?
- What was the stock price yesterday?

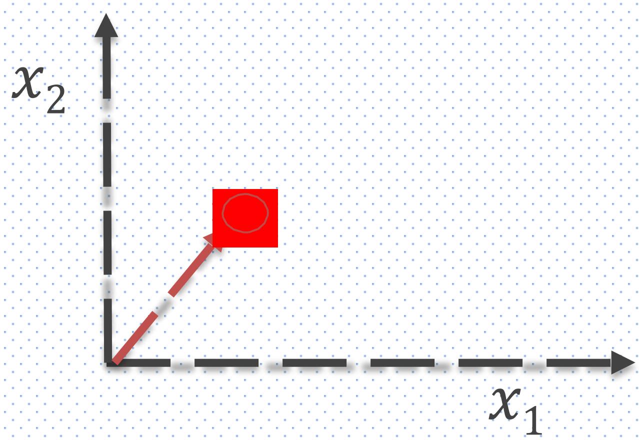
Instances as feature vectors

Feature functions - feature extractors

- Deterministic (for the most part)
- Convert the examples to a collection of attributes
  - Very often easy to think of them as vectors
- Important part of the design of a learning based solution

Instances as feature vectors

- Features functions convert inputs to vectors
  - Fixed mapping
- The instance space X ***is*** a N-dimensional vector space (e.g. $\mathcal{R}^N$ or $\{0,1\}^N$)
  - *Each dimension is one feature*
- Each $x \in X$ is a feature vector
  - Each $x = [x_1, x_2, \ldots, x_N]$ is a point in the vector space



Feature functions produce feature vectors

- When designing feature functions, think of them as templates
  - Feature: *"The second letter of the name"*
    - N**a**oki $a \rightarrow [1\ 0\ 0\ 0\ \ldots]$
    - A**b**e $b \rightarrow [0\ 1\ 0\ 0\ \ldots]$
    - M**a**nning $a \rightarrow [1\ 0\ 0\ 0\ \ldots]$
    - S**c**rooge $c \rightarrow [0\ 0\ 1\ 0\ \ldots]$

Question: What is the length of this feature vector?

26 (One dimension per letter)

  - Feature: *"The length of the name"*
    - Naoki $\rightarrow 5$
    - Abe $\rightarrow 3$

Good features are essential

- Good features decide how well a task can be learned
  - E.g.: A bad feature function the badges game
    - "Is there a day of the week that begins with the last letter of the first name?"
- Much effort goes into designing features
  - Or maybe learning them
- Will touch upon general principles for designing good features
  - But feature definition largely domain specific
  - Comes with experience

On using supervised learning

We should be able to decide:

✓ What is our instance space?

What are the inputs to the problem? What are the features?

2. What is our label space?

   What is the prediction task?

3. What is our hypothesis space?

   What function should the learning algorithm search over?

4. What is our learning algorithm?

   How do we learn from labeled data?

5. What is our loss function or evaluation metric?

   What is success?

The Label Space Y

X : Instance Space

The set of examples that need to be classified

$$y = f(x)$$

Y: Label Space

The set of all possible labels

The goal of learning: Find this target function learning is search over functions

Eg: The set of all possible names, documents, sentences, images, emails etc.

Eg: {Spam, Not-Spam}, {+,-}, etc.

The Label Space Y

- *Classification:* The outputs are categorical
  - Binary classification: Two possible labels
  - Multiclass classification: K possible labels
  - Structured classification: Graph valued outputs

The label Space Y

- The output space can be numerical
  - Regression
    - Y is the set (or a subset) of real numbers
  - Ranking
    - Labels are ordinal
    - That is, there is an ordering over the labels
    - Eg: A Yelp 5-star review is only slightly different from a 4-star review, but very different from a 1-star review

On using supervised learning

We should be able to decide:

✓ What is our instance space?

What are the inputs to the problem? What are the features?

✓ What is our label space?

What is the prediction task?

3. What is our hypothesis space?

   What function should the learning algorithm search over?

4. What is our learning algorithm?

   How do we learn from labeled data?

5. What is our loss function or evaluation metric?

   What is success?

The Hypothesis Space

X : Instance Space

The set of examples that need to be classified

$$y = f(x)$$

Y: Label Space

The set of all possible labels

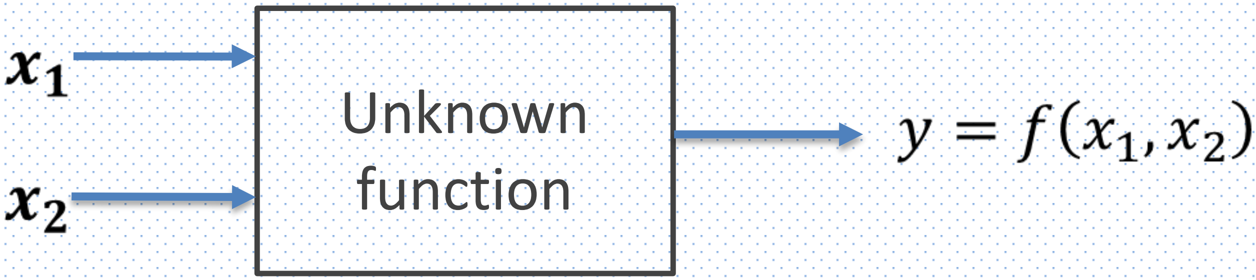
The goal of learning:
Find this target function

e.g: The set of all possible names, documents, sentences, images, emails etc.

e.g: {Spam, Not-Spam}, {+,-}, etc.

The hypothesis space is the set of functions we consider for this search

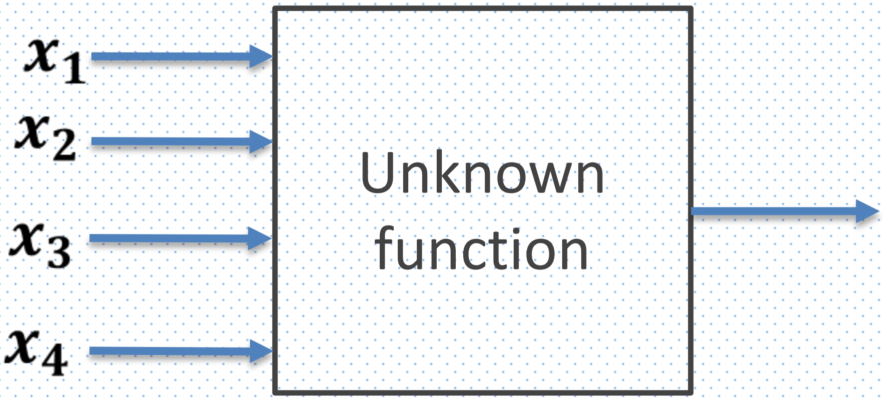
Example of search over functions



x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Can you learn this function? What is it?

Example of search over functions



$$y = f(x_1, x_2, x_3, x_4)$$

x_1	x_2	x_3	x_4	y
0	0	1	0	0
0	1	0	0	0
0	0	1	1	1
1	0	0	1	1
0	1	1	0	0
1	1	0	0	0
0	1	0	1	0

Can you learn this function? what is it?

Is learning possible at all?

- There are $2^{16} = 65536$ possible Boolean functions over 4 inputs
  - Why? There are 16 possible outputs. Each way to fill these 16 slots is a different function, giving 2^{16} functions
- We have seen only 7 outputs
- *How could we possibly know the rest without seeing every label?*
  - Think of an adversary filling in the labels every time you make a guess at the function

x_1	x_2	x_3	x_4	y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	?
1	0	0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	?
1	1	0	0	0
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

Is learning possible at all?

