

# **Introduction**

**Jun Zhu**

[dcszj@mail.tsinghua.edu.cn](mailto:dcszj@mail.tsinghua.edu.cn)

<http://bigml.cs.tsinghua.edu.cn/~jun>

Sate Key Lab of Intelligent Tech. & Systems,  
Tsinghua University

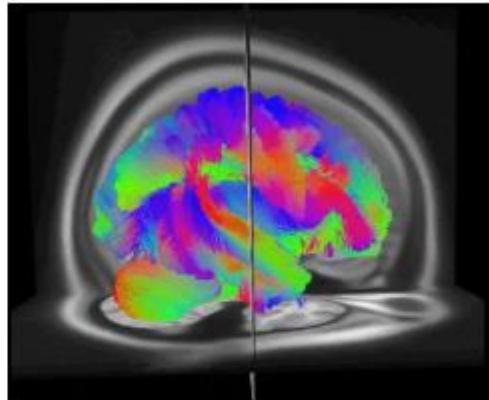
# Goals of this Lecture ...

- Show that machine learning (ML) is cool
- Get you excited about ML
- Give an overview of basic problems & methods in ML
- Help you distinguish hype and science
- Entice you to take further study on ML, write a thesis on ML, dedicate your life to ML ...

# The age of Big Data



CERN Collider  
 $320 \times 10^{12}$  bytes/second



Personal Connectome  
 $10^{18}$  bytes/human

facebook

1 billion  
messages/day

twitter

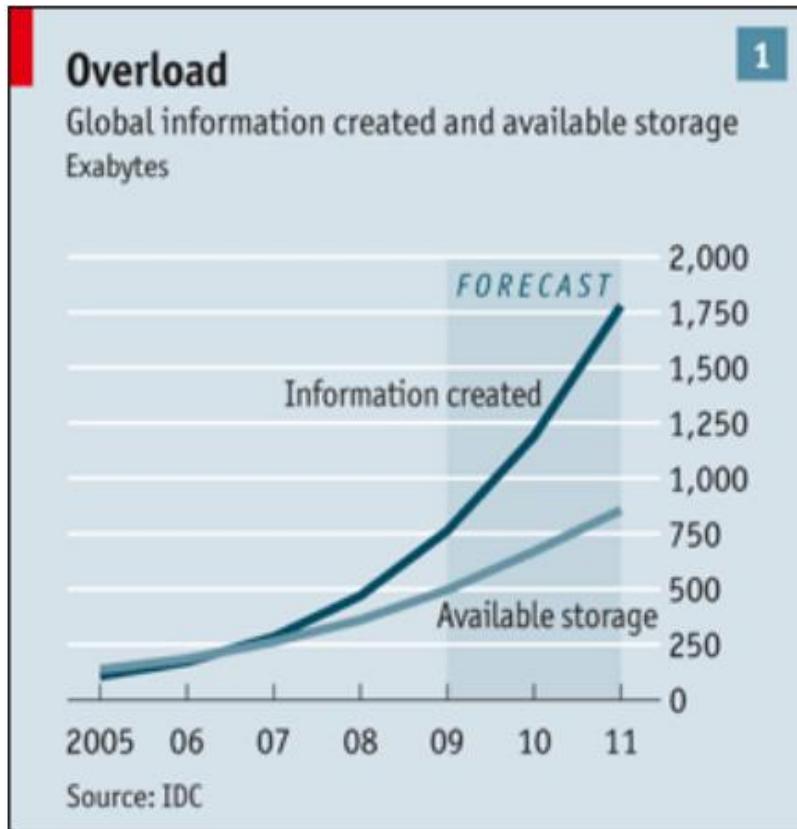


200 million  
tweets/day

"Every day, people create the equivalent of 2.5 **quintillion** bytes of data from sensors, mobile devices, online transactions, and social networks; so much that 90 percent of the world's data has been generated in the past two years."

*The Huffington Post: Arnal Dayaratna: IBM Releases Big Data*

# The age of Big Data



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)

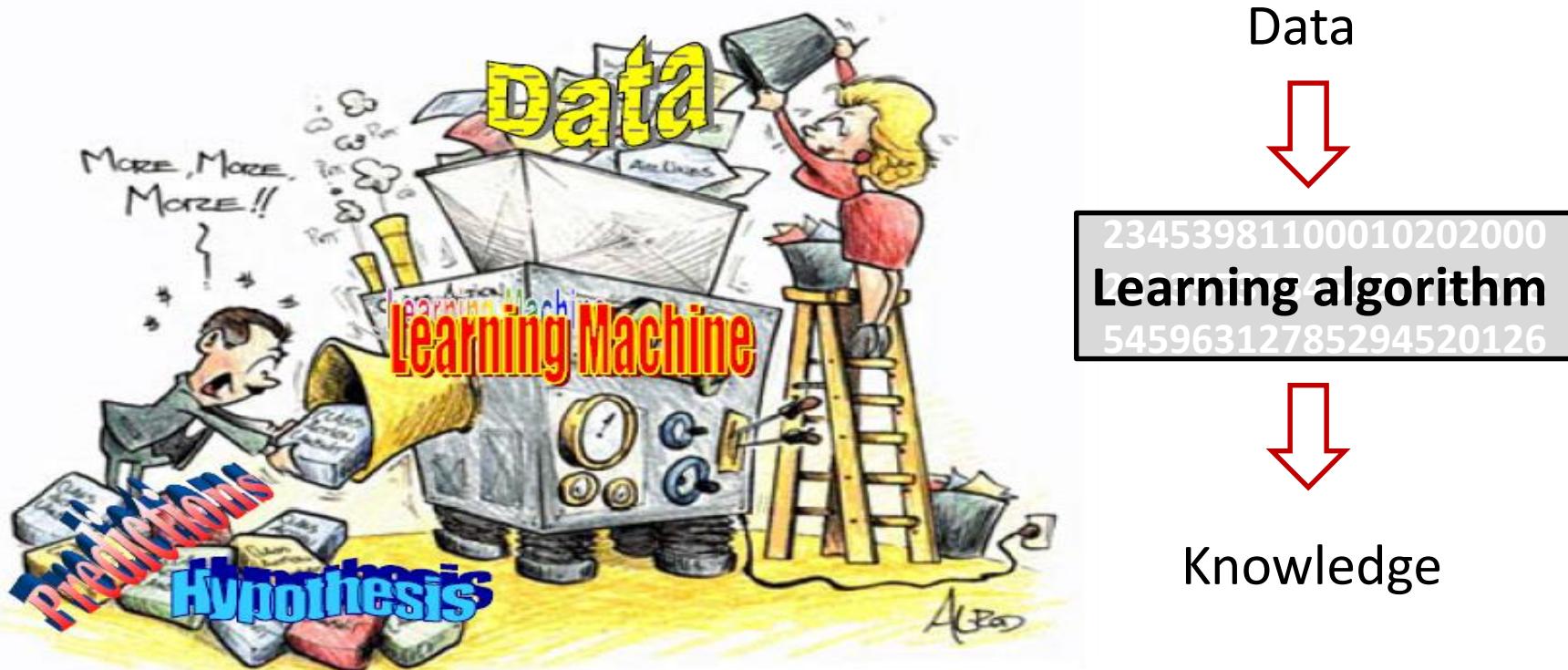
Data → Knowledge

From Data to Knowledge ...

# What is Machine Learning?

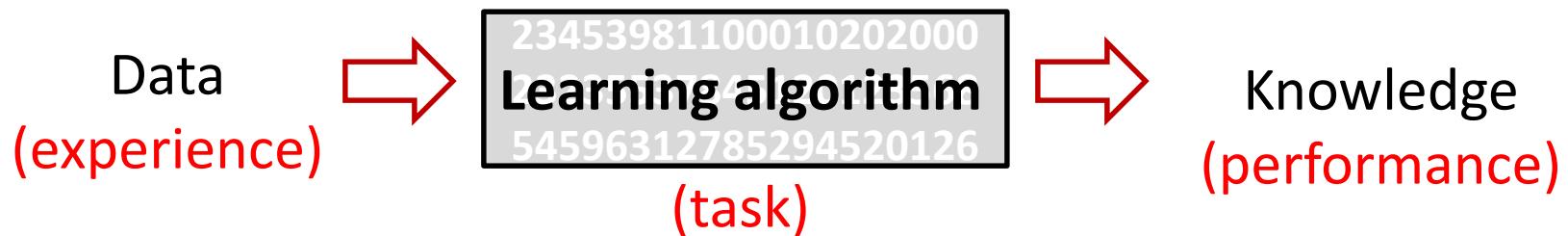


**Machine learning**, a branch of [artificial intelligence](#), is a scientific discipline concerned with the design and development of [algorithms](#) that take as input empirical [data](#), and yield patterns or predictions thought to be features of the [underlying mechanism](#) that generated the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

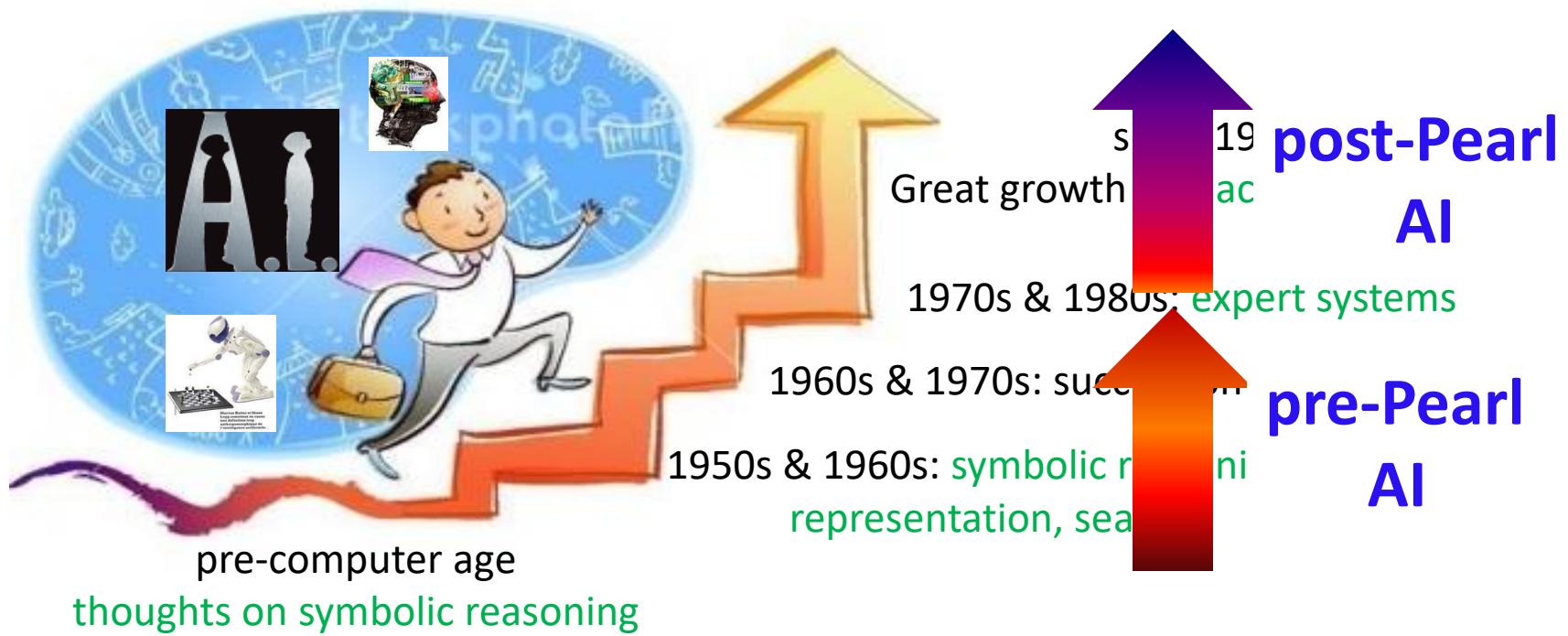


# What is machine learning?

- Study of algorithms that
  - (automatically) improve their performance
  - at some task
  - with experience



# (Statistical) Machine Learning in AI



[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."



## **Heuristics, Probability and Causality**

A Tribute to Judea Pearl

“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 ...

Editors

Rina Dechter  
Hector Geffner Joseph  
Y. Halpern

# Machine learning in Action

- Document classification



# Spam Filter

垃圾邮件通知目录摘要/Spam notification Abstract

回复 回复全部 转发 删除 邮件标签

垃圾邮件通知目录摘要 /Spam notification Abstract ★

**postmaster**  
发给 dcszj Dear Jun:

dcszj@tsinghua.edu.cn 2015 ASE Data Science conference will be held at Stanford University, August 18-20, 2015. You can find more details via the conference website (<http://www.scienceengineeringacademy.org/99999>). We would like to invite you to our program committee. Could you please confirm your availability by March 6, 2015.

以下是邮件安全网关的提醒。  
目前你收到的疑似垃圾邮件有 7 封。这些可疑邮件我们将为您只保留 5 天，请尽快处理。

Thanks,  
Justin

2014-06-09 06:02 详细信息

移动 → 移动到垃圾邮件文件夹，然后在垃圾邮件中接收。

**Spam / not Spam**

Dear dcszj@tsinghua.edu.cn,

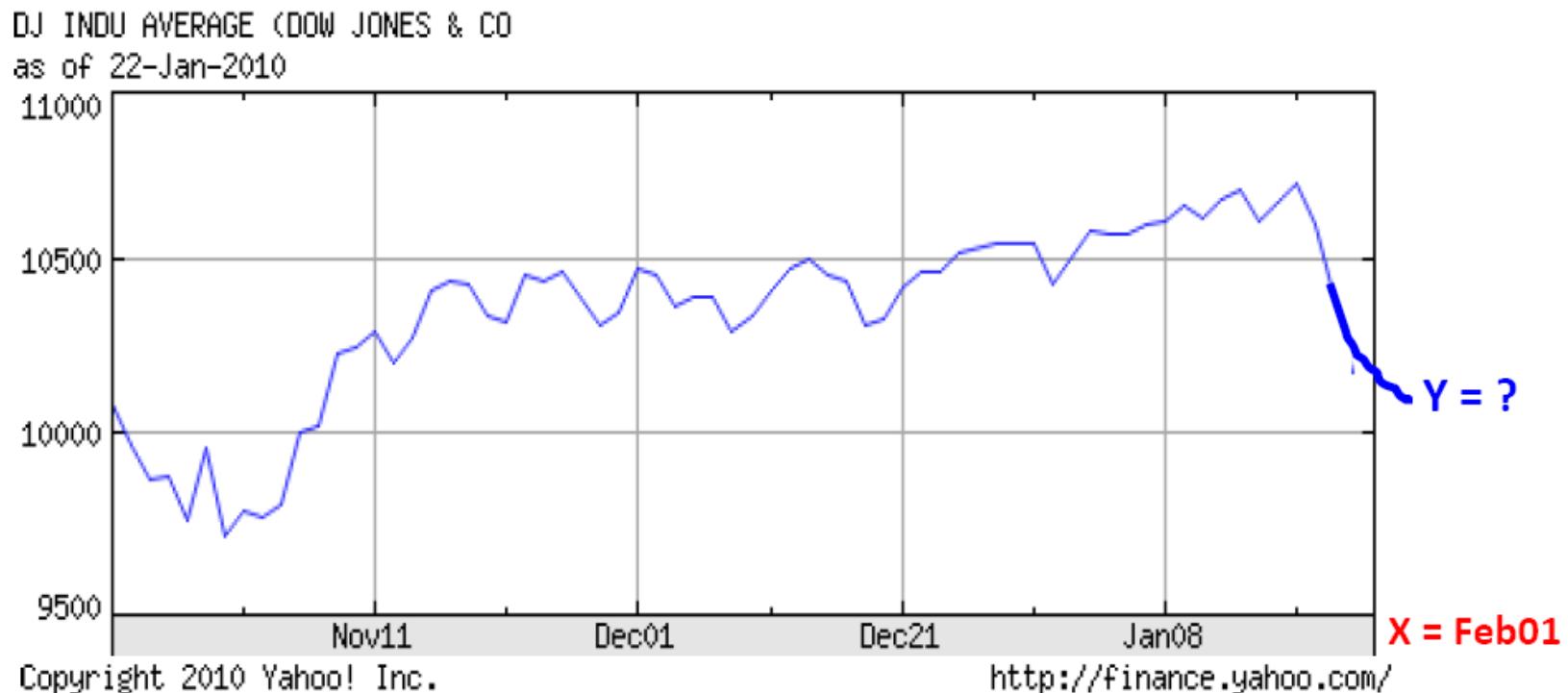
The mails in the following list are the spams our gateway has isolated for you. If there is any email you need, you can click the "移动/Move" button after the listed mail to send it back to the Inbox and then Receive your mails again.

The total number of spams you received is 7. Please check these mails as soon as possible because we will reserve them for 5 days only.

