Statistical Learning

Mathieu Lagrange & Diana Mateus
mathieu.lagrange@univ-nantes.fr

CENTRALE NANTES

https://mathieulagrange.github.io

Outline

Outline

1. Motivation
2. Definition
3. Typology
4. Supervised learning

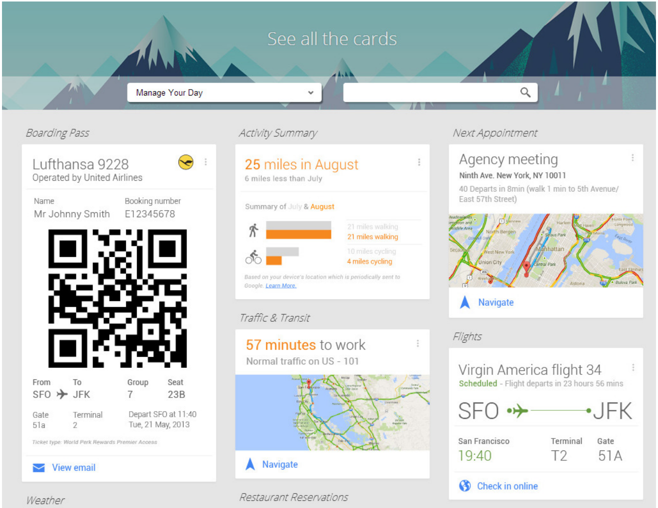
Outline

Outline

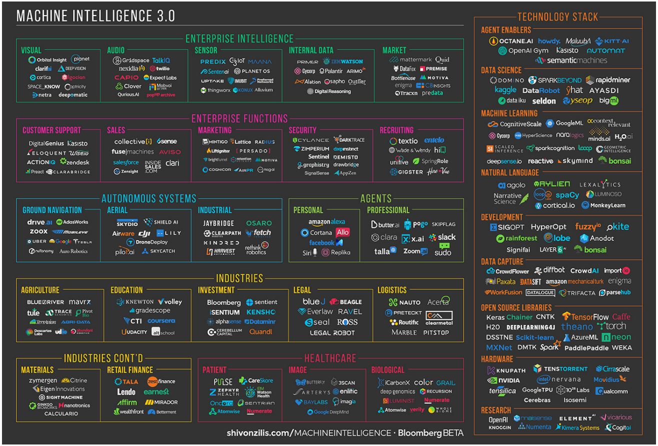
Statistical learning is everywhere

- human computer interaction
  - voice recognition / synthesis
  - handwriting recognition
- recommendation
  - web search
  - goods, and digital media
- data access
  - indexing
  - classification

Usage


See all the cards
Manage Your Day
Boarding Pass
Lufthansa 9228
Operated by United Airlines
Name
Mr Johnny Smith
Booking number
E12345678
From
SFO
To
JFK
Group
7
Seat
23B
Gate
51a
Terminal
2
Depart SFO at 11:40
Tue, 21 May, 2013
Ticket type: World Perk Rewards Premier Access
View email
Weather
Activity Summary
25 miles in August
6 miles less than July
Summary of July & August
21 miles walking
21 miles walking
10 miles cycling
4 miles cycling
Based on your device's location which is periodically sent to Google. Learn More.
Traffic & Transit
57 minutes to work
Normal traffic on US - 101
Navigate
Restaurant Reservations
Next Appointment
Agency meeting
Ninth Ave. New York, NY 10011
40 Departs in 8min (walk 1 min to 5th Avenue/ East 57th Street)
Navigate
Flights
Virgin America flight 34
Scheduled - Flight departs in 23 hours 56 mins
SFO
JFK
San Francisco
19:40
Terminal
T2
Gate
51A
Check in online

Jobs


MACHINE INTELLIGENCE 3.0
ENTERPRISE INTELLIGENCE
VISUAL
AUDIO
SENSOR
INTERNAL DATA
MARKET
ENTERPRISE FUNCTIONS
CUSTOMER SUPPORT
SALES
MARKETING
SECURITY
RECRUITING
AUTONOMOUS SYSTEMS
GROUND NAVIGATION
AERIAL
INDUSTRIAL
AGENTS
PERSONAL
PROFESSIONAL
INDUSTRIES
AGRICULTURE
EDUCATION
INVESTMENT
LEGAL
LOGISTICS
INDUSTRIES CONT'D
MATERIALS
RETAIL FINANCE
HEALTHCARE
PATIENT
IMAGE
BIOLOGICAL
TECHNOLOGY STACK
AGENT ENABLERS
DATA SCIENCE
MACHINE LEARNING
NATURAL LANGUAGE
DEVELOPMENT
DATA CAPTURE
OPEN SOURCE LIBRARIES
HARDWARE
RESEARCH
shivonzilis.com/MACHINEINTELLIGENCE · Bloomberg BETA