- There are $2^{16} = 65536$ possible Boolean functions over 4 inputs
  - Why? There are 16 possible outputs. Each way to fill these 16 slots is a different function, giving 2^{16} functions
- We h

How could we possibly learn anything?

- *How could we possibly know the rest without seeing every label?*
  - Think of an adversary filling in the labels every time you make a guess at the function

x_1	x_2	x_3	x_4	y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
		1	0	0
		1	1	?
		0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	?
1	1	0	0	0
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

Solution: Restrict the search space

- A *hypothesis space* is the set of all possible functions we consider :
  – We were looking at the space of ***all*** Boolean functions
  – Instead choose a hypothesis space that is smaller than the space of all functions

- How do we pick a hypothesis space?
  – Using some prior knowledge (or by guessing)

- What if the hypothes is : «*space is so small that nothing in it agrees with the data*»?

  – **We need a hypothesis space that is flexible enough**

Example: Hypothesis space 2

- m-of-n rules
  - Pick a subset with n variables.
  - Y = 1 if at least m of them are 1

Example:

If at least 2 of $\{x_1\ x_3\ x_4\}$ are 1, then the output is 1, otherwise, the output is 0.

Is there a consistent hypothesis in this space?

x_1	x_2	x_3	x_4	y
0	0	1	0	0
0	1	0	0	0
0	0	1	1	1
1	0	0	1	1
0	1	1	0	0
1	1	0	0	0
0	1	0	1	0

Views of learning

- *Learning is the removal of remaining uncertainty*
    - If we knew that the unknown function is a simple conjunction, we could use the training data to figure out which one it is
- *Requires guessing a good, small hypothesis class*
    - And we could be wrong
    - We could find a consistent hypothesis and still be incorrect with a new example!

On using supervised learning

We should be able to decide:

✓ What is our instance space?

What are the inputs to the problem? What are the features?

✓ What is our label space?

What is the prediction task?

✓ What is our hypothesis space?

What function should the learning algorithm search over?

4. What is our learning algorithm?

   How do we learn from labeled data?

5. What is our loss function or evaluation metric?

   What is success?

Resources

- G.Petasis and A.Krithara, "M.Sc. Course in Data Science Lecture 1", Univ. of Peleoponese – NCSR "Demokritos"
- Hal Daumé, A Course in Machine Learning
- Shai Shalev-Shwartz and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms
- Christopher Bishop, Pattern Recognition and Machine Learning. Springer 2007.

Preparing and Architecting for Machine Learning(*)

Key Findings

- Machine learning (ML) — *a subset of artificial intelligence* (AI) — is more than a technique for analyzing data. It's a system that is *fueled by data*, with the *ability to learn* and *improve* by using algorithms that provide *new insights without being explicitly programmed to do so*.

- *Preparing data* for ML pipelines is challenging when end-to-end *data and analytic architectures* are not refined to interoperate with underlying analytic platforms.

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

Key Findings

- ML is best-suited for dealing with *big data*. Organizations overwhelmed with data are using multiple ML frameworks to *increase operational* efficiencies and achieve *greater business* agility.

- *Technical professionals* are using machine learning to add elements of *intelligence* to software development and IT operations (DevOps) to *gain operational* efficiencies.

- The ML *compute and storage cluster* — which is the heart of the ML system — will *vary* based on learning method, learning application and need for automation.

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

Recommendations

- To modernize your organization's *business intelligence* and *analytics* capabilities to support machine learning:

- Update the data organization layer in *end-to-end analytics* architectures to *support data preparation* for ML algorithms.

- Incorporate a development life cycle that supports *learning models* when the organization plans to aggressively build *custom ML algorithms and applications*.

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

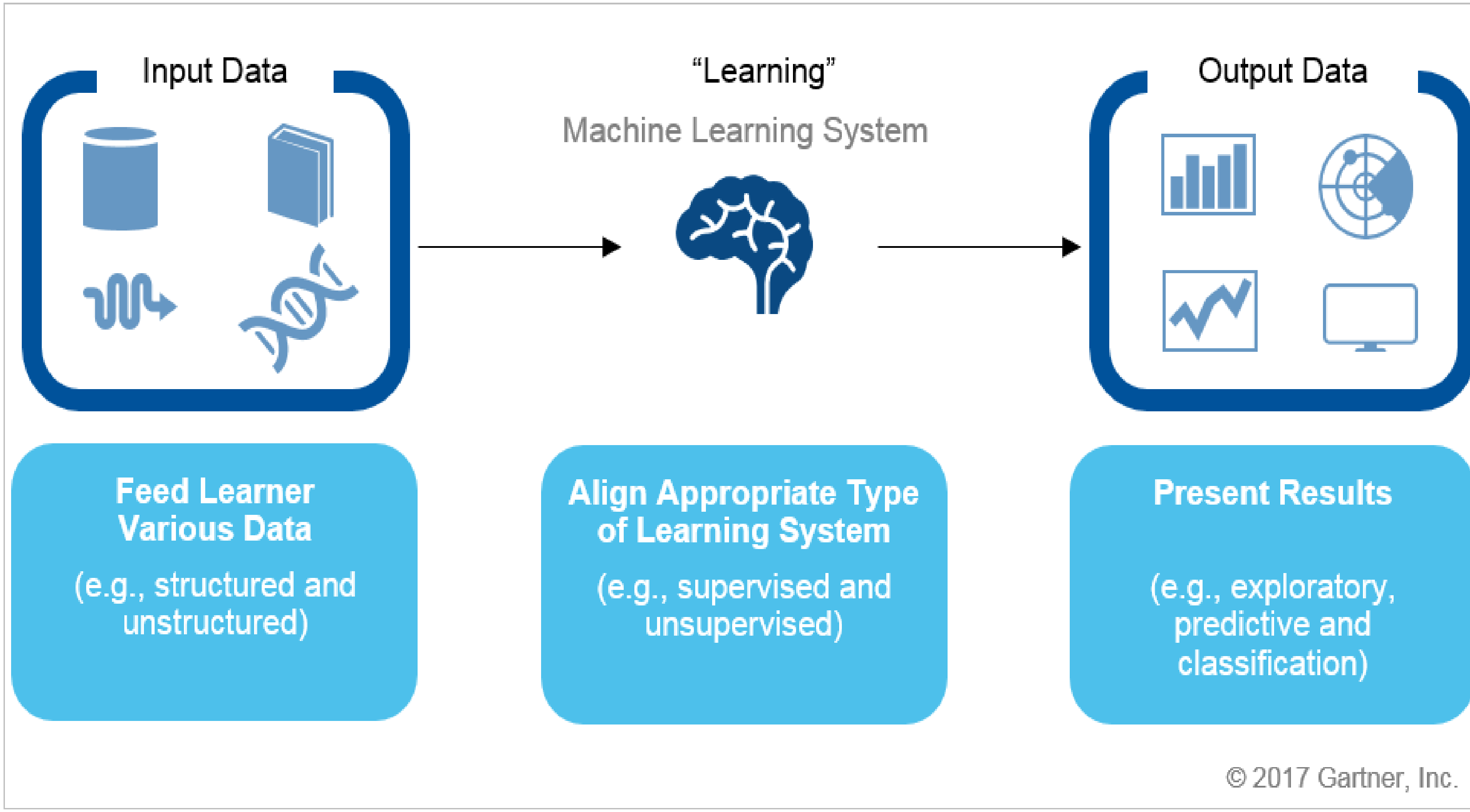
Recommendations

- *Choose an ML platform* that supports and interoperates with multiple ML frameworks when the organization plans to leverage service providers or commercial off-the-shelf solutions. As AI and ML gain momentum, *more frameworks will be packaged with solutions and service providers*.