发信人/Sender	邮件主题/Subject	收信时间/Received Time	邮件大小/Size	选择移动/Move
"The 31st International Conference on Machine Learning (ICML 2014)"	Registration Alert: The 31st International Conference on Machine Learning (ICML 2014) (1504183) BEIJING, - Yunhong Zhou (69984085)	2014-06-09 04:41:13	8.284 KB	<a href="#">移动/Move</a>
"The 31st International Conference on Machine Learning (ICML 2014)"	Registration Alert: The 31st International Conference on Machine Learning (ICML 2014)	2014-06-09 03:21:00	8.253 KB	<a href="#">移动/Move</a>

# Regression

- Stock market prediction

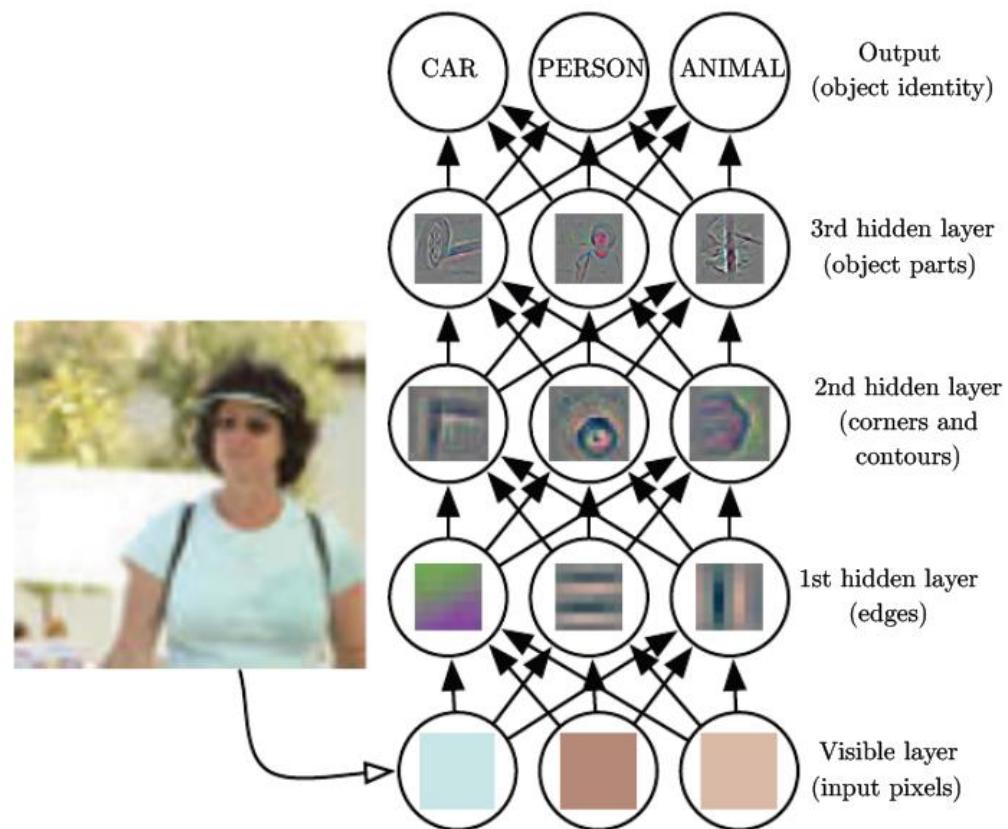


# Computer Vision

- Image Classification, Face recognition, Scene understanding, Action/behavior recognition, Image tagging and search, Optical character recognition (OCR)



ImageNet Challenge: 1000 categories, 1.2 million images for training



# Speech Recognition

- A classic problem in AI, very difficult!
  - “Let’s talk about how to wreck a nice beach”
  - small vocabulary is – easy
  - challenges: large vocabulary, noise, accent, semantics

iPhone 4S



amazon echo

Always ready, connected, and fast. [Just ask.](#)



# Natural Language Processing

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

The image displays two side-by-side screenshots of the Google Translate mobile application. Both screenshots show a two-way translation interface with Chinese on the left and English on the right.

**Top Screenshot:** The source language is set to "Chinese – detected". The input text is "今天星期二" (Jīntiān xīngqī'èr), which is translated into English as "Today is tuesday".

**Bottom Screenshot:** The source language is also set to "Chinese – detected". The input text is "机器学习" (Jīqì xuéxí), which is translated into English as "robotic leanring".

Both screenshots include standard translation controls: microphone, speaker, and refresh icons at the top, and "Edit" and "Feedback" buttons at the bottom.

# Control

- Cars navigating on their own



- DAPA urban challenge
- Tsinghua Mobile Robot V (THMR-V):

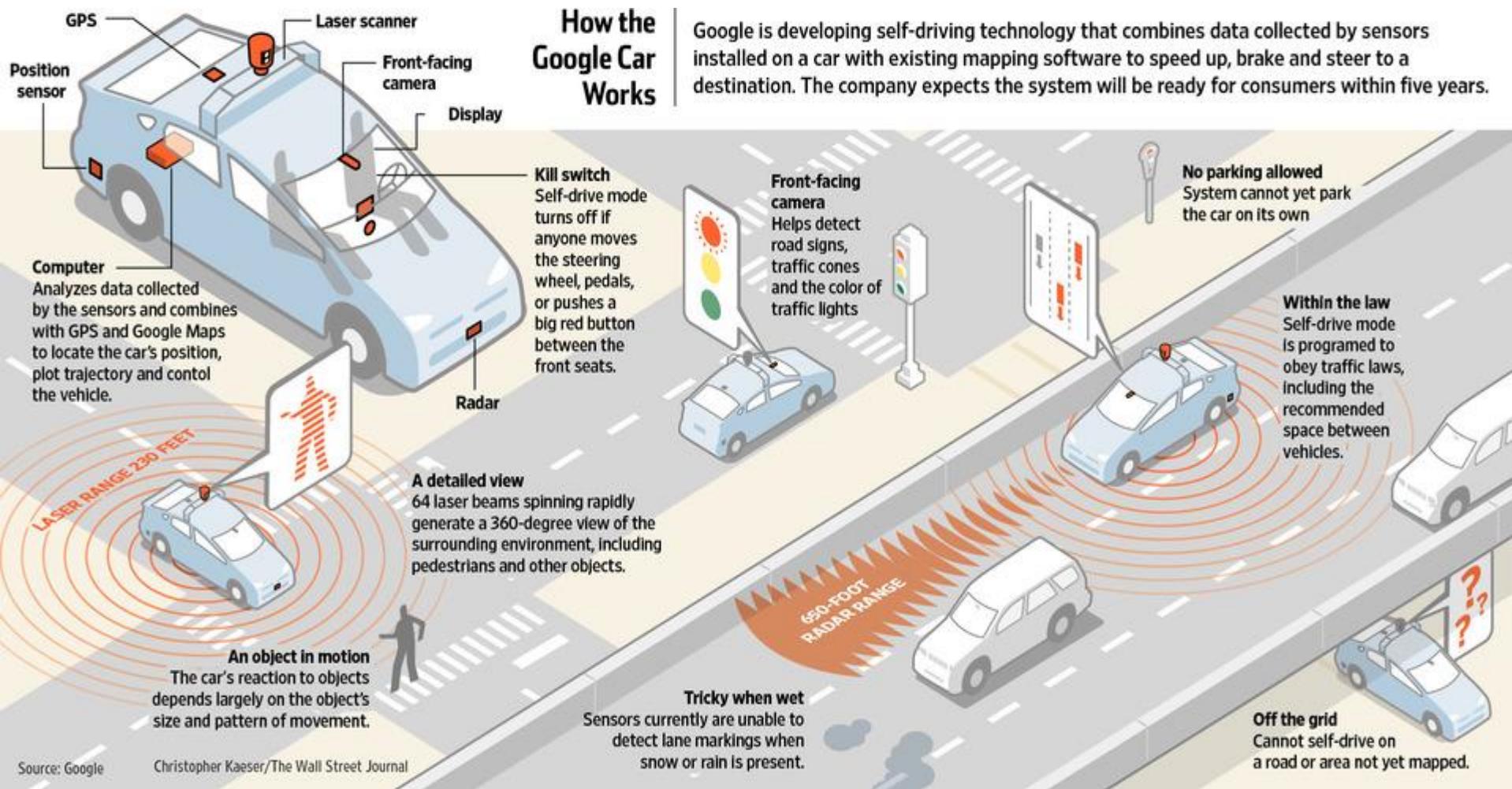




- The first license, Nevada, 2012
- Nevada, Florida, California, Michigan, allows testing on public roads

# Control (cont'd)

- How the Google car works

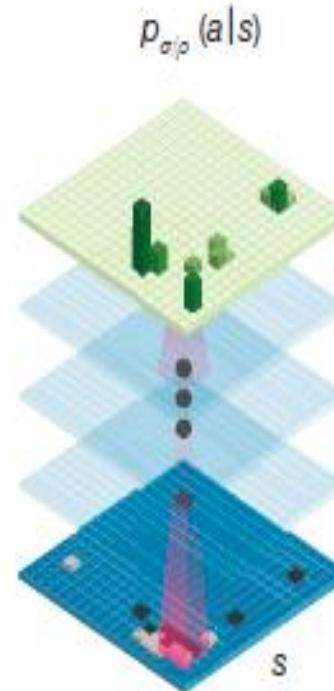


# AlphaGO

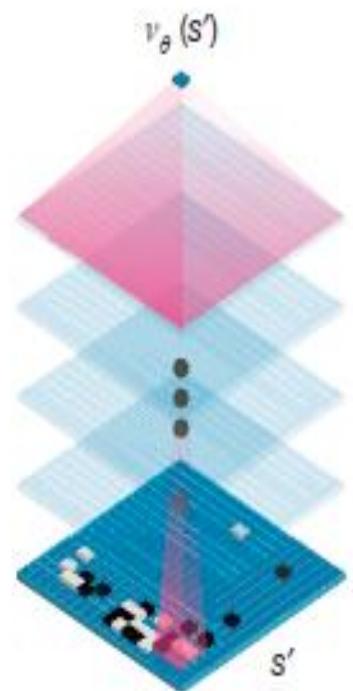
- March, 2016: AlphaGO beats Sedol Lee at 4:1



Policy network

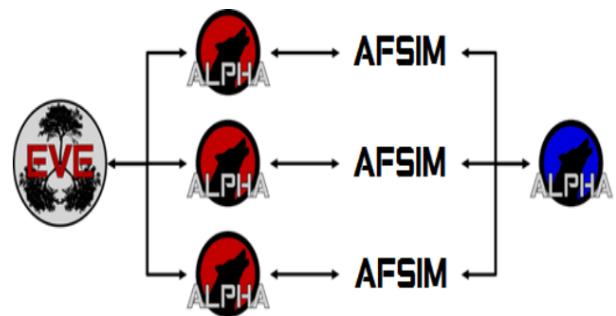
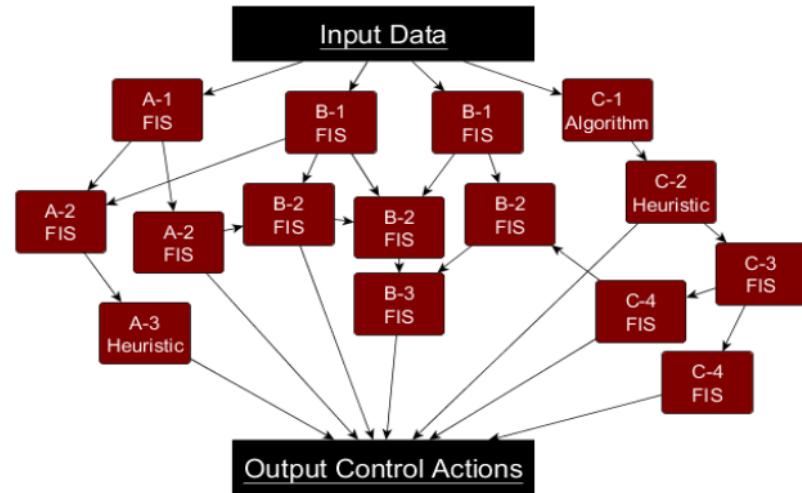
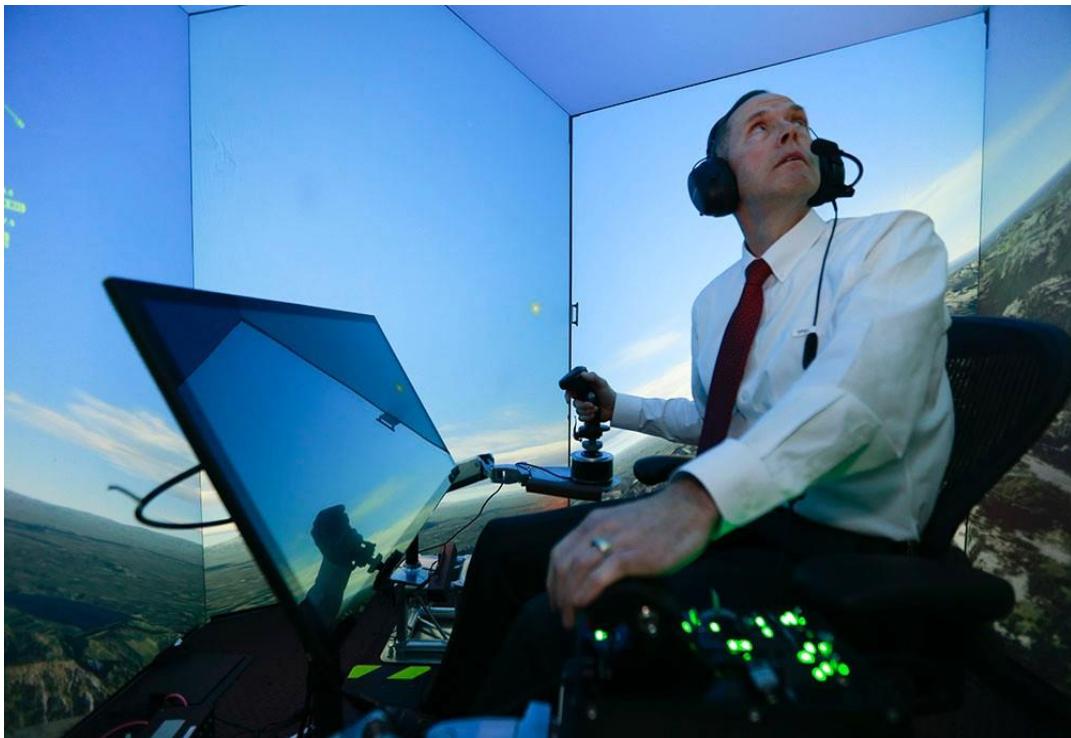


Value network



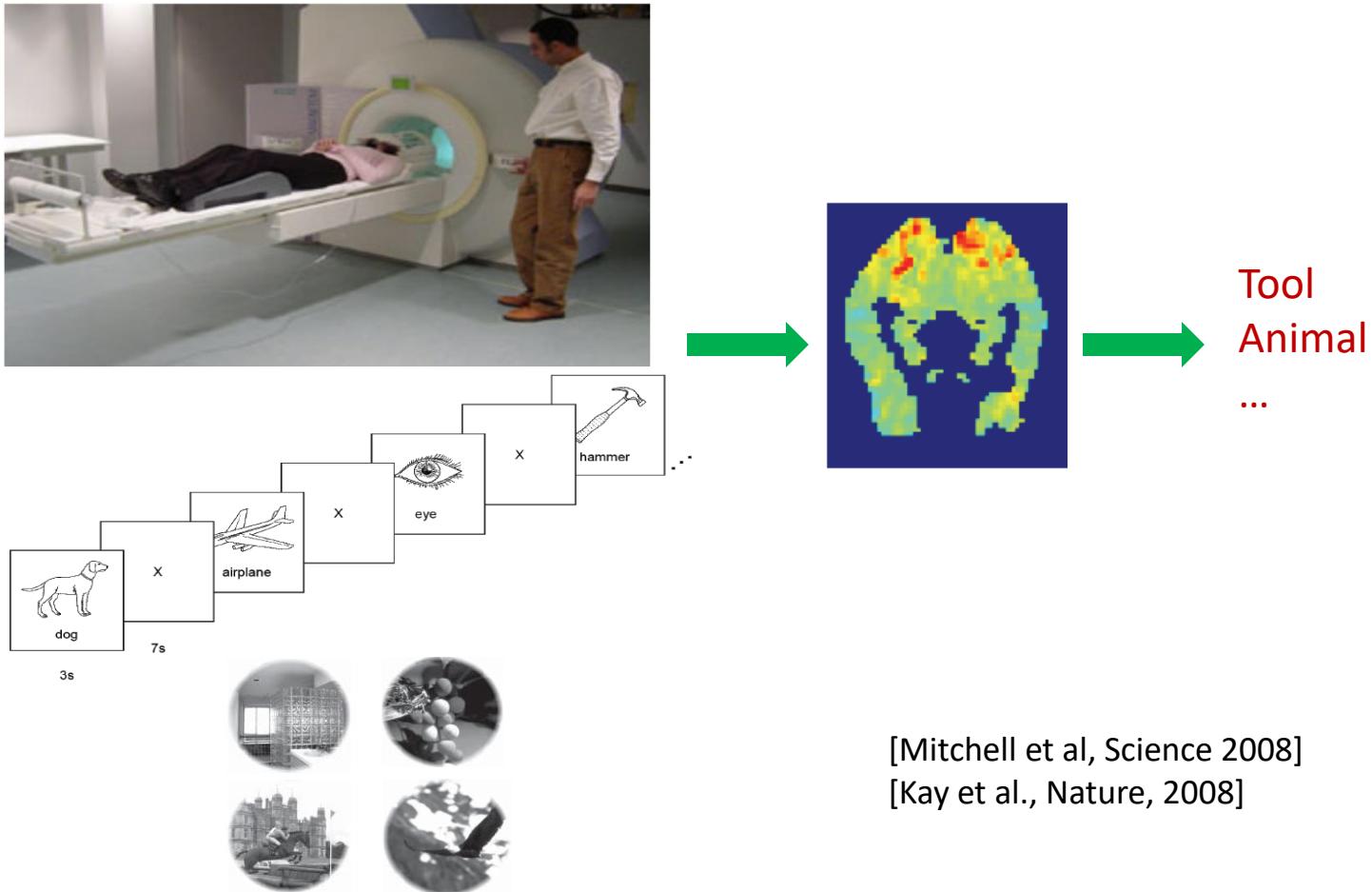
# Alpha

- June, 2016: Alpha beats Gene Lee in combat simulation



# Science

- Decoding thoughts from brain activity



# Science (cont'd)

- Bayesian models of inductive learning and reasoning [Tenenbaum et al., Science 2011]
  - Challenge:
    - *How can people generalize well from sparse, noisy, and ambiguous data?*
  - Hypothesis:
    - *If the mind goes beyond the data given, some more abstract background knowledge must generate and delimit the possible hypotheses*
  - Bayesian models make **structured abstract knowledge** and **statistical inference** cooperate
  - Examples
    - Word learning [Xu & Tenenbaum, Psychol. Rev. 2007]
    - Causal relation learning [Griffiths & Tenenbaum, 2005]
    - Human feature learning [Austerweil & Griffiths, NIPS 2009]
    - ...

