VBM683
Machine Learning

Pinar Duygulu

Slides are adapted from
Dhruv Batra, David Sontag, Aykut Erdem

Quotes

- “If you were a current computer science student what area would you start studying heavily?”
  - Answer: Machine Learning.
  - “The ultimate is computers that learn”
  - Bill Gates, Reddit AMA

- “Machine learning is the next Internet”
  - Tony Tether, Director, DARPA

- “Machine learning is today’s discontinuity”
  - Jerry Yang, CEO, Yahoo

Slide Credit: Pedro Domingos, Tom Mitchel, Tom Dietterich

Two definitions of learning

(1) Learning is the acquisition of knowledge about the world.

Kupfermann (1985)

(2) Learning is an adaptive change in behavior caused by experience.

Shepherd (1988)

Slide by Bernhard Schölkopf

What is Machine Learning?

- [Arthur Samuel, 1959]
  - Field of study that gives computers
  - the ability to learn without being explicitly programmed

- [Kevin Murphy] algorithms that
  - automatically detect patterns in data
  - use the uncovered patterns to predict future data or other outcomes of interest

- [Tom Mitchell] algorithms that
  - improve their performance (P)
  - at some task (T)
  - with experience (E)

What is Machine Learning?

- If you are a Scientist

- If you are an Engineer / Entrepreneur
  - Get lots of data
  - Machine Learning
  - ???
  - Profit!

Acquisitions

Google snaps up object recognition startup DNNr

Google has ac
Toronto, who

by Josh Lowensohn

2 / 0

Google has acqui
research compan

« Search needs a shake-up

Songbirds use grammar rules »

Machine Learning Startup Acquired by ai-one

Press Release

For Immediate Release: August 4, 2011

San Diego artificial intelligence startup acquired by leading
ning SDKs as market for advanced

oday that it acquired Auto-Semantics, a local start-up
es to corporate IT departments. The acquisition is the
nd acquisitions by ai-one that consolidates its
ging market for machine learning technologies.

Topic: Cloud

Follow via:

Microsoft acquires legal-focused machine-learning vendor Equivio

Summary: *Microsoft has purchased Equivio, maker of a machine-learning platform for the legal industry, for an undisclosed amount.*

By Mary Jo Foley for All About Microsoft | January 20, 2015 -- 16:24 GMT (08:24 PST)

Follow @maryjofoley 95.6K followers Get the ZDNet Cloud newsletter now

Microsoft has purchased Equivio, an eDiscovery/compliance vendor with a specialization in text analysis, for an undisclosed amount.

Microsoft officials announced the acquisition of the Israeli company -- its first acquisition of 2015 using more of its offshore cash -- on January 20.

Update: The Wall Street Journal reported back in October last year that Microsoft planned to buy Equivio for $200 million.

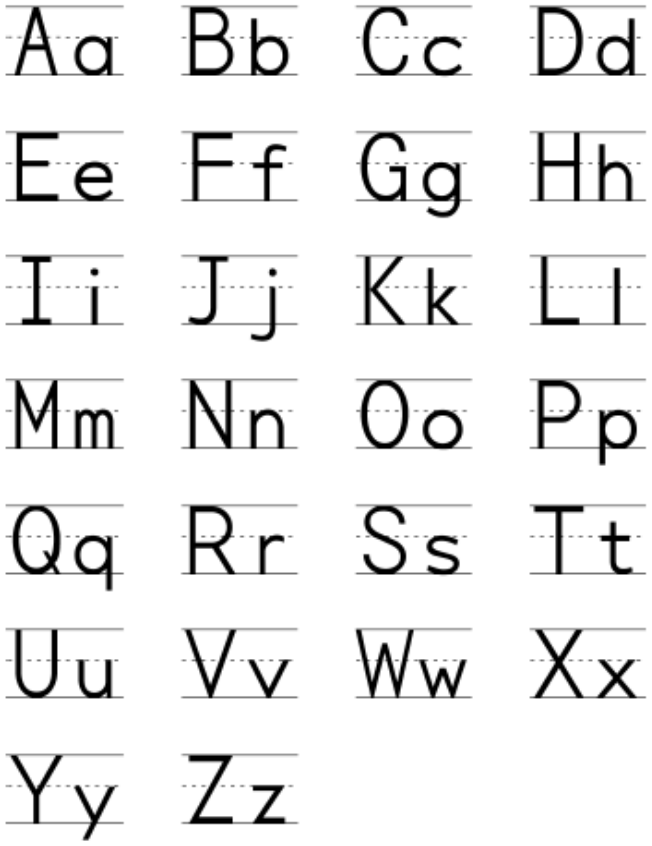
Update No. 2: A Microsoft spokesperson said the $200 million estimate was inflated and incorrect, but declined to provide a different figure.