- *Focus on storage and compute clusters* to support machine learning capabilities. Choose the *public cloud* when you don't have the appropriate staff for engineering infrastructures for ML. The cloud is a great place for designing ML capabilities because of its elastic capabilities for *scaling algorithms*.

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)



The Basics of Machine Learning Technology

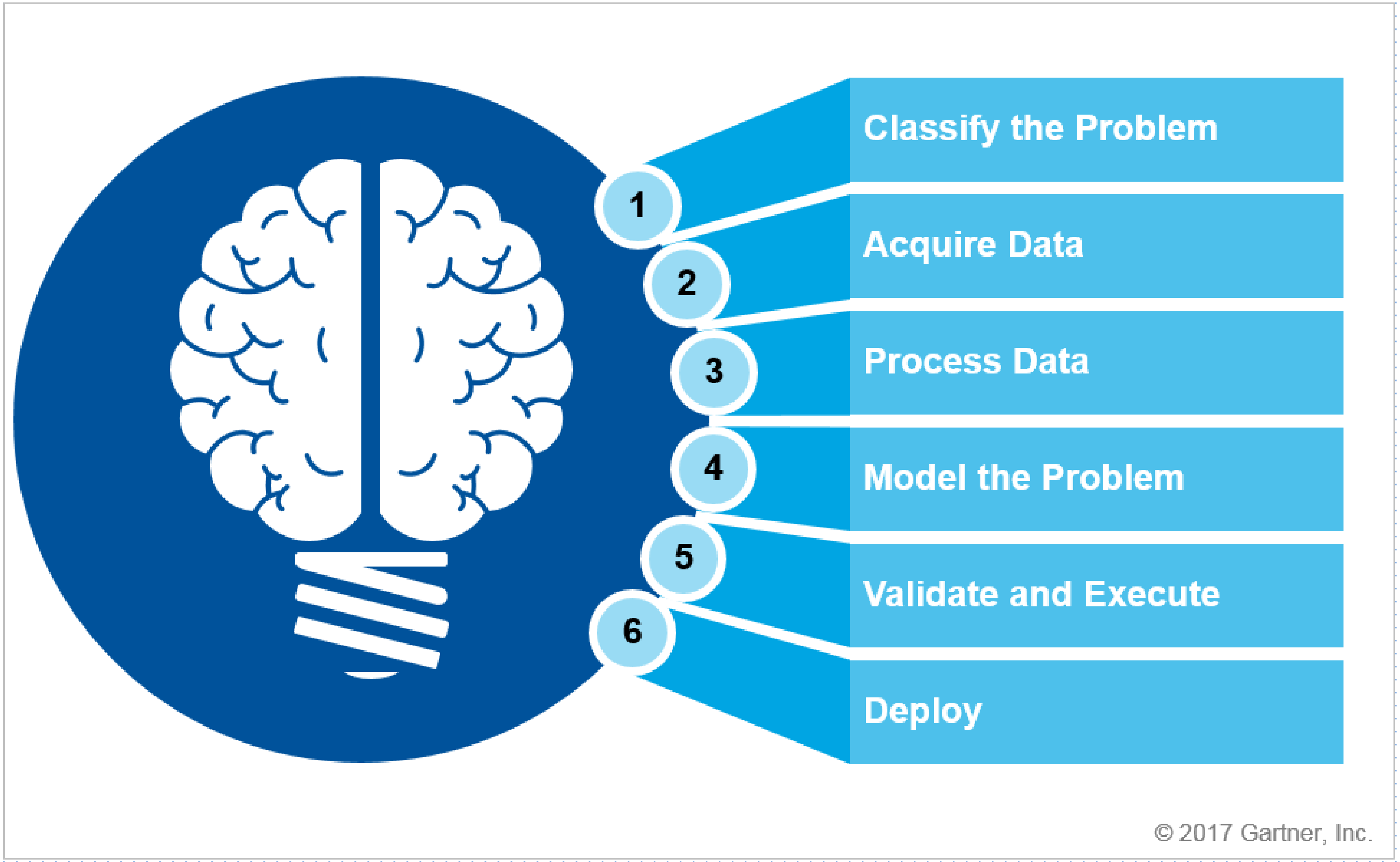
(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

Type	Model/ Algorithm or Task	Description	Usage Examples in Business
Supervised	Neural network	Computations are structured in terms of interconnected groups, much like the neurons in a brain. Neural networks are used to model complex relationships between inputs and outputs to find patterns in data or to capture a statistical structure among variables with unknown relationships. They may also be used to discover unknown inputs (unsupervised).	- **Predicting financial results** - **Fraud detection**
Supervised	Classification and/or regression	Computations are structured in terms of categorized outputs or observations based on defined classifications. Classification models are used to predict new outputs based on classification rules. Regression models are generally used to predict outputs from training data.	- **Spam filtering** - **Fraud detection**
Supervised	Decision tree	Computations are particular representations of possible solutions to a decision based on certain conditions. Decision trees are great for building classification models because they can decompose datasets into smaller, more manageable subsets.	- **Risk assessment** - **Threat management systems** - **Any optimization problem where an exhaustive search is not feasible**
Unsupervised	Cluster analysis	Computations are structured in terms of groups of input data (clusters) based on how similar they are to one another. Cluster analysis is heavily used to solve exploratory challenges where little is known about the data.	- **Financial transactions** - **Streaming analytics in IoT** - **Underwriting in insurance**
Unsupervised	Pattern recognition	Computations are used to provide a description or label to input data, such as in classification. Each input is evaluated and matched based on a pattern identified. Pattern recognition can be used for supervised learning as well.	- **Spam detection** - **Biometrics** - **Identity management**
Unsupervised	**Association rule learning**	**Computations are rule-based in order to determine the relationship between different types of input or variables and to make predictions.**	- **Security and intrusion detection** - **Bioinformatics**

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)



Stages of the Machine Learning Process

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

Machine Learning Taxonomy for Use Cases (Cheat Sheet)

Exploratory
(Used Often in Applied Machine Learning)

Business question: What business challenge would you like to explore?

Example 1: What other factors contribute to a consumer's default on a bank loan that might help better predict creditworthiness?

Goal: Determine patterns in data/groupings

Predictive
(Used Often in Traditional Machine Learning)

Business question: What business challenge would you like to predict?

Example 1: When will our insurance claim occur, and what new factors will drive the next occurrence? Or, what will our customers buy next?

