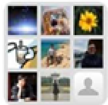
[80245013 Machine Learning, Fall, 2020]

Machine Learning

Jun Zhu
dcszj@mail.tsinghua.edu.cn
http://ml.cs.tsinghua.edu.cn/~jun
Institute for AI
Tsinghua University

September 15, 2020



Machine-Learning-fall-2020

该二维码7天内(9月22日前)有效，重新进入将更新

A bit about Jun…

- Jun Zhu, Professor, Depart. of Computer Science. I received Ph.D. in 2009. My research interest includes machine learning, Bayesian methods, and data mining
- I did post-doc at the Machine Learning Department in CMU with Prof. Eric P. Xing. Before that I was invited to visit CMU for twice. I was also invited to visit Stanford for joint research (with Prof. Li Fei-Fei)
- 2015-2018: Adjunct Associate Professor at CMU
- Published 100+ papers on the top-tier ML conferences and journals, including JMLR, TPAMI, ICML, NIPS, etc.
- Served as Area Chairs for ICML, NIPS, UAI, AAAI, IJCAI; Associate Editor-in-Chief for PAMI, AI Journal
- Research is supported by National 973, NSFC, "Tsinghua 221 Basic Research Plan for Young Talents".
- IEEE AI's 10 to Watch; MIT TR35 China (pioneers)
- Homepage: http://ml.cs.tsinghua.edu.cn/~jun

Contact Information

- Jun Zhu
  - Institute for AI, Department of Computer Science, Tsinghua U.

  - Office: Rm 4-513, FIT Building
  - E-mail: dcszj@tsinghua.edu.cn
  - Phone: 62772322, 18810502646
  - Office hours: Thursday afternoon 3:30pm-5:00pm
    - Better to make an appointment in advance

A bit about Jie…

- Jie Tang, Professor, Department of Computer Science of Tsinghua University. My research interests include social network, data mining, and machine learning.
- I have been visiting scholar at Cornell U. (working with John Hopcroft, Jon Kleinberg), UIUC (working with Jiawei Han), CUHK (with Jeffrey Yu), and HKUST (with Qiong Luo).

- I was awarded with the CCF Young Scientist Award, NSFC Excellent Young Scholar, Newton Advanced Fellowships Award, IBM Innovation Faculty Award, and New Star of Beijing S&T.
- Have published more than 200 paper on major international conf/journals, including KDD (19), IJCAI/AAAI (16), IEEE Trans. (21), ICML, Machine Learning
- #Citation: 16,000+ and H-index: 63
- Have a notable system, AMiner.org for academic researcher network analysis. The system has attracted 8.32 million users from 220 countries/regions.
- **Homepage:** http://keg.cs.tsinghua.edu.cn/jietang/

Contact Information

- Jie Tang
  - Software Division, Department of Computer Science, Tsinghua U.
  - Office: Rm 1-308, FIT Building
  - E-mail: jietang@tsinghua.edu.cn
  - Phone: 62788788-20, 13911215746
  - Open hours: Tuesday afternoon 2:00pm-5:00pm

Teaching Assistants

- You Qiaoben (Head TA)
  - E-mail: qby17@mails.tsinghua.edu.cn
  - Phone: 62795869, 18810690095
  - Reinforcement learning, Deep learning

- Jiezhong Qiu
  - PhD student
  - qiujz16@mails.tsinghua.edu.cn
  - More information will be provided by Jie

雨课堂 Rain Classroom

Resources

- Mainly class slides/notes
- Recommended text books
  - Christopher M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2007.
  - Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. Deep Learning. 2016.
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman. Elements of Statistical Learning. 2nd Edition, Springer, 2009.
- Further readings:
  - Conferences:
    - Theory: ICML, NIPS, UAI, COLT, AISTATS, AAAI, IJCAI
    - App: KDD, SIGIR, WWW, ACL
  - Journals:
    - JMLR, PAMI, MLJ

Prerequisites

- Knowledge of probability, linear algebra, statistics and algorithms
  - Calculus:
    - Derivative, integral of multivariate functions
  - Linear Algebra
    - Matrix inversion, eigen-decomposition, ...
  - Basic Probability and Statistics
    - Probability distributions, Mean, Variance, Conditional probabilities, Bayes rule, ...

- Knowledge of programming languages, e.g., C/C++, Java, matlab, Python

- **Homework 0**: take the Self-Evaluation
  - Minimum & modest background tests (available at course webpage)

Potential achievements

- Able to understand the underlying principles of classical ML algorithms
- Able to apply right ML algorithms to the applications at your hand
- Able to design effective ML algorithms to solve new problems

Overview of Class

Introduction	3 units	
Unsupervised learning	6 units	HW1 out
Supervised learning	6 units	
Reinforcement Learning	6 units	HW1 due HW2 out
Convolutional neural network	3 units	
Auto-Encoders	3 units	HW2 due HW3 out
Recurrent neural network	3 units	
Representation Learning	6 units	HW3 due HW4 out
GAN and AutoML	6 units	
		HW4 due in 2 weeks

Grading

- Participation (10%)
  - Will collect information from Rain-classroom (10 points)
- Homeworks (40%)
  - 4 homeworks (10 points each time)
- Project (50%)
  - <=3 students to form a team
  - Apply machine learning to solve a real problem
    - Choose one task at Kaggle (http://www.kaggle.com/competitions)
  - Submit materials:
    - a proposal (6^{th} week), a mid-term report (9^{th} week), a final report (18^{th} week), and the implementation code (18^{th} week)
  - All reports should be in NIPS format, written in English: (http://nips.cc/Conferences/2014/PaperInformation/StyleFiles)
  - Poster presentation (16^{th} week)

Some example Kaggle tasks

All Competitions

Active | Completed | InClass — All Categories ▾ | Default Sort ▾

OSIC Pulmonary Fibrosis Progression
Predict lung function decline
Featured • 22 days to go • Code Competition • 1701 Teams

RSNA-STR Pulmonary Embolism Detection
Classify Pulmonary Embolism cases in chest CT scans
Featured • a month to go • Code Competition • 56 Teams

Lyft Motion Prediction for Autonomous Vehicles
Build motion prediction models for self-driving vehicles
Featured • 2 months to go • Code Competition • 379 Teams

Mechanisms of Action (MoA) Prediction
Can you improve the algorithm that classifies drugs based on their biological activity
Research • 3 months to go • Code Competition • 1130 Teams

Cornell Birdcall Identification
Build tools for bird population monitoring
Research • a day to go • Code Competition • 1375 Teams

Google Landmark Recognition 2020
Label famous (and not-so-famous) landmarks in images
Research • 15 days to go • Code Competition • 634 Teams

OpenVaccine: COVID-19 mRNA Vaccine Degradation Predicti
Urgent need to bring the COVID-19 vaccine to mass production
Research • 21 days to go • 473 Teams

Halite by Two Sigma
Collect the most halite during your match in space
Featured • a day to go • Simulation Competition • 1140 Teams

Conway's Reverse Game of Life 2020
Reverse the arrow of time in the Game of Life
Playground • 3 months to go • Code Competition • 62 Teams

Hash Code Archive: Drone Delivery
Can you help coordinate the drone delivery supply chain?
Playground • 3 months to go • Code Competition • 18 Teams

Predict Future Sales
Final project for "How to win a data science competition" Coursera course
Playground • 4 months to go • 8711 Teams

Titanic: Machine Learning from Disaster
Start here! Predict survival on the Titanic and get familiar with ML basics
Getting Started • Ongoing • 18615 Teams

House Prices: Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting
Getting Started • Ongoing • 4855 Teams

Featured Code Competition

Lyft Motion Prediction for Autonomous Vehicles

Build motion prediction models for self-driving vehicles

$30,000 Prize Money

Lyft · 379 teams · 2 months to go (2 months to go until merger deadline)

Overview | Data | Notebooks | Discussion | Leaderboard | Rules | **Join Competition**

Overview

- Description
- Evaluation
- Timeline
- Prizes
- Code Requirements

Autonomous vehicles (AVs) are expected to dramatically redefine the future of transportation. However, there are still significant engineering challenges to be solved before one can fully realize the benefits of self-driving cars. One such challenge is building models that reliably predict the movement of traffic agents around the AV, such as cars, cyclists, and pedestrians.

The ridesharing company Lyft started Level 5 to take on the self-driving challenge and build a full self-driving system (they're hiring!). Their previous competition tasked participants with identifying 3D objects, an important step prior to detecting their movement. Now, they're challenging you to predict the motion of these traffic agents.

In this competition, you'll apply your data science skills to build motion prediction models for self-driving vehicles. You'll have access to the largest Prediction Dataset ever released to train and test your models. Your knowledge of machine learning will then be required to predict how cars, cyclists,and pedestrians move in the AV's environment.

Lyft's mission is to improve people's lives with the world's best transportation. They believe in a future where self-driving cars make transportation safer, environment-friendly and more accessible for everyone. Their goal is to accelerate development across the industry by sharing data with researchers. As a result of your participation, you can have a hand in propelling the industry forward and helping people around the world benefit from self-driving cars sooner.

This is a Code Competition. Refer to Code Requirements for details.

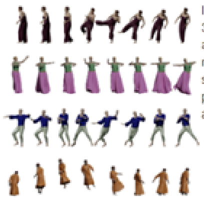
NeurIPS Competitions

- Website: https://neurips.cc/Conferences/2020/CompetitionTrack
- Many are research oriented
- Early due dates
- Datasets can be used

2020 ChaLearn 3D+Texture Garment Reconstruction

(May 13 - October 3)

Sergio Escalera (CVC and University of Barcelona), Meysam Madadi (CVC), Hugo Bertiche (University of Barcelona)



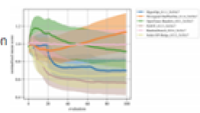
In this competition we plan to push the research to better understand human dynamics in 2D and 3D, with special attention to garments. We provide a large-scale dataset (more than 2M frames) of animated garments with variable topology and type and special care to garment dynamics and realistic rendering. The dataset contains paired RGB images with 3D garment vertices in a sequence. We designed three tracks so participants can compete to develop the best method to perform 3D garment reconstruction and texture estimation in a sequence from (1) 3D garments, and (2) RGB images.

Black-Box Optimization for Machine Learning

(July 1-October 15)

Ryan Turner (Twitter), David Eriksson (Uber AI), Serim Park (Twitter), Mike Mccourt (SigOpt), Zhen Xu (4Paradigm), Isabelle Guyon (ChaLearn), Eero Laaksonen (Valohai) and Juha Kili (Valohai)

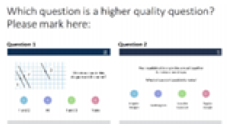
This challenge is about the optimization of black-box functions arising when tuning ML models. The configuration of the search space (function inputs) will be provided to the algorithms, but everything else about the objectives remains hidden. Submissions built from open source Bayesian optimization and/or evolutionary algorithms packages are highly encouraged.



Diagnostic Questions: Predicting Student Responses and Measuring Question Quality

(July 15-October 23)

Simon Woodhead (Eedi), Craig Barton (Eedi), José Miguel Hernández-Lobato (University of Cambridge), Richard Turner (University of Cambridge), Jack Wang (Rice University), Richard G. Baraniuk (Rice University), Angus Lamb (Microsoft Research), Evgeny Saveliev (Microsoft Research), Camilla Longden (Microsoft Research), Pashmina Cameron (Microsoft Research), Yordan Zaykov (Microsoft Research), Simon Peyton-Jones (Microsoft Research), Chen Zhang (Microsoft Research)



In the personalisation of education, the questions used to assess students' learning is paramount, we aim to learn as much as we can from each interaction with the student. Diagnostic questions are designed to elicit not just if a student understands a concept but why they do not understand it. In this competition, participants will aim to predict students' answers to diagnostic questions, provide a measure of question quality, and determine which sequence of diagnostic questions best predicts students' answers.

- If the end date is later than the end of this semester, report the position in the leaderboard;
- Otherwise, follow the standard partition or ask TAs to define a train/test split and compare your methods with 1 or 2 baselines.

Other Projects

- Self-defined topics
  - Need to propose as early as possible to filter out improper ones

- Other candidates
  - Chinese handwritten characters generation and recognition
  - Adversarial attacks and defense of deep learning
  - Deepfake detection challenge
  - Reinforcement learning
  - More to come

About final report

- We expect to see
  - Problems (what?)
  - Motivations (why?)
  - Techniques (how?)
  - Results & Analysis (did you verify what you claimed above?)
  - Conclusions

- The final report should look like a NeurIPS technical paper
  - Style file: https://neurips.cc/Conferences/2019/PaperInformation/StyleFiles

Questions?

[80245013 Machine Learning, Fall, 2020]

Introduction to Machine Learning

Jun Zhu

dcszj@mail.tsinghua.edu.cn

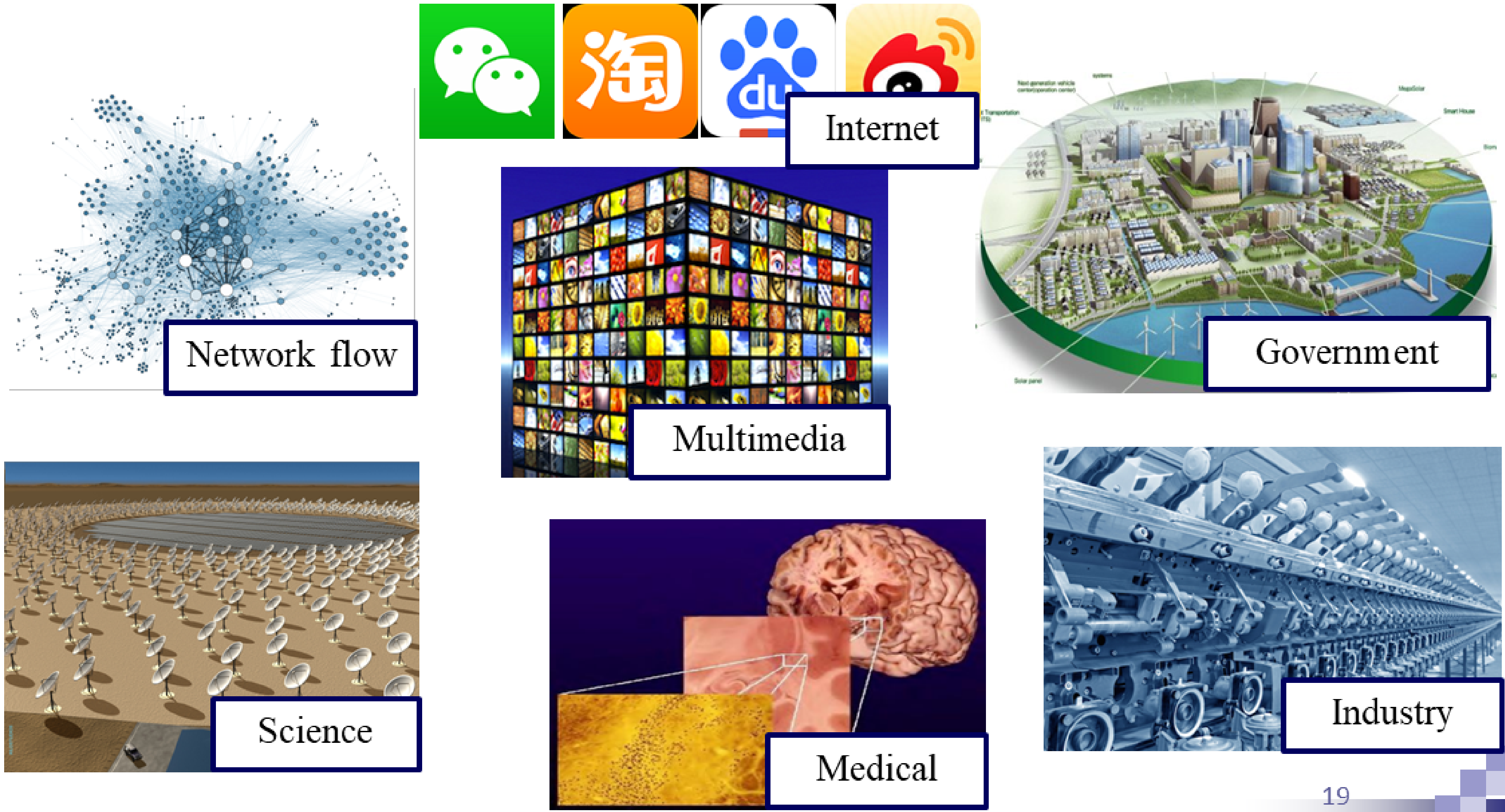
http://ml.cs.tsinghua.edu.cn/~jun

Institute for AI Tsinghua University

September 15, 2020

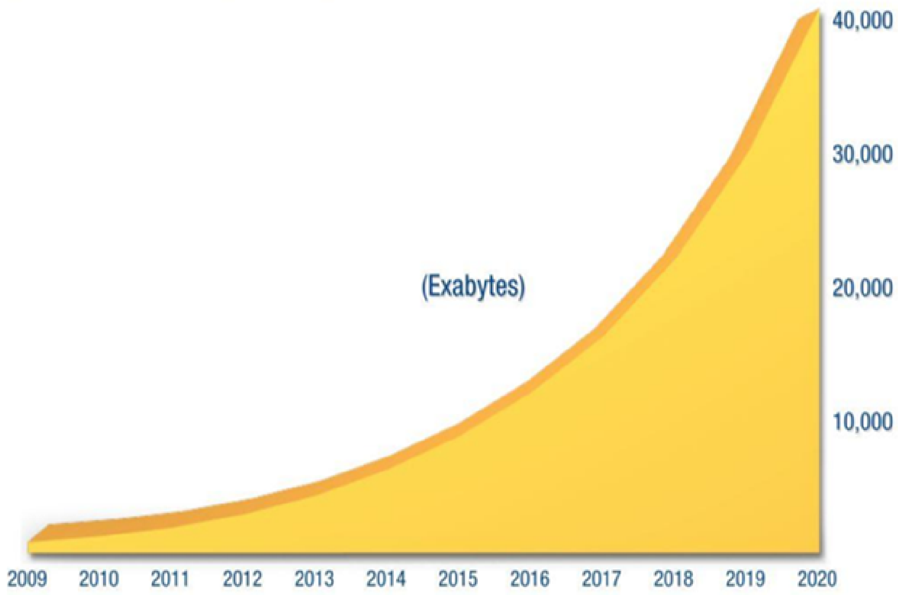
The Age of Big Data

Data is more deeply woven into the fabric of our lives than ever before. We aspire to use data to solve problems, improve well-being, and generate economic prosperity.