Investment

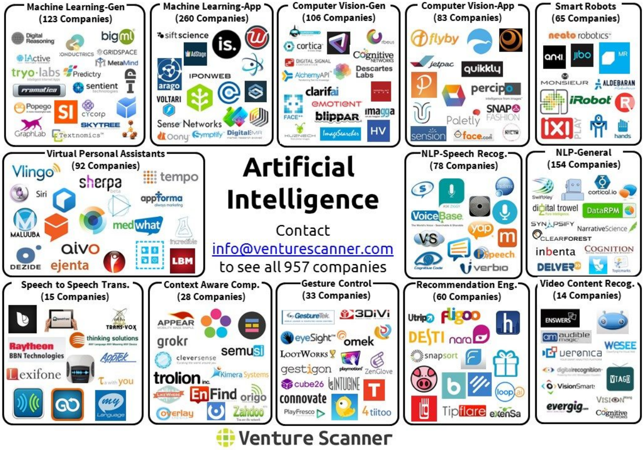
Artificial Intelligence: Most Active Corporate Investors

2011-2016YTD (as of 6/15/2016)

Investor	Rank	Select Investments
Intel Capital	1	DataRobot, prelert, lumiata, MAANA, incoming, saffron, PERFANT, EMOTIENT, Reflektion, Parallel Machines, MindMeld, smartzip, api.ai, indisys, COLDLIGHT
Google Ventures	2	Building Robotics, clarifai, KENSHO, FRAMED, Zephyr Health, Unbabel, tamr, MindMeld, Orbital Insight, Predilytics
GE Ventures	3	Sight Machine, ARTERYS, AYASDI, BITSTEW Systems, MedAware, PingThings, stem, Predixion, MAANA
Samsung Ventures	4	vicarious, sentiance, Maluuba, iDiBON, jibo, MindMeld, Automated Insights
Bloomberg Beta	4	deep genomics, context relevant, AVISO, howdy., DOMINO, Orbital Insight, DigitalGenius, DIFFBOT
In-Q-Tel	6	MindMeld, CYLANCE, celect, INTERSET, Digital Reasoning, DOMINO
Tencent	7	DIFFBOT, CloudMedx, Scaled Inference, iCarbonX, skymind
Nokia Growth Partners	8	rocketfuel, WorkFusion, rapidminer, indix
Microsoft Ventures	8	Building Robotics, NEURA, insidesales, CrowdFlower
Qualcomm Ventures	8	clarifai, Predilytics, Welltok, tempo
Salesforce Ventures	8	DigitalGenius, insidesales, 6sense
AXA Strategic Ventures	8	NEURA, BIOBEATS, medlanes, pricemethod
New York Life Insurance Company	8	context relevant, DataRobot, Skycure, aptricity