FOUNDED
2011

OVERVIEW
DeepMind is a cutting edge artificial intelligence company. We combine the best techniques from machine learning and systems neuroscience to build powerful general-purpose learning algorithms. Founded by Demis Hassabis, Shane Legg and Mustafa Suleyman, the company is based in London and supported by some of the most iconic technology entrepreneurs and investors of the past decade. Our first commercial ...

What is Machine Learning?

- Let's say you want to solve Character Recognition
- Hard way: Understand handwriting/characters



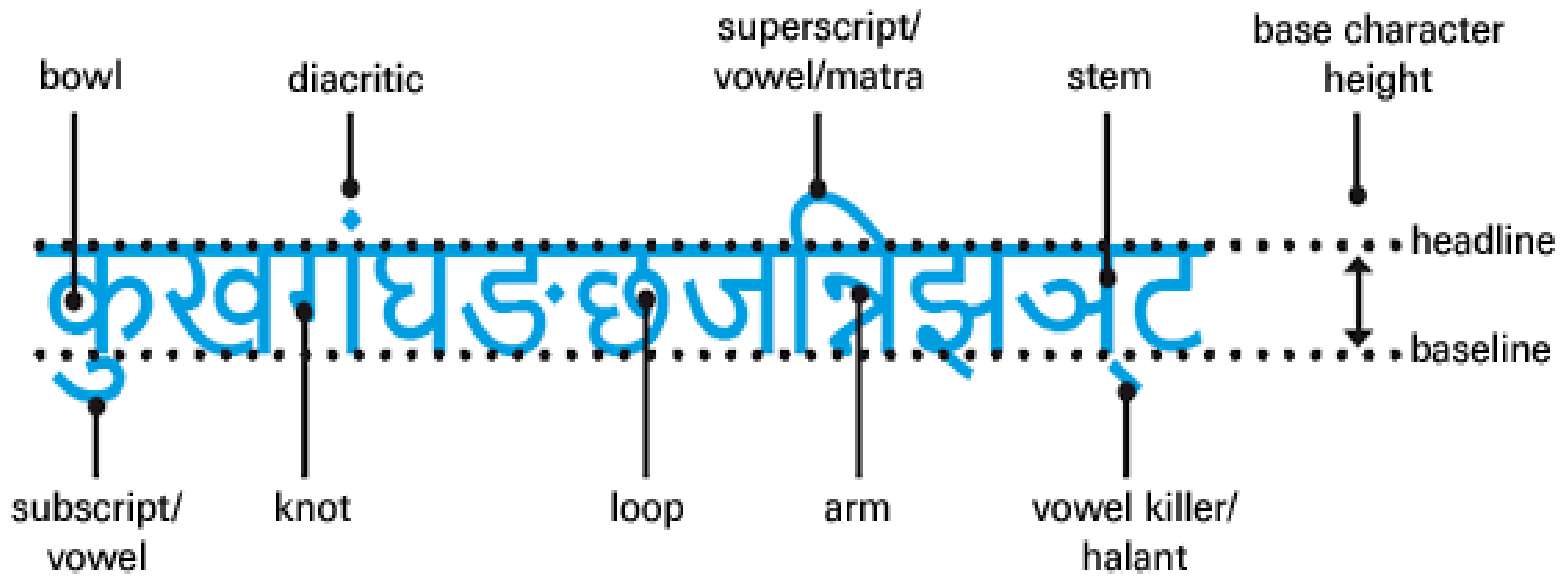
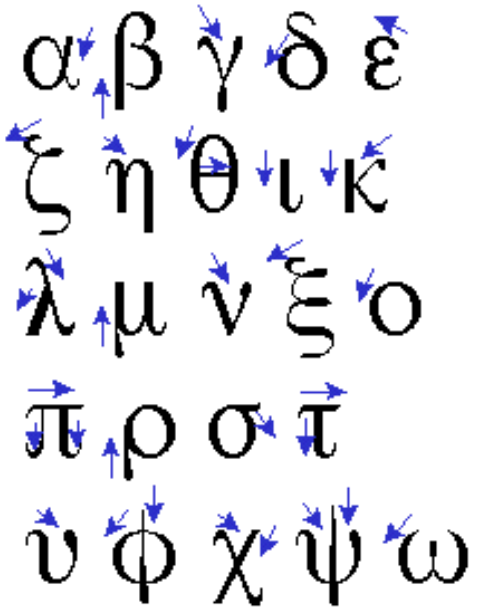


Image Credit: http://www.linotype.com/6896/devanagari.html

What is Machine Learning?