The Age of Big Data

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

40,000 Exabytes by 2020
(IDC)

200 million in government funding
(White house initiative)

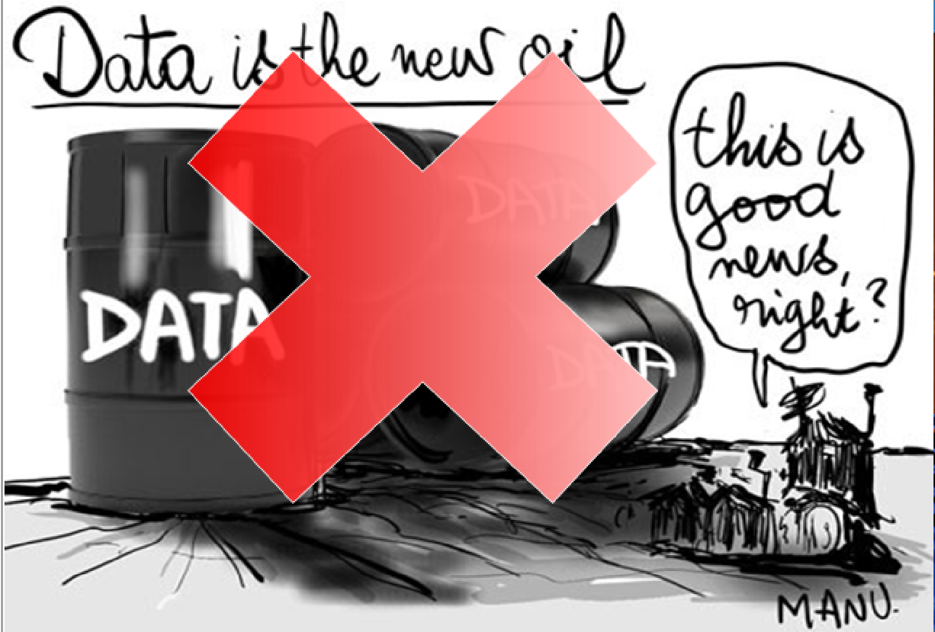
jobs shortage of 200,000 data experts by 2018
(Bloomberg)

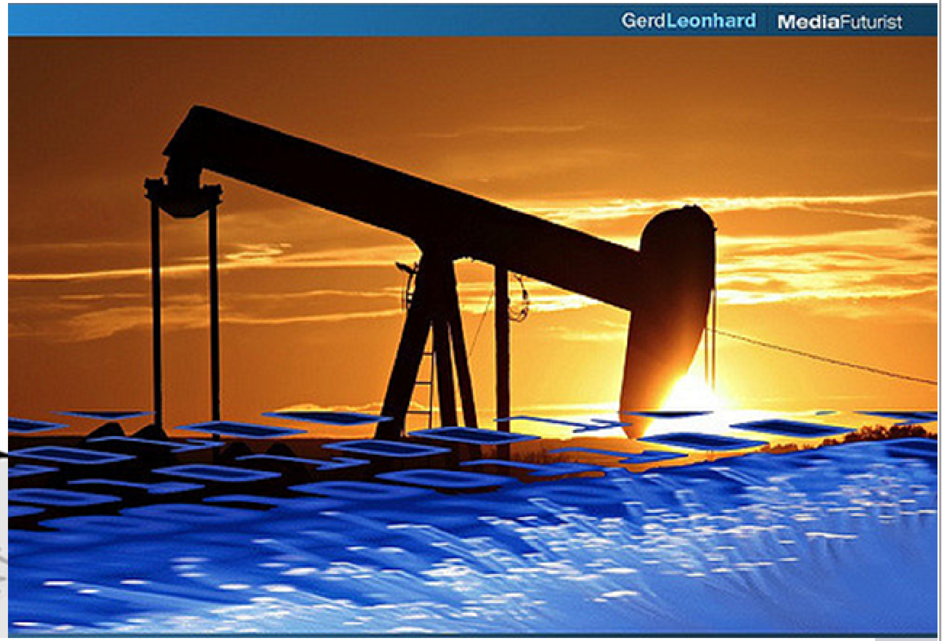
"the sexiest job of the 21st century."
(Harvard Business Review)

Big Data ≠ Big Knowledge

Data is oil

- Data is oil
- Machine learning is to refine the oil (data) and produce the petrol (knowledge).





雨课堂 Rain Classroom

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Introduction The 10 Technologies Past Years

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. →

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous. →

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child? →

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts. →

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people. →

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases

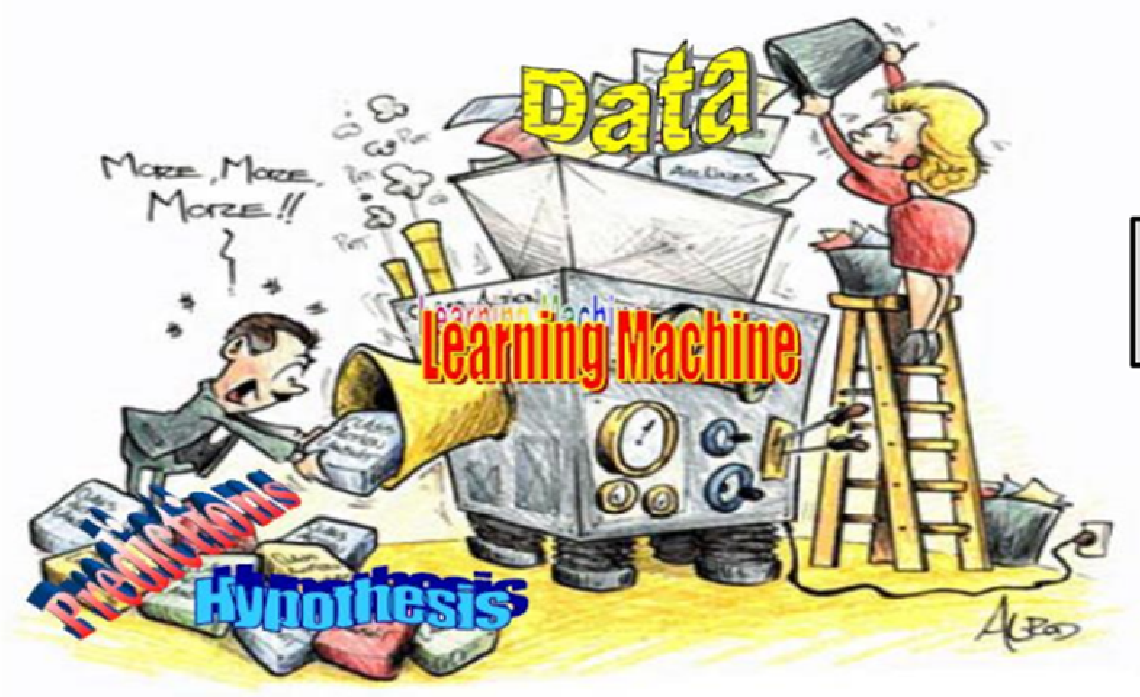
Supergrids

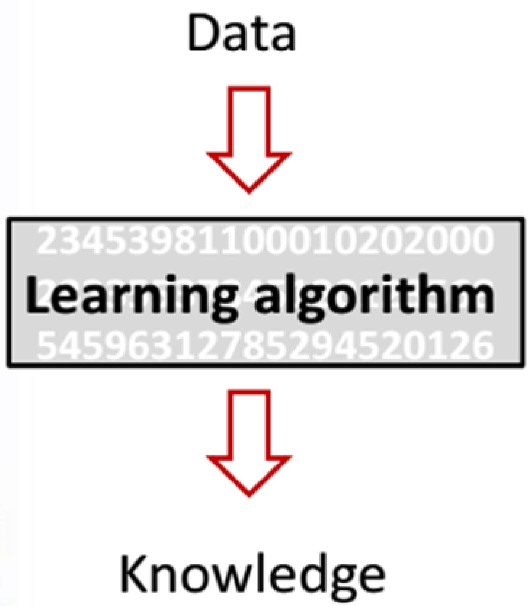
A new high-power circuit breaker could finally make highly efficient DC power grids practical

What is Machine Learning?

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that take as input empirical data, and yield patterns or predictions thought to be features of the underlying mechanism that generated the data

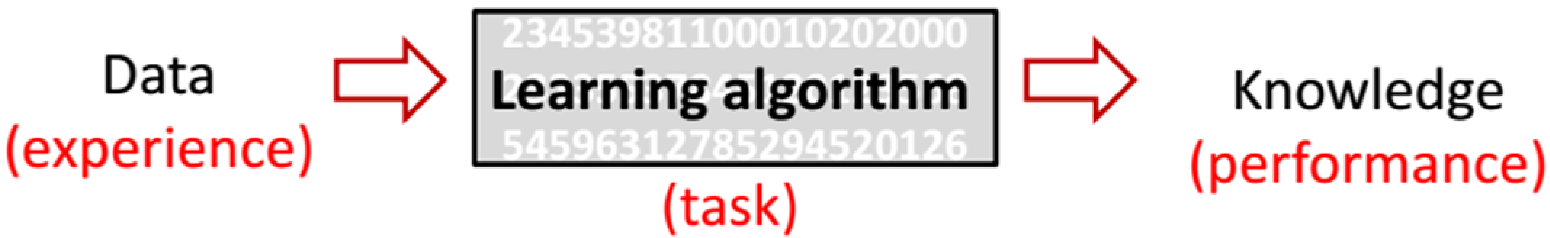
WIKIPEDIA
The Free Encyclopedia





What is machine learning?

- Study of algorithms with respect to <T,P,E> that
  - (automatically) improve their performance P
  - at some task T
  - with experience E



雨课堂 Rain Classroom

(Statistical) Machine Learning in AI

s 19

Great growth ac

post-Pearl AI

1970s & 1980s: expert systems

1960s & 1970s: suc

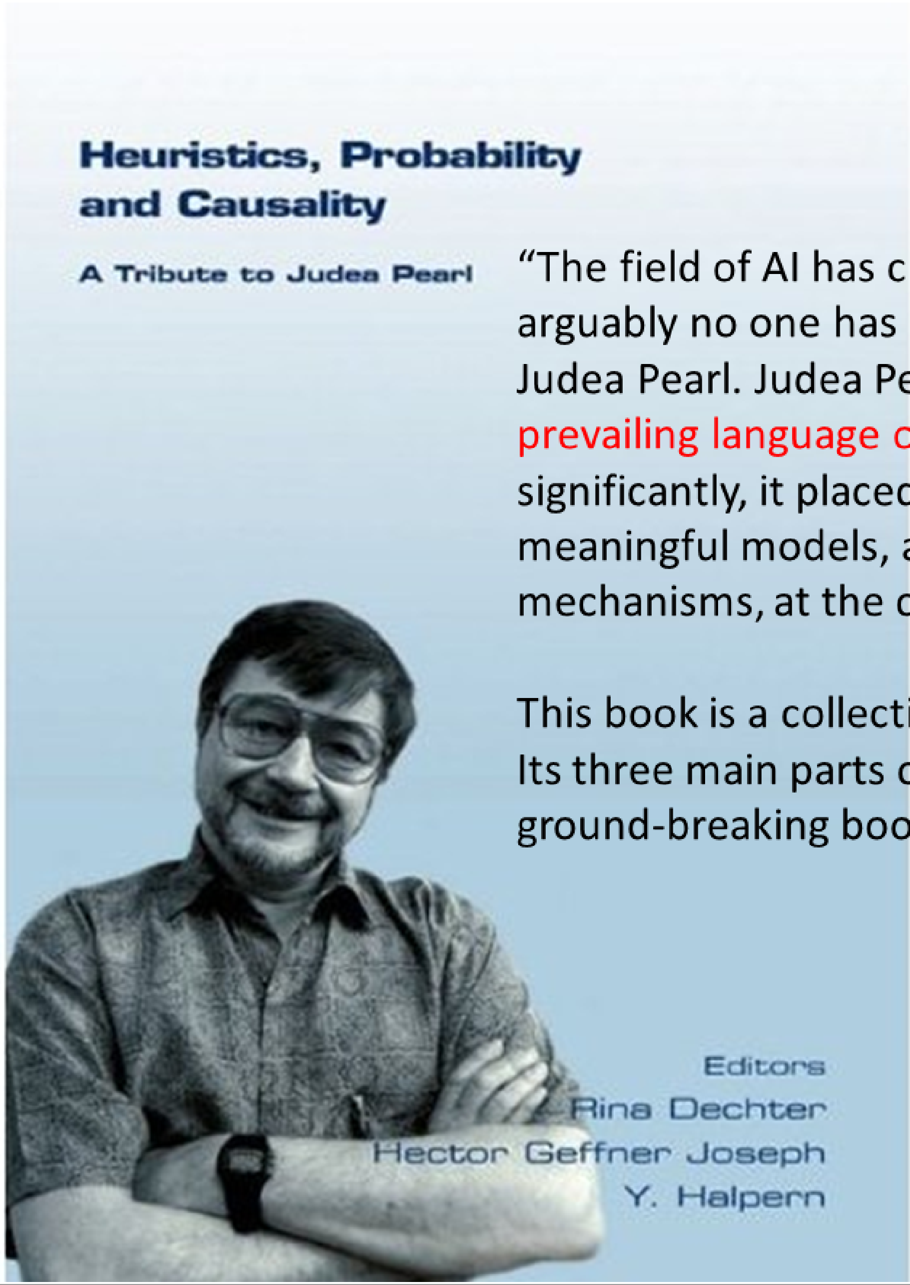
pre-Pearl AI

1950s & 1960s: symbolic r ni
representation, sea

pre-computer age
thoughts on symbolic reasoning

[Judea Pearl, Turing Award 2011]

- For "innovations that enabled remarkable advances in the partnership between humans and machines that is the foundation of Artificial Intelligence (AI)"
- "His work serves as the standard method for handling uncertainty in computer systems, with applications from medical diagnosis, homeland security and genetic counseling to natural language understanding and mapping gene expression data."
- "Modern applications of AI, such as robotics, self-driving cars, speech recognition, and machine translation deal with uncertainty. Pearl has been instrumental in supplying the rationale and much valuable technology that allow these applications to flourish."



"The field of AI has changed a great deal since the 80s, and arguably no one has played a larger role in that change than Judea Pearl. Judea Pearl's work made probability the prevailing language of modern AI and, perhaps more significantly, it placed the elaboration of crisp and meaningful models, and of effective computational mechanisms, at the center of AI research …"

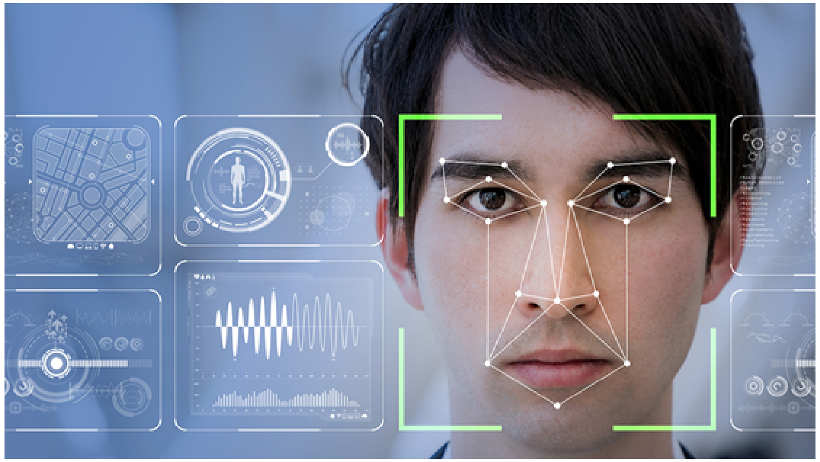
This book is a collection of articles in honor of Judea Pearl. Its three main parts correspond to the titles of the three ground-breaking books authored by Judea …

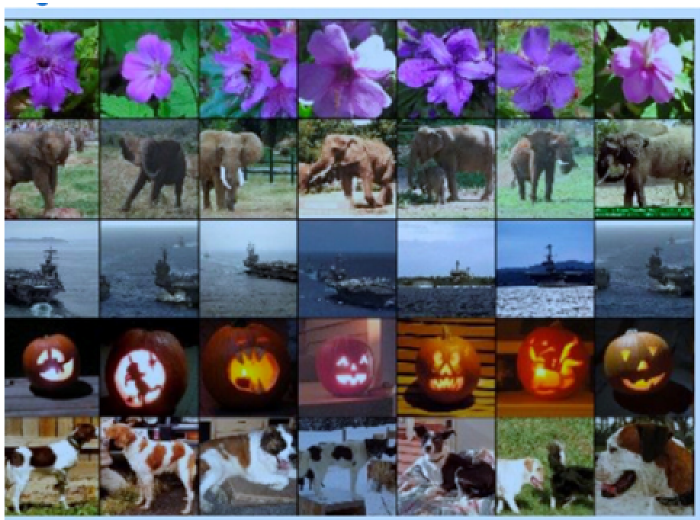
"Machine learning will become a calculus"

--- Tom Mitchell

Machine Learning in Action

- Computer Vision
  - Face recognition
  - Scene understanding
  - Action/behavior recognition
  - Image tagging and search



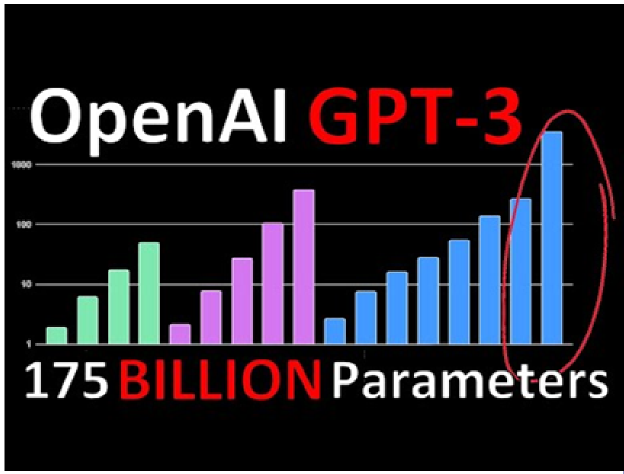


ınd Geoffrey Hinton, 2012. Reproduced with permission.