Goal: Prediction

Unsupervised Learning
(No Prior Knowledge of Output; Used to Classify Future Output)

Common ML algorithms include:

- Clustering
- K-means
- Genetic algorithms
- And more

Supervised Learning
(Based on Training Data; Very Familiar With the Data; Knowledge of Output)

Common ML algorithms include:

- Neural networks
- Decision trees
- Bayesian networks
- And more

Machine Learning Guide

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

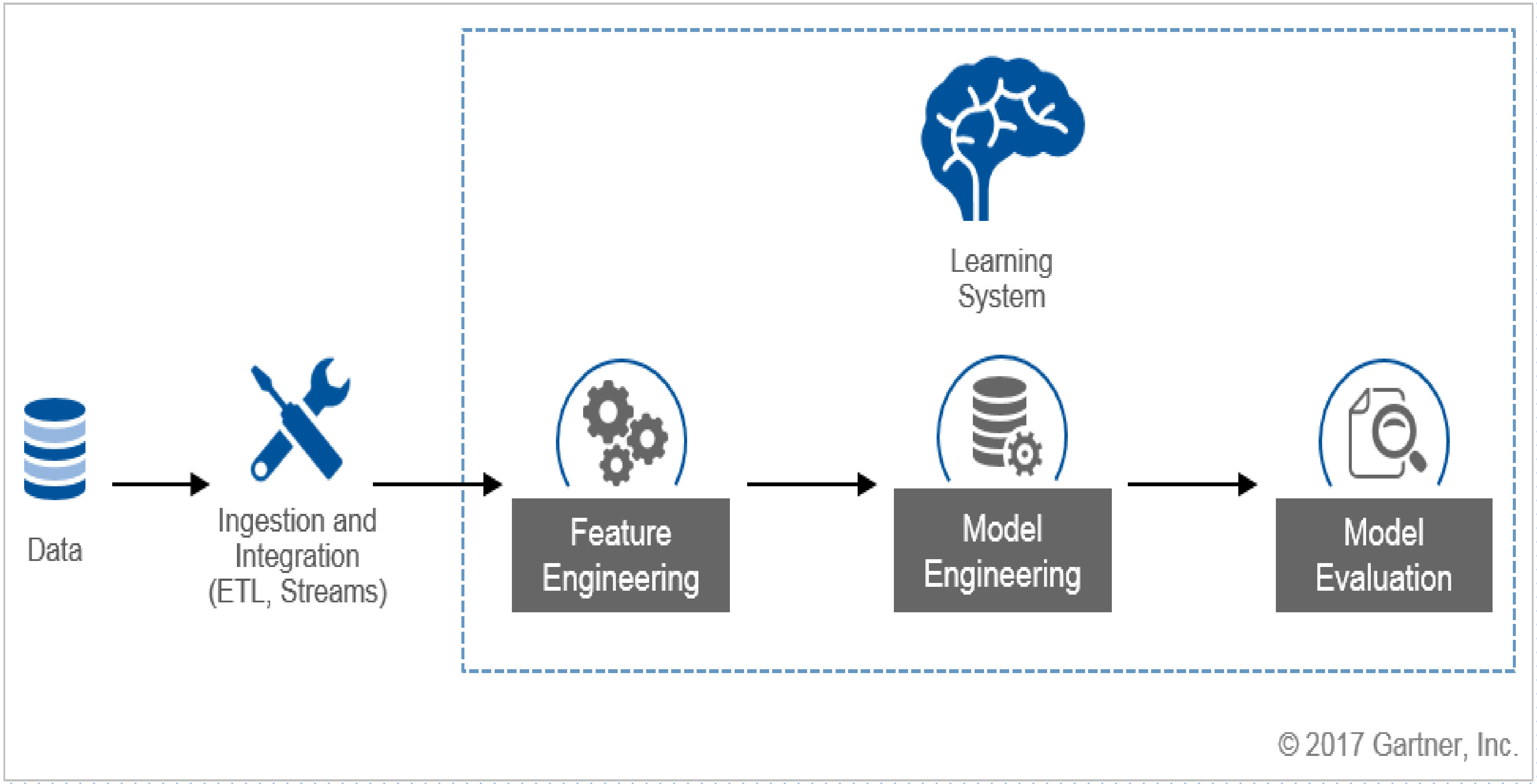
Preparing and Architecting for Machine Learning(*)

Task (Proportion of Effort)	Subtasks	Business	Stakeholder: Data Scientist	Stakeholder: IT/ Operations
1. Problem Understanding (5% to 10%)	a) Determine Objective	X	X	
	b) Define Success Criteria	X	X	
	c) Assess Constraints	X	X	X
2. Data Understanding (10% to 25%)	a) Assess Data Situation	X	X	X
	b) Obtain Data (Access)		X	X
	c) Explore Data	X	X	X
3. Data Preparation (20% to 40%)	a) Filter Data		X	X
	b) Clean Data		X	X
	c) Feature Engineering	X	X	
4. Modeling (20% to 30%)	a) Select Model Approach		X	
	b) Build Models		X	
5. Evaluation of Results (5% to 10%)	a) Select Model		X	
	b) Validate Model		X	
	c) Explain Model	X	X	
6. Deployment (5% to 15%)	a) Deploy Model		X	X
	b) Monitor and Maintain	X	X	X
	c) Terminate	X	X	X

Data Science Life Cycle

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

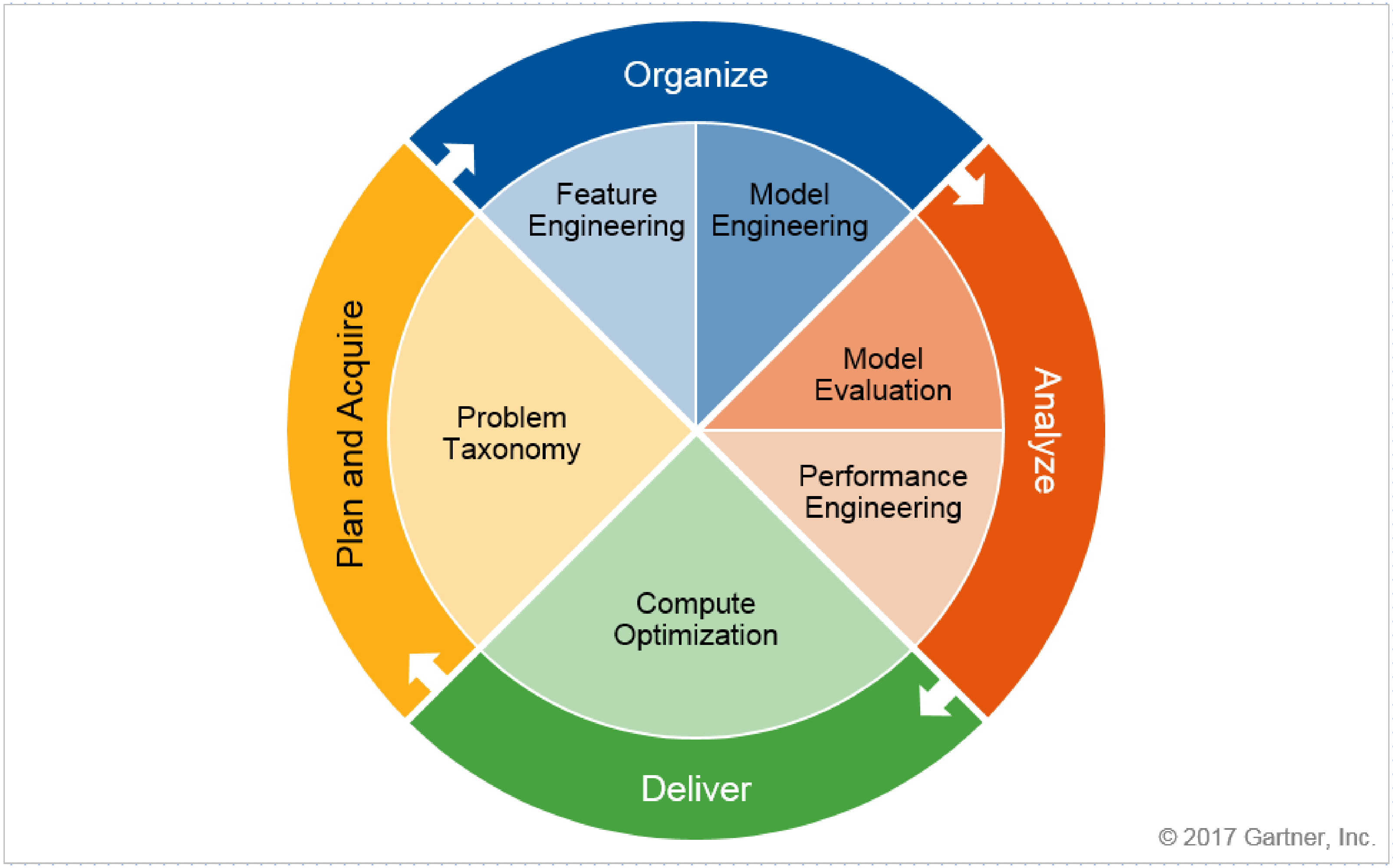
Preparing and Architecting for Machine Learning(*)