- Let's say you want to solve Character Recognition

- Hard way: Understand handwriting/characters
  - Latin
  - Devanagri
  - Symbols: http://detexify.kirelabs.org/classify.html



What is Machine Learning?

- Let’s say you want to solve Character Recognition
- Hard way: Understand handwriting/characters
- Lazy way: Throw data!

Example: Netflix Challenge

- Goal: Predict how a viewer will rate a movie

- 10% improvement = 1 million dollars

(C) Dhruv Batra

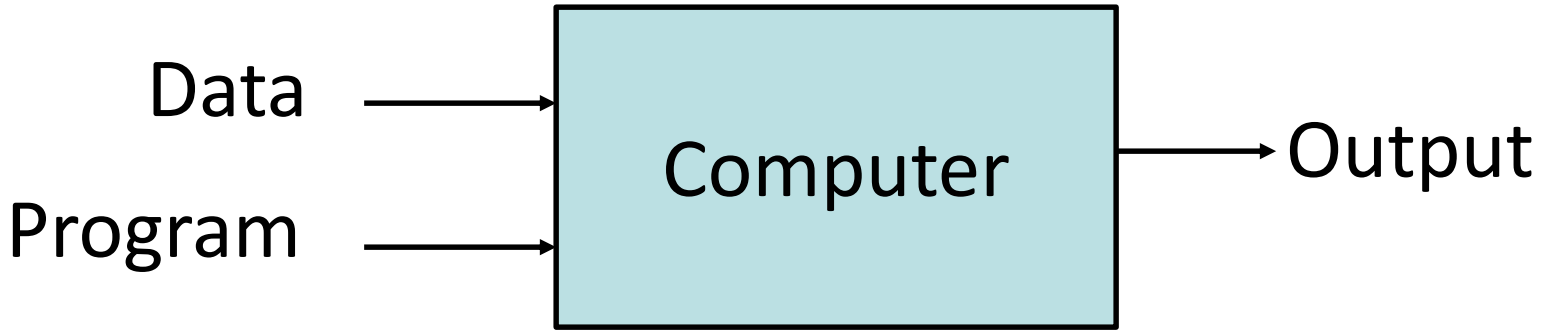
Slide Credit: Yaser Abu-Mostapha

Essence of Machine Learning:

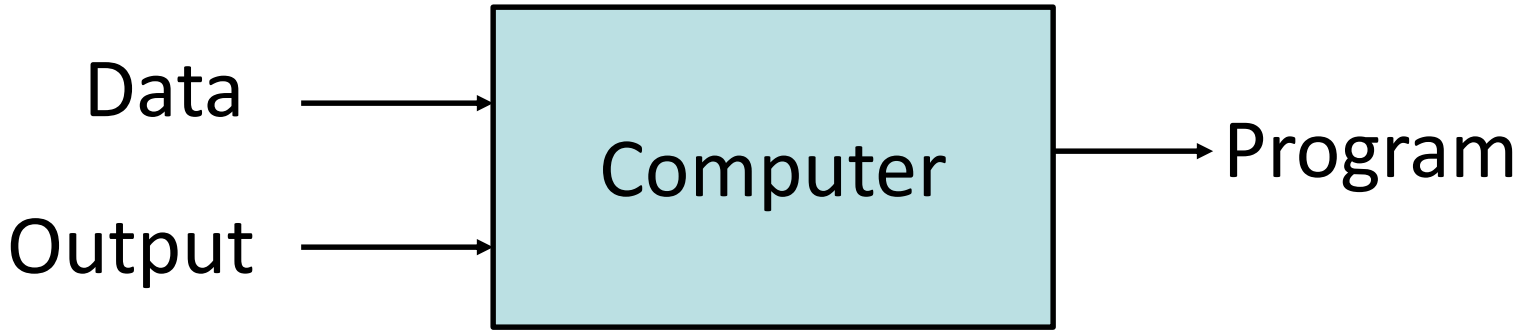
- A pattern exists
- We cannot pin it down mathematically
- We have data on it

Comparison

- **Traditional Programming**



- **Machine Learning**



Slide Credit: Pedro Domingos, Tom Mitchel, Tom Dietterich

Why Study Machine Learning?