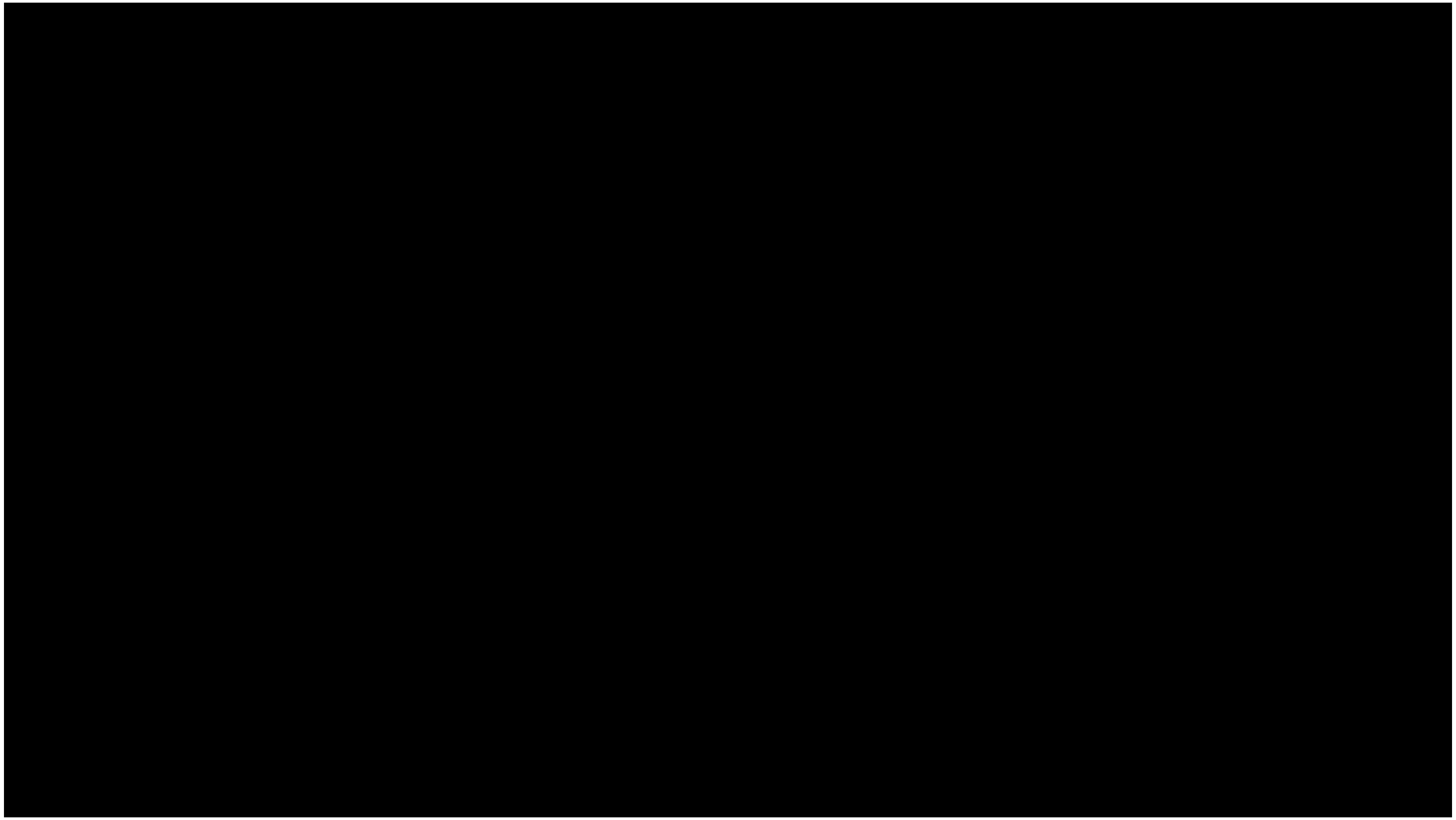
Natural Language Processing

- Machine translation
- Information Extraction
- Information Retrieval, question
- answering
- Text classification, spam

微软小冰

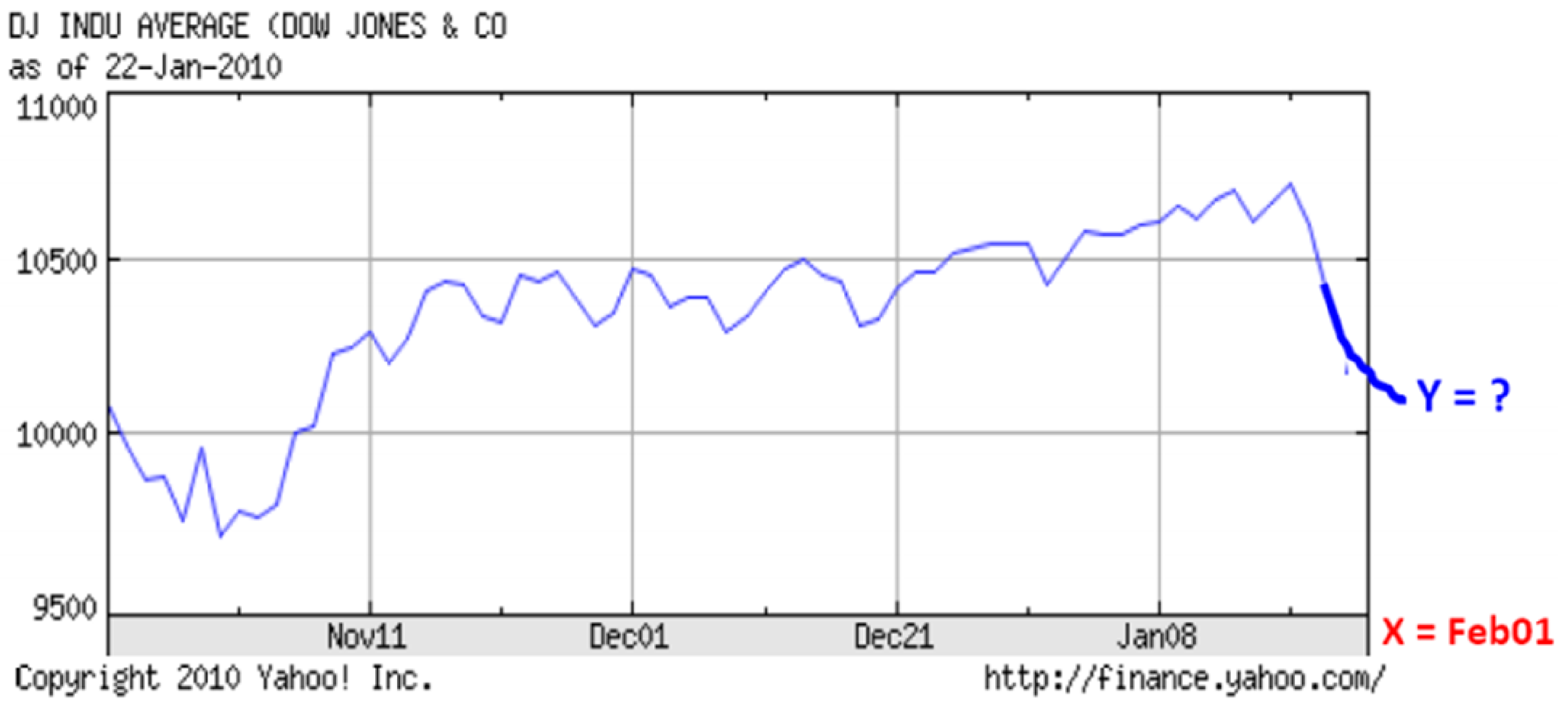


GPT-3 Demo



Financial Analysis

- Stock market prediction

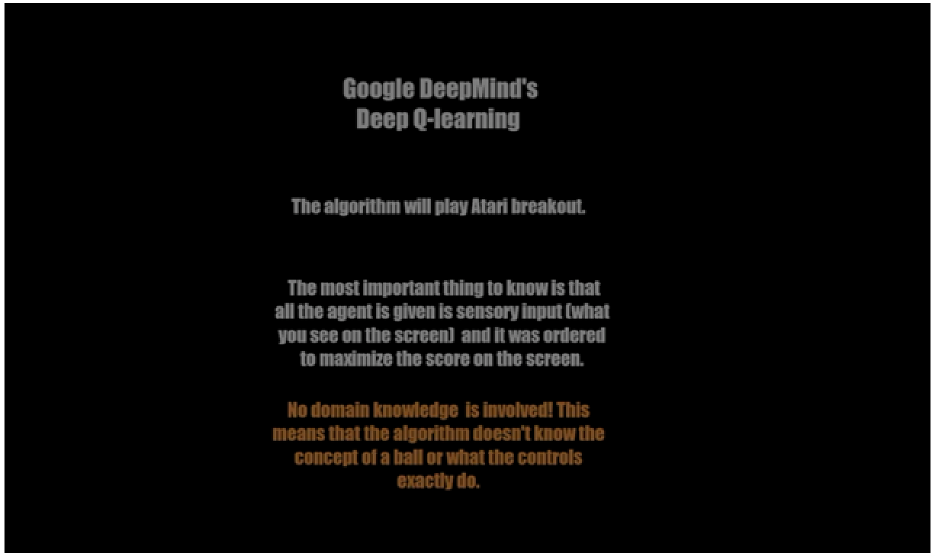


Control

- Cars navigating on their own

Video Games

- Deepmind Deep Q-learning on Atari
  - Mnih et al. Human-level control through deep reinforcement learning. Nature, 518(7540): 529-533, 2015

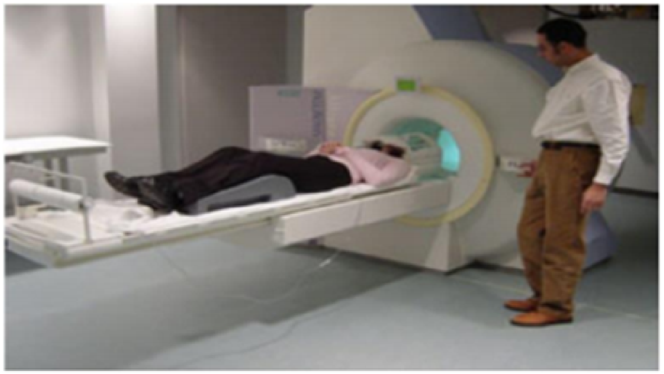


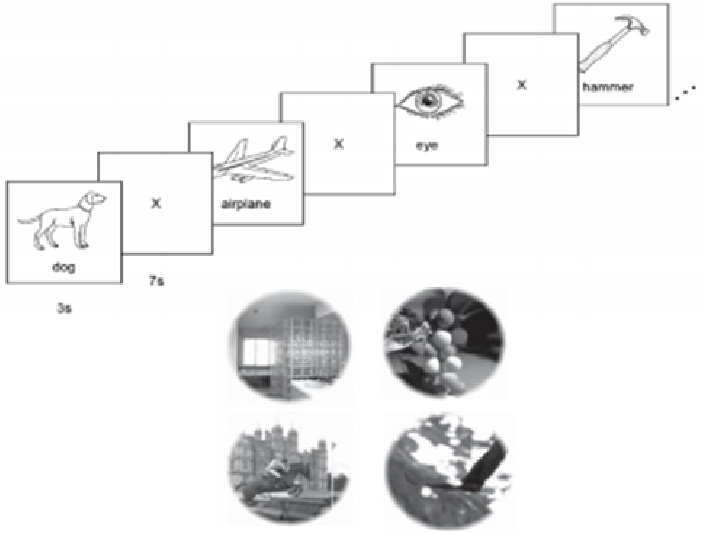
The game of Go

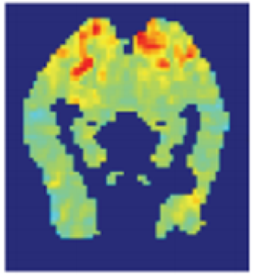
- Deepmind Deep Q-learning on Go
  - Silver et al. Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587): 484–489, 2016

Science

- Decoding thoughts from brain activity



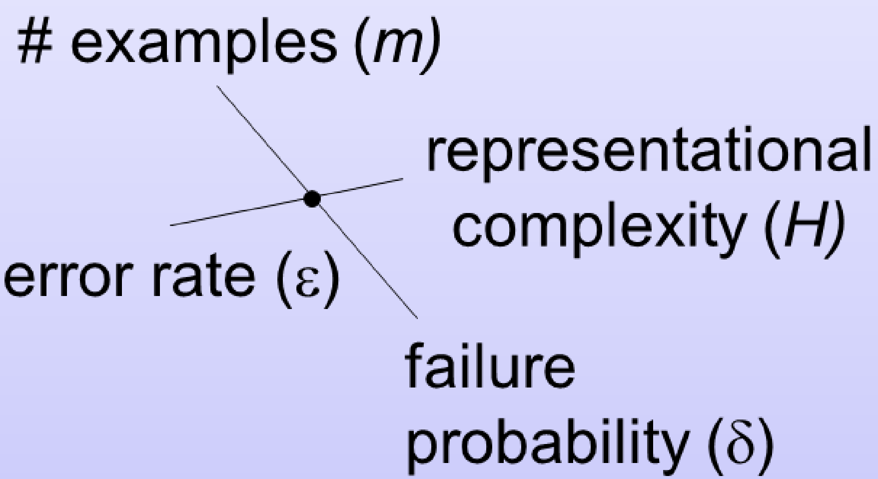




Tool
Animal
...

Machine Learning – theory

PAC Learning Theory
(for supervised concept learning)



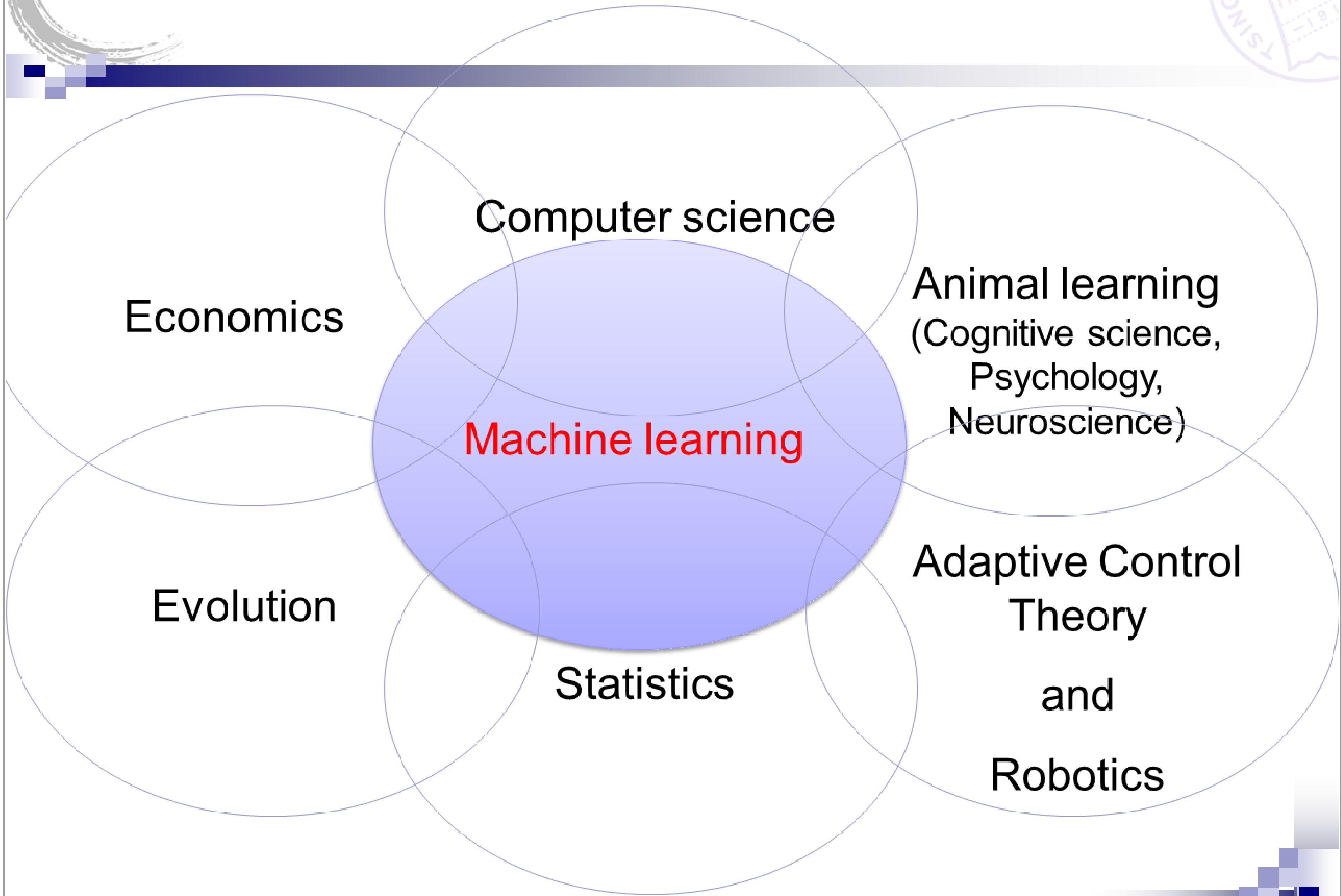
$$m \geq \frac{1}{\epsilon}(\ln|H| + \ln(1/\delta))$$

Other theories for

- Reinforcement skill learning
- Semi-supervised learning
- Active student querying
- …

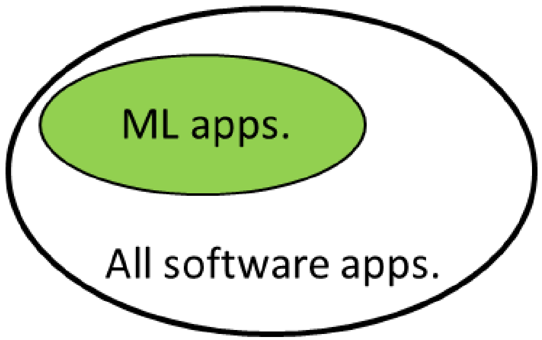
… also relating:

- # of mistakes during learning
- convergence rate
- asymptotic performance
- bias, variance
- VC dimension


Computer science
Economics
Animal learning
(Cognitive science,
Psychology,
Neuroscience)
Machine learning
Evolution
Adaptive Control
Theory
and
Robotics
Statistics

Growth of Machine Learning in CS

- Machine learning already the preferred approach to
  - Speech recognition, natural language process
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - …



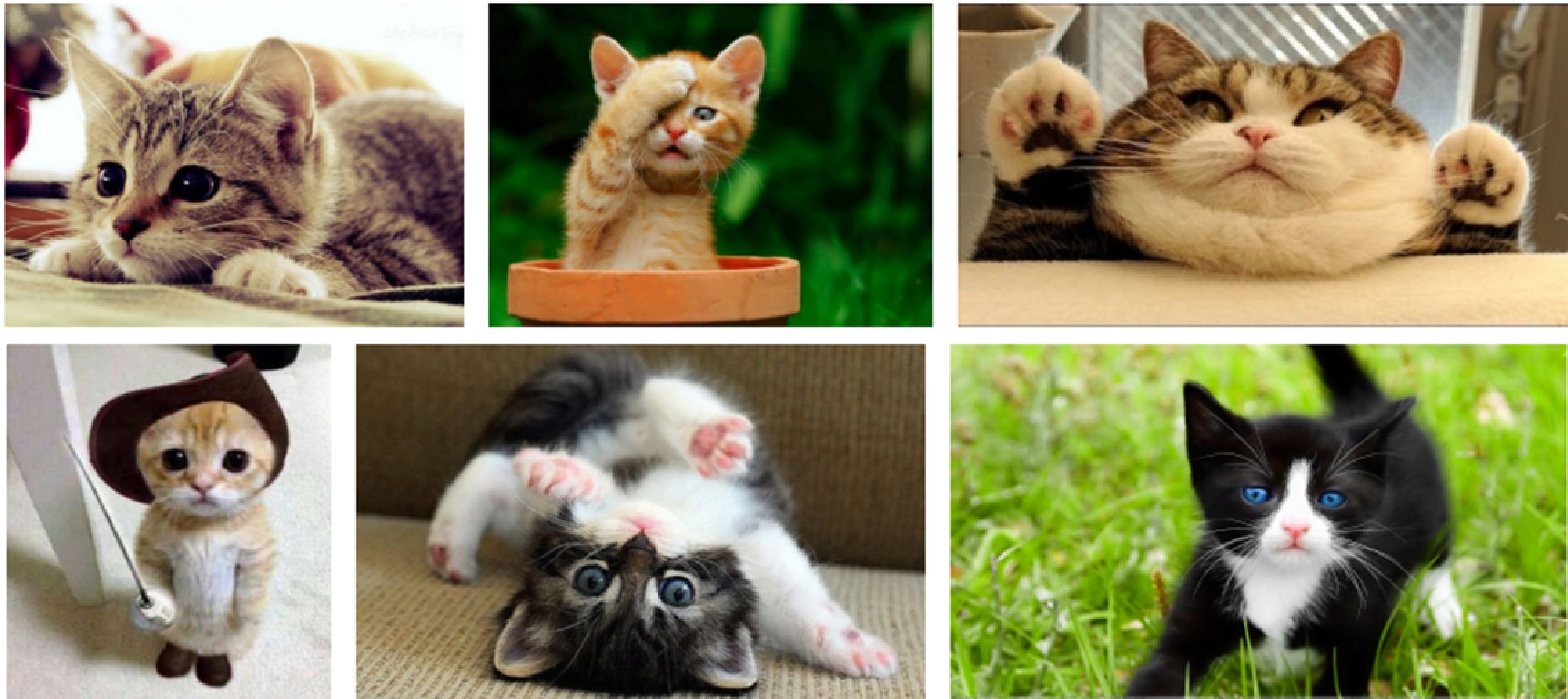
- This ML niche is growing
  - Improved machine learning algorithms
  - Increased data capture, networking, new sensors
  - Software too complex to write by hand
  - Demand for self-customization to user, environment

Machine Learning in CS

- How can we solve a specific problem?
  - As computer scientists we write a program that encodes a set of rules that are useful to solve the problem
  - In many cases is very difficult to specify those rules, e.g., whether there is a cat in the image

Why use learning?