Life Cycle for Developing Learning Systems

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

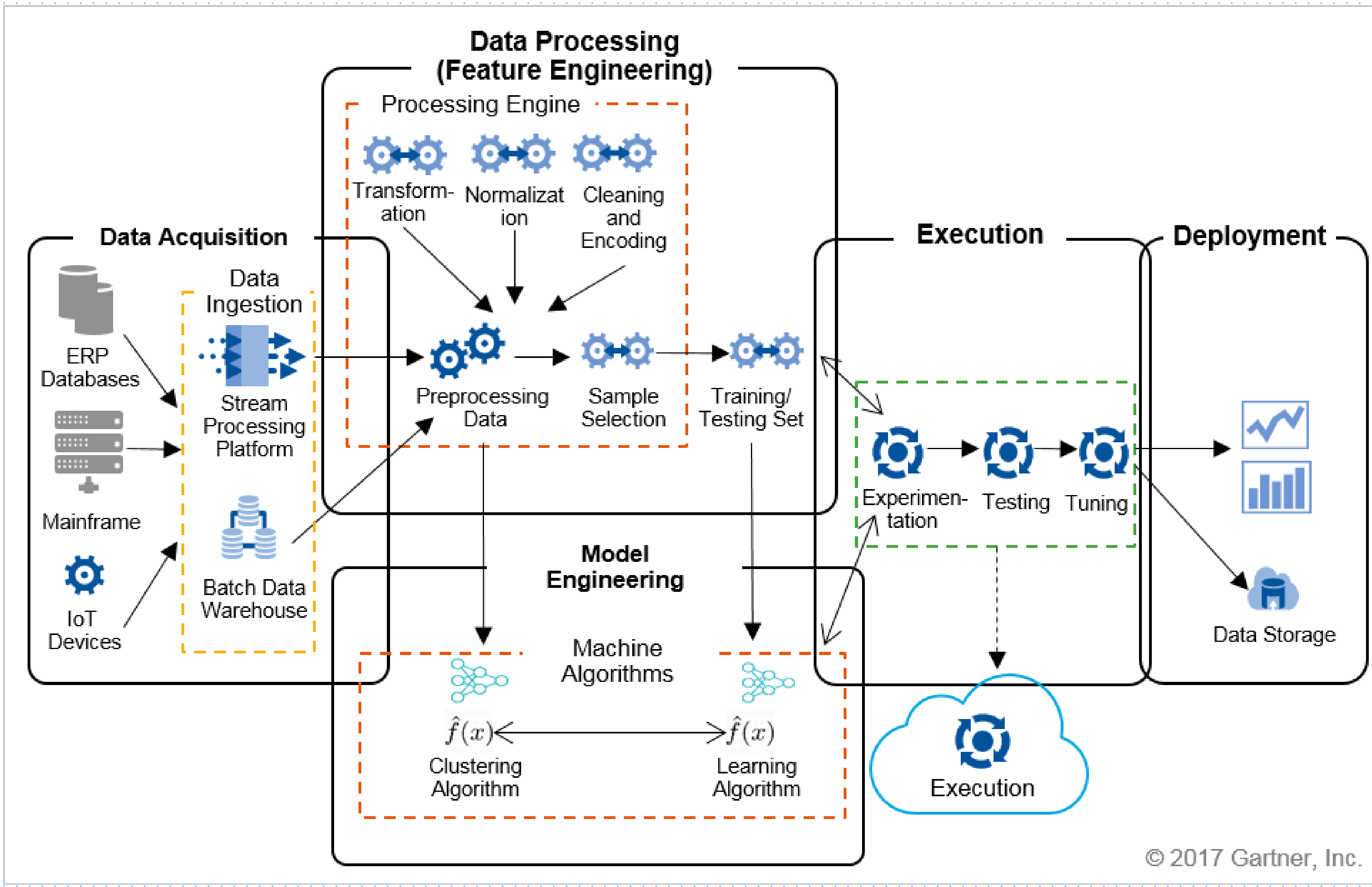
Preparing and Architecting for Machine Learning(*)



Continuous ML Model and Control Framework

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

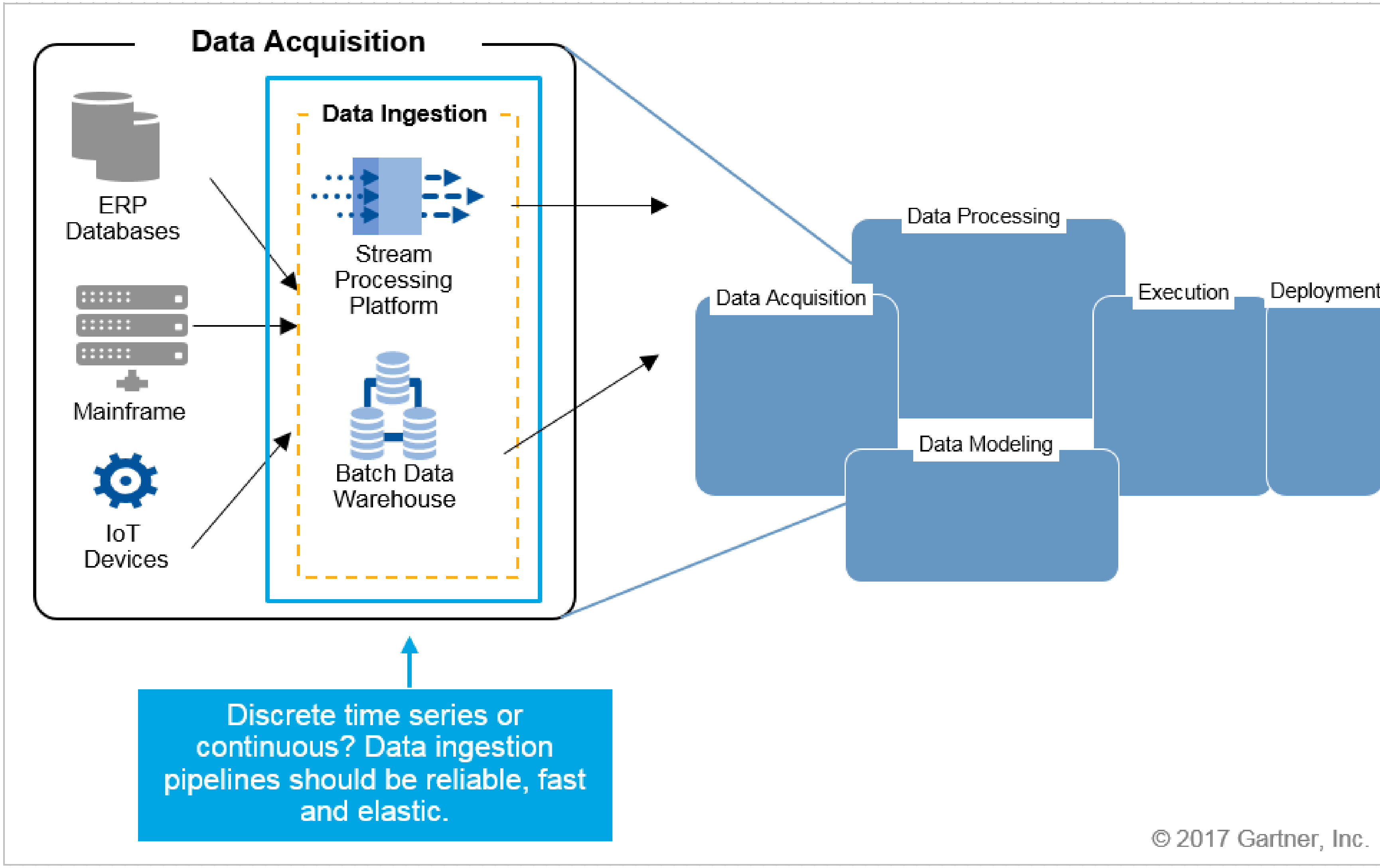
Preparing and Architecting for Machine Learning(*)