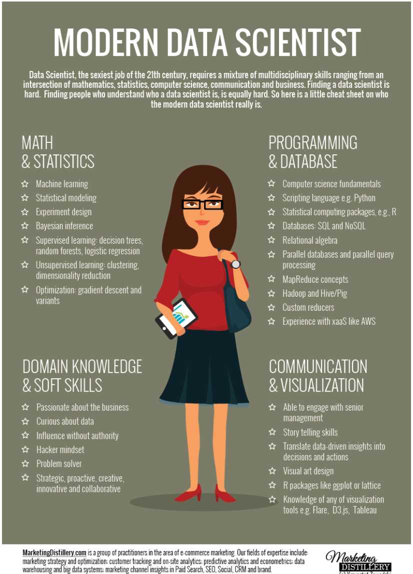
CB Insights www.cbinsights.com

Investment

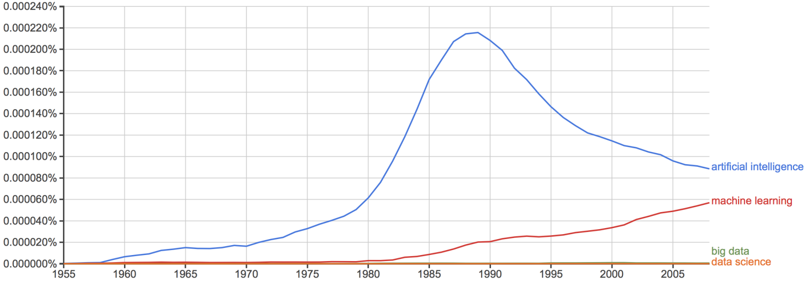

Machine Learning-Gen (123 Companies)
Machine Learning-App (260 Companies)
Computer Vision-Gen (106 Companies)
Computer Vision-App (83 Companies)
Smart Robots (65 Companies)
Virtual Personal Assistants (92 Companies)
Artificial Intelligence
Contact
info@venturescanner.com
to see all 957 companies
NLP-Speech Recog. (78 Companies)
NLP-General (154 Companies)
Speech to Speech Trans. (15 Companies)
Context Aware Comp. (28 Companies)
Gesture Control (33 Companies)
Recommendation Eng. (60 Companies)
Video Content Recog. (14 Companies)
Venture Scanner

Big data <-> Big business

Skills


MODERN DATA SCIENTIST
Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.
MATH & STATISTICS
Machine learning
Statistical modeling
Experiment design
Bayesian inference
Supervised learning: decision trees, random forests, logistic regression
Unsupervised learning: clustering, dimensionality reduction
Optimization: gradient descent and variants
PROGRAMMING & DATABASE
Computer science fundamentals
Scripting language e.g. Python
Statistical computing packages, e.g., R
Databases: SQL and NoSQL
Relational algebra
Parallel databases and parallel query processing
MapReduce concepts
Hadoop and Hive/Pig
Custom reducers
Experience with xaaS like AWS
DOMAIN KNOWLEDGE & SOFT SKILLS
Passionate about the business
Curious about data
Influence without authority
Hacker mindset
Problem solver
Strategic, proactive, creative, innovative and collaborative
COMMUNICATION & VISUALIZATION
Able to engage with senior management
Story telling skills
Translate data-driven insights into decisions and actions
Visual art design
R packages like ggplot or lattice
Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau
MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.
Marketing DISTILLERY

Naming trends



Artificial Intelligence

The traditional problems (or goals) of AI research include

- reasoning, knowledge representation,
- planning, learning,
- natural language processing, perception and
- the ability to move and manipulate objects.

General intelligence is among the field's long-term goals.

AI Methods

Approaches include statistical methods, computational intelligence, and traditional symbolic AI.
Used tools are:

- search and mathematical optimization : find interesting solutions
- statistical modeling
- artificial neural networks

Machine learning

Machine learning

- is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
- More largely used for systems able to make complex predictions.
- term coined by Arthur Lee Samuel in 1959
- The Samuel Checkers-playing Program appears to be the world's first self-learning program and as such a very early demonstration of the fundamental concept ofÂ artificial intelligence.

ML Methods

- It consists in the construction of algorithms that can learn from and make predictions on data
- such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions.

Evaluation

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.

[Tom M. Mitchell]

When considering ML ?

ML can be useful when

- **Data:** is large and heterogeneous
- **Uncertainty:** data is noisy, unstructured, redundant, ...
- **Limited knowledge:** complex and not well understood phenomena

Problem solved

- does this item belongs to A or B ? (closed set classification)
- does this item belongs to A ? (closed set classification)
- is this item very A or only a bit ? (regression)
- how my data is structured ? (clustering)
- is this item very different from the usual ones ? (anomaly detection)

Types of learning

- **Reinforcement learning:** the world is completely described (explicit reward)
- **Supervised learning:** the relation between the items and the corresponding supervisory signal is known for some items
- **Unsupervised learning:** Discover the structure and regularities of the items by observing them (and potentially living with them)

Types of learning

- **Reinforcement learning:** The machine predicts a scalar reward given once in a while (A few bits for some samples)
- **Supervised learning:** The machine predicts a category or a few numbers for each input (10 to 10,000 bits per sample)
- **Unsupervised learning:** The machine predicts any part of its input for any observed part, eg predicting future frames in videos (Millions of bits per sample)