# More others ...

- Many more
  - Natural language processing
  - Speech recognition
  - Computer vision
  - Robotics
  - Computational biology
  - Social network analysis
  - Sensor networks
  - Health care
  - Protest ??
  - ...

# Machine learning in Action

- Machine learning for protest?

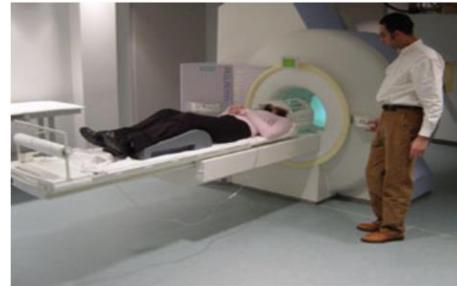


CMU ML students and post-docs at G-20 Pittsburgh Summit 2009

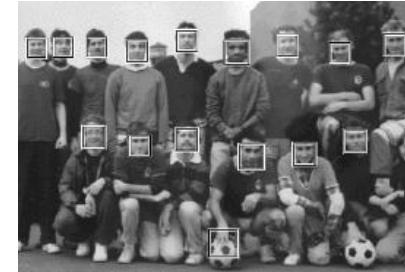
# Machine Learning – practice



document classification



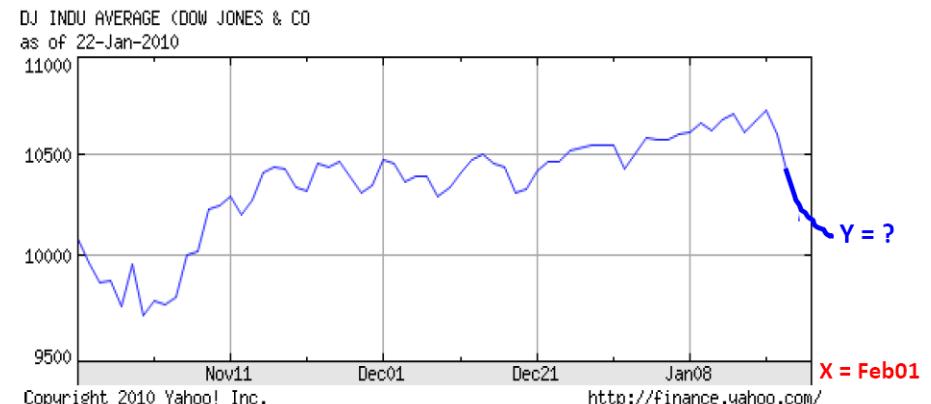
decoding brain signal



face recognition

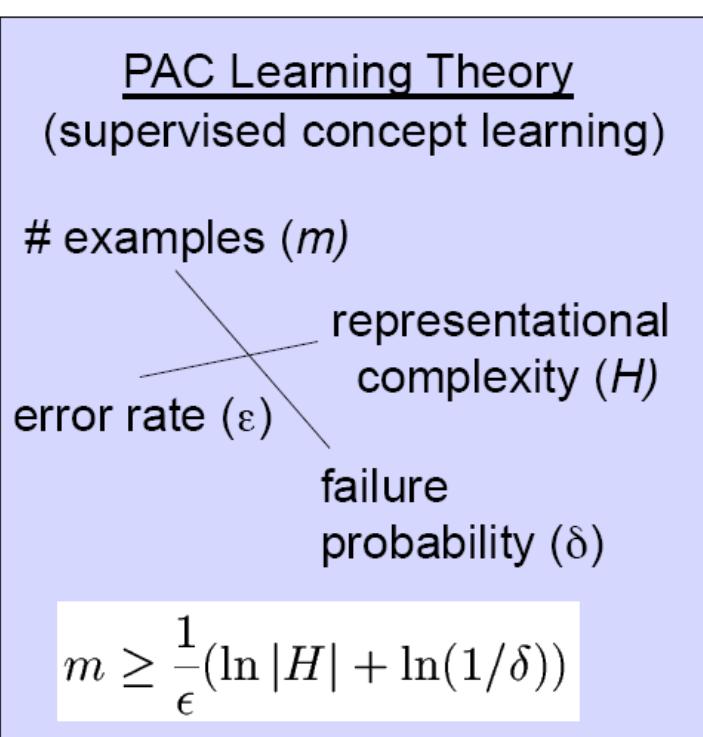


robot control



stock market prediction

# Machine Learning – theory



Other theories for

- semi-supervised learning
- reinforcement skill learning
- active learning
- ...

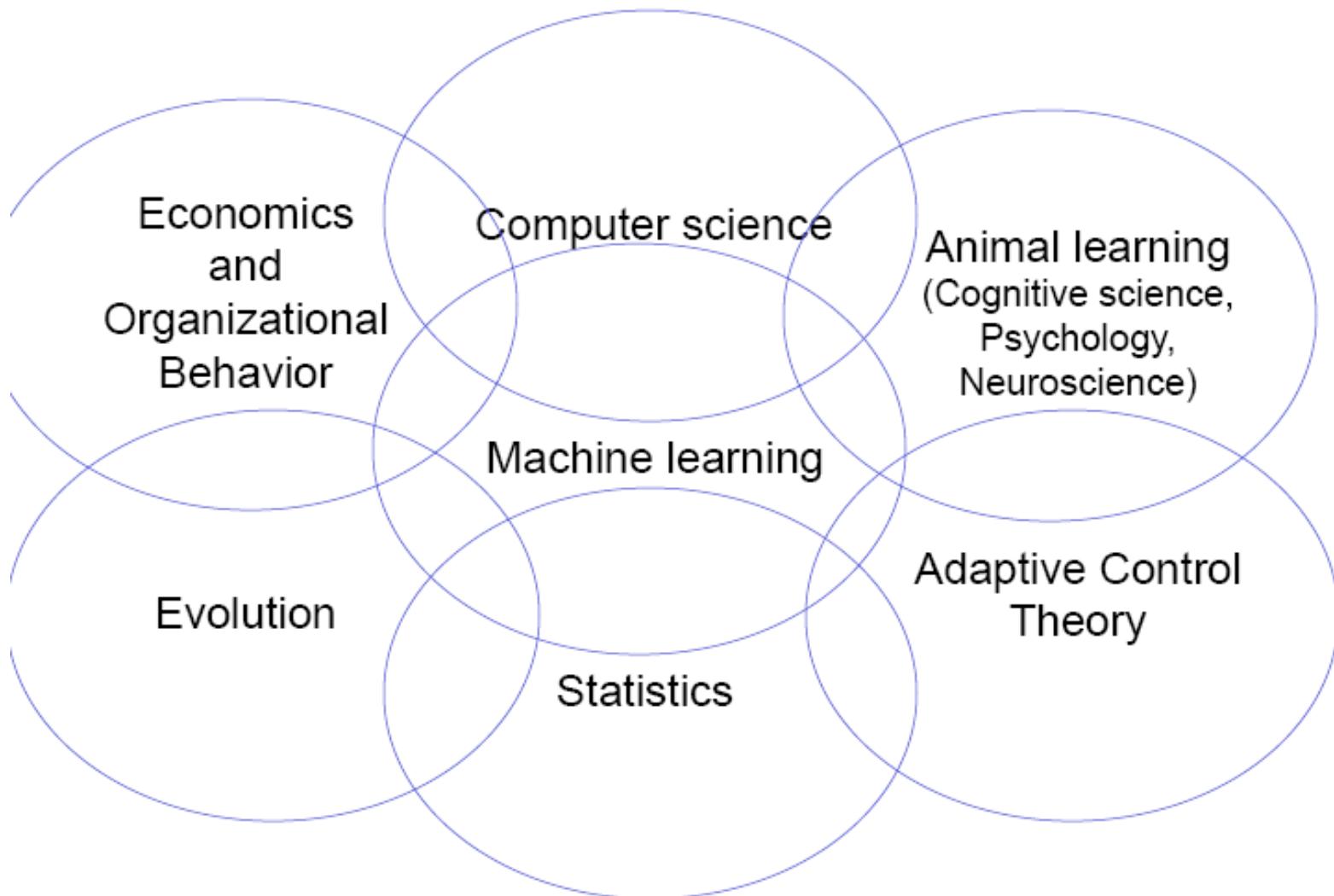
... also relating to

- # mistakes during training
- asymptotic performance
- convergence rate
- bias, variance tradeoff
- ...

[Leslie G. Valiant, 1984; Turing Award, 2010]

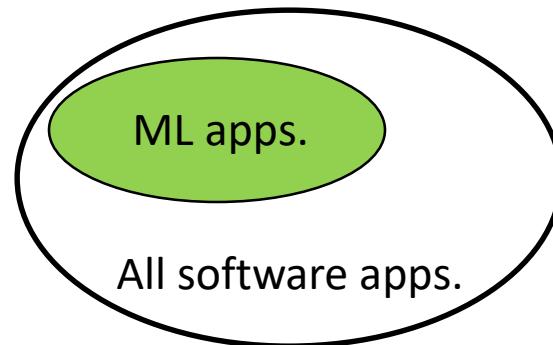
“For transformative contributions to the theory of computation, including the theory of probably approximately correct (PAC) learning, the complexity of enumeration and of algebraic computation, and the theory of parallel and distributed computing.”





# 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 (**why?**)



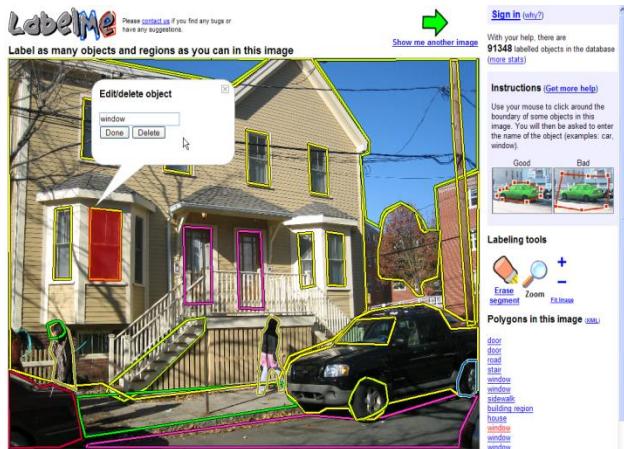
# Growth of Machine Learning in CS

- Machine learning already the preferred approach to
    - Speech recognition, natural language processing
    - 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
- Huge amount of data ...**

  - Web: estimated Google index 45 billion pages
  - Transaction data: 5-50 TB/day
  - Satellite image feeds: ~1TB/day/satellite
  - Biological data: 1-10TB/day/sequencer
  - TV: 2TB/day/channel;
  - YouTube 4TB/day uploaded
  - Photos: 1.5 billion photos/week uploaded

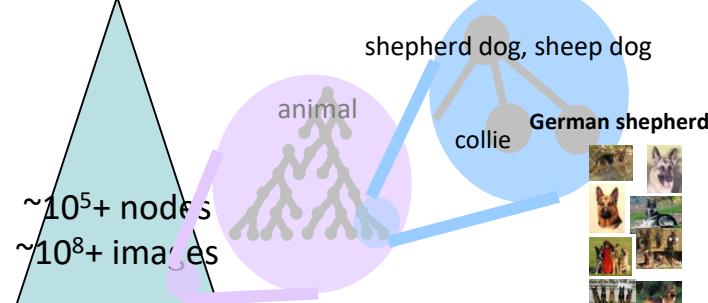
# ML has a long way to go ...

- Very large-scale learning in rich media



IMAGENET

~ $10^5$ + nodes  
~ $10^8$ + images



$10^5$   
images  
 $10^{1-2}$   
categories



$10^5$   
images  
 $10^{2-3}$   
categories



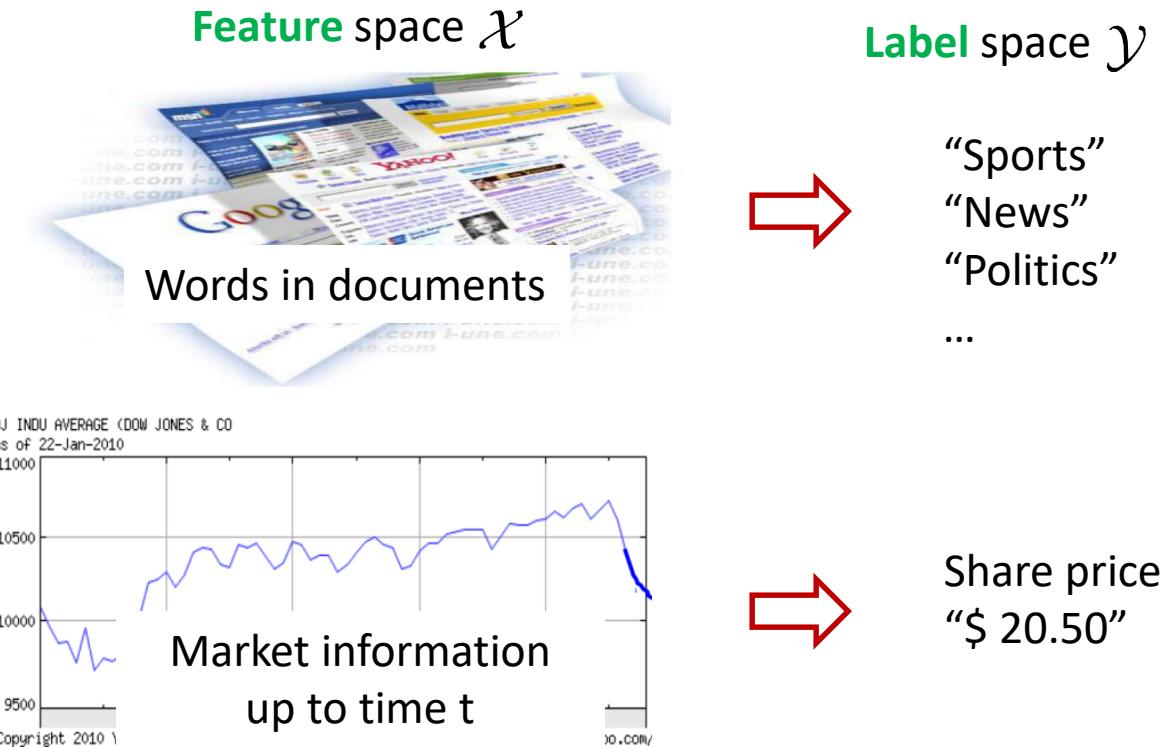
$10^{6-7}$   
images  
 $10^{3-4}$   
categories

# 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 ...

# Supervised Learning

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



- “Experience” or training data:

$$\{\langle x_d, y_d \rangle\}_{d=1}^D, x_d \in \mathcal{X}, y_d \in \mathcal{Y}$$

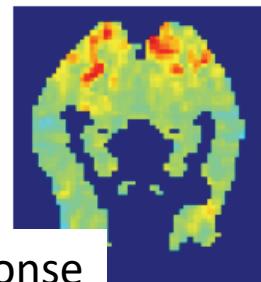
# Supervised Learning – classification

Feature space  $\mathcal{X}$



Label space  $\mathcal{Y}$

“Sports”  
“News”  
“Politics”  
...