- It is very hard to write programs that solve problems like recognizing a cat
  - What distinguishes a cat from a dog?
  - How does our brain do it?



Take home message

- **Machine learning** is to design and develop algorithms that take as input empirical data, and yield patterns or predictions thought to be features of the underlying mechanism that generated the data

- It turns a small amount of input knowledge into a large amount of output knowledge.

投票 最多可选3项

Which are the essential components to develop a machine learning?

A Data

B Algorithm

C Human specified rules

D Evaluation

Machine Learning Tasks

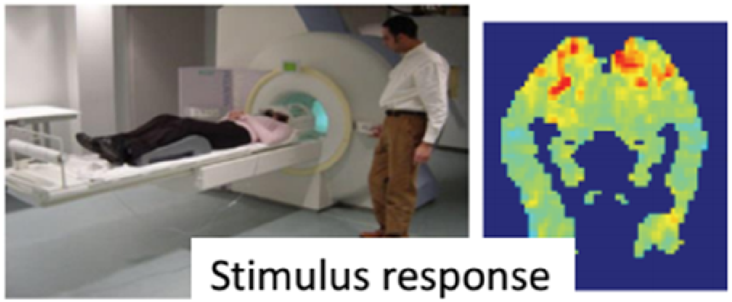
- Broad categories
  - Supervised learning
    - Classification, Regression
  - Unsupervised learning
    - Density estimation, Clustering, Dimensionality reduction
  - Semi-supervised learning
  - Active learning
  - Reinforcement learning
  - Transfer learning
  - Many more …

Machine Learning Task– classification

Feature space $\mathcal{X}$

Label space $\mathcal{Y}$

“Sports”
“News”
“Politics”
…



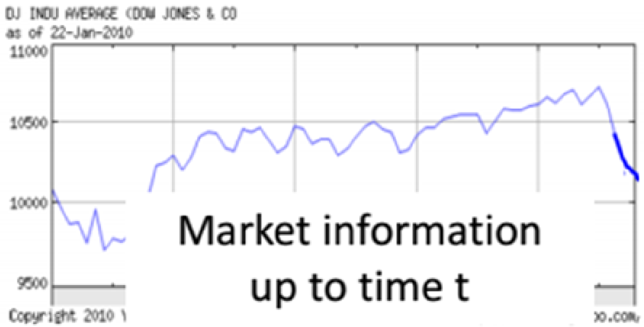
“Tool”
“Animal”
…

Discrete Labels

Machine Learning Task– Regression

Feature space $\mathcal{X}$

Label space $\mathcal{Y}$



Market information
up to time t

Share price
“$ 20.50”

(session, location, time …)

Temperature
“42° F”

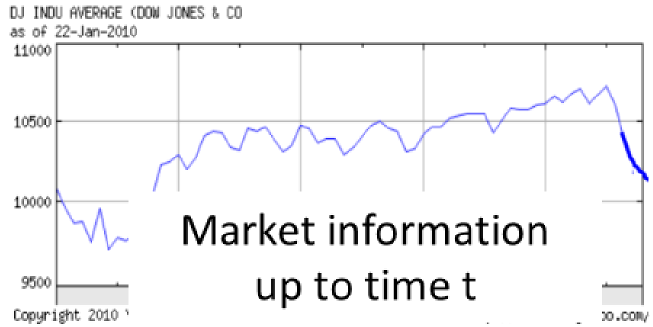
Continuous Labels

雨课堂 Rain Classroom

Supervised Learning

- Task: learn a predictive function $h: \mathcal{X} \rightarrow \mathcal{Y}$

Feature space $\mathcal{X}$



Label space $\mathcal{Y}$

"Sports"
"News"
"Politics"
…

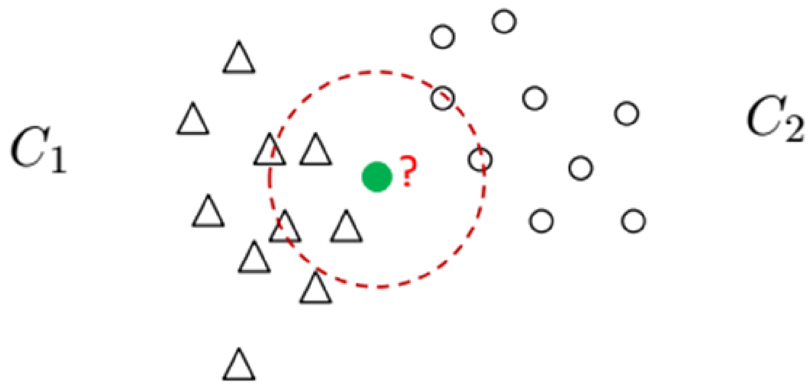
Share price
"$ 20.50"

knowledge

- "Experience" or training data:

$$\{< \mathrm{x}_d, y_d >\}_{d=1}^{D}, \ \mathrm{x}_d \in \mathcal{X}, y_d \in \mathcal{Y}$$

How to learn a classifier?



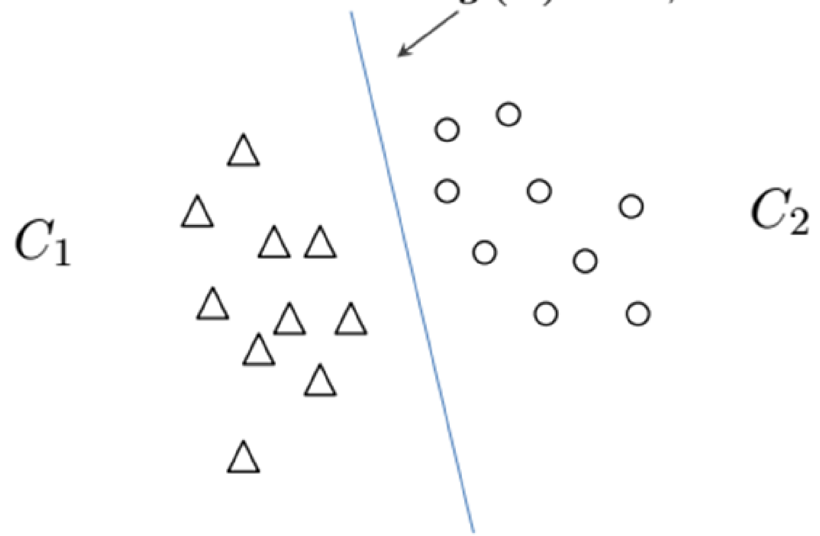
K-NN: a Non-parametric approach

Distance metric matters!

How to learn a classifier?

- Parametric (model-based) approaches:

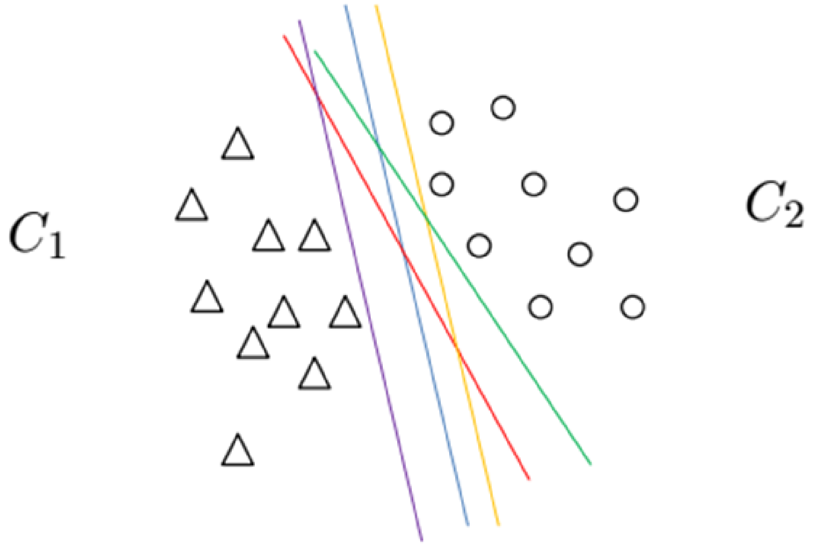
$$g(x) = 0, \;\; where \;\; g(x) = w^\top x + w_0$$



a good decision boundary

$$y^* = \begin{cases} C_1 & if \;\; g(x) > 0 \\ C_2 & if \;\; g(x) < 0 \end{cases}$$

How to learn a classifier?



Many good decision boundaries！！

Which one should we choose?

单选题 1分

设置

此题未设置答案，请点击右侧设置按钮

Which one is the best boundary?

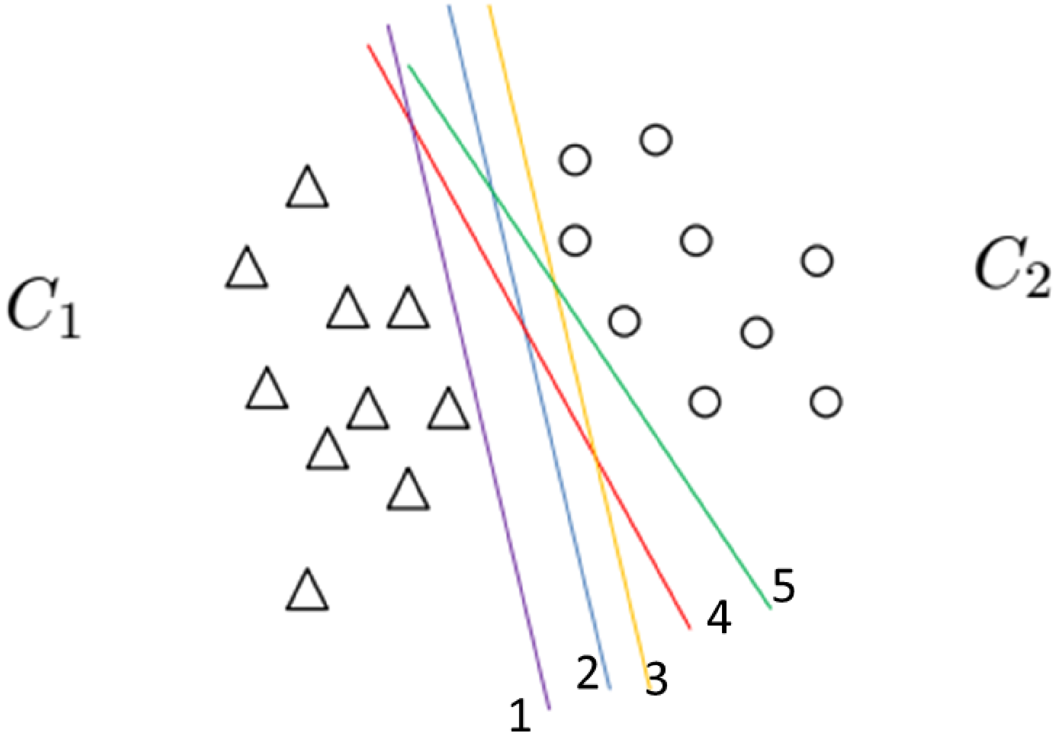
A 1

B 2

C 3

D 4

E 5

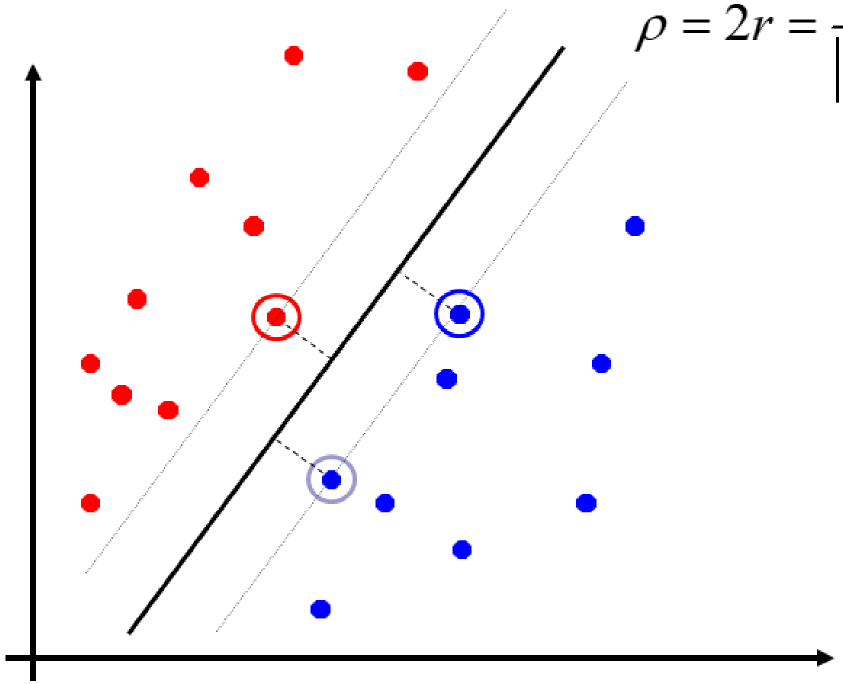


提交

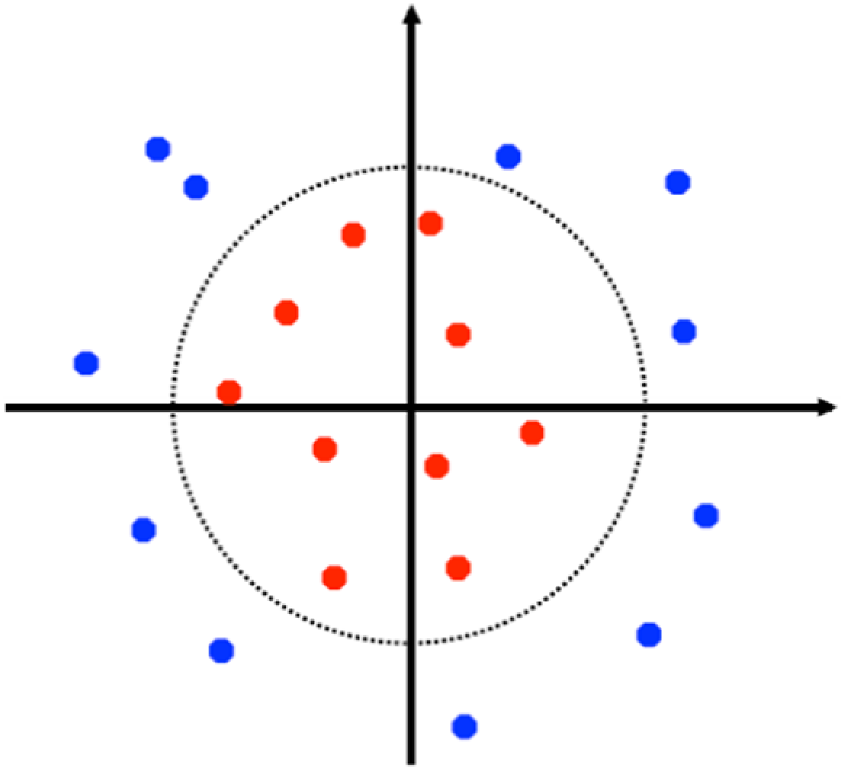
Max-margin Classification

- Maximizing the margin is good according to intuition and PAC theory.
- Implies that only support vectors matter; other training examples are ignorable.

$$\rho = 2r = \frac{2}{\|\mathbf{w}\|}$$



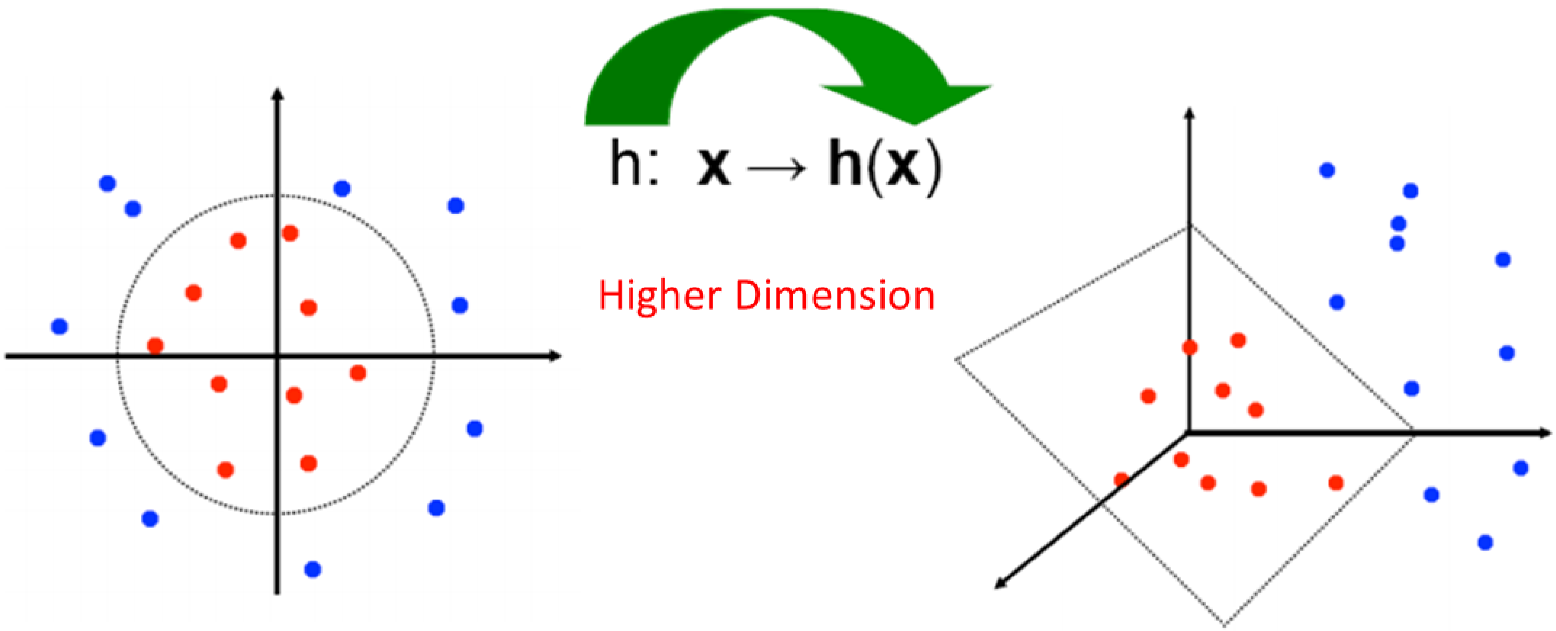
How to learn a classifier?



How about non-linearity?