Yan LeCun Unsupervised learning seminar 2016

Types of learning

Unsupervised Learning is the "Dark Matter" of AI

ML trends: Unsupervised Learning

Unsupervised learning is the only form of learning that can provide enough information

- to train large neural nets with billions of parameters.
- Supervised learning would take too much labeling effort
- Reinforcement learning would take too many trials

Supervised learning

- Let $y \in A$ be the labels assigned to some items $x \in \mathcal{R}^d$
- n couples are available for training: $(x_i, y_i)_{i \leq n}$
- they are assumed to be iid samples $(X_i, Y_i)_{i \leq n}$ from non observed distributions (X, Y)
- from a given x, the system predicts an estimate $\tilde{y}$
- parameters of the system are optimized such that $\tilde{y}_i \approx y_i$

Generalization

- We wish that the precision obtained on the training set is preserved over unseen data
- this is called generalization capabilities

Learning

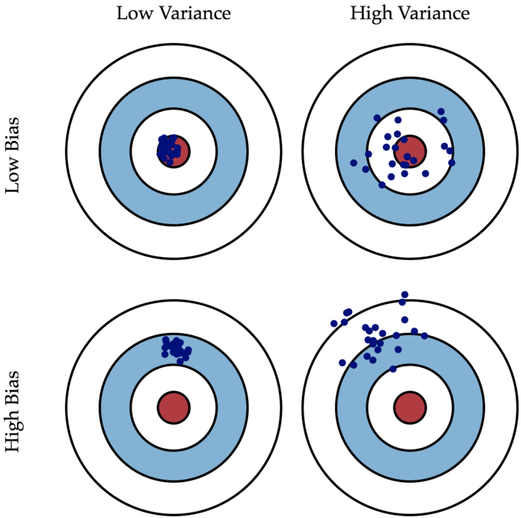
- the learning system computes $\tilde{y} = \tilde{f}(x)$
- $\tilde{f}$ is chosen among a class $\mathcal{H}$
- assuming that the y paired with x is unique, $y = f(x)$
- The system then compute an approximation $\tilde{f}$ of f.

Risk

The important questions are:

- how do we measure and major the difference between the empirical error $\tilde{R}_e(\tilde{f})$ and the generalization error $R(\tilde{f})$?
- how do we ensure that $R(\tilde{f})$ is low ?

Bias and Variance explained



Bias Variance Tradeoff

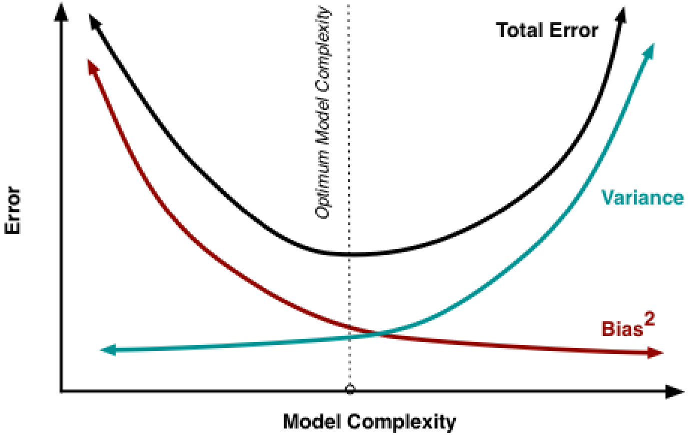
Given

- $f_a = \underset{h \in \mathcal{H}}{\operatorname{argmin}} R(h)$
- $f = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \tilde{R}_e(h)$

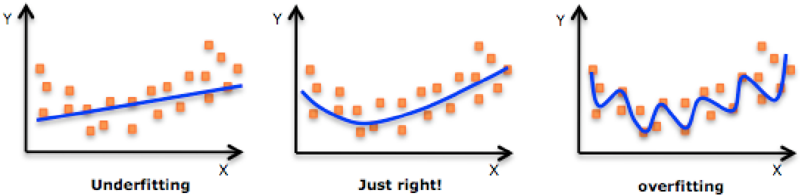
we have

- $R(\tilde{f}_a) \leq R(\tilde{f}) \leq R(\tilde{f}_a) + 2 \max_{h \in \mathcal{H}} |R(h) - \tilde{R}_e(h)|$