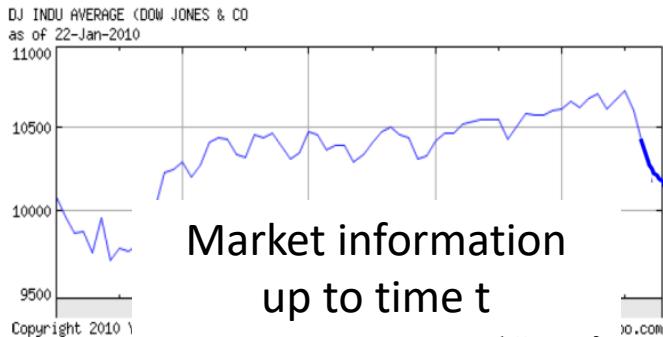


“Tool”  
“Animal”  
...

Discrete Labels

# Supervised Learning – regression

Feature space  $\mathcal{X}$

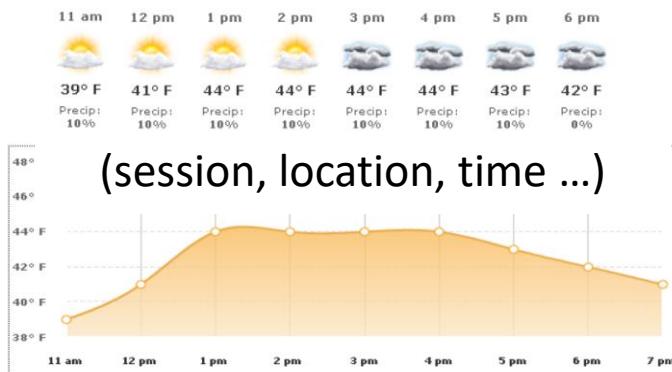


Label space  $\mathcal{Y}$

Share price  
“\$ 20.50”



(session, location, time ...)

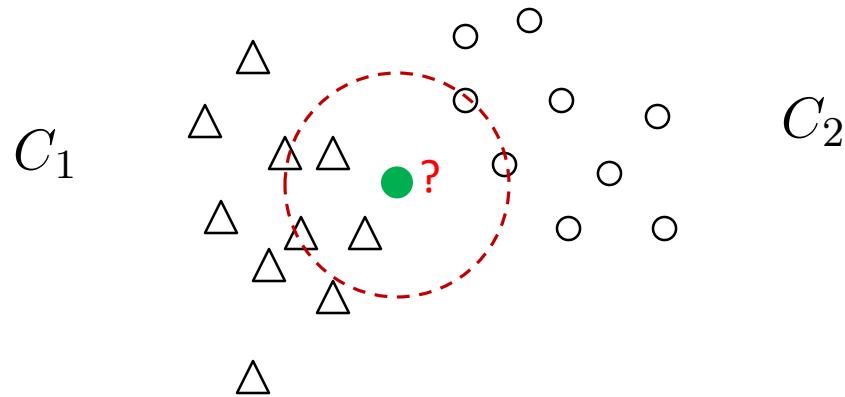


Temperature  
“42° F”



Continuous Labels

# How to learn a classifier?

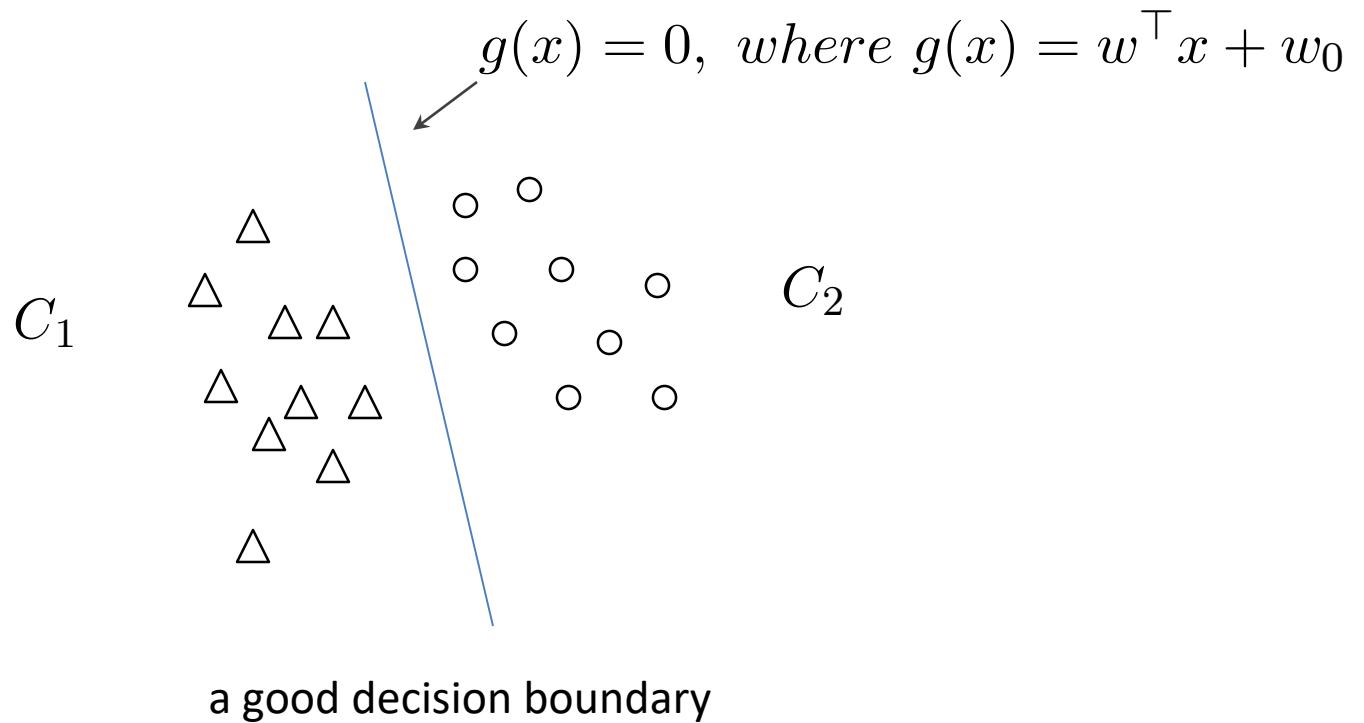


**K-NN**: a Non-parametric approach

*Distance metric matters!*

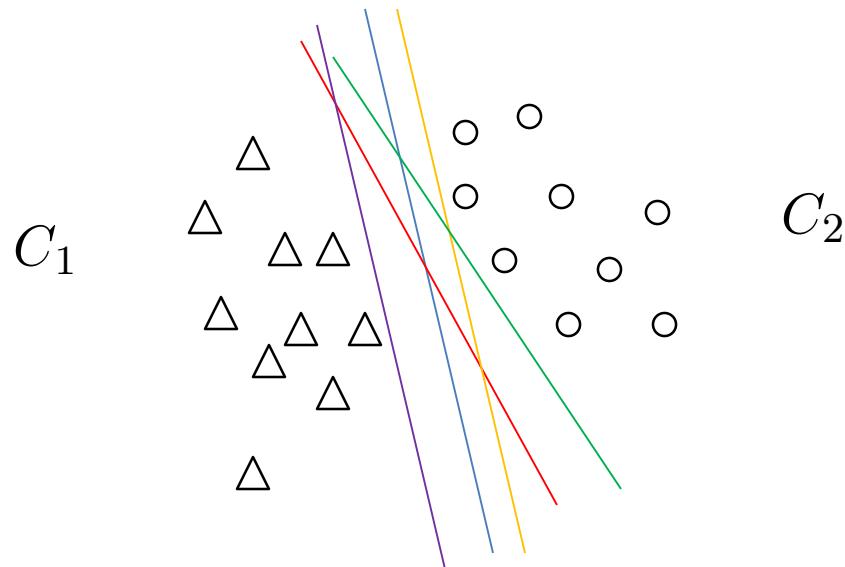
# How to learn a classifier?

Parametric (model-based) approaches:



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

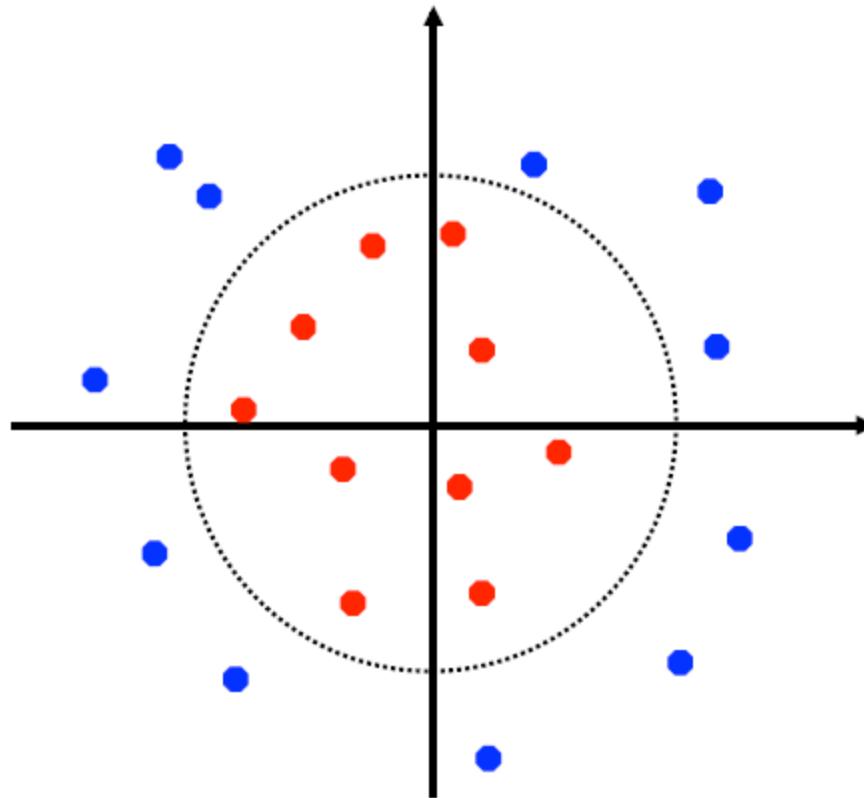
# How to learn a classifier?



Many good decision boundaries

*which one should we choose?*

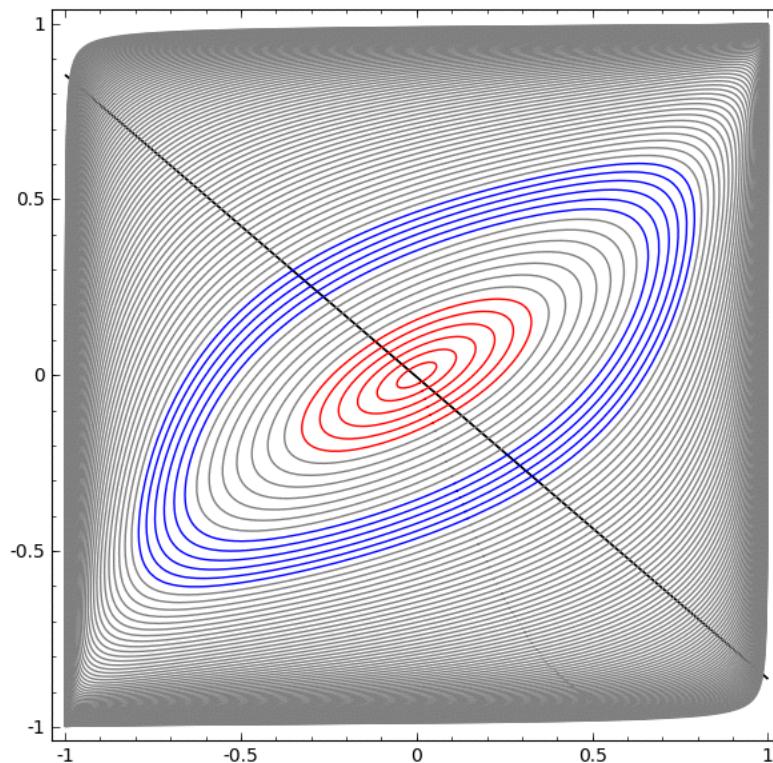
# How to learn a classifier?



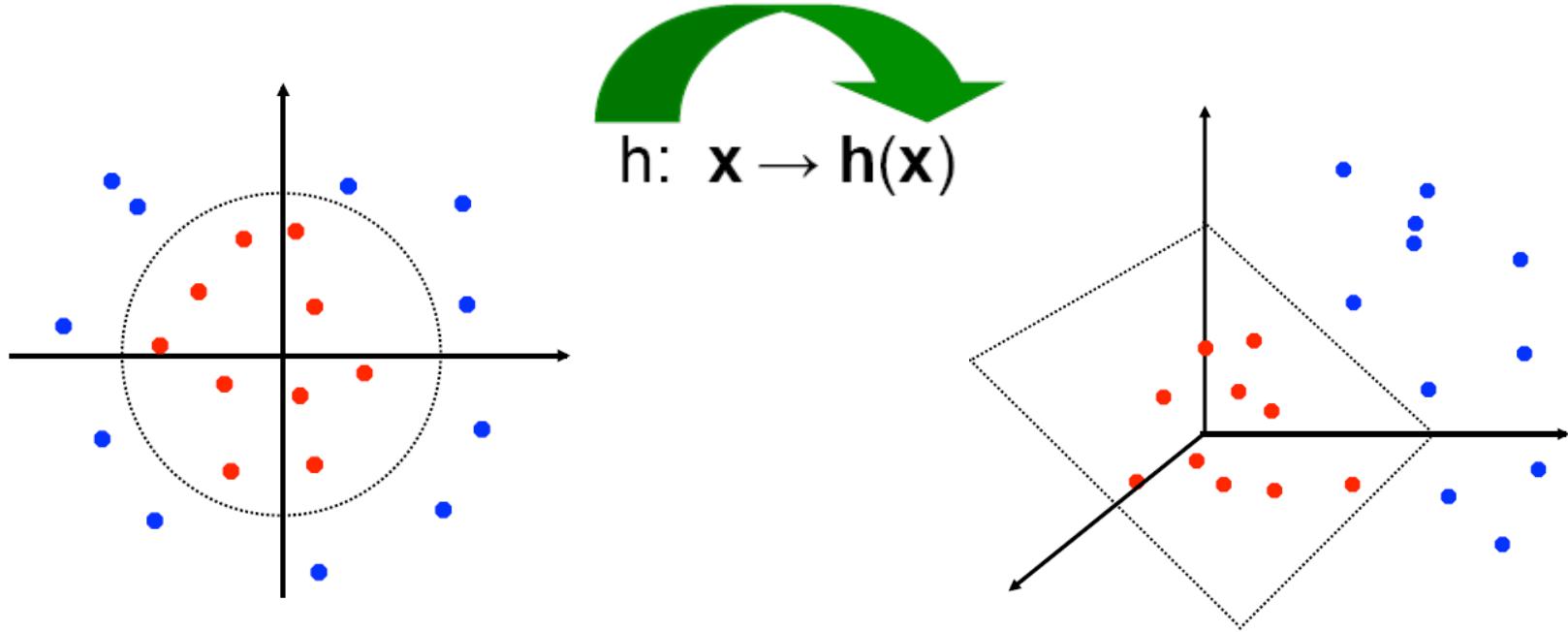
*How about non-linearity?*

# How to learn a classifier?

- 2D mapping is insufficient



# How to learn a classifier?

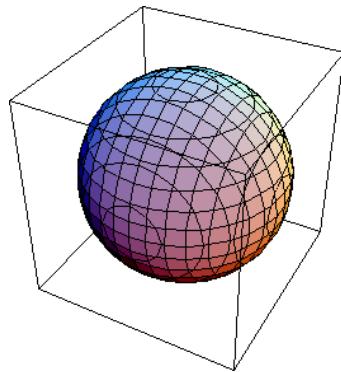
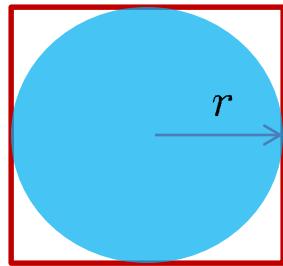