52

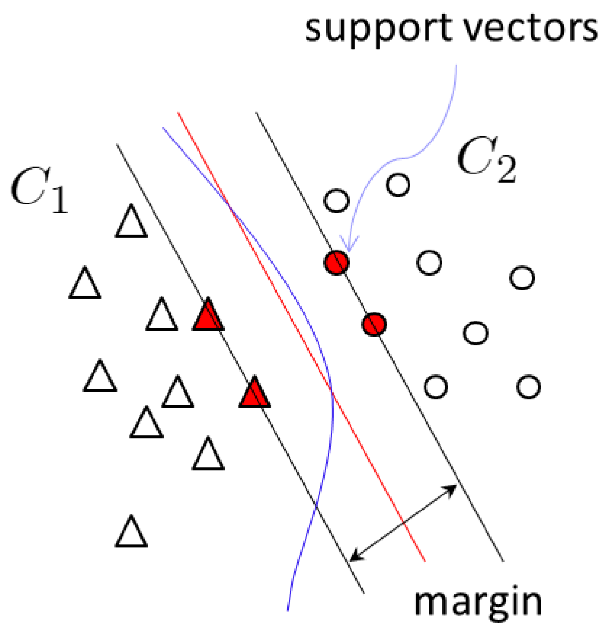
How to learn a classifier?



The higher dimension, the better?

How to learn a classifier?

- Support vector machines (SVM) – basics
  - SVM is among the most popular/successful classifiers
  - It provides a *principled way* to learn a *robust* classifier (i.e., a *decision boundary*)



- SVM
  - chooses the one with *maximum margin principle*
  - has sound *theoretical guarantee*
  - extends to *nonlinear decision boundary* by using *kernel* trick
  - learning problem efficiently solved using convex optimization techniques

How to learn a classifier?

- Naïve Bayes classifier – basics
  - an representative method from the very important family of *probabilistic graphical models* and *Bayesian methods*

A joint distribution： $p(x,y) = p(y)p(x|y)$

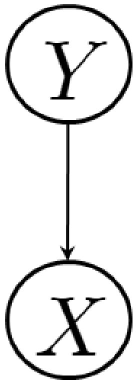
Inference using Bayes rule：

prior likelihood

$$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(y)p(x|y)}{p(x)}$$

evidence

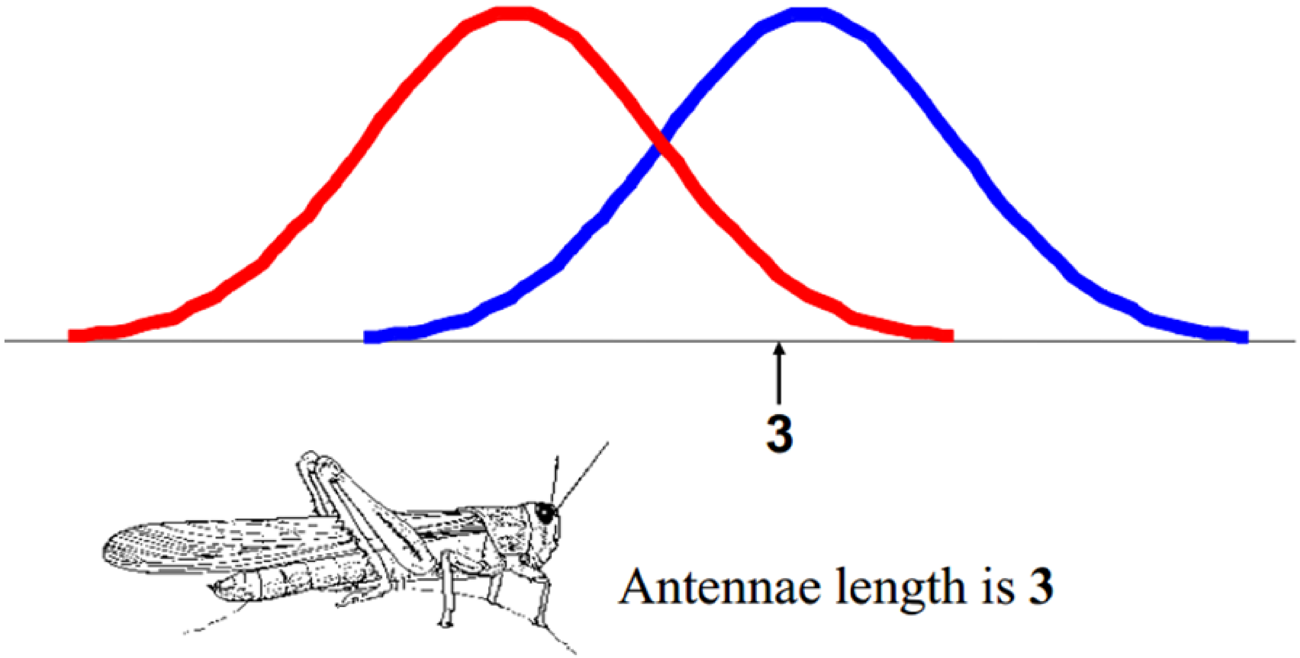
Prediction rule： $y^* = \arg\max_{y \in \mathcal{Y}} p(y|x)$



  - fundamental building blocks for *Bayesian networks*
  - nice illustrative example of Bayesian methods

Classification?

- Classify another insect we find. Its antennae are 3 units long
- Is it more probable that the insect is a **Grasshopper** or a **Katydid**?

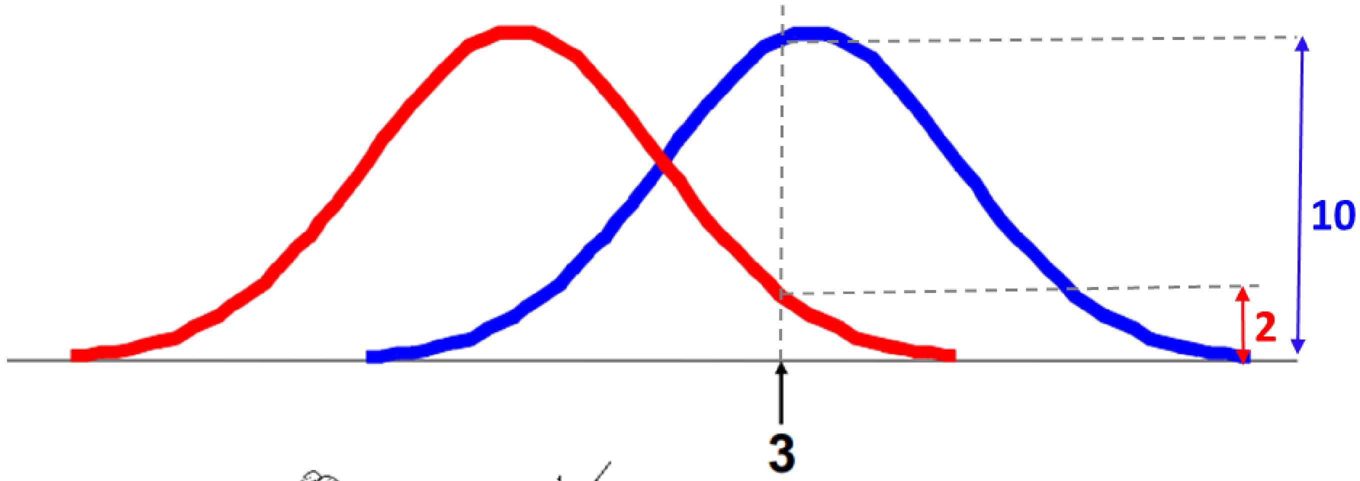


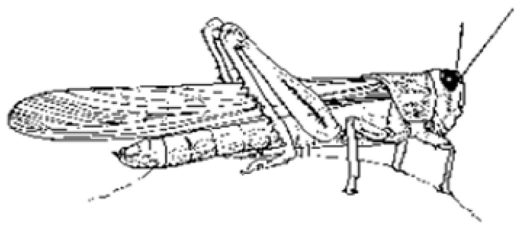
[Courtesy of E. Keogh]

Classification Probability

P(**Grasshopper** | **3**) = 10 / (10 + 2) = 0.833

P(**Katydid** | **3**) = 2 / (10 + 2) = 0.166





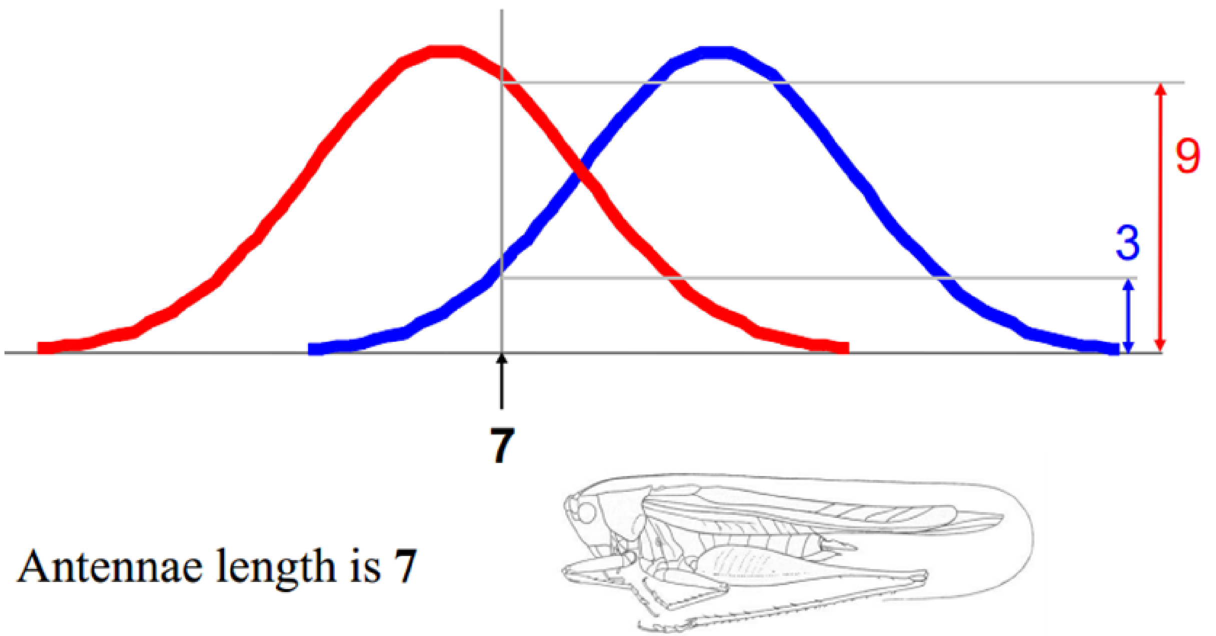
Antennae length is **3**

[Courtesy of E. Keogh]

Classification Probability

P(**Grasshopper** | **7**) = 3 / (3 + 9) = 0.250

P(**Katydid** | **7**) = 9 / (3 + 9) = 0.750

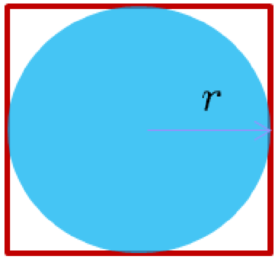


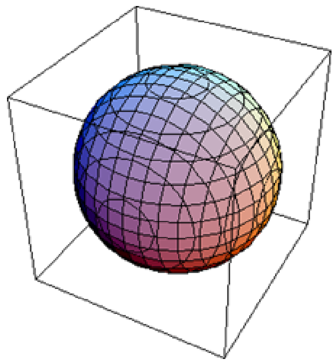
[Courtesy of E. Keogh]

雨课堂 Rain Classroom

How to learn a classifier?

- Curse of dimensionality
  - A high dimensional space is always almost empty





d dimensional space

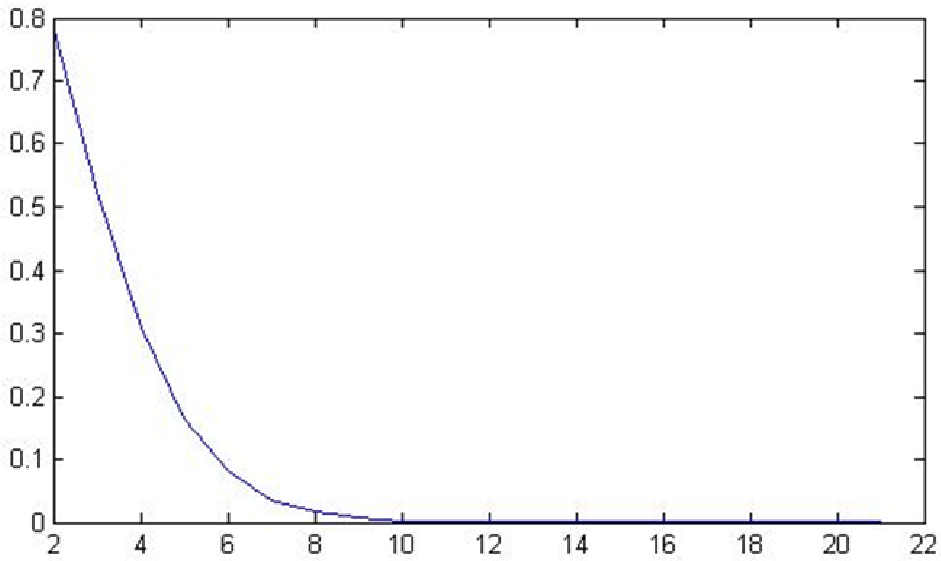
$$\frac{\pi r^2}{(2r)^2} = \frac{\pi}{4}$$

$$\frac{\dfrac{2r^3\pi^{3/2}}{3\Gamma(3/2)}}{(2r)^3} = \frac{\pi^{3/2}}{12\Gamma(3/2)}$$

$$\frac{\dfrac{2r^d\pi^{d/2}}{d\Gamma(d/2)}}{(2r)^d} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma(d/2)} \xrightarrow{d \to \infty} 0$$

How to learn a classifier?

- Curse of dimensionality
  - A high dimensional space is always almost empty



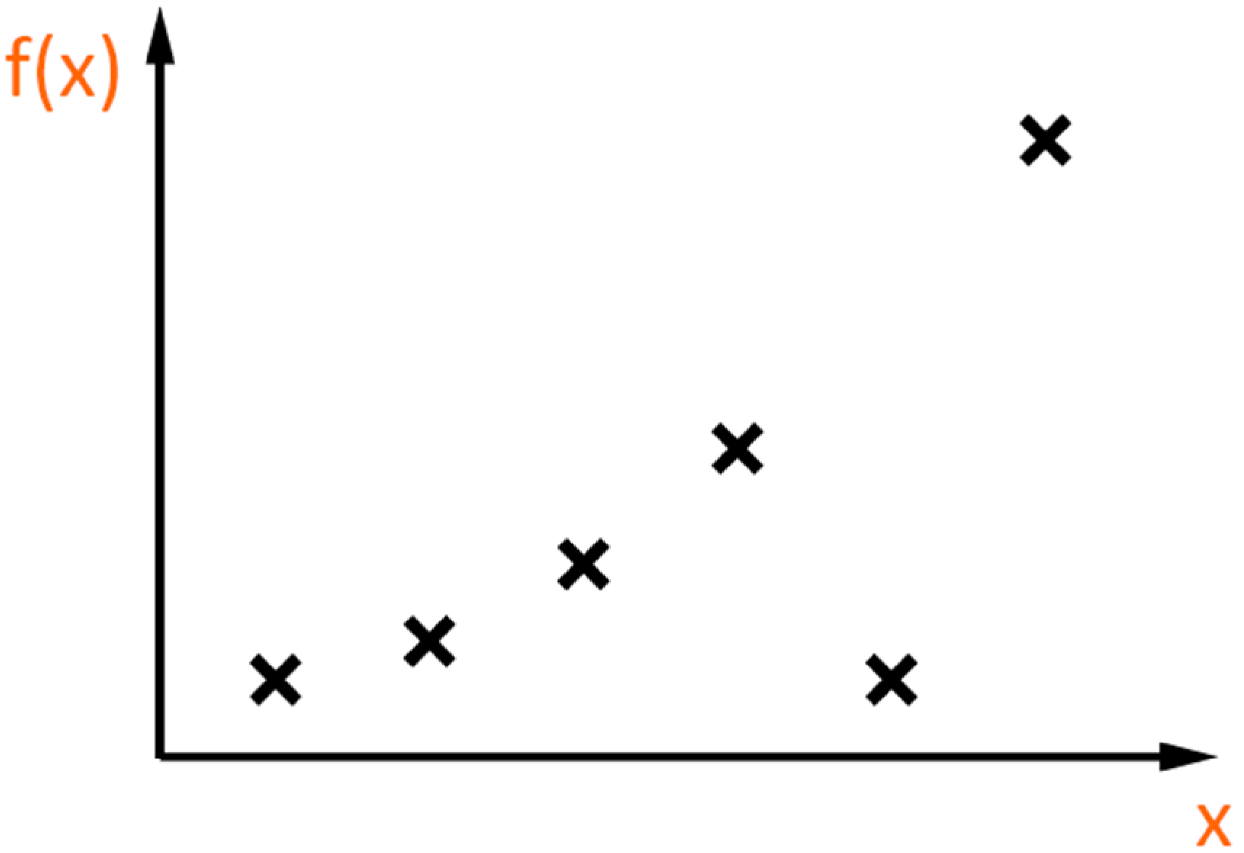
when one wants to learn pattern from data in high dimensions no matter how much data you have it always seems less!

How to learn a classifier?

- Curse of dimensionality
  - A high dimensional space is always almost empty
  - … in high dimensions no matter how much data you have it always seems less!
- The blessing of dimensionality
  - *… real data highly concentrate on low-dimensional, sparse, or degenerate structures in the high-dimensional space.*
- But no free lunch: *Gross errors and irrelevant measurements are now ubiquitous in massive cheap data.*

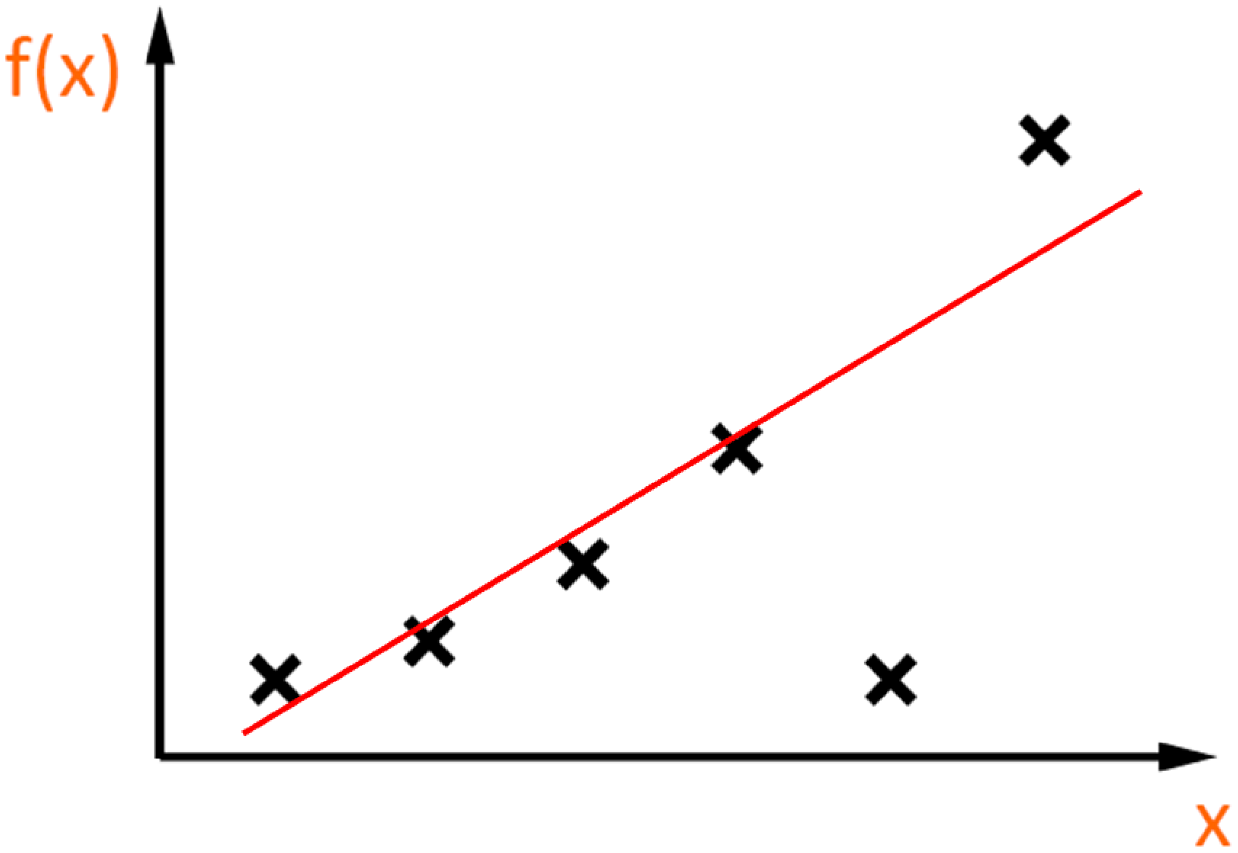
Are complicated models preferred?