Engineering Better Computing Systems

- Develop systems
  - too difficult/expensive to construct manually
  - because they require specific detailed skills/knowledge
  - ***knowledge engineering bottleneck***

- Develop systems
  - that adapt and customize themselves to individual users.
  - Personalized news or mail filter
  - Personalized tutoring

- Discover new knowledge from large databases
  - Medical text mining (e.g. migraines to calcium channel blockers to magnesium)
  - ***data mining***

Slide Credit: Ray Mooney

Why Study Machine Learning?
Cognitive Science

- Computational studies of learning may help us understand learning in humans
  - and other biological organisms.

  - Hebbian neural learning
    - “Neurons that fire together, wire together.”

Why Study Machine Learning?

The Time is Ripe

- More compute power
- More data
- Better algorithms /models

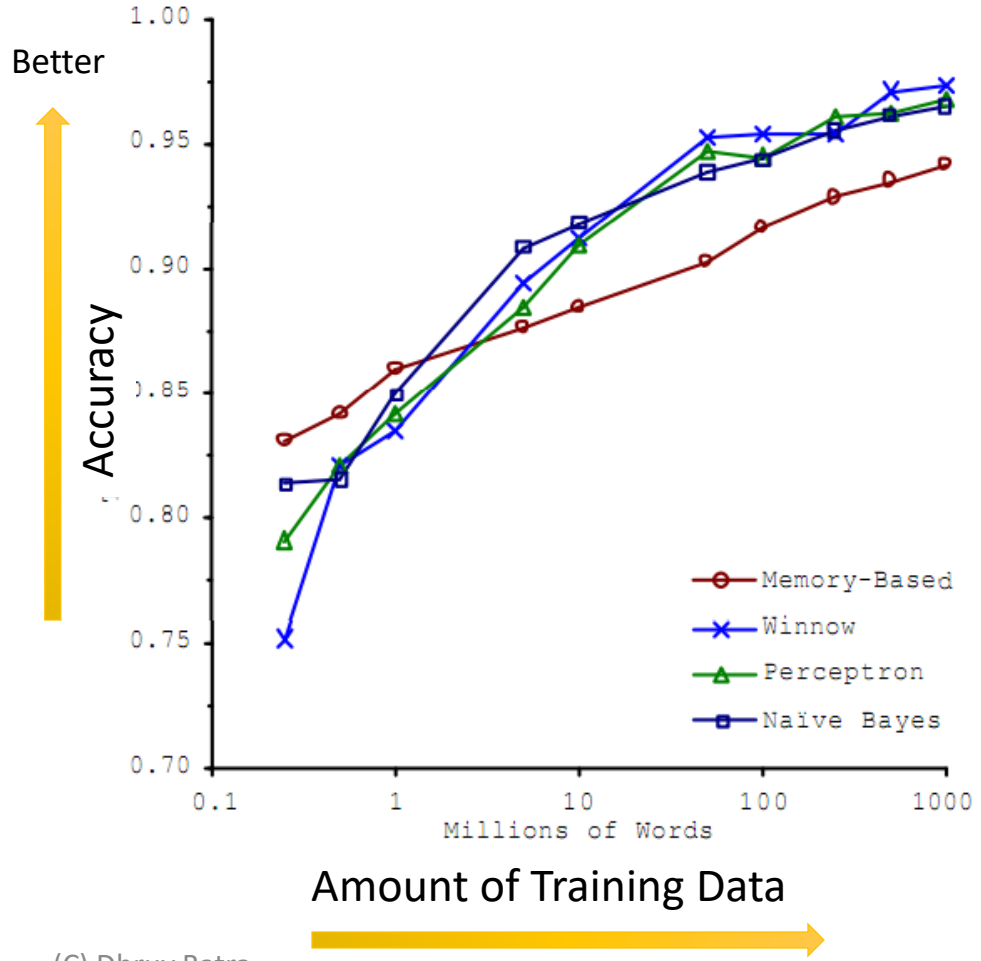
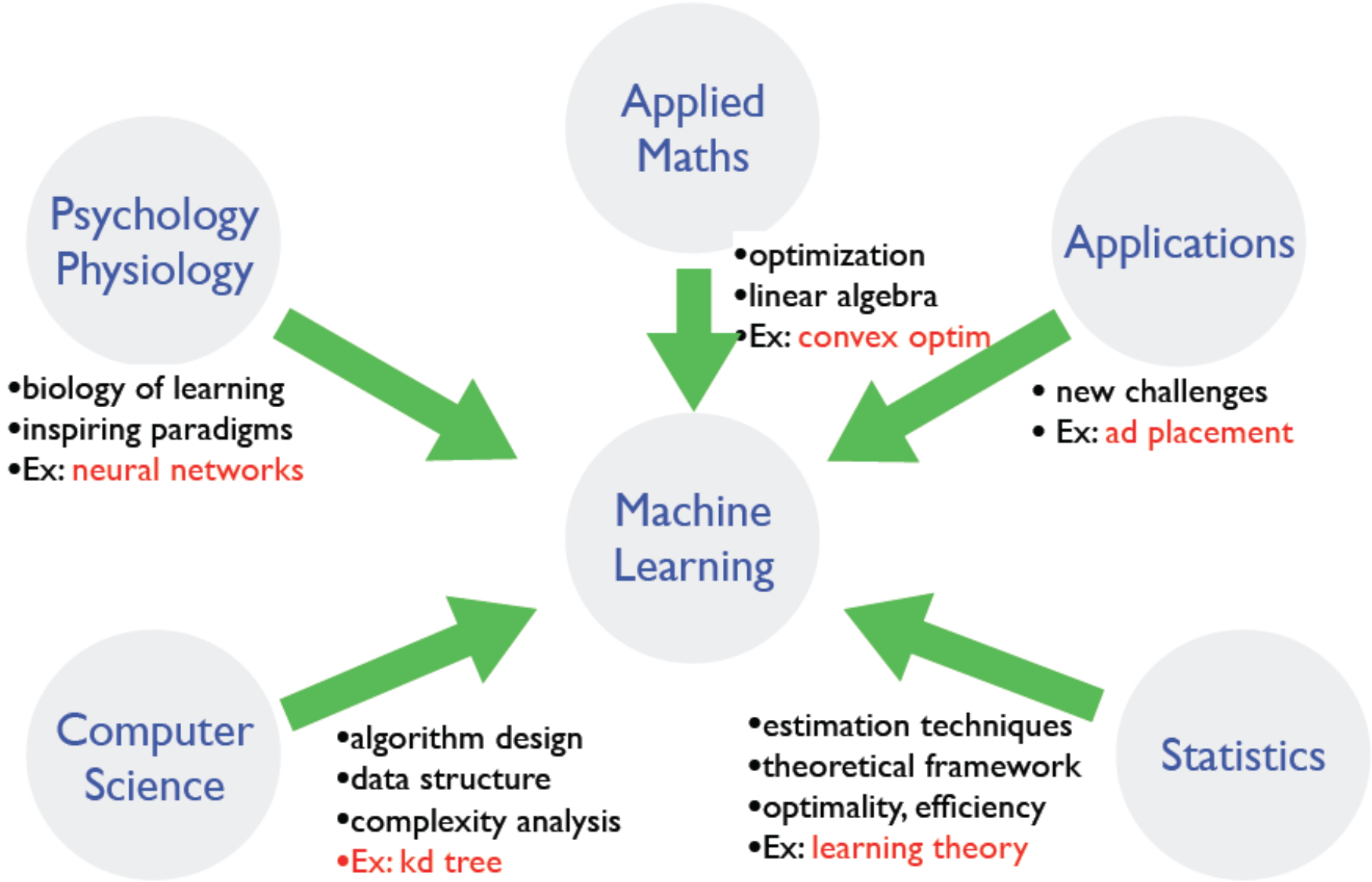