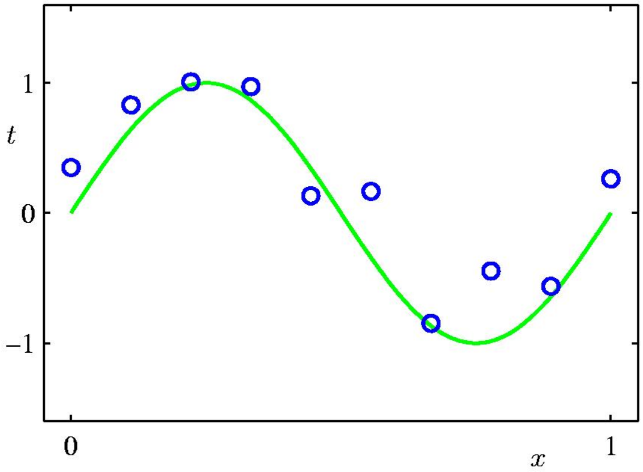
Bias Variance Tradeoff



Illustration



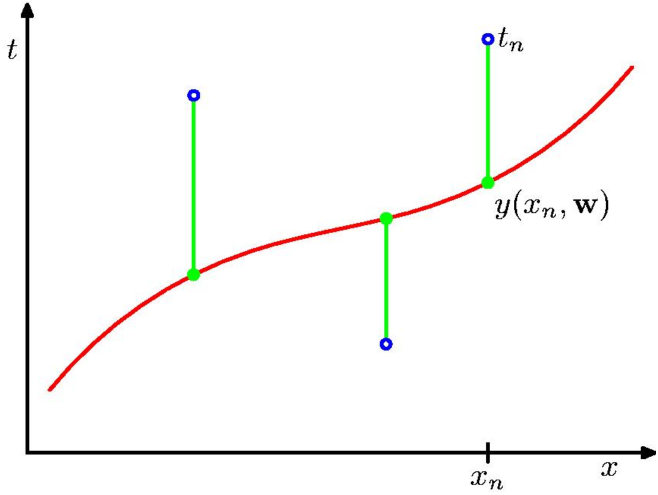
Toy example: polynomial curve fitting



Toy example: polynomial curve fitting

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

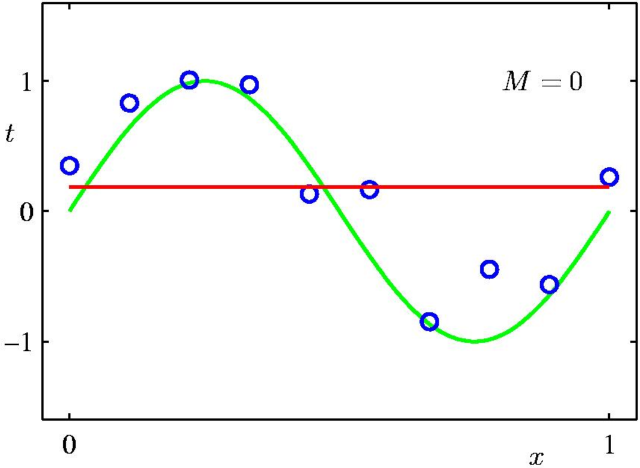
Toy example: polynomial curve fitting



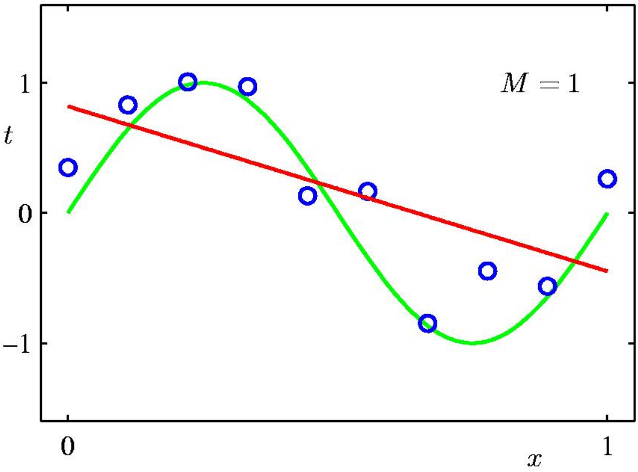
Toy example: polynomial curve fitting

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

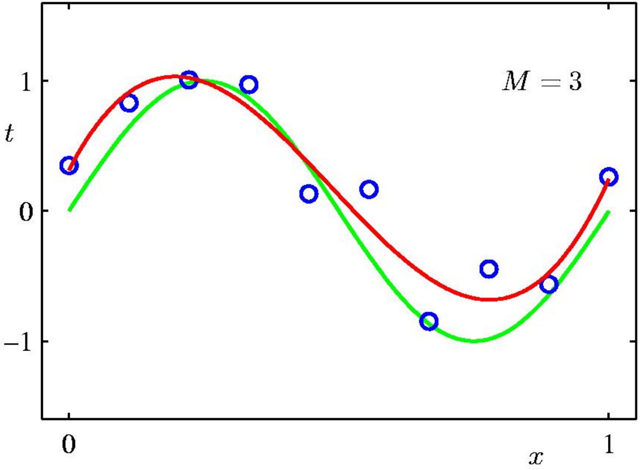
Toy example: polynomial curve fitting



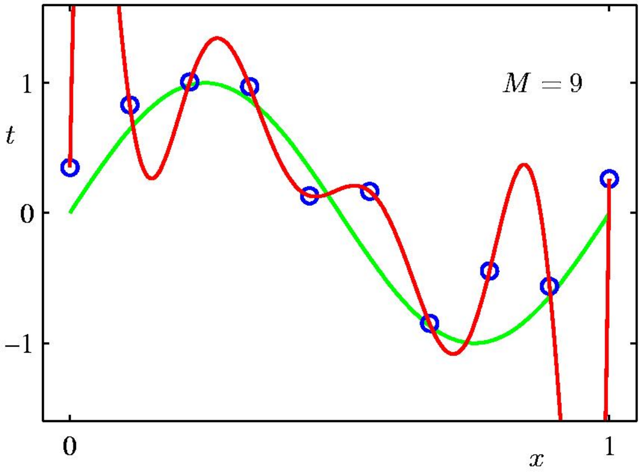
Toy example: polynomial curve fitting



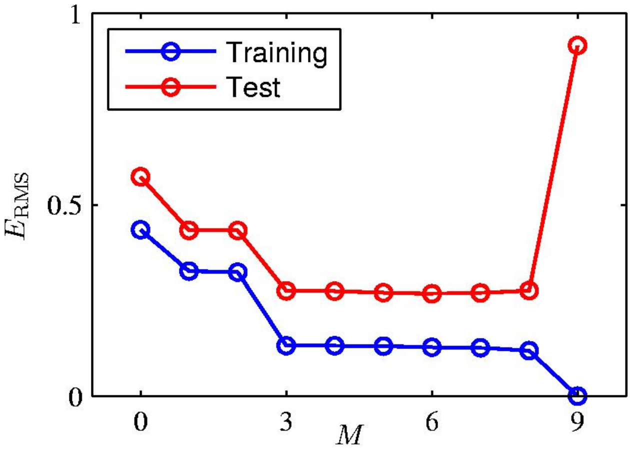
Toy example: polynomial curve fitting



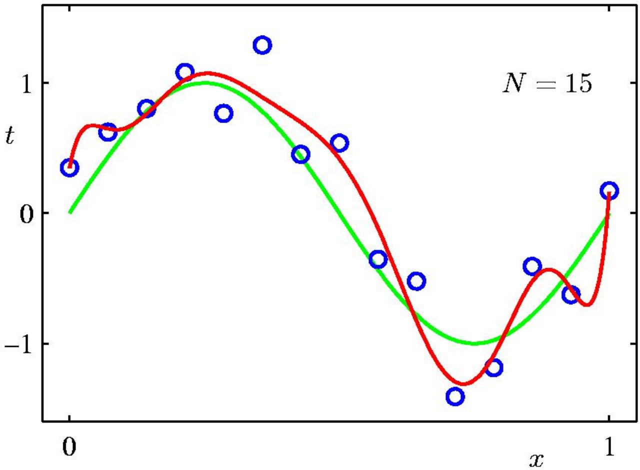
Toy example: polynomial curve fitting



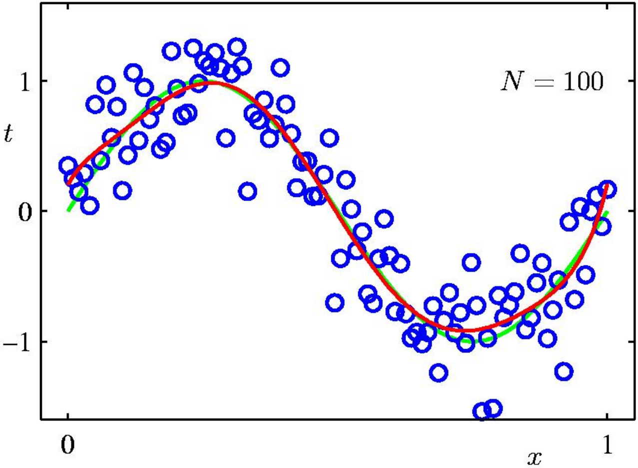
Toy example: polynomial curve fitting