Machine Learning Architecture

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

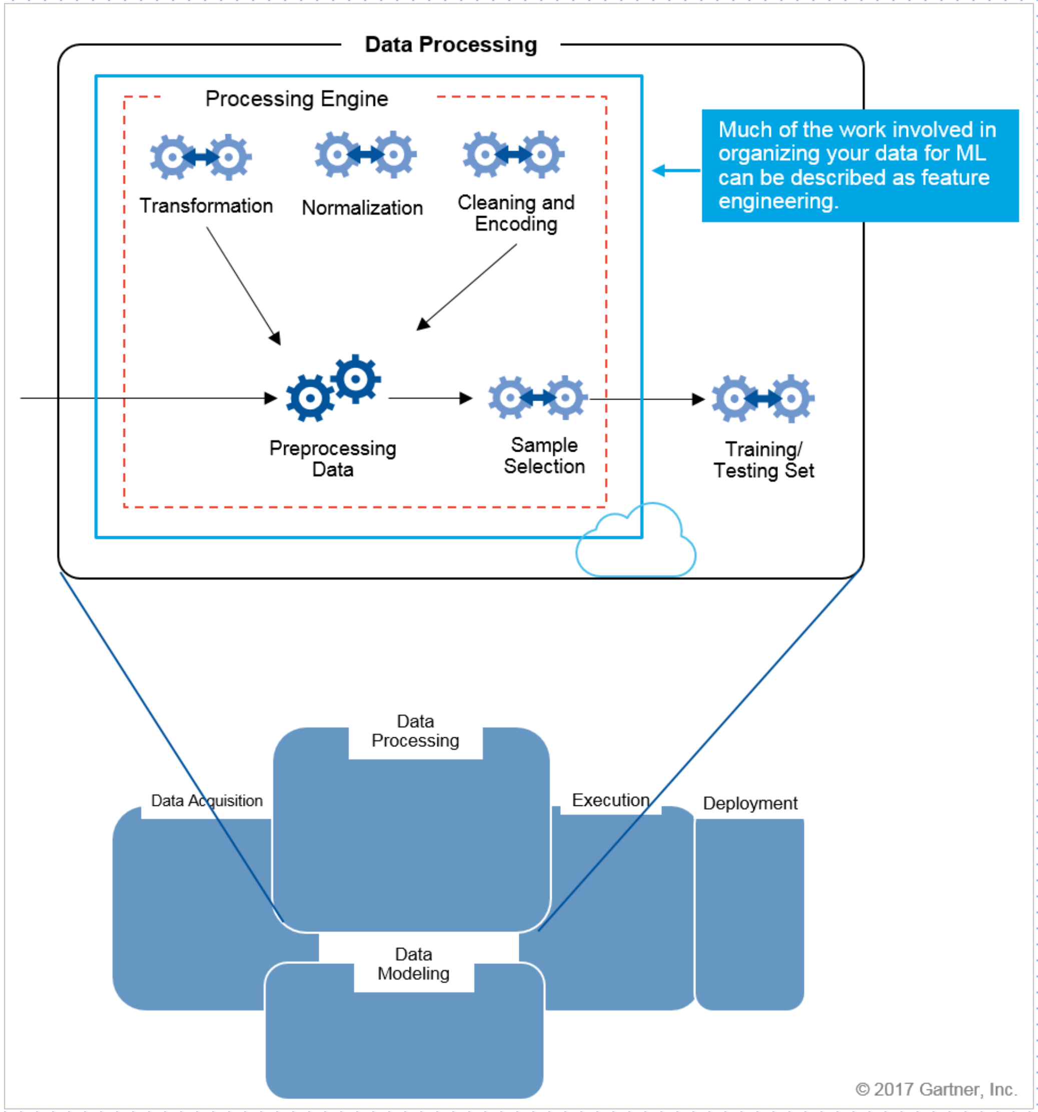
Preparing and Architecting for Machine Learning(*)



ML Architecture: Data Acquisition

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

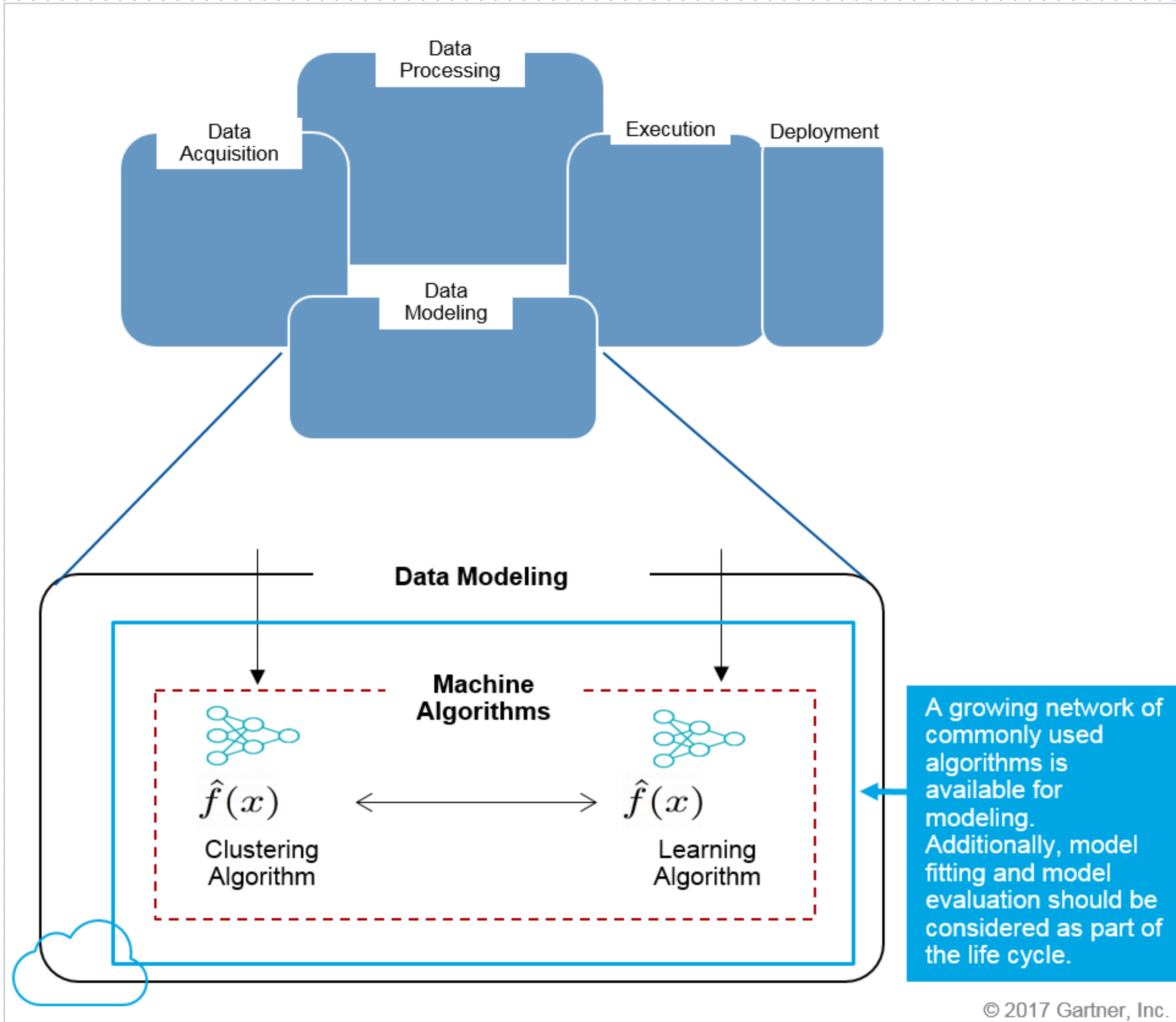
Preparing and Architecting for Machine Learning(*)