- A simple curve fitting task



Are complicated models preferred?

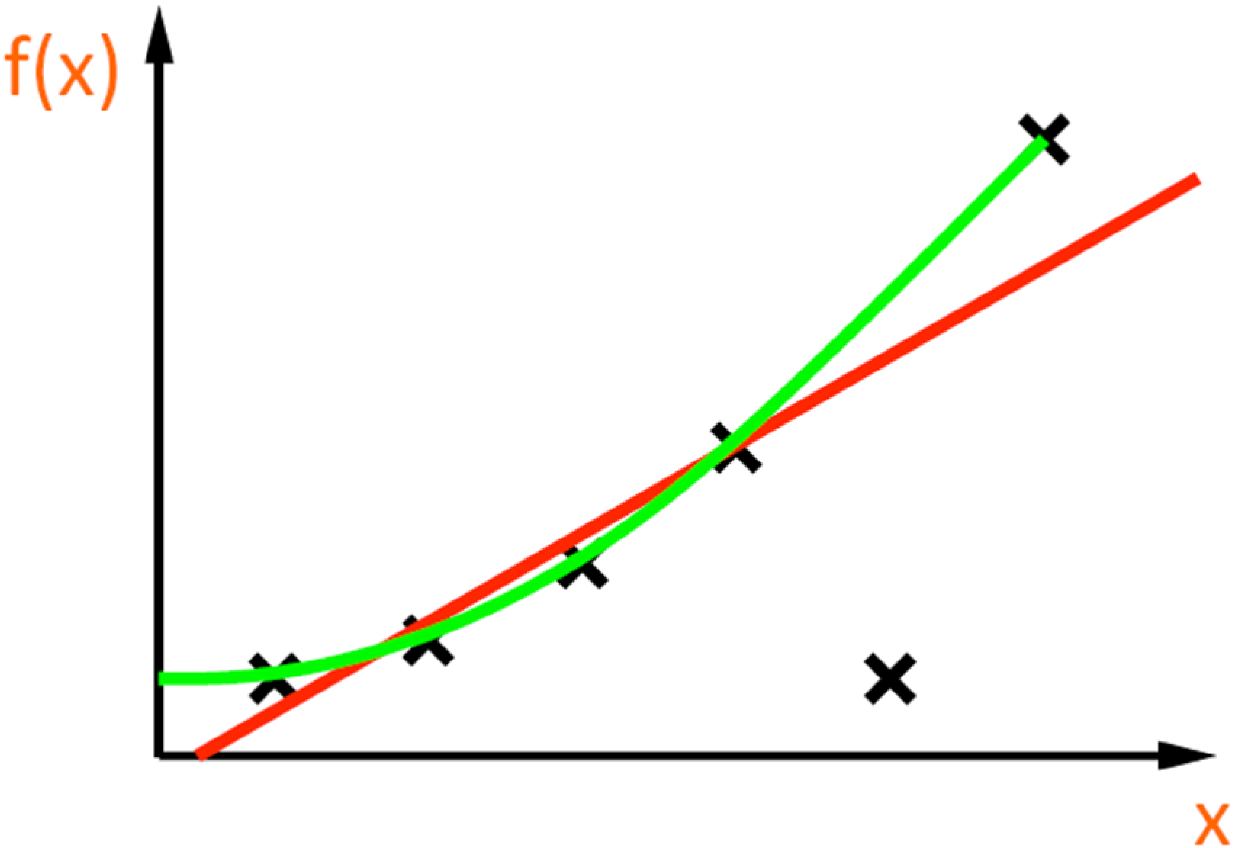
- Order = 1



63

Are complicated models preferred?

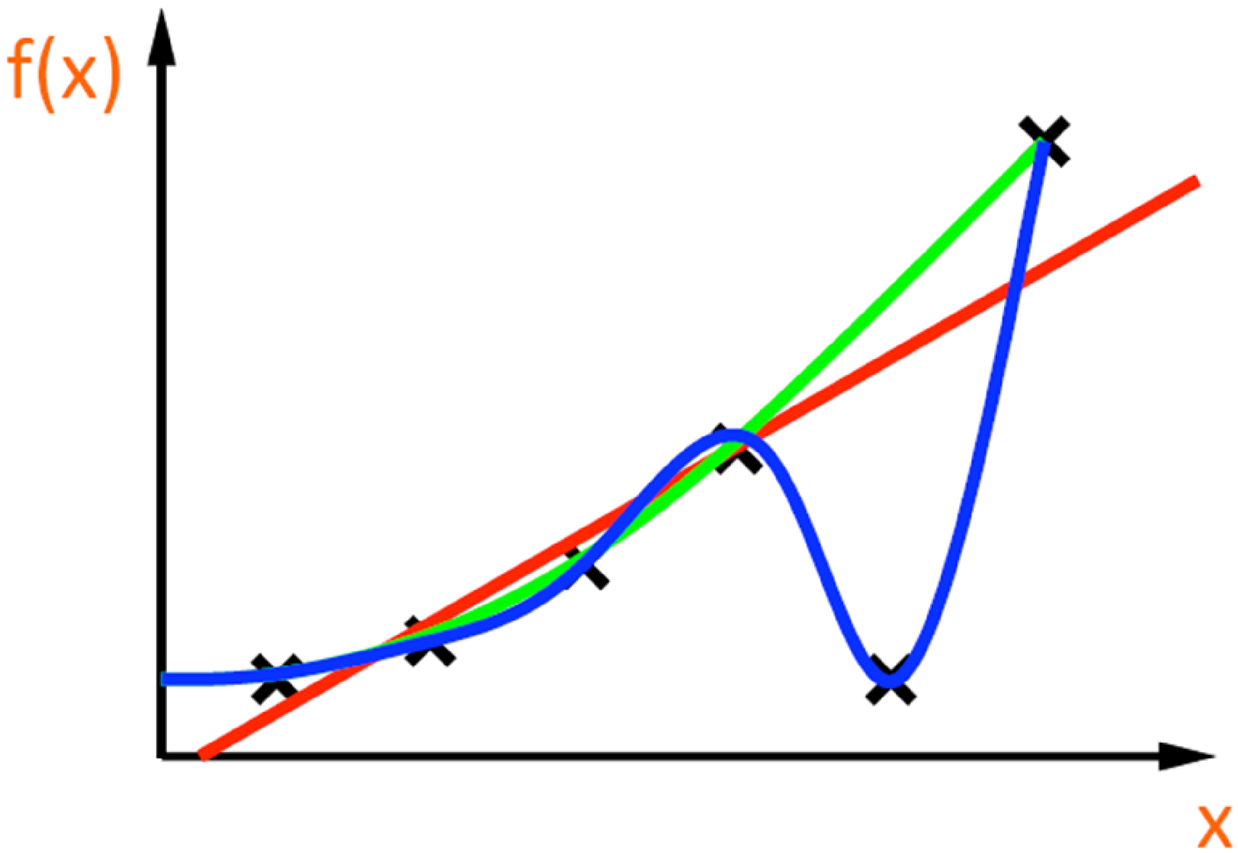
- Order = 2



64

Are complicated models preferred?

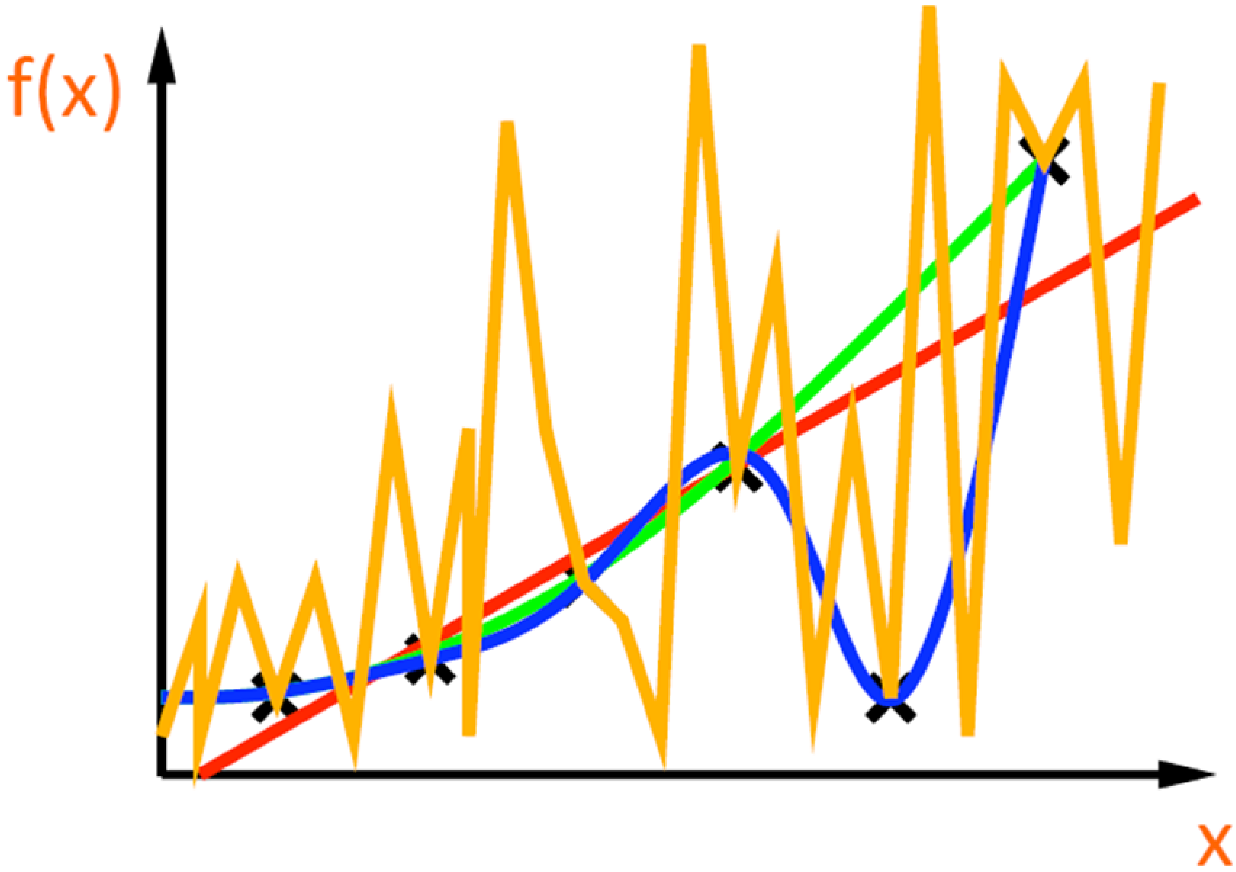
- Order = 3



雨课堂 Rain Classroom

Are complicated models preferred?

- Order = 9?



66

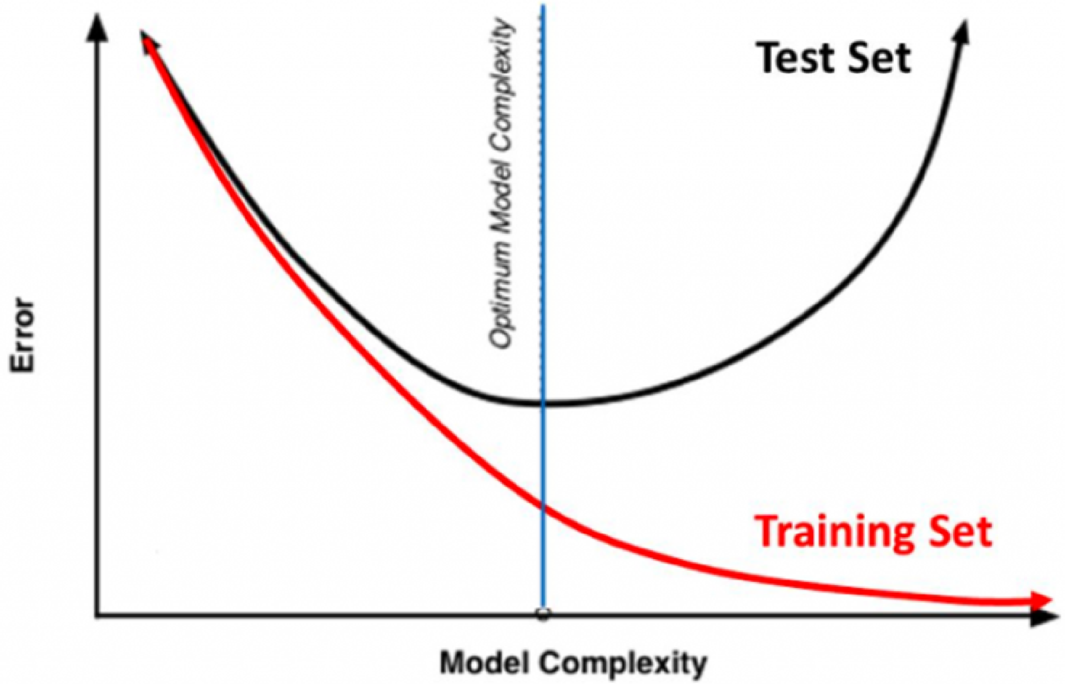
Are complicated models preferred?

Advice: use ML & sophisticated models when necessary

- Issues with model selection!!

It's Generalization that counts

- The fundamental goal of machine learning is to generalize beyond the examples in the training set
- We have to use training error as a surrogate for test error



Unsupervised Learning

- Task: learn an explanatory function
- Aka "Learning without a teacher" $f(x),\ x \in \mathcal{X}$

Feature space $\mathcal{X}$

Words in documents

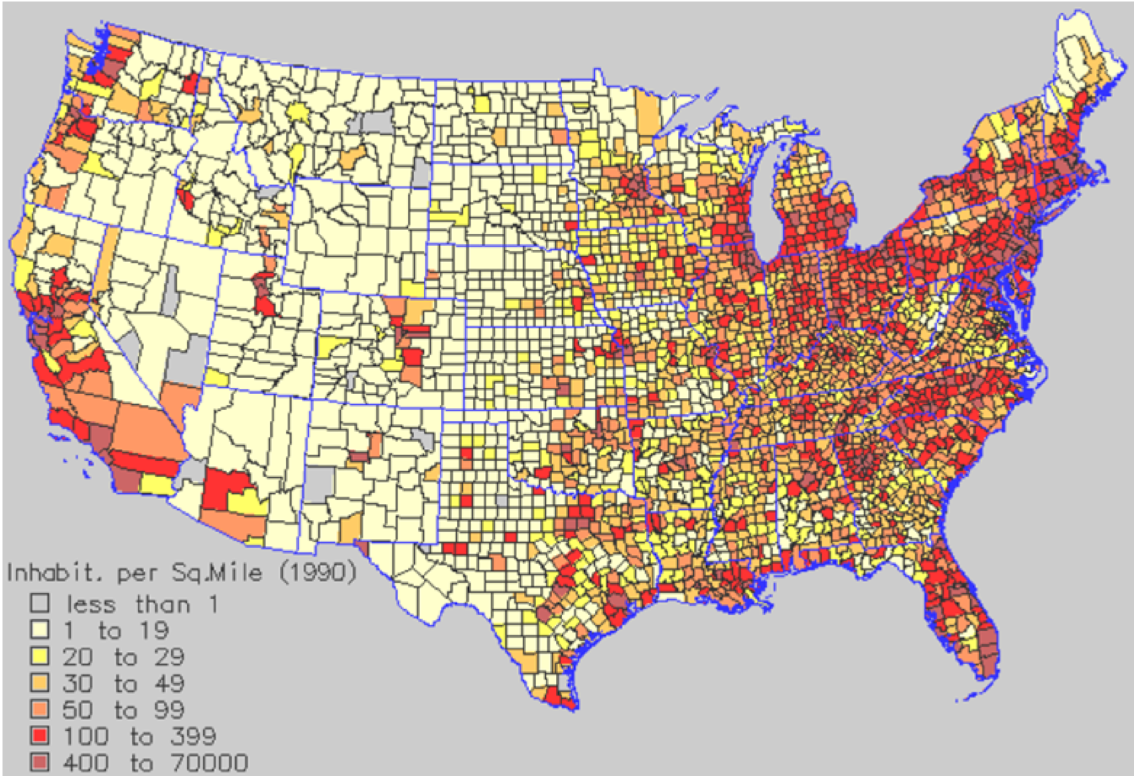
$\Rightarrow$

Word distribution
(probability of a word)