*How about non-linearity?*

*The higher dimension, the better?*

# How to learn a classifier?

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



d dimensional space

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

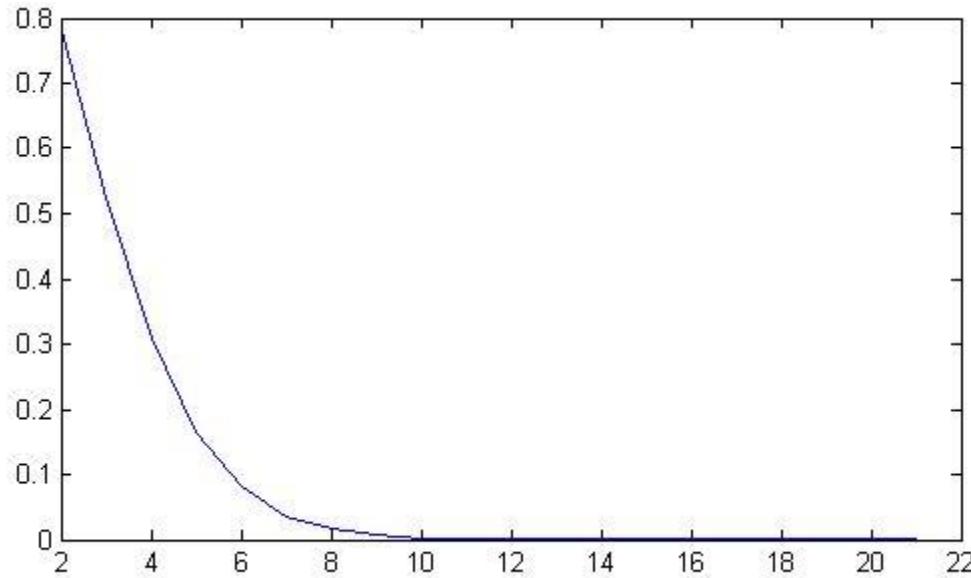
$$\frac{\frac{2r^3\pi^{3/2}}{3\Gamma(3/2)}}{(2r)^3} = \frac{\pi^{3/2}}{12\Gamma(3/2)}$$

$$\frac{\frac{2r^d\pi^{d/2}}{d\Gamma(d/2)}}{(2r)^d} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma(d/2)}$$

$\uparrow d \rightarrow \infty$

# How to learn a classifier?

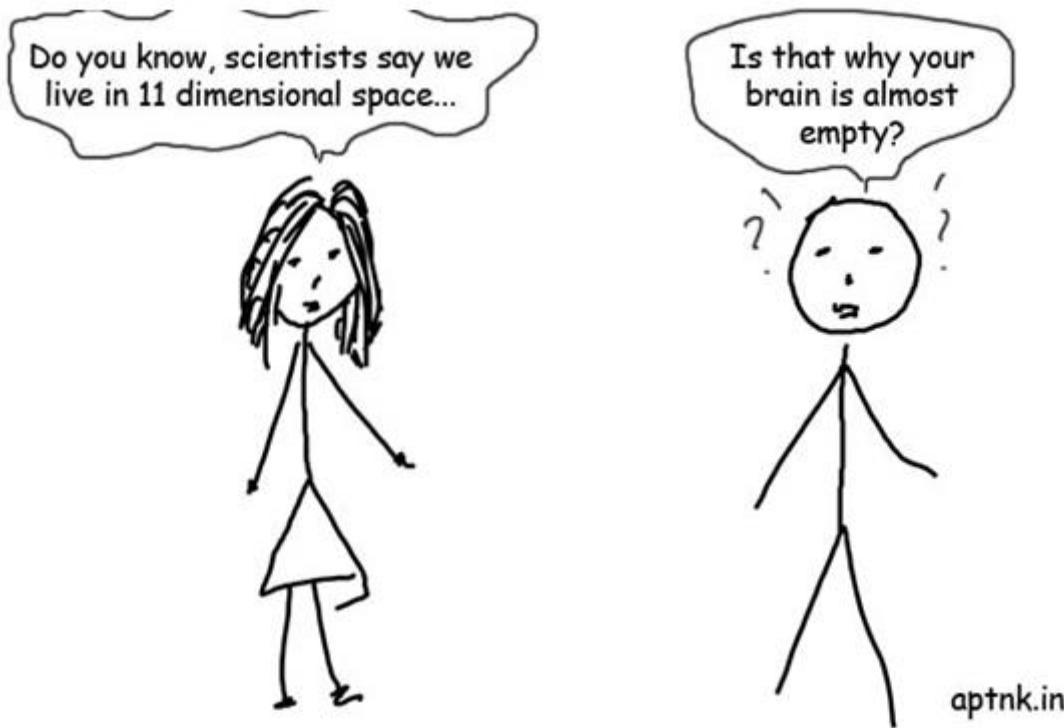
- Curse of dimensionality
  - A high dimensional space is always almost empty



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



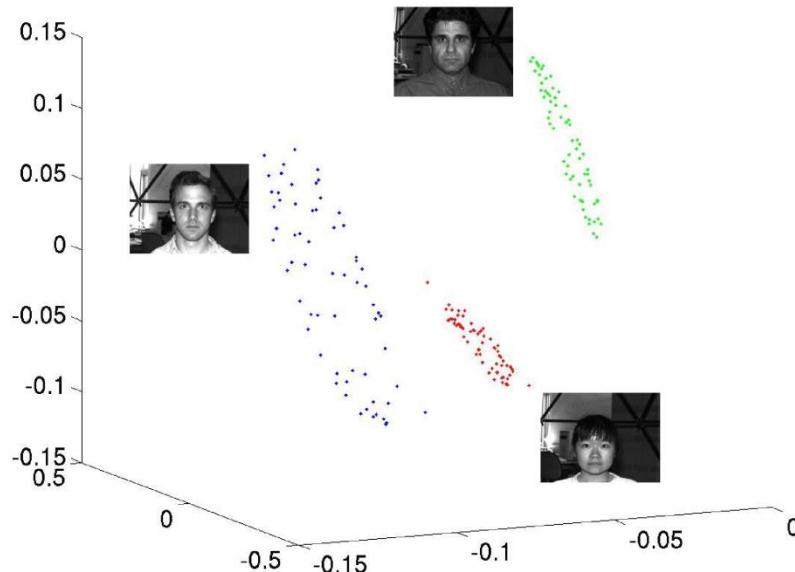
when one wants to learn pattern from data in high dimensions no matter how much data you have it always seems less! A high dimensional space is always almost empty<sup>48</sup>

# 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.*

# How to learn a classifier?

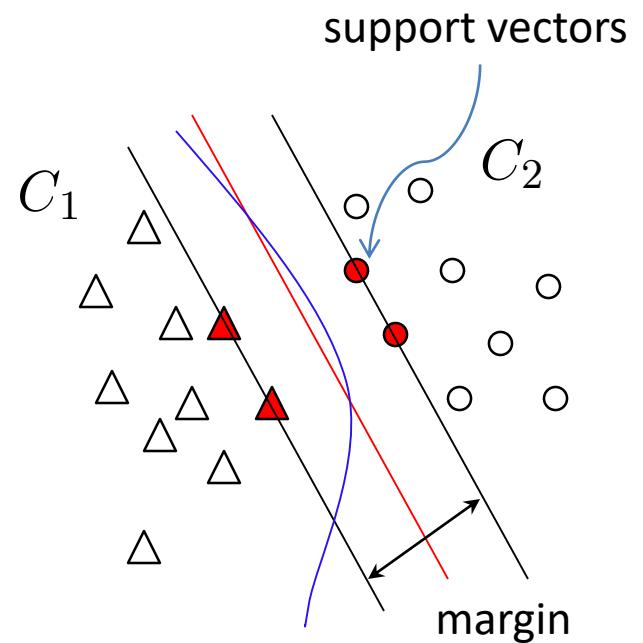
- The blessing of dimensionality
  - ... *real data highly concentrate on low-dimensional, sparse, or degenerate structures in the high-dimensional space.*



Images of the same face under varying illumination lie approximately on a low (nine)-dimensional subspace, known as the **harmonic plane** [Basri & Jacobs, PAMI, 2003].

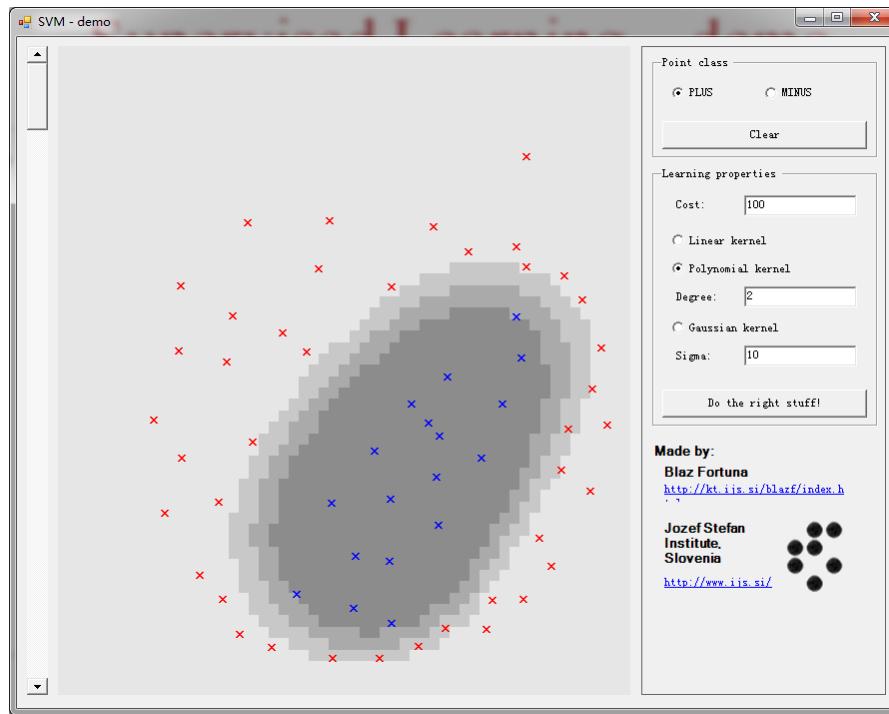
# How to learn a classifier?

- 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?

- Support vector machines (SVM) – demo



Good ToolKits: [1] SVM-Light: <http://svmlight.joachims.org/>  
[2] LibSVM: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

# How to learn a classifier?

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

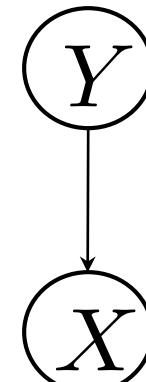
A joint distribution:  $p(x, y) = p(y)p(x|y)$

Inference using Bayes rule: prior likelihood

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

↑                      ↑  
prior                  likelihood  
evidence

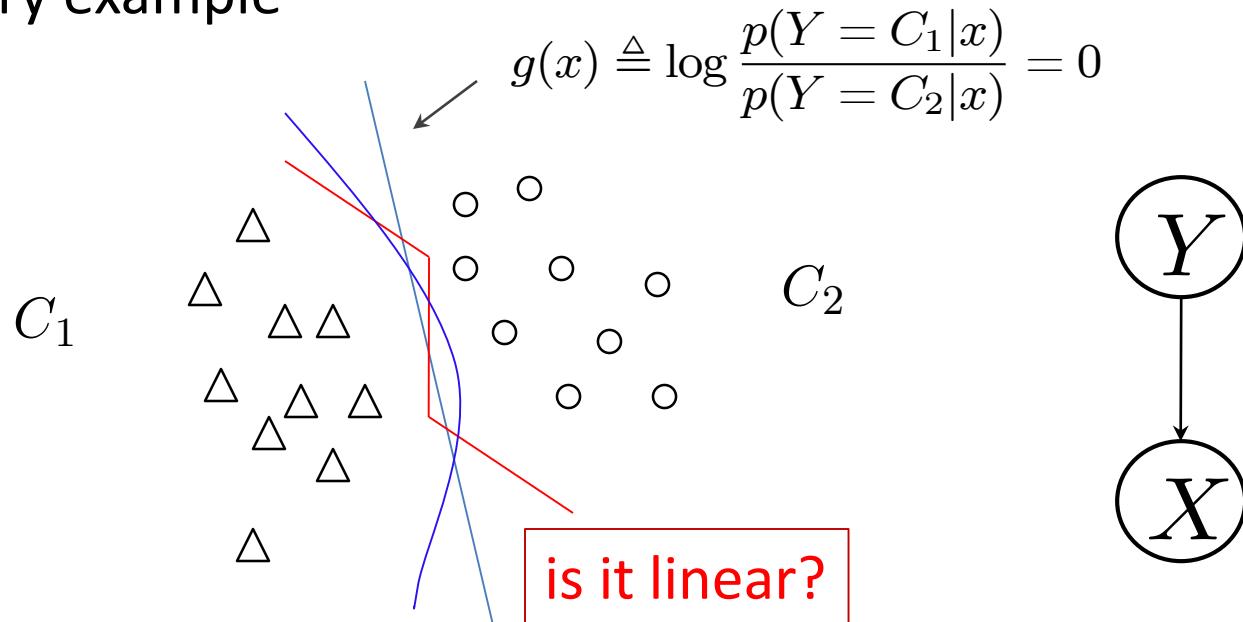
Prediction rule:  $y^* = \arg \max_{y \in \mathcal{Y}} p(y|x)$



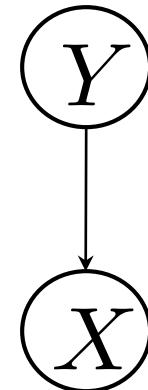
- fundamental building blocks for *Bayesian networks*
- nice illustrative example of Bayesian methods

# How to learn a classifier?

- Naïve Bayes classifier – basics
  - binary example



$$y^* = \begin{cases} C_1 & \text{if } p(Y = C_1|x) > 0.5 \\ C_2 & \text{if } p(Y = C_1|x) < 0.5 \end{cases}$$



It is for generalized linear models (GLMs)

# How to learn a classifier?

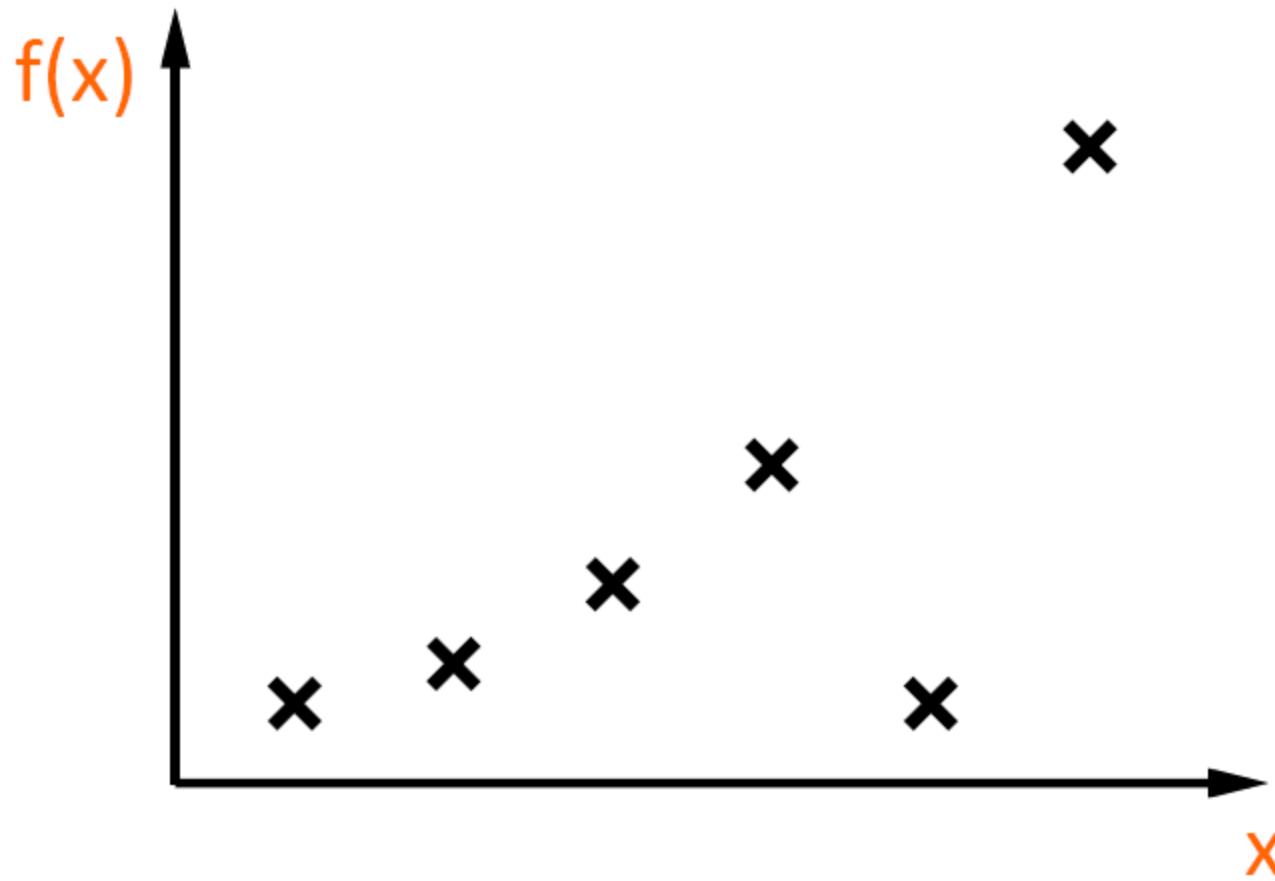
- Many other classifiers
  - K-nearest neighbors
  - Decision trees
  - Logistic regression
  - Boosting
  - Random forests
  - Mixture of experts
  - Maximum entropy discrimination (a nice combination of max-margin learning and Bayesian methods)
  - ...