ML Architecture: Data Processing

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

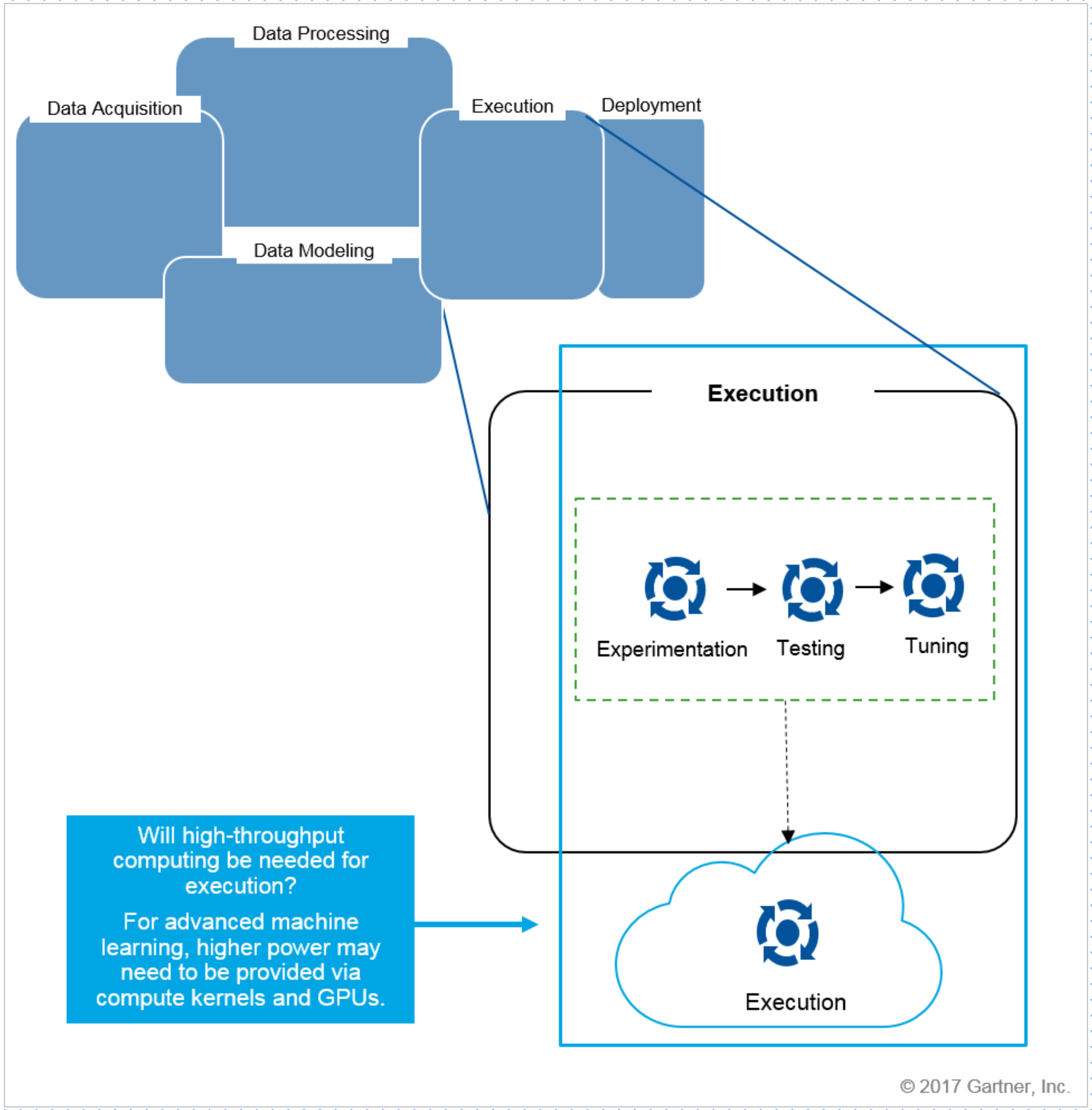
Preparing and Architecting for Machine Learning(*)



ML Architecture: Data Modeling

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

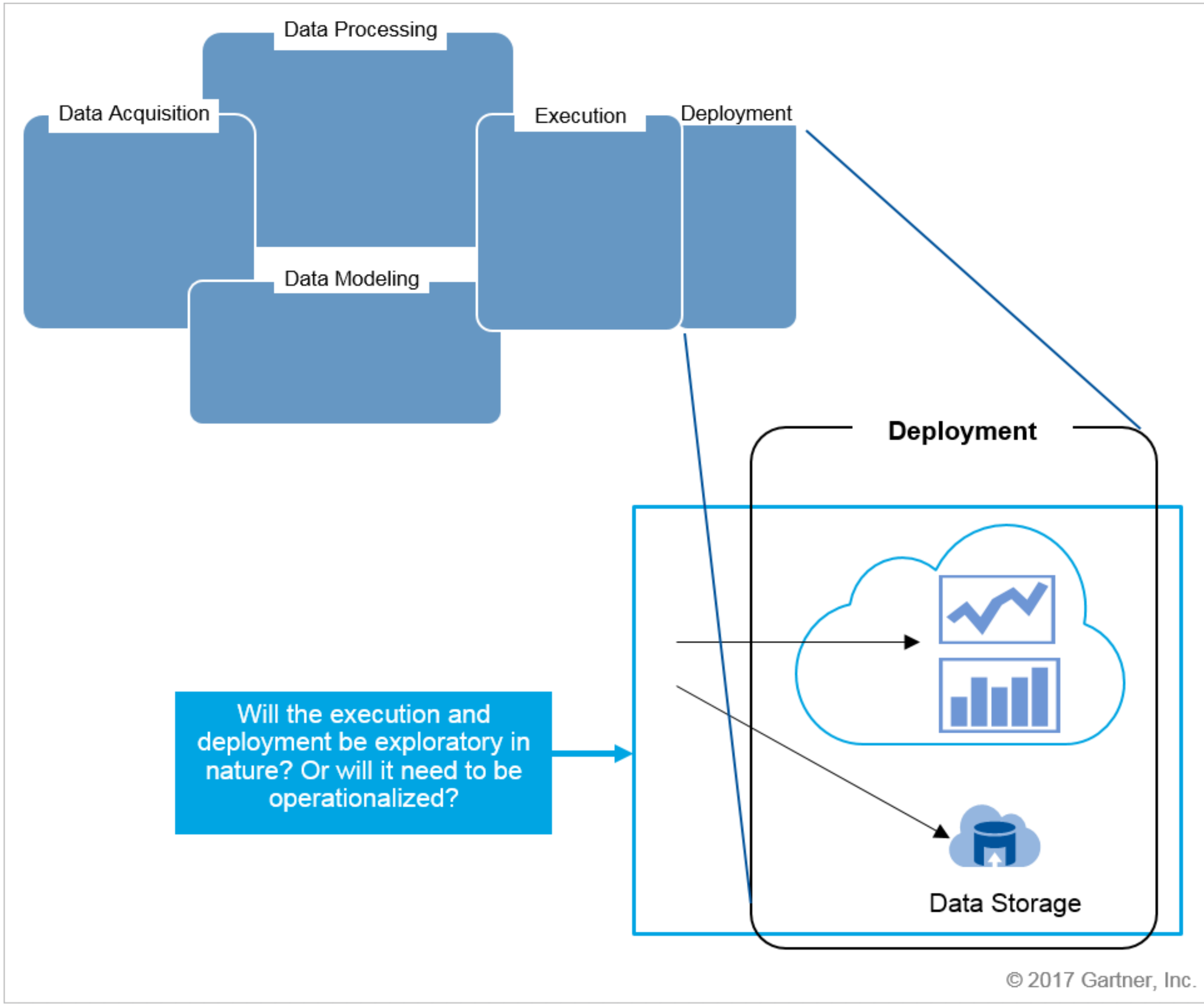
Preparing and Architecting for Machine Learning(*)