Toy example: polynomial curve fitting



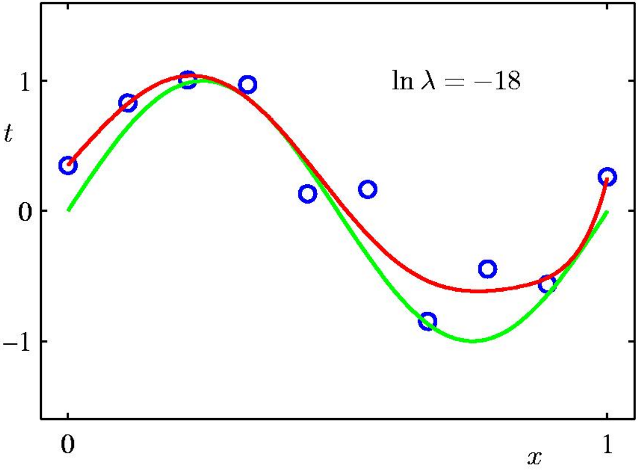
Toy example: polynomial curve fitting



Toy example: polynomial curve fitting

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

Toy example: polynomial curve fitting

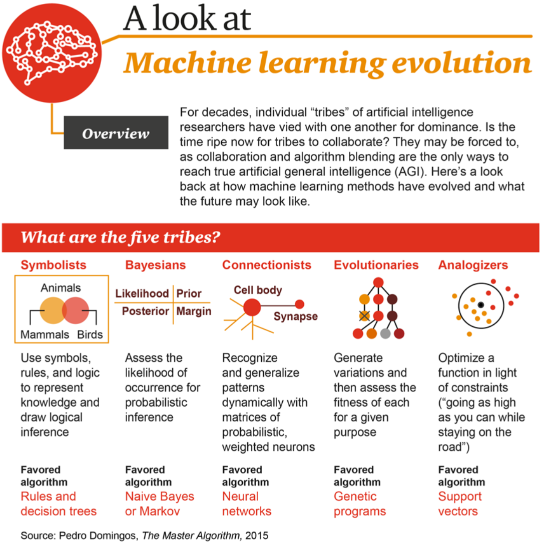


Conclusion

Machine learning

- is a fascinating discipline
- with now flourishing effective techniques that shape humanity
- more is needed with respect to the understanding of the underlying concepts...

Tribes


A look at
Machine learning evolution
Overview
For decades, individual "tribes" of artificial intelligence researchers have vied with one another for dominance. Is the time ripe now for tribes to collaborate? They may be forced to, as collaboration and algorithm blending are the only ways to reach true artificial general intelligence (AGI). Here's a look back at how machine learning methods have evolved and what the future may look like.
What are the five tribes?
Symbolists
Animals
Mammals Birds
Use symbols, rules, and logic to represent knowledge and draw logical inference
Favored algorithm
Rules and decision trees
Bayesians
Likelihood Prior
Posterior Margin
Assess the likelihood of occurrence for probabilistic inference
Favored algorithm
Naive Bayes or Markov
Connectionists
Cell body
Synapse
Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons
Favored algorithm
Neural networks
Evolutionaries
Generate variations and then assess the fitness of each for a given purpose
Favored algorithm
Genetic programs
Analogizers
Optimize a function in light of constraints ("going as high as you can while staying on the road")
Favored algorithm
Support vectors
Source: Pedro Domingos, The Master Algorithm, 2015