Advice #1:

All models are wrong, but some are useful. – G.E.P. Box

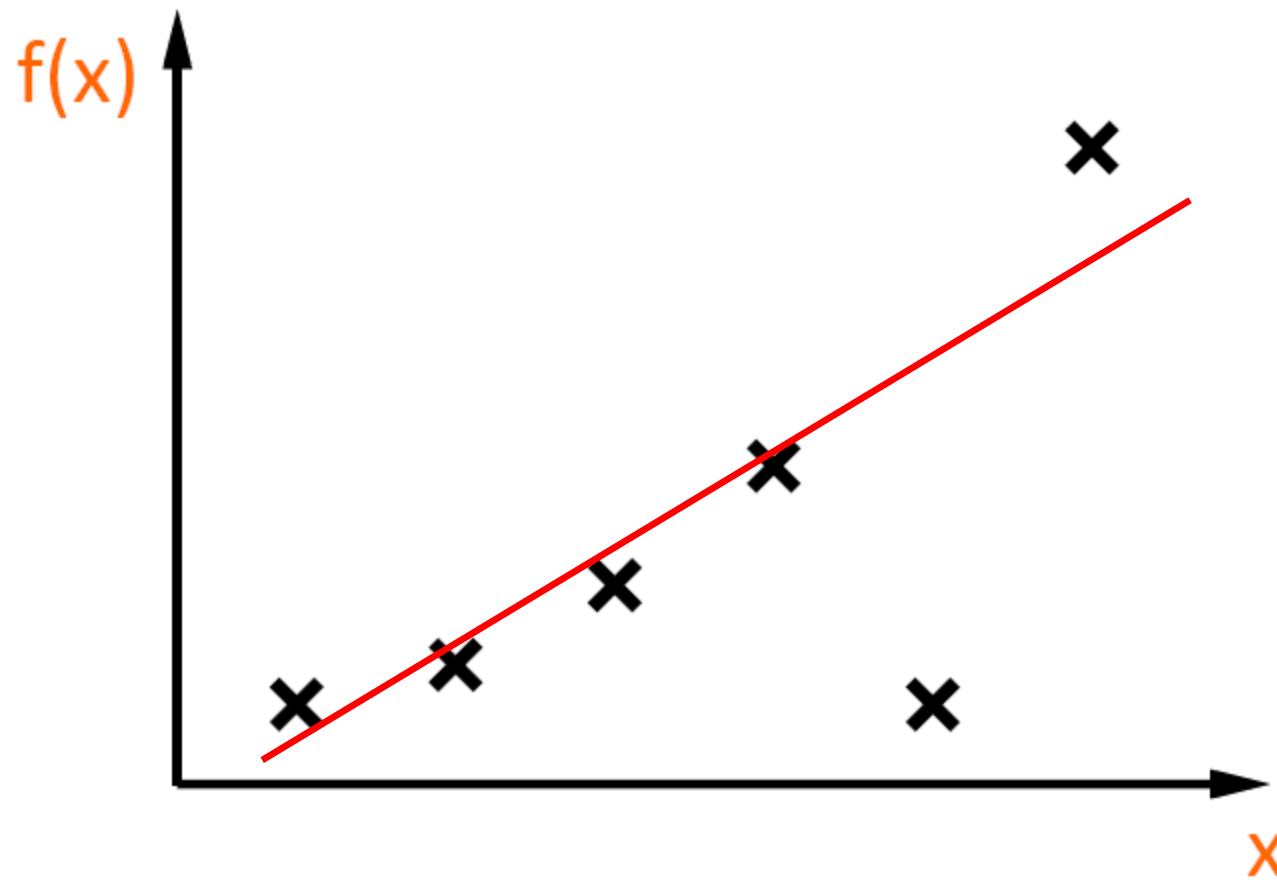
# Are complicated models preferred?

- A simple curve fitting task



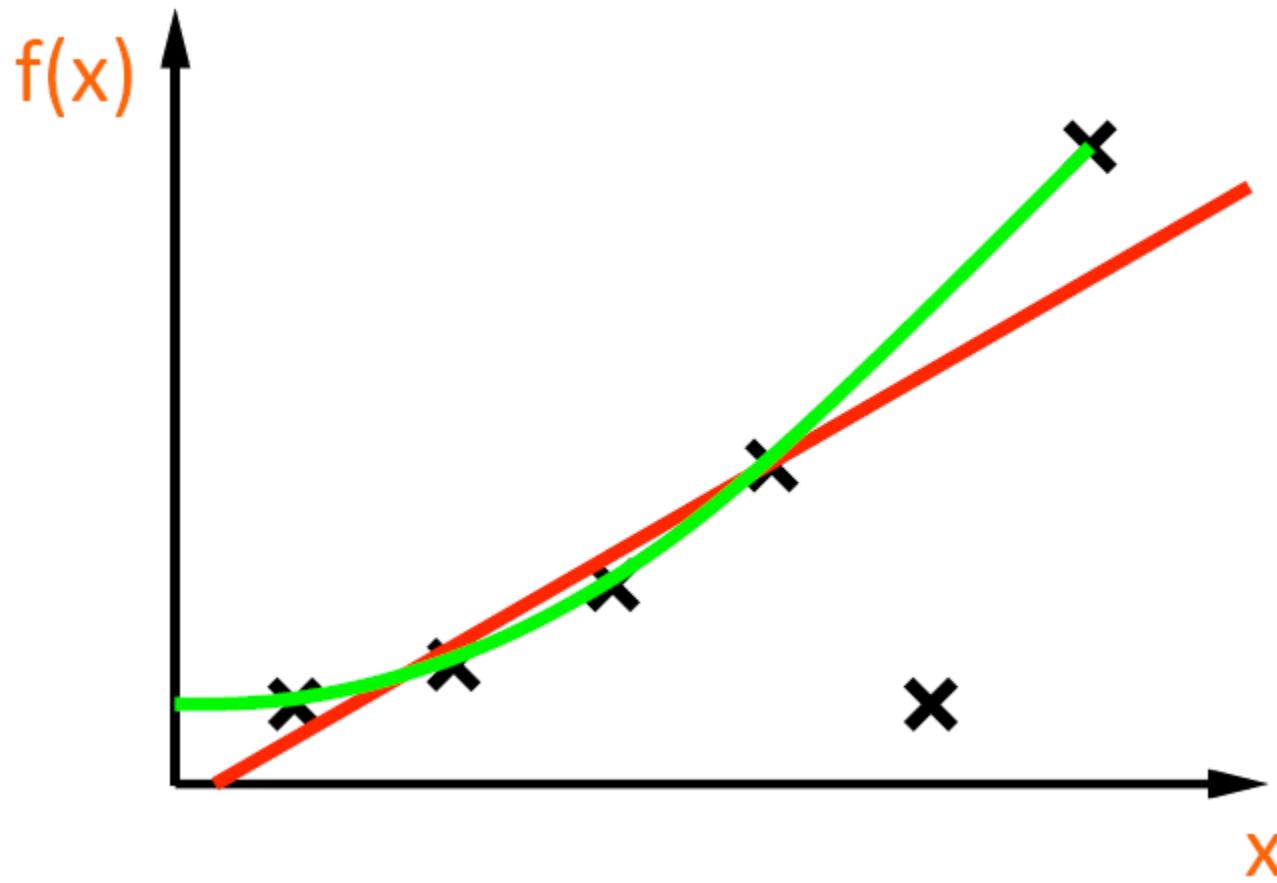
# Are complicated models preferred?

- Order = 1



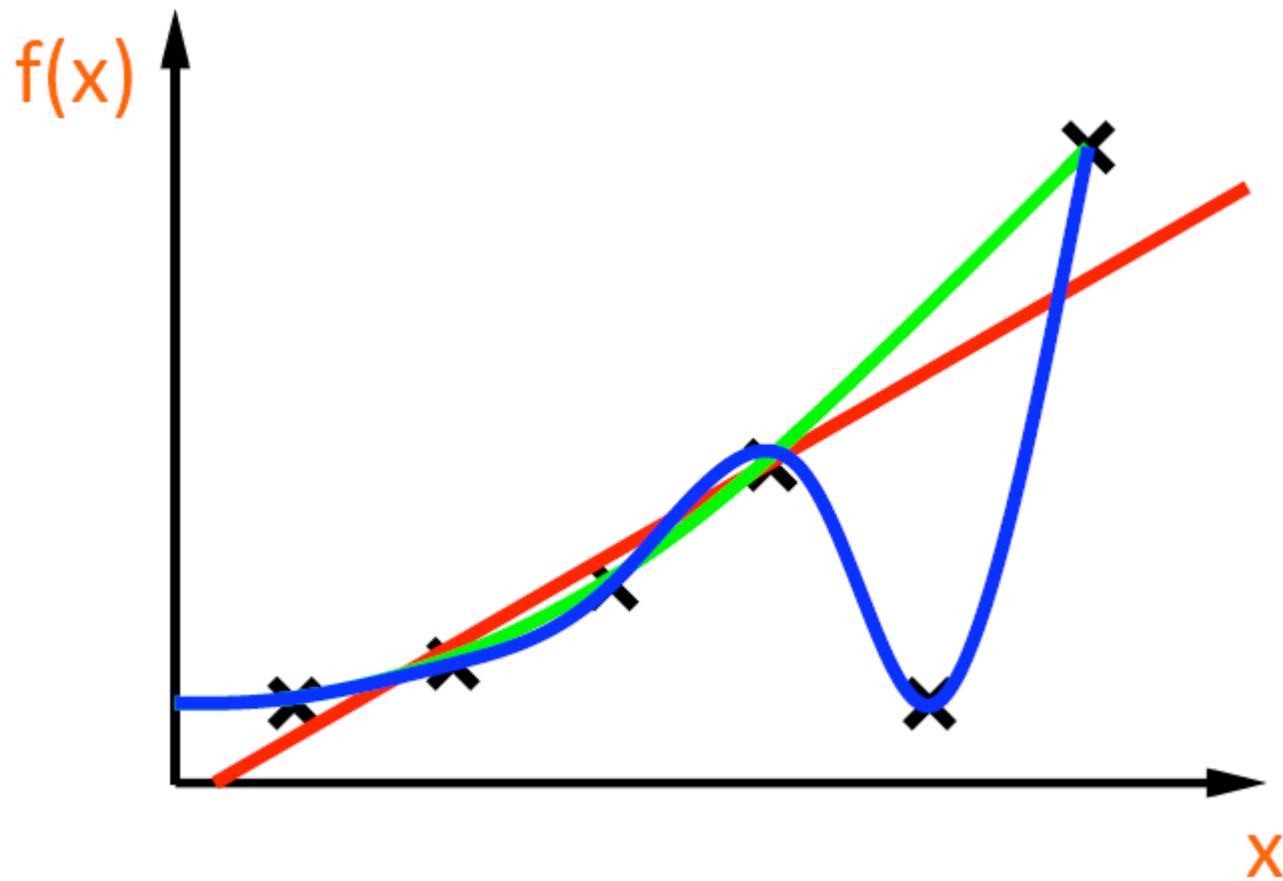
# Are complicated models preferred?

- Order = 2



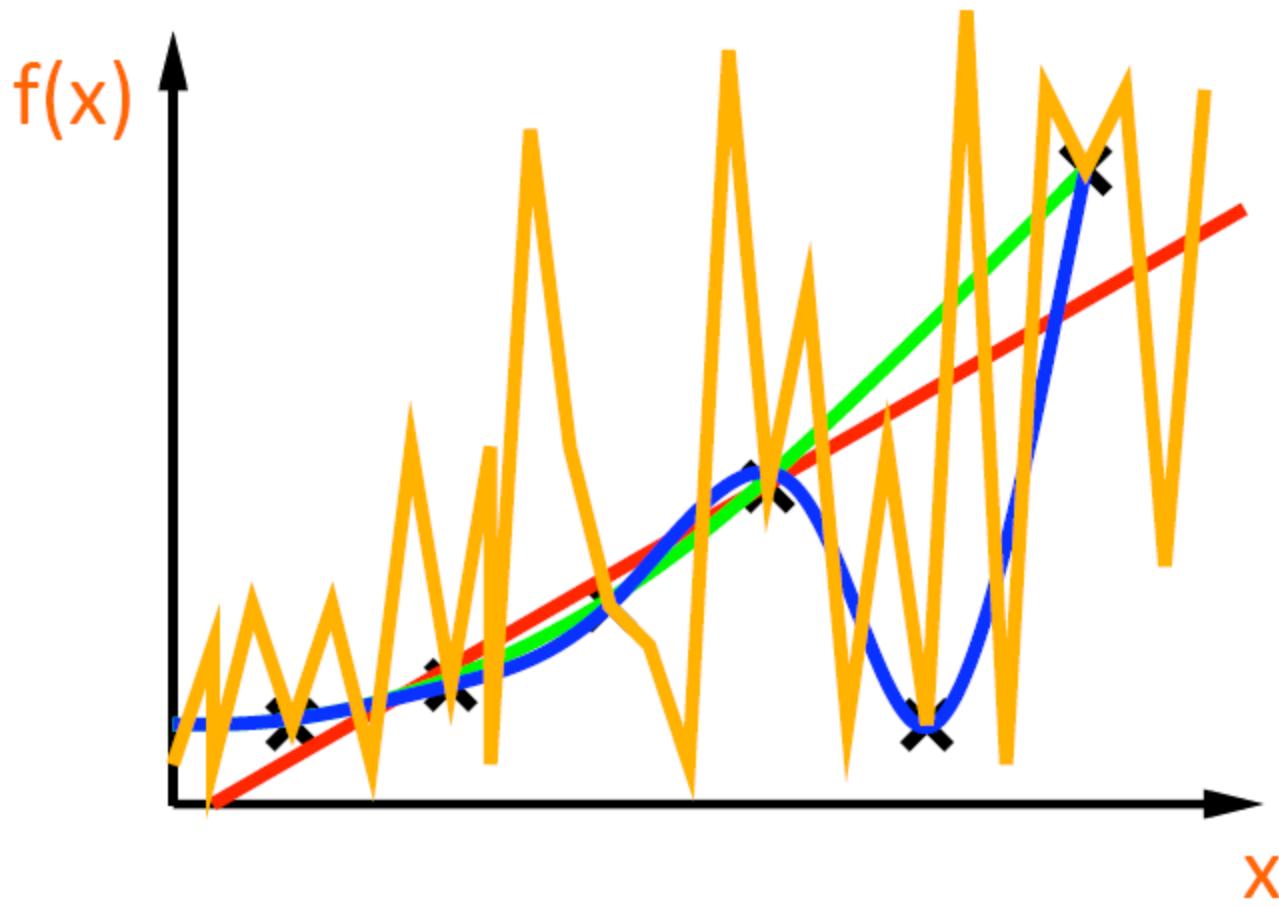
# Are complicated models preferred?

- Order = 3



# Are complicated models preferred?

- Order = 9?



# Are complicated models preferred?

Advice #2: use ML & sophisticated models when necessary



- Issues with model selection!!