ML Architecture: Execution

(*) Gartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

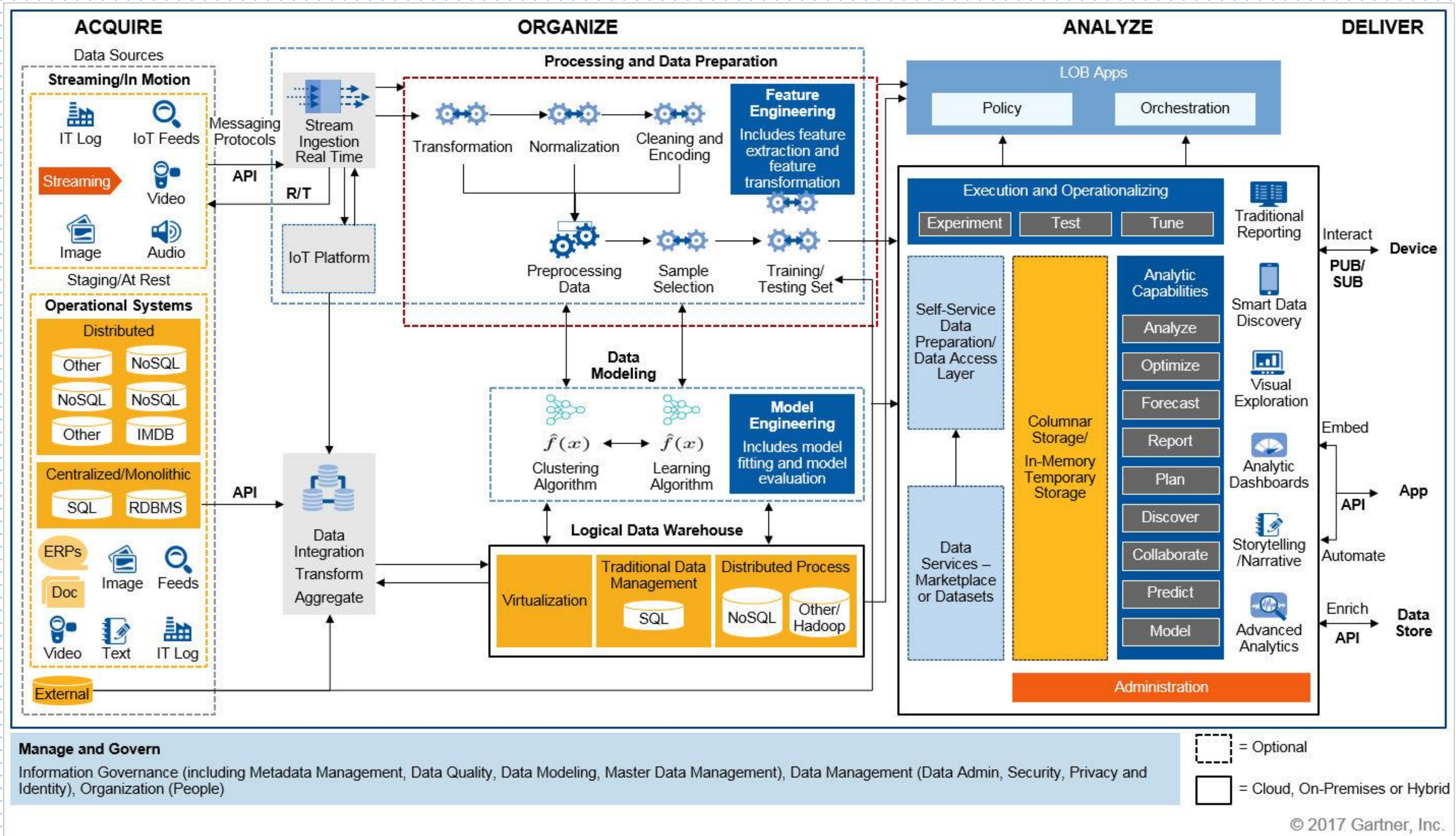
Preparing and Architecting for Machine Learning(*)



ML Architecture: Deployment

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)



End-to-End ML and Analytics Architecture

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

Recommendations

Gartner recommends that technical professionals take the following steps to prepare and initiate ML capabilities*:*

- *Build a taxonomy for classifying the problems or challenges to be solved by ML.* ML algorithms can be overwhelming because there are many to choose from. Organizations often spend too much time ***debugging models*** that don't fit the data, business problem or challenge they are trying to address. Start by ***categorizing*** to help reduce capabilities and to avoid overwhelming users.

- *Evaluate self-service platforms that support data preparation and applied machine learning*. *There are a variety of ML platforms that support proprietary deep learning frameworks, but don't support common frameworks offered by the open-source community (such as Google TensorFlow, Caffe, Torch, Deeplearning4j and so on). Gartner recommends evaluating self-service ML platforms against their capability to interoperate with multiple deep learning frameworks*.

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

Preparing and Architecting for Machine Learning(*)

Recommendations

Gartner recommends that technical professionals take the following steps to prepare and initiate ML capabilities:

- *Offer ML as a toolkit to data scientists rather than allowing them to build their own customized algorithms.* *There are extensive toolkits available, and they will likely support your use case or business challenge. Developing customized algorithms can be a nontrivial undertaking and can expand your architecture with unconventional integration to third-party tools. Gartner recommends offering toolkits to be exploited by data science teams to avoid potential integration challenges.*

- *Use the public cloud to start your initiative because it can elastically scale to accommodate any requirement.* *Amazon, Microsoft, IBM, Google and many other cloud providers offer ML capabilities that can be leveraged to achieve self-service capabilities.*

(*) Cartner Group, Analyst: Carlton E. Sapp, Published: 17 January 2017, ID: G00317328

End of Lecture

Thank YOU!!