- No training/test split

69

Unsupervised Learning – density estimation

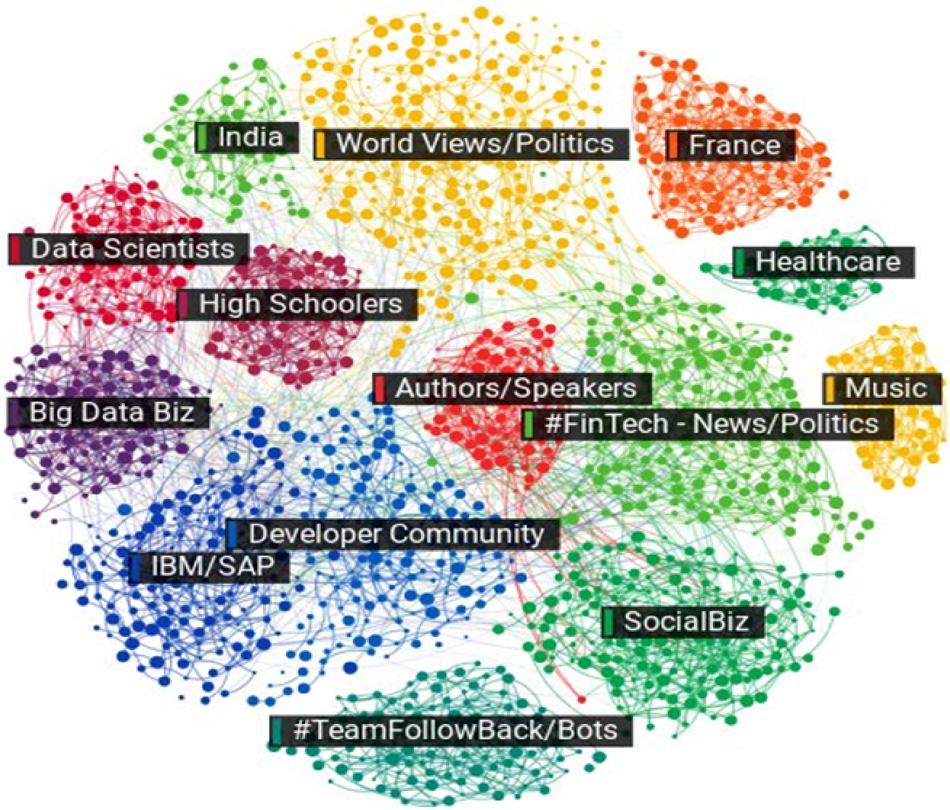


Feature space $\mathcal{X}$
geographical information of a location

Density function
$f(x), \ x \in \mathcal{X}$

70

Unsupervised Learning – clustering

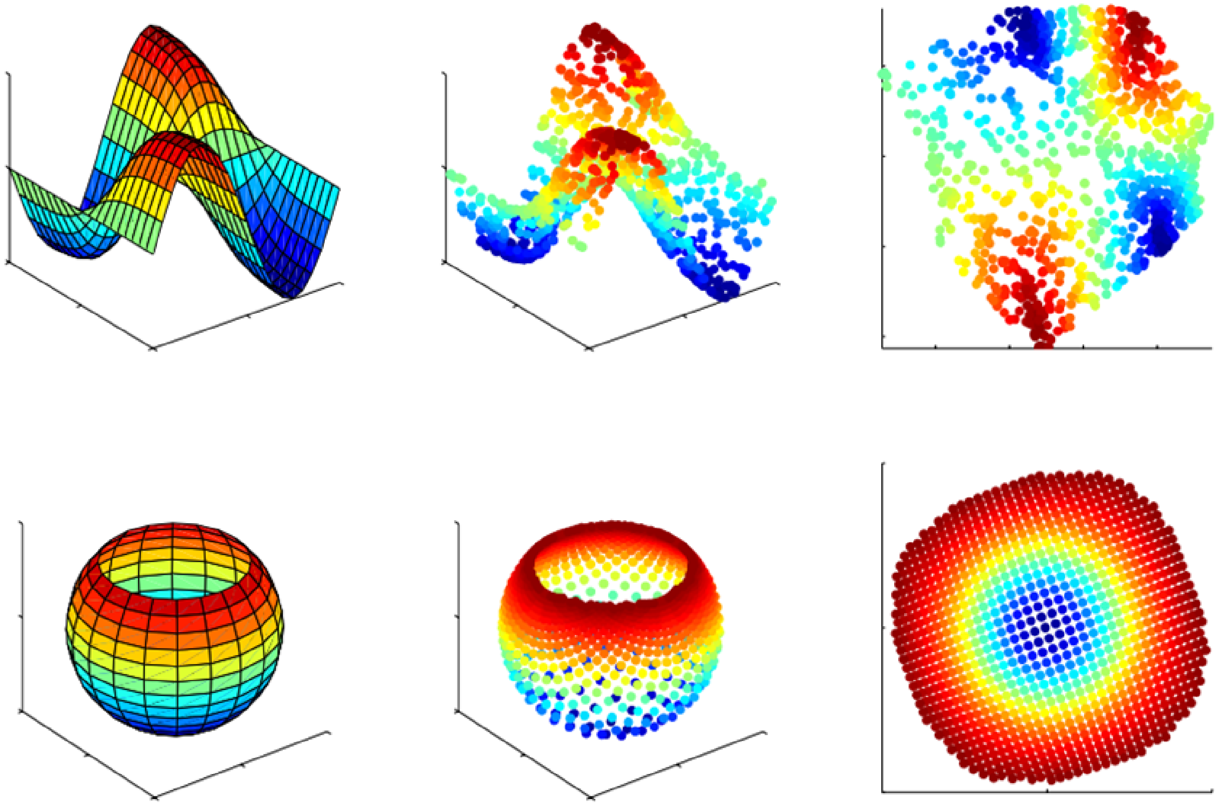


Feature space $\mathcal{X}$
Attributes (e.g., pixels & text)

Cluster assignment function
$f(x), \ x \in \mathcal{X}$

71

Unsupervised Learning – dimensionality reduction



Feature space $\mathcal{X}$
pixels of images

Coordinate function in 2D space
$f(x), \ x \in \mathcal{X}$

72

雨课堂 Rain Classroom

Summary: what is machine learning

- Machine Learning seeks to develop theories and computer systems for

dealing with

- complex, real world data, based on the system's own experience with data, and (hopefully) under a unified model or mathematical framework, that

have nice properties.

73

雨课堂 Rain Classroom

Summary: what is machine learning

- Machine Learning seeks to develop theories and computer systems for
  - representing;
  - classifying, clustering, recognizing, organizing;
  - reasoning under uncertainty;
  - predicting;
  - and reacting to
  - …
- complex, real world data, based on the system's own experience with data, and (hopefully) under a unified model or mathematical framework, that

have nice properties.

Summary: what is machine learning

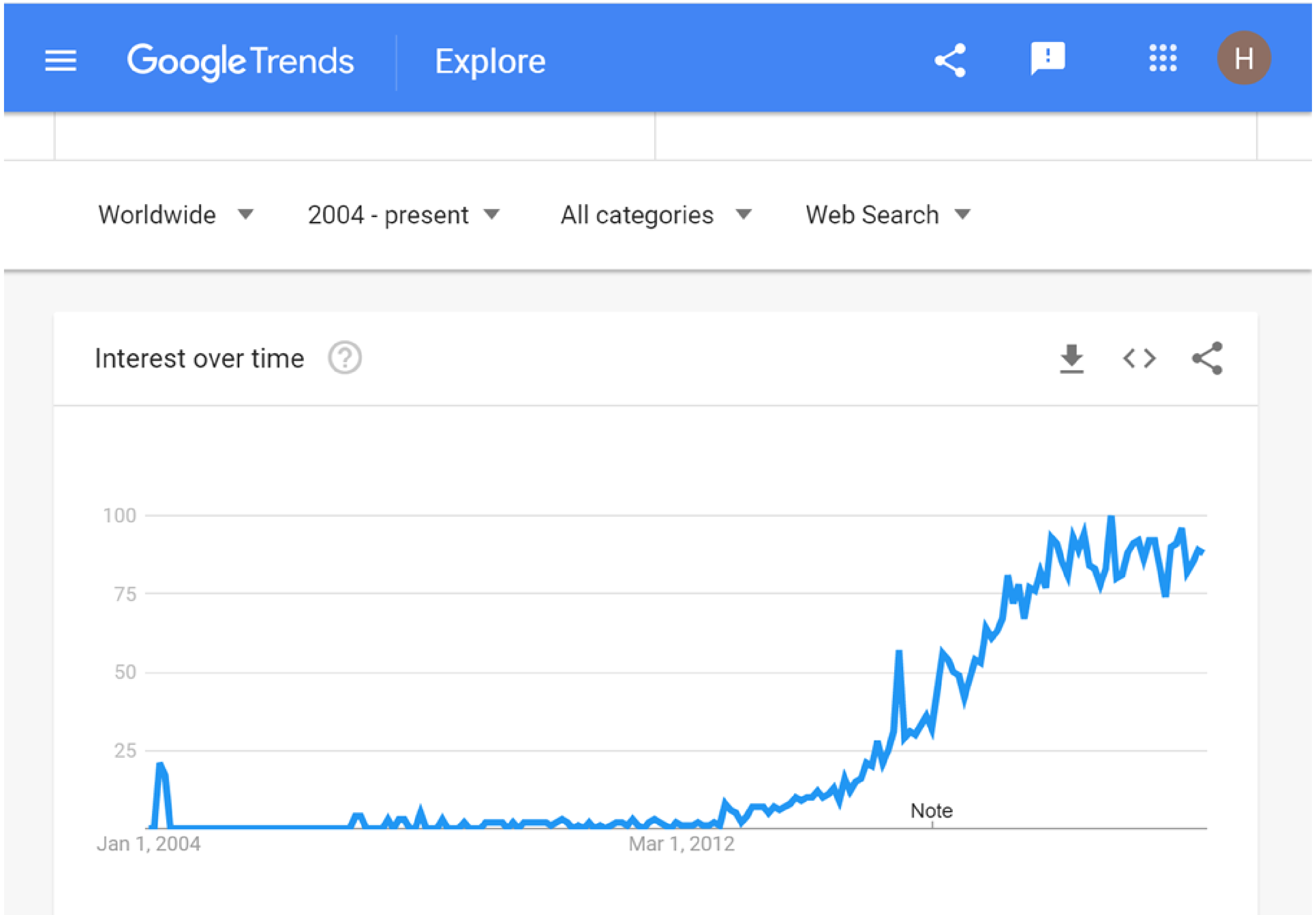
- Machine Learning seeks to develop theories and computer systems for
  - representing;
  - classifying, clustering, recognizing, organizing;
  - reasoning under uncertainty;
  - predicting;
  - and reacting to
  - …
- complex, real world data, based on the system's own experience with data, and (hopefully) under a unified model or mathematical framework, that
  - can be formally characterized and analyzed;
  - can take into account human prior knowledge;
  - can generalize and adapt across data and domains;
  - can operate automatically and autonomously;
  - and can be interpreted and perceived by human.
- ML covers algorithms, theory and very exciting applications
- It's going to be fun and challenging ☺

Recent Progress

Representation Learning (Deep Learning)

雨课堂 Rain Classroom

Deep Learning is popular



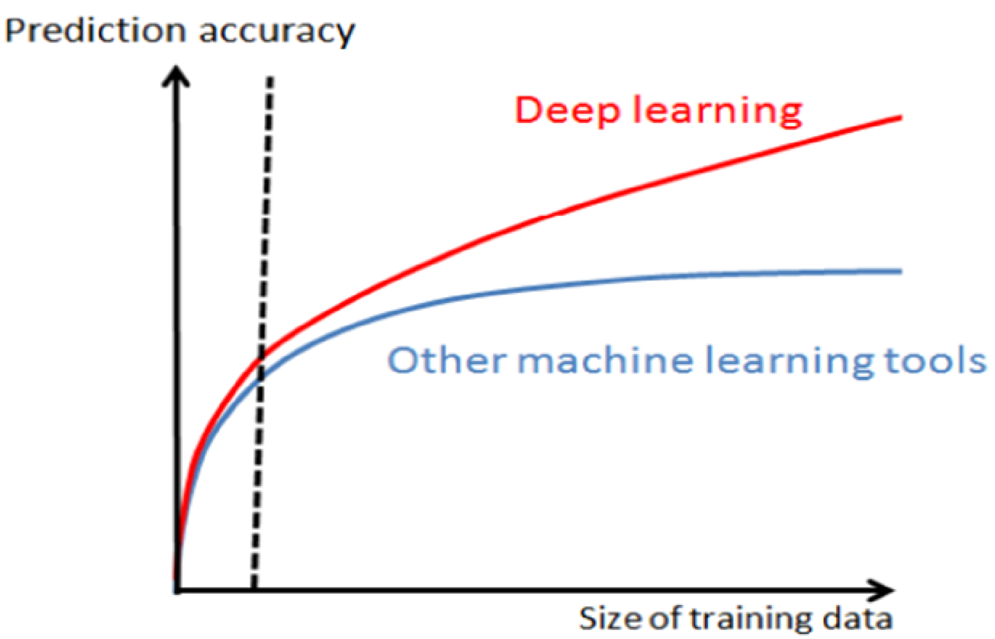
ACM A.M. Turing Award 2019

- Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award
- Bengio, Hinton and LeCun Ushered in Major Breakthroughs in Artificial Intelligence

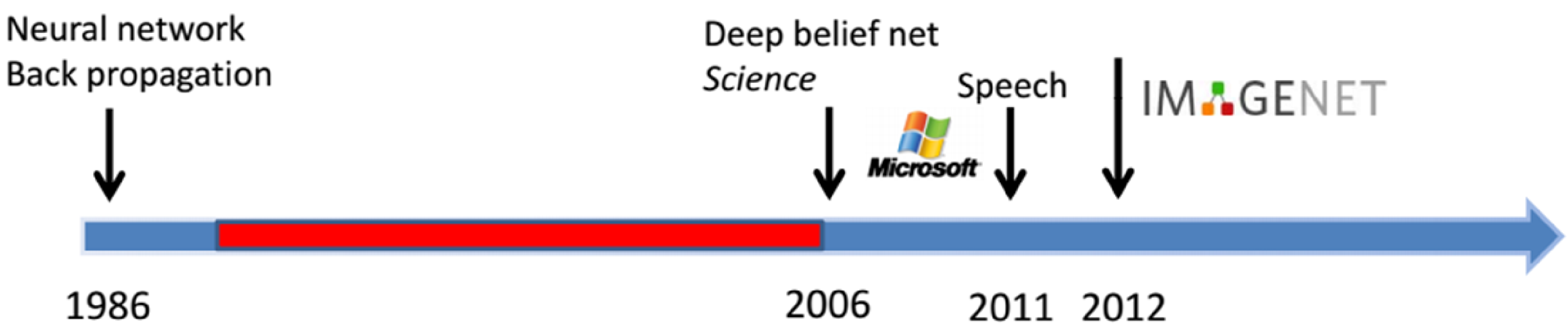
78

Machine Learning with Big Data

- Machine learning with small data: overfitting, reducing model complexity (capacity)
- Machine learning with big data: underfitting, increasing model complexity, optimization, computation resource



Examples from ImageNet



- ImageNet 2013 – image classification challenge

Rank	Name	Error rate	Description
1	NYU	0.11197	Deep learning
2	NUS	0.12535	Deep learning
3	Oxford	0.13555	Deep learning

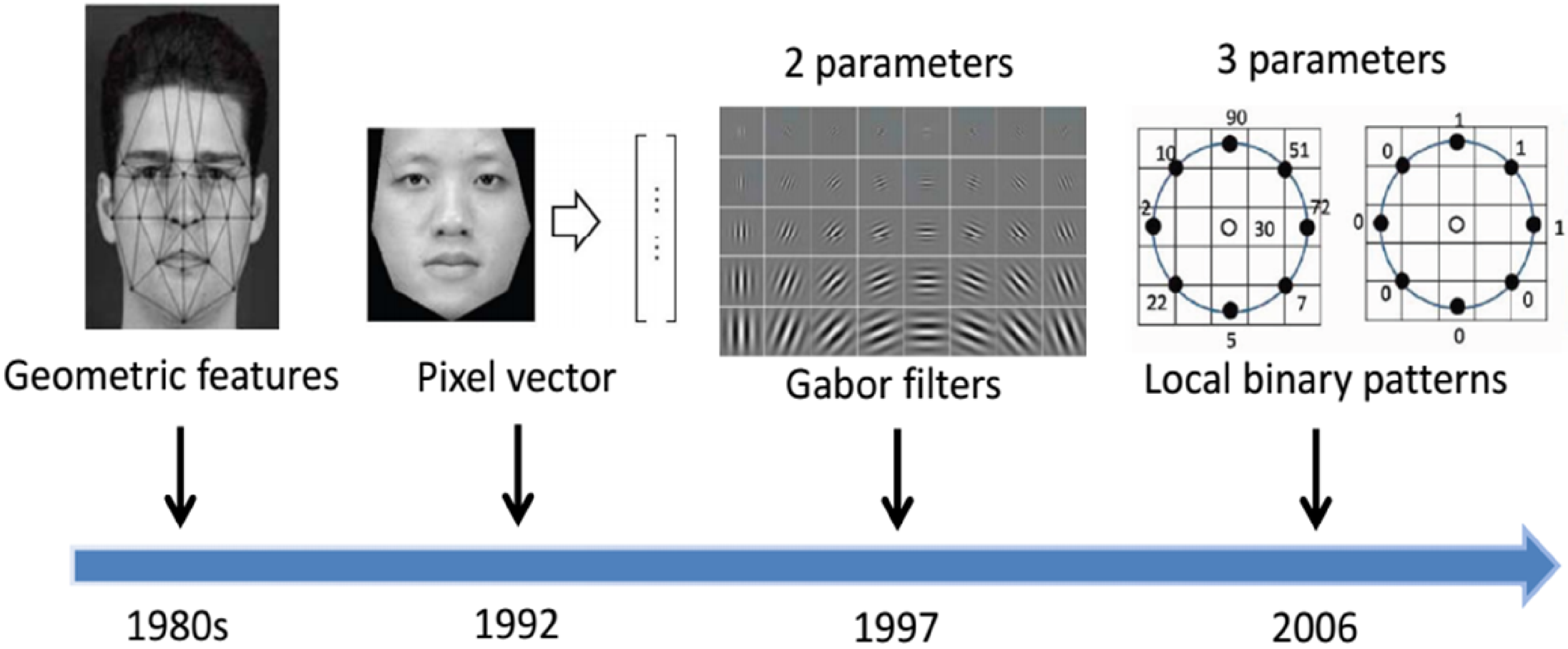
MSRA, IBM, Adobe, NEC, Clarifai, Berkley, U. Tokyo, UCLA, UIUC, Toronto .... Top 20 groups all used deep learning

- ImageNet 2013 – object detection challenge

Rank	Name	Mean Average Precision	Description
1	UvA-Euvision	0.22581	Hand-crafted features
2	NEC-MU	0.20895	Hand-crafted features
3	NYU	0.19400	Deep learning

Feature Engineering

- Handcrafted Features for Face Recognition



Then., two things happened...