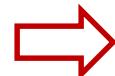
# Unsupervised Learning

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

Feature space  $\mathcal{X}$



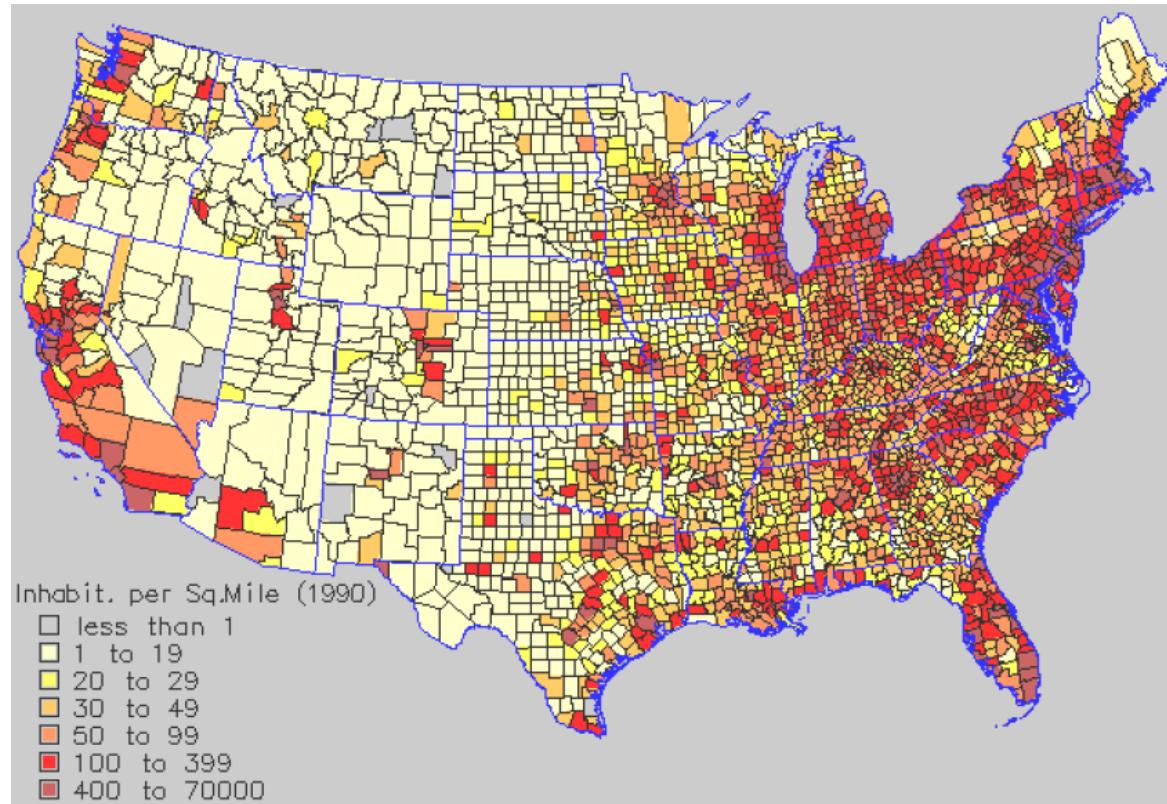
Words in documents



Word distribution  
(probability of a word)

- No training/test split

# Unsupervised Learning – density estimation



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

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

# Unsupervised Learning – clustering

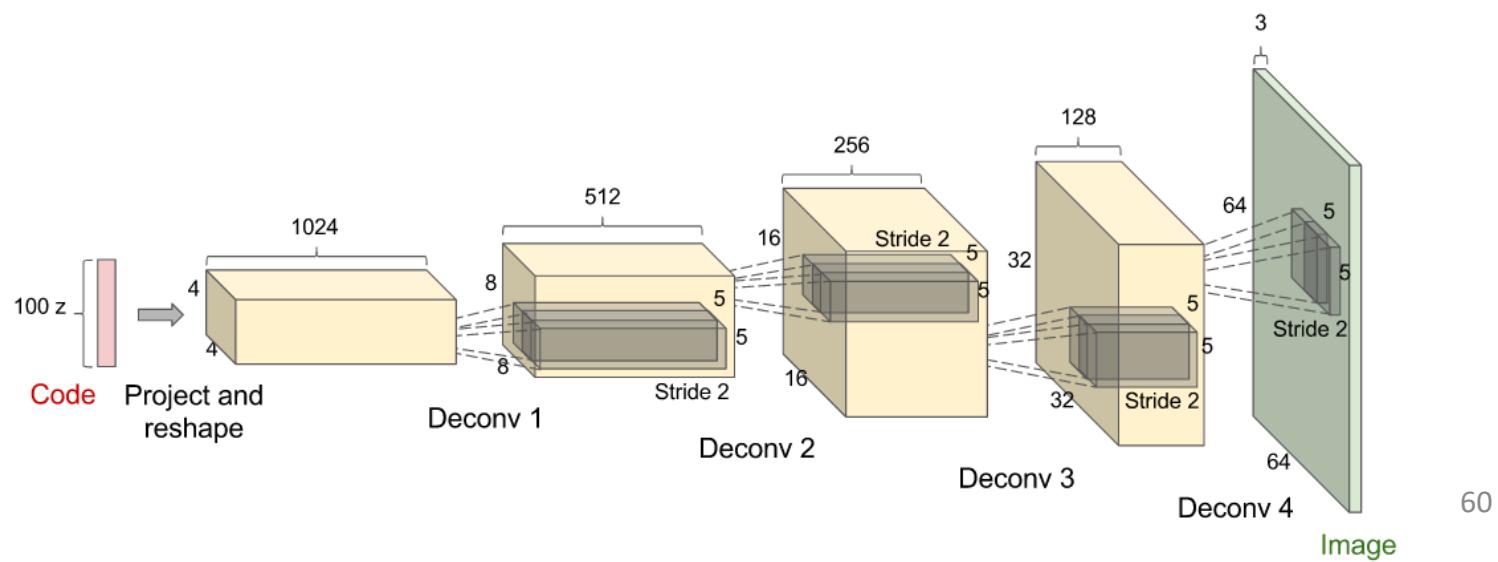
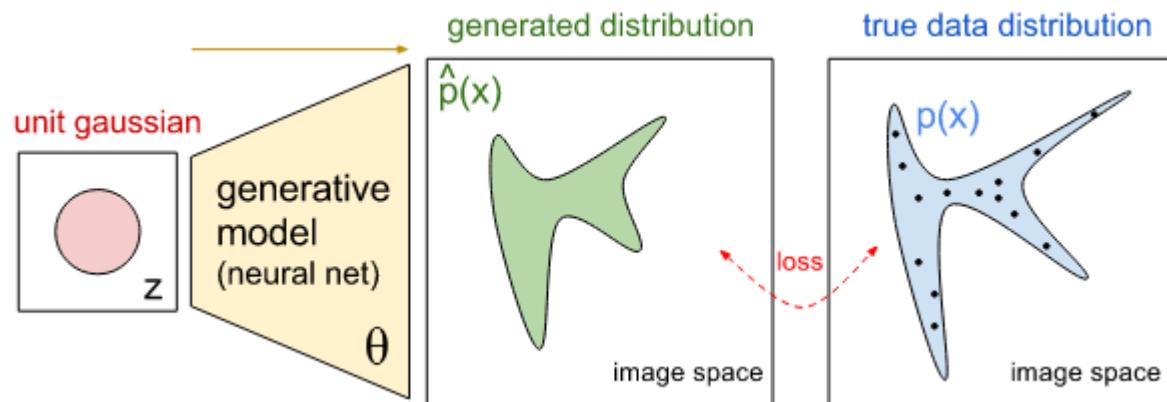


**Feature** space  $\mathcal{X}$   
Attributes (e.g., pixels & text) of images

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

# Deep Generative Models

- Learn a generative model



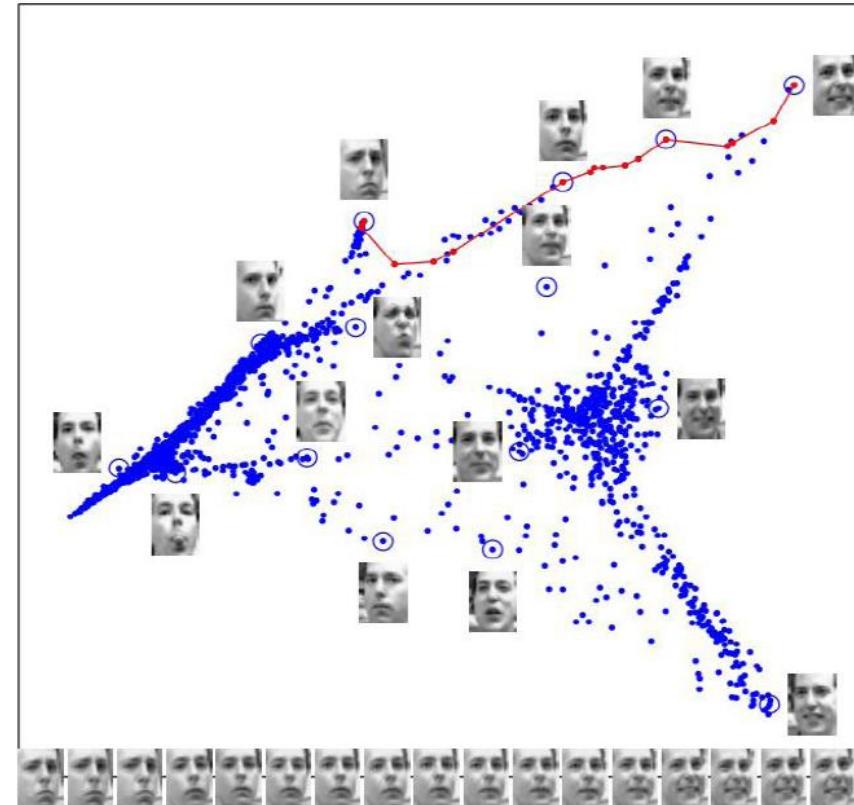
# Unsupervised Learning – dimensionality reduction

Images have thousands or millions of pixels

Can we give each image a coordinate, such that similar images are near each other ?

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

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



# 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.

# 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

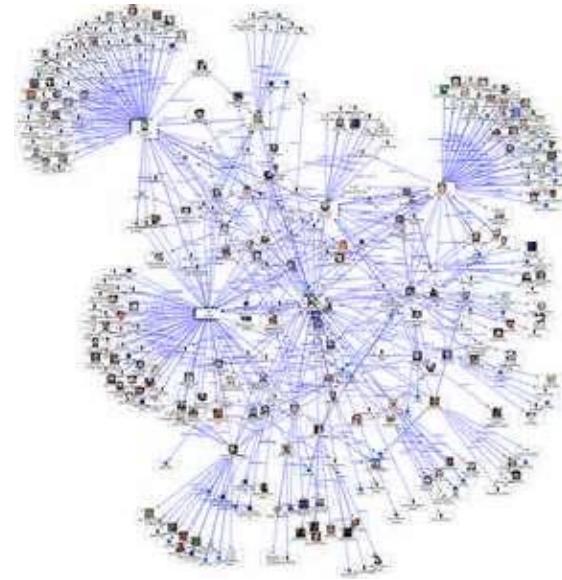
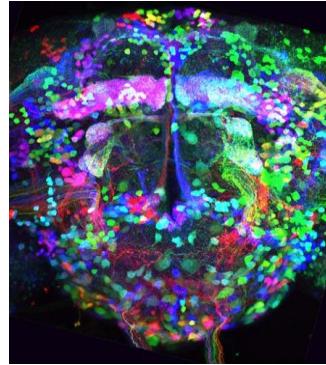
have nice properties.

# 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)
- Big Learning

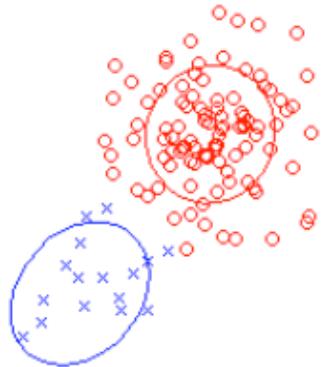


**Ideal paradigm that  
computers help solve big  
data problems**

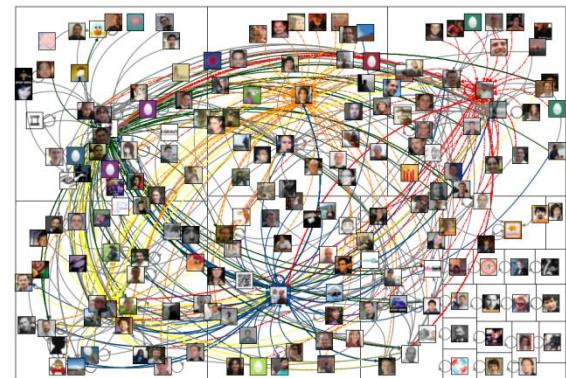
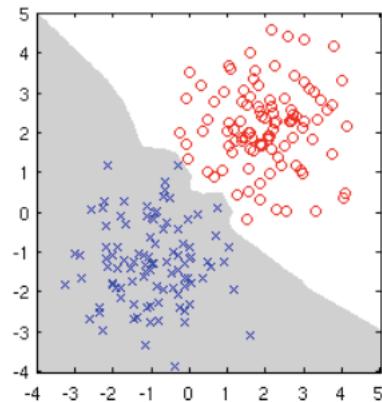
**Input Data**



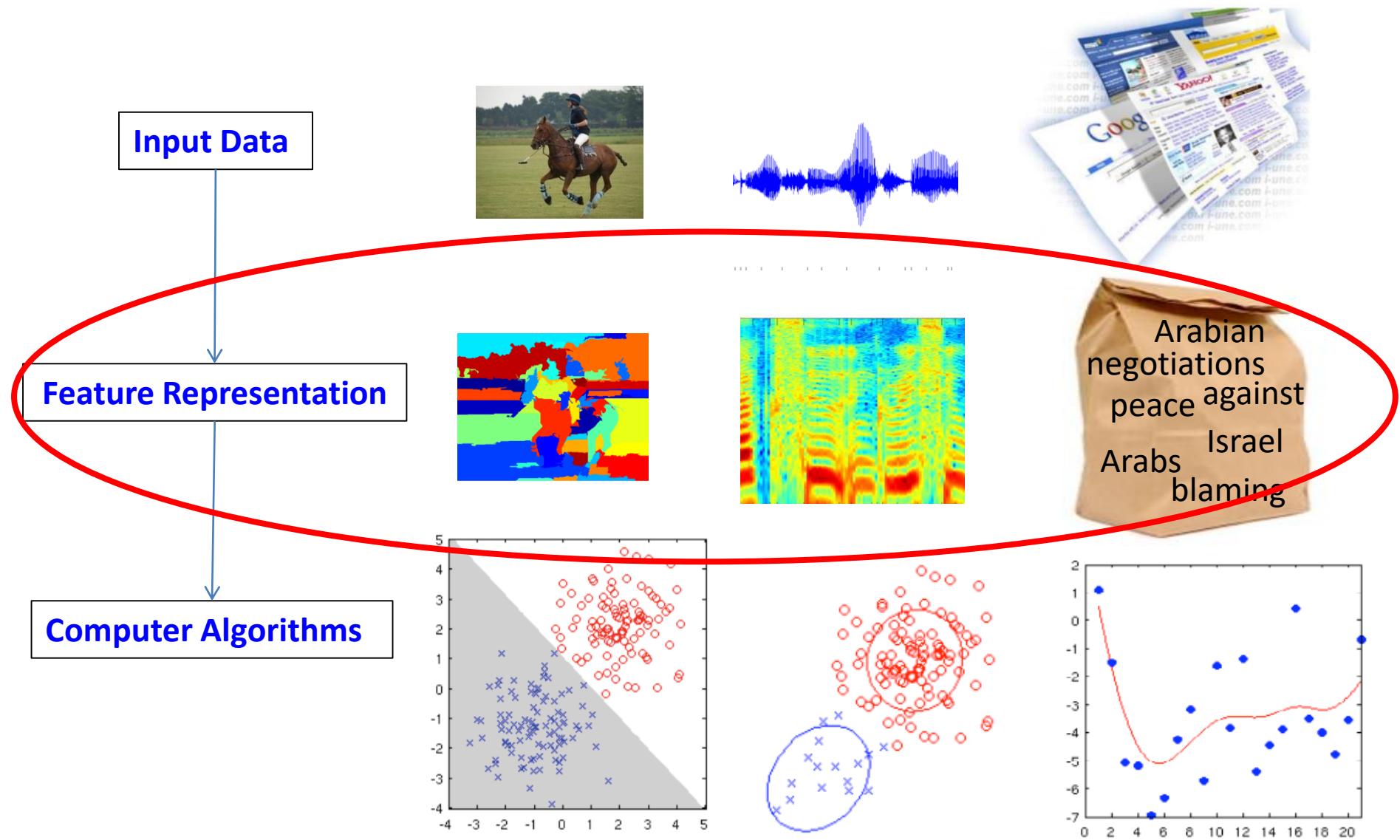
**Inference, Decision, Reasoning**



Topic 1					
	squirrel, nature, animal, wildlife, rabbit, cute, bunny, interestingness				
Topic 2					
	wolf, alaska, animal, nature, wildlife, africa, squirrel				
Topic 3					
	hawk, bird, flying, wildlife, wings, nature, fabulous, texas				
Topic 4					
	ocean, boat, animal, wildlife, diving, sea, sydney, pacific, blue				
Topic 5					
	zebra, zoo, animal, stripes, africa, mammal, black, white, nature, eyes				



# A Conventional Data Analysis Pipeline



# Representation Learning

“Lovely welcoming staff, good rooms that give a good nights sleep, downtown location”  
Meramees Hostel

 SheikhSahib 10 contributions London  
Jul 7, 2009 | Trip type: Friends getaway

This hotel is just off the side streets of Talat Harb, one of the main arteries to downtown Cairo. It is walking distance to the Nile, riverfront hotels, Egyptian Museum, and there are many eateries in the area at night when it is still bustling. Only a short cab ride away from the Old Fatimid Cairo.

The staff are young and very friendly and able to sort out things like mobile chargers, internet, and they have skype installed on their computers which is brilliant. The rooms are nicer than the Luna (nearby) and much quieter as well.

My ratings for this hotel

- Value
- Rooms
- Location
- Cleanliness

Service

Date of stay February 2009

Visit was for Leisure

Travelled with With Friends

Member since July 03, 2009

Would you recommend this place?