- **The ImageNet dataset [Fei-Fei et al. 2012]**
  - 1.2 million training samples
  - 1000 categories

- **Fast Graphical Processing Units (GPU)**
  - Capable of 1 trillion operations/second

雨课堂 Rain Classroom

Deep Learning = Learning Hierarchical Representations

- **Traditional Pattern Recognition:** Fixed/Handcrafted Feature Extractor

Feature Extractor → Trainable Classifier →

- **Mainstream Modern Pattern Recognition:** Unsupervised mid-level features

Feature Extractor → Mid-Level Features → Trainable Classifier →

- **Deep Learning:** Representations are hierarchical and trained

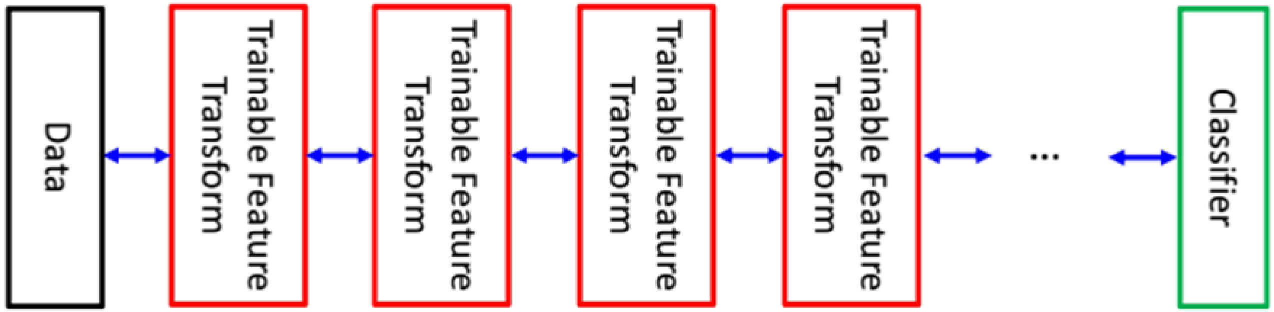
Low-Level Features → Mid-Level Features → High-Level Features → Trainable Classifier →

2020/9/15

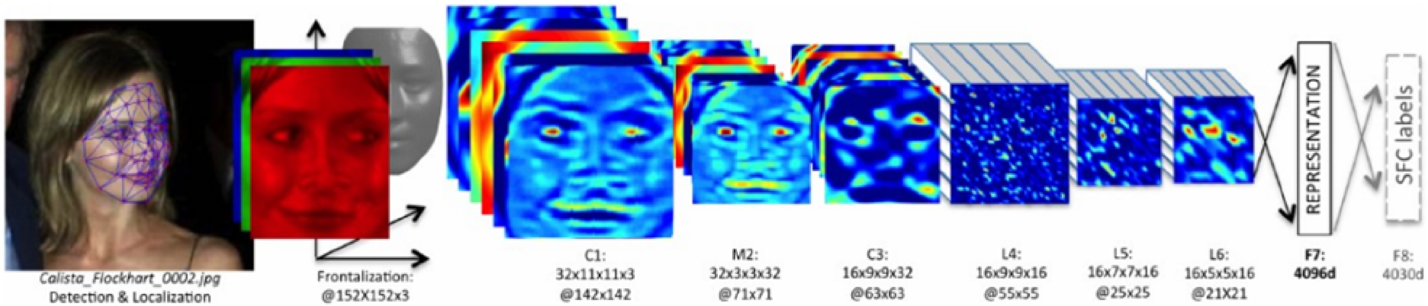
Deep Learning Means Feature Learning

- Deep learning is about learning hierarchical feature representations

$$\mathbf{y} = F(\mathbf{W}^k(F(\mathbf{W}^{k-1}F(\cdots F(\mathbf{W}^0 \cdot \mathbf{x})))))$$



- Deep learning: Hierarchical; nonlinear



The deeper, the better

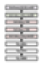
- The deeper network can cover more complex problems
  - Receptive field size ↑
  - Non-linearity ↑
- However, training the deeper network is more difficult because of vanishing/exploding gradients problem

CVPR Best Paper Award :

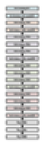
Deep Residual Learning for Image Recognition
Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun

Revolution of Depth

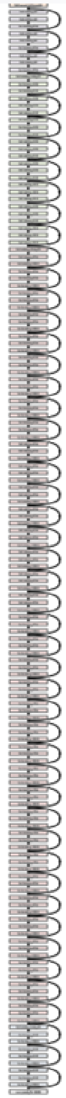
AlexNet, 8 layers
(ILSVRC 2012)

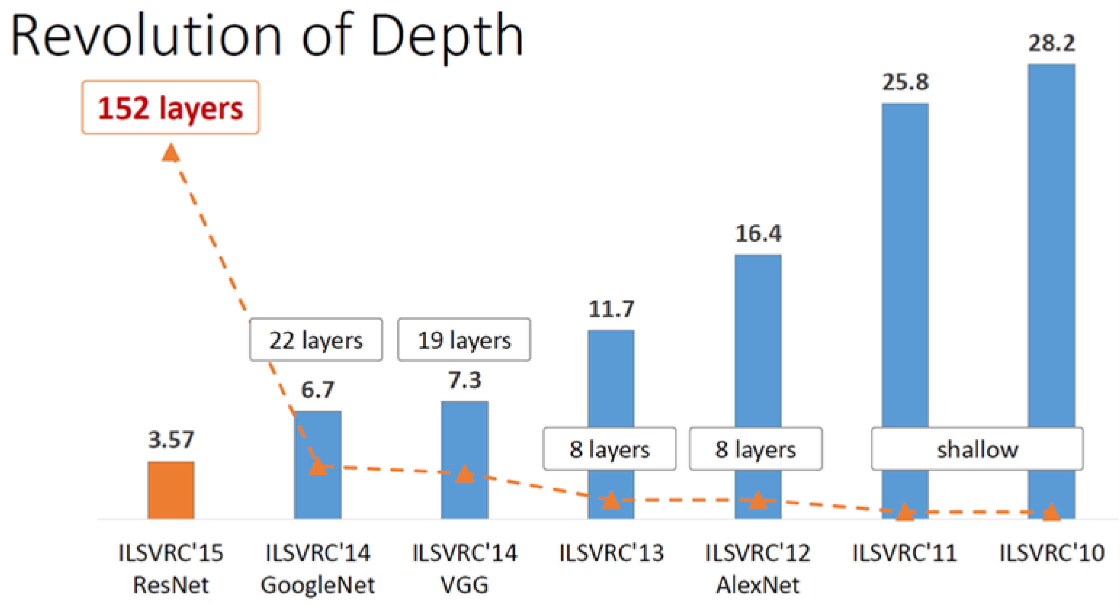


VGG, 19 layers
(ILSVRC 2014)



esNet, 152 layers
(ILSVRC 2015)

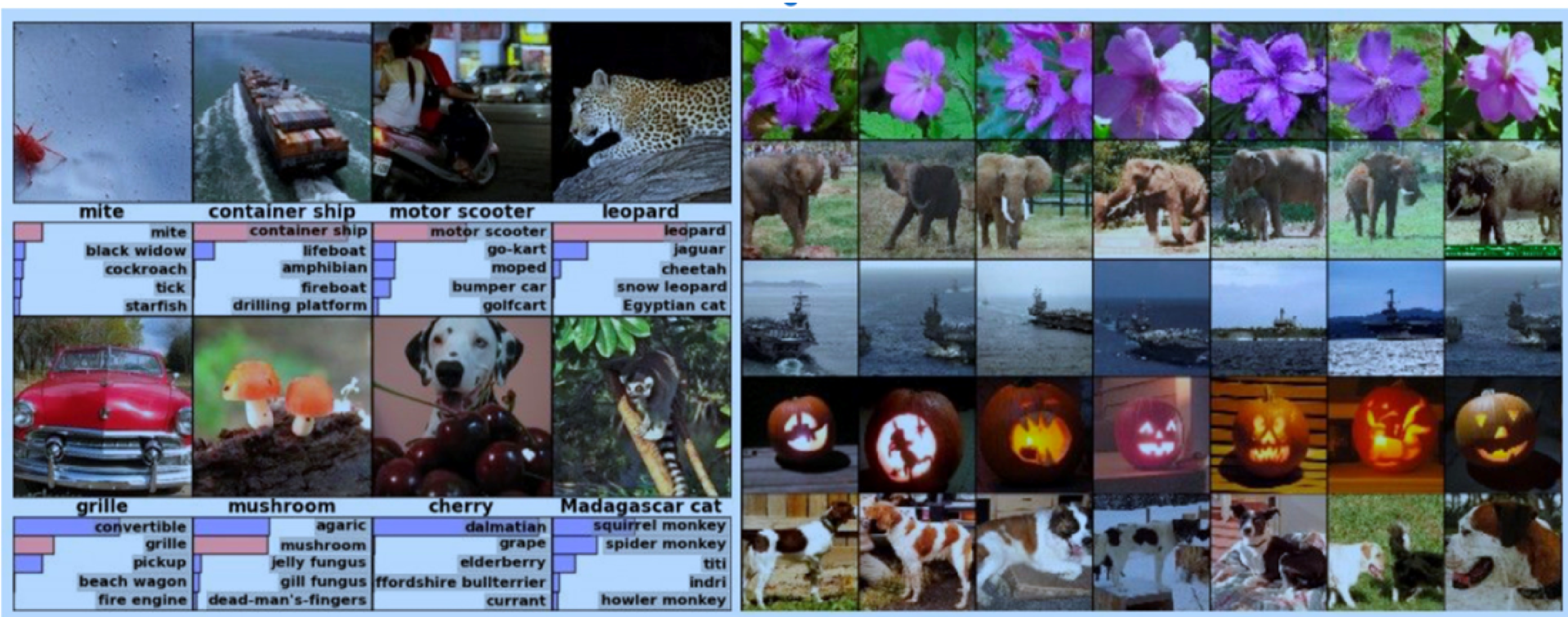




ImageNet Classification top-5 error (%)

Fast-forward to today: ConvNets are everywhere

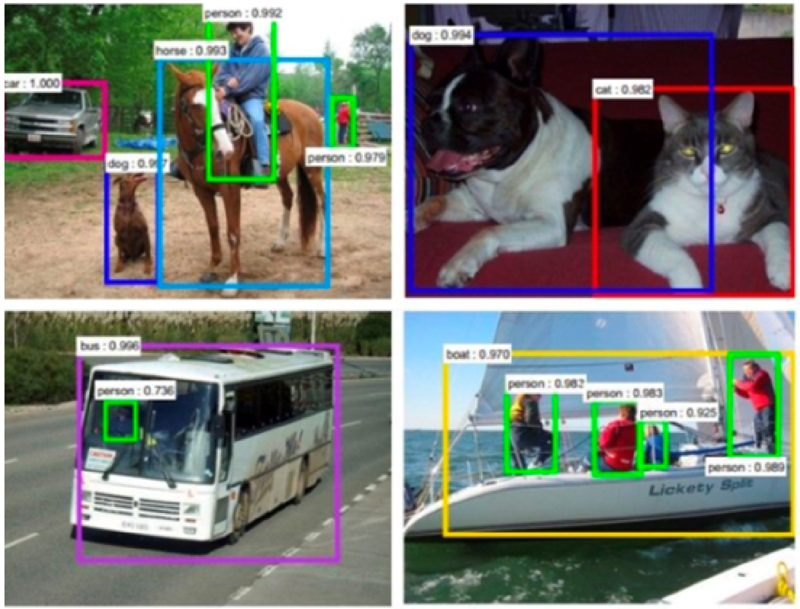
Classification   Retrieval



Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Fast-forward to today: ConvNets are everywhere

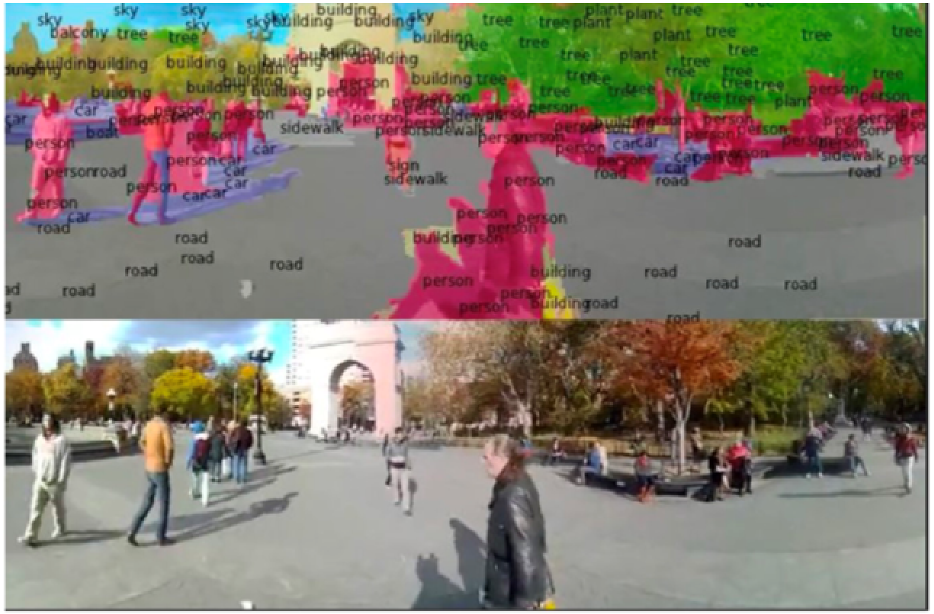
Detection



Figures copyright Shaoqing Ren, Kaiming He, Ross Girschick, Jian Sun, 2015. Reproduced with permission.

[Faster R-CNN: Ren, He, Girshick, Sun 2015]

Segmentation

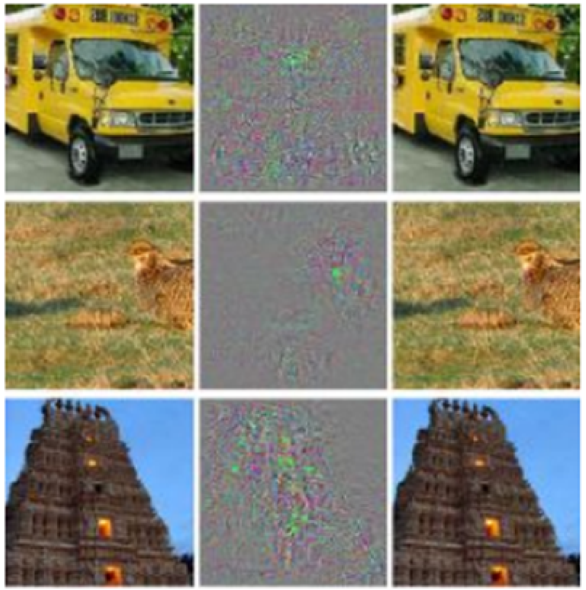


Figures copyright Clement Farabet, 2012. Reproduced with permission.

[Farabet et al., 2012]

Some counter-intuitive properties

- Stability w.r.t small perturbations to inputs
  - Imperceptible non-random perturbation can arbitrarily change the prediction (adversarial examples exist!)



(a)

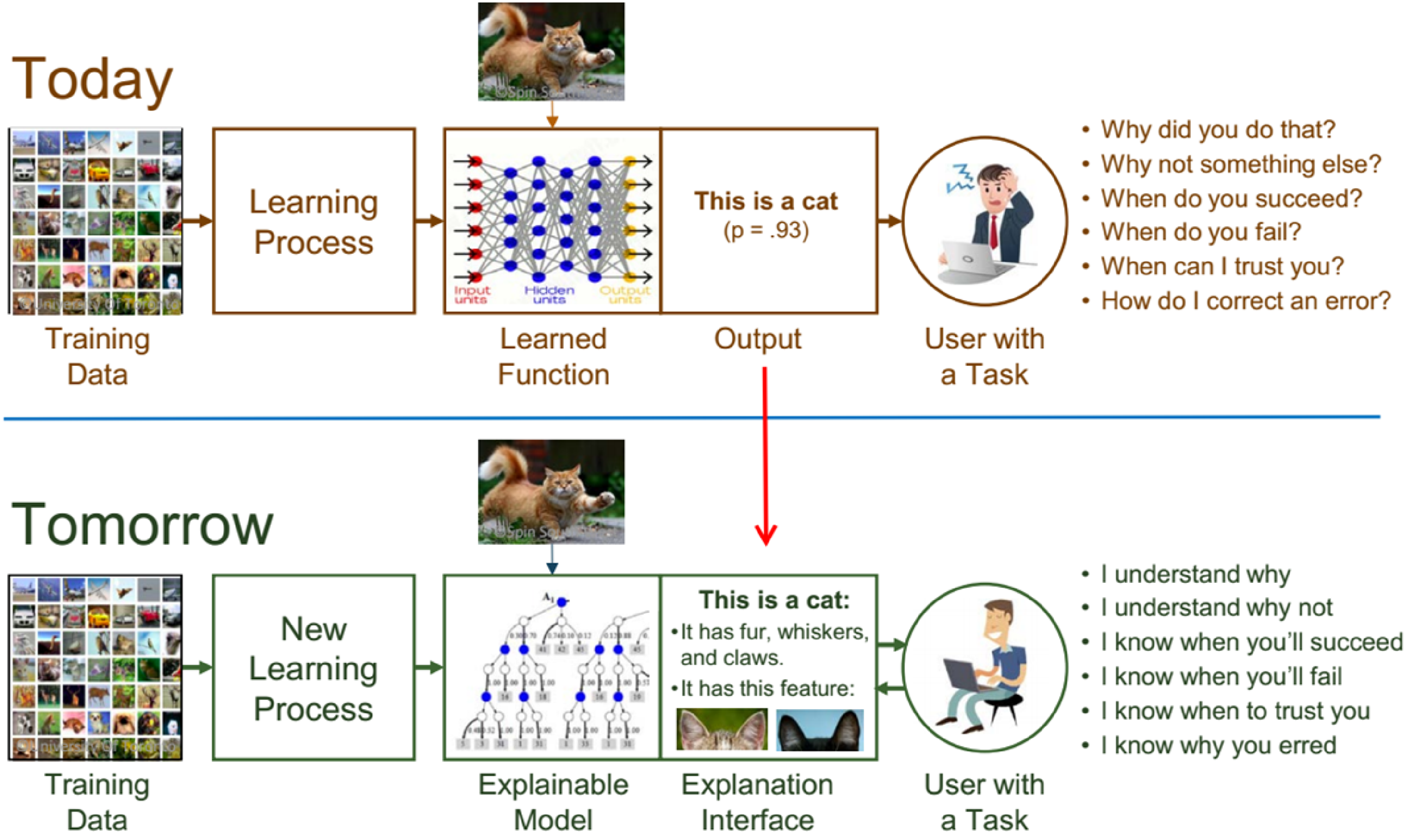
(b)

10x of differences

[Szegedy et al., Intriguing properties of neural nets, 2013

90

Explainable AI – What Are We Trying To Do?



More discussion

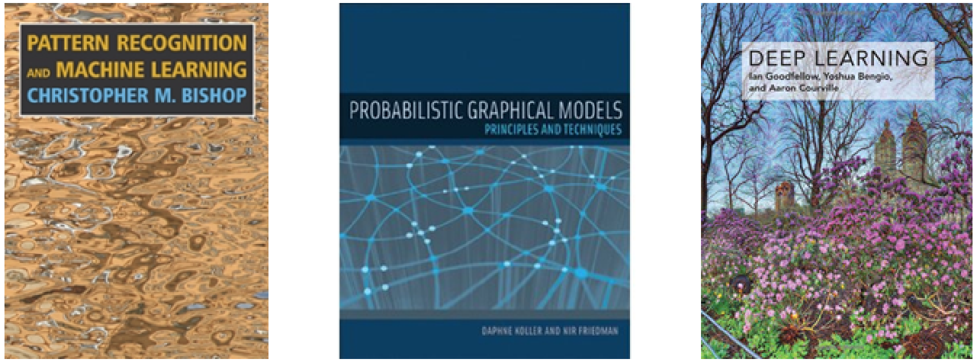
- **Learning = Representation +Evaluation + optimization**

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

[1] Domingos, Pedro. "A few useful things to know about machine learning." Communications of the ACM 55, no. 10 (2012): 78-87.

Resources for Further Learning

- Text books:
  - Pattern Recognition and Machine Learning
  - Probabilistic Graphical Models (http://pgm.stanford.edu/)
  - Deep Learning



- Public lectures:
  - CMU :
    - http://www.cs.cmu.edu/~guestrin/Class/10708-F08/projects.html
  - Stanford:
    - http://cs228.stanford.edu/
    - http://cs228t.stanford.edu/
  - UPenn:
    - http://www.seas.upenn.edu/~cis620/

Resources for Further Learning

- Top-tier Conferences
  - International Joint Conference on Artificial Intelligence (IJCAI)
  - AAAI Annual Conference (AAAI)
  - International Conference on Machine Learning (ICML)
  - Advances in Neural Information Processing Systems (NIPS)
  - Uncertainty in Artificial Intelligence (UAI)
  - IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
  - International Conference on Computer Vision (ICCV)
  - Annual Meeting of the Association for Computational Linguistics (ACL)

Resources for Further Learning

- Top-tier Journals
  - Journal of Machine Learning Research (JMLR)
  - Machine Learning (MLJ)
  - IEEE Trans. on Pattern Recognition and Machine Intelligence (PAMI)
  - Artificial Intelligence
  - Journal of Artificial Intelligence Research (JAIR)
  - International Journal of Computer Vision (IJCV)

Q&A