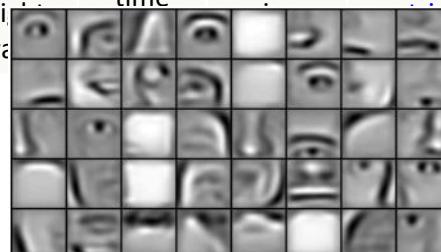


Learning  
Algorithms

E.g., Topic Models

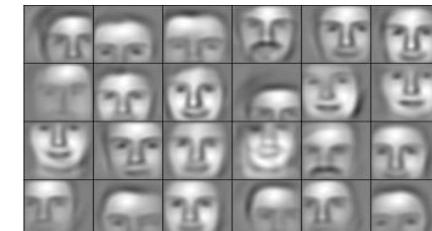
Axis's of a semantic representation space:

T1	T2	T3	T4	T5	T6	T7
told <b>dirty</b> room front asked hotel <b>bad</b> <b>small</b> <b>worst</b> <b>poor</b> called <b>rude</b>	place hotel room days time day night people stay water rooms food	hotel food bar day pool time service holida y room people night wa	area staff pool breakfast day view location service walk time	beach pool resort food island kids trip service day staff time	beach resort pool ocean island kids <b>good</b> restaurant s <b>enjoyed</b> <b>loved</b>	great good nice lovely beautiful excellent wonderful comfortable beach friendly fresh amazing



Learning  
Algorithms

E.g., Deep Networks



[Figures from (Lee et al., ICML2009)]

# History of neural networks



Pitts



Rosenblatt



Minsky

Papert

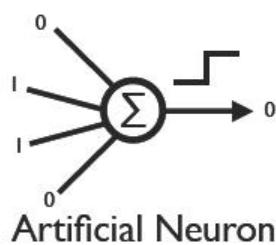


Ackley

Hinton

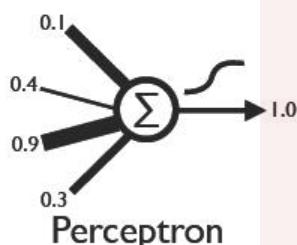
Sejnowski

1943



Artificial Neuron

1960



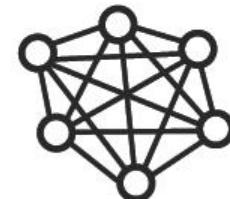
Perceptron

1969



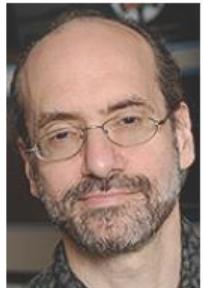
Perceptrons

1985



Boltzmann Machine

# History of neural networks



Smolensky

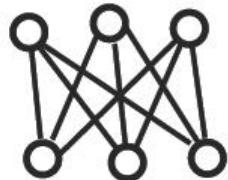


Hinton

Hinton et al.

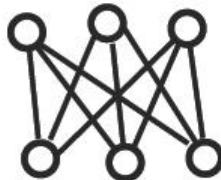


1986



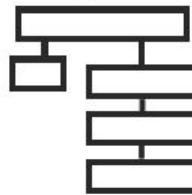
Harmoniums  
(Restricted Boltzmann Machine)

2002



Contrastive  
Divergence

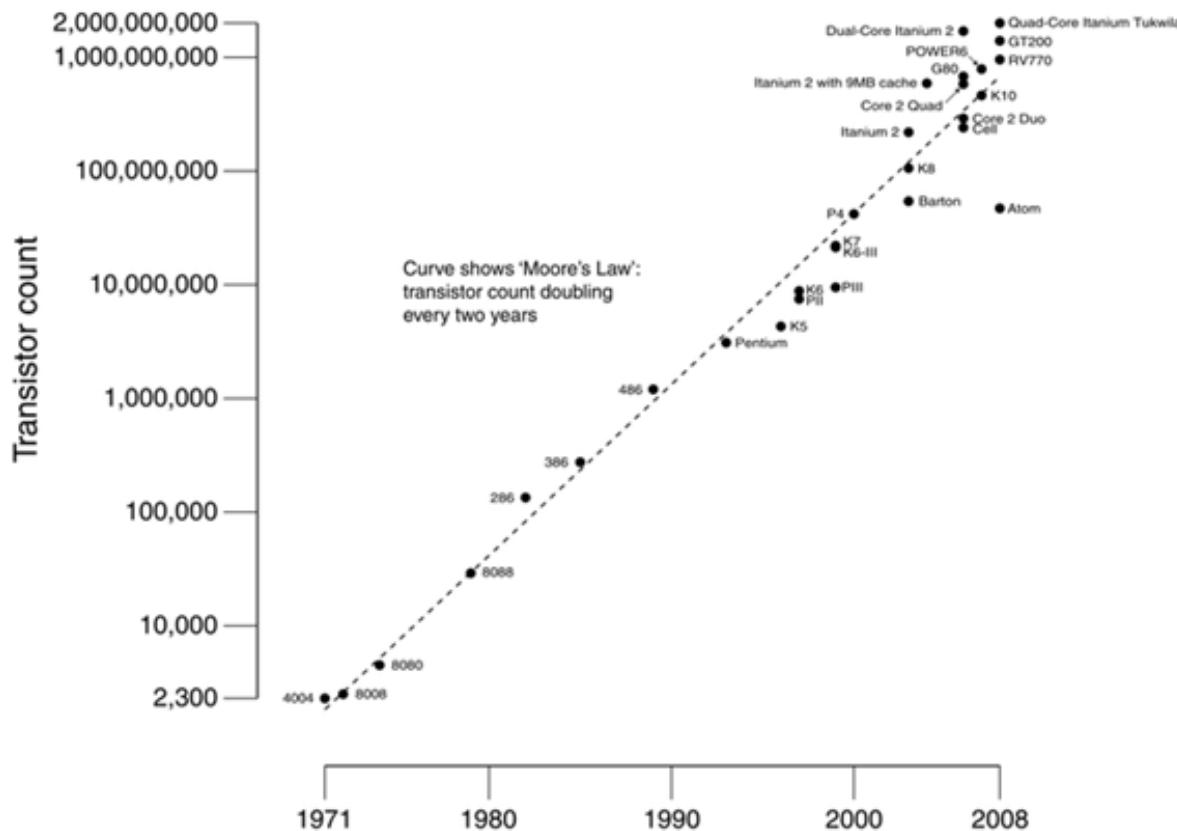
2006



Deep Belief  
Networks

# Overfitting in Big Data

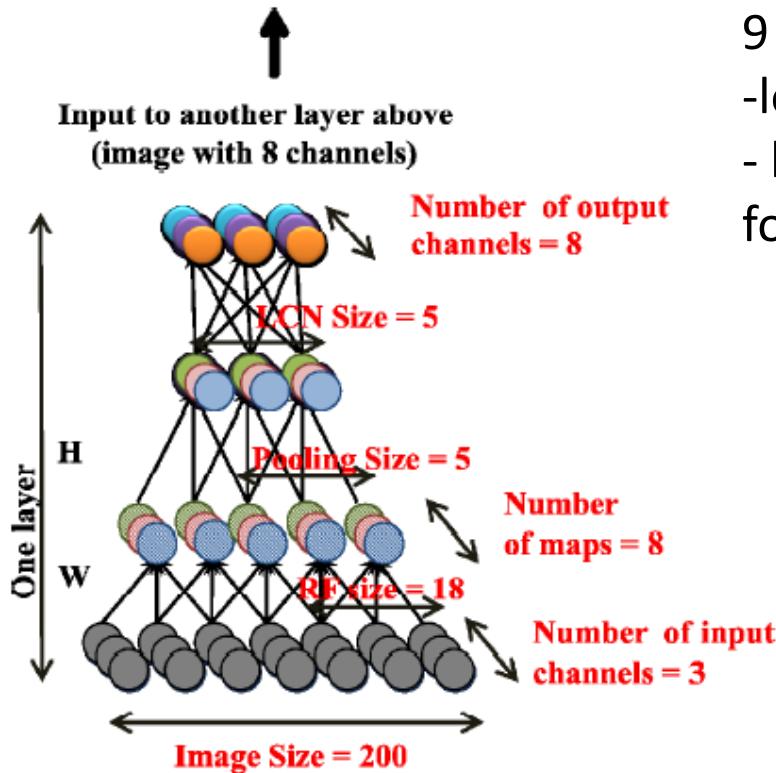
- “with more data overfitting is becoming less of a concern”?



# Overfitting in Big Data

**“Big Model + Big Data + Big/Super Cluster”**

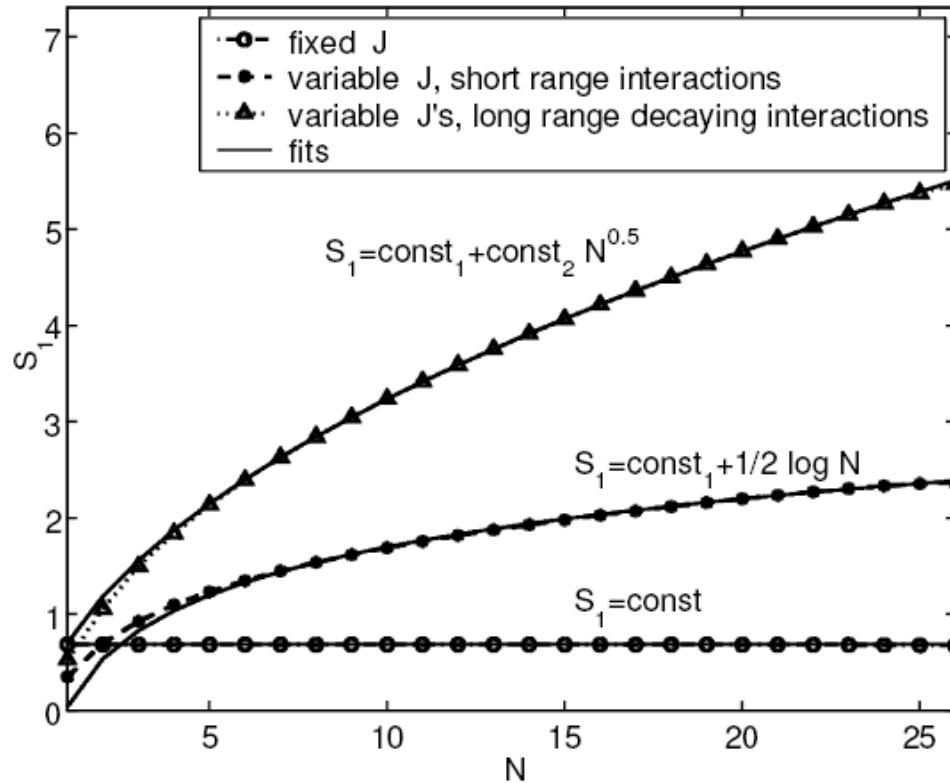
## Big Learning



- 9 layers sparse autoencoder with:
- local receptive fields to scale up;
  - local L2 pooling and local contrast normalization for invariant features
    - 1B parameters (connections)
    - 10M 200x200 images
    - train with 1K machines (16K cores) for 3 days
  - able to build high-level concepts, e.g., cat faces and human bodies
  - 15.8% accuracy in recognizing 22K objects (70% relative improvements)

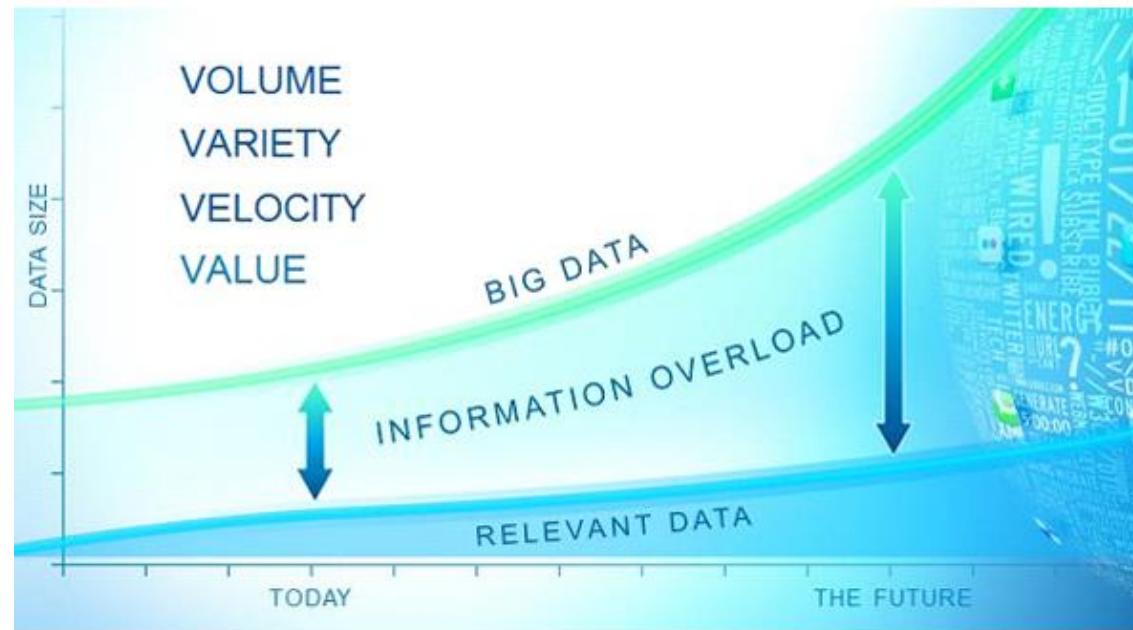
# Overfitting in Big Data

- **Predictive information** grows slower than the amount of Shannon entropy (Bialek et al., 2001)



# Overfitting in Big Data

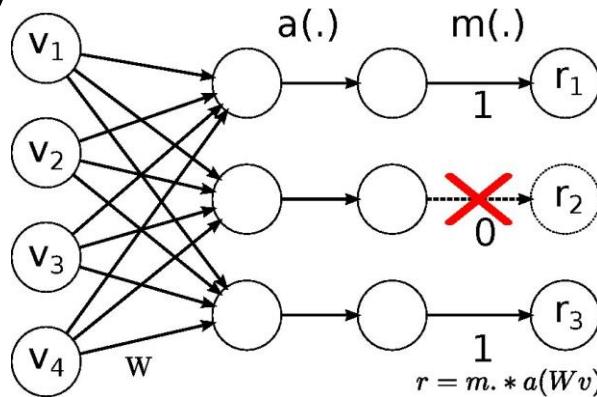
- **Predictive information** grows slower than the amount of Shannon entropy (Bialek et al., 2001)



**Model capacity grows faster than the amount of predictive information!**

# Overfitting in Big Data

- Surprisingly, regularization to prevent overfitting is ***increasingly important***, rather than increasingly irrelevant!
- Increasing research attention, e.g., dropout training (Hinton, 2012)



- More theoretical understanding and extensions
  - MCF (van der Maaten et al., 2013); Logistic-loss (Wager et al., 2013); Dropout SVM (Chen, Zhu et al., 2014)

# Why Big Data could be a Big Fail?



**Michael I. Jordan**

UC Berkeley

Pehong Chen Distinguished Professor

NAS, NAE, NAAS Fellow

ACM, IEEE, IMS, ASA, AAAI Fellow



- When you have large amounts of data, your appetite for hypotheses tends to get even larger
- If it's growing faster than the statistical strength of the data, then many of your inferences are likely to be false. They are likely to be white noise.
- Too much hype: “The whole big-data thing came and went. It died. It was wrong”

## Therefore ...

- Computationally efficient Bayesian models are becoming increasingly relevant in Big data era
  - **Relevant:** high capacity models need a protection
  - **Efficient:** need to deal with large data volumes

# Big Learning with Bayesian Methods

- Basics, Algorithms, Systems, Examples
  - Big Learning with Bayesian Methods, J. Zhu, J. Chen, & W. Hu, arXiv 1411.6370, preprint, 2014

# Resources for Further Learning

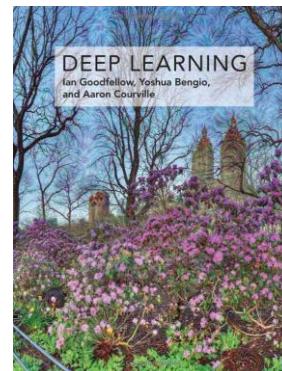
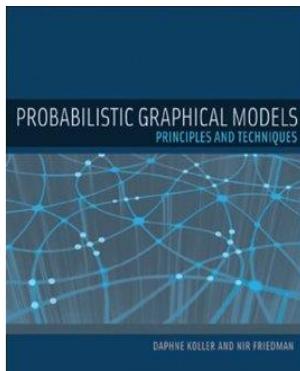
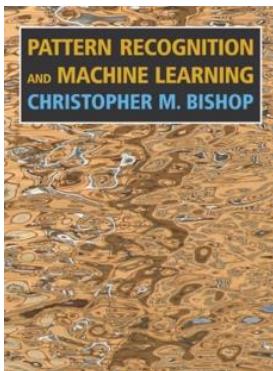
- Top-tier Conferences:
  - International Conference on Machine Learning (ICML)
  - Advances in Neural Information Processing Systems (NIPS)
  - Uncertainty in Artificial Intelligence (UAI)
  - International Joint Conference on Artificial Intelligence (IJCAI)
  - AAAI Annual Conference (AAAI)
  - Artificial Intelligence and Statistics (AISTATS)
- 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)
  - Neural Computation

# Hot Topics from ICML & NIPS

- Hot topics:
  - Deep Learning with Rich Model Architecture
  - Probabilistic Latent Variable Models & Bayesian Nonparametrics
  - Sparse Learning in High Dimensions
  - Large-scale Optimization and Inference
  - Online learning
  - Reinforcement Learning
  - Learning Theory
  - Interdisciplinary Research on Machine Learning, Cognitive Science , etc.

# 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/>

# Thanks!

**Jun Zhu**

[dcszj@mail.tsinghua.edu.cn](mailto:dcszj@mail.tsinghua.edu.cn)

<http://bigml.cs.tsinghua.edu.cn/~jun>