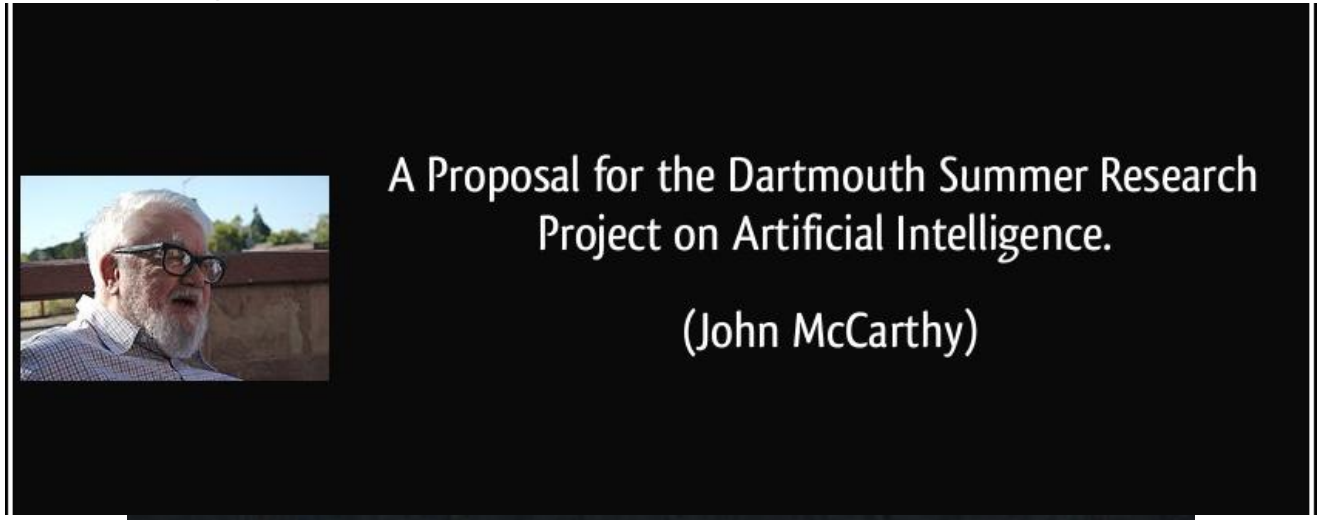
Figure Credit: Banko & Brill, 2011

Where does ML fit in?



Slide Credit: Fei Sha

A Brief History of AI



A Brief History of AI

- "We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire."

- The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

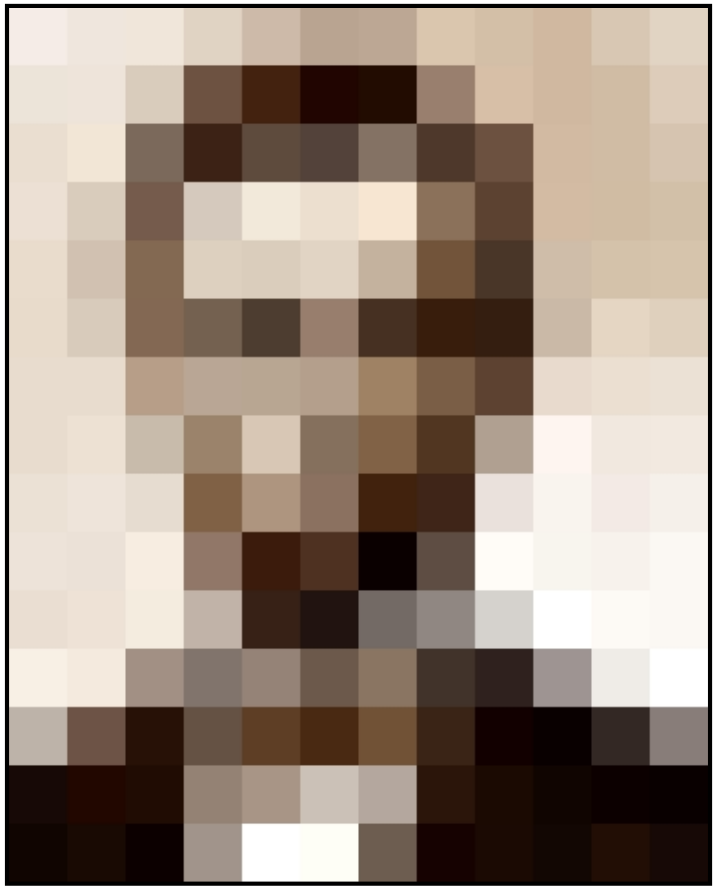
- An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

- We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

Why is AI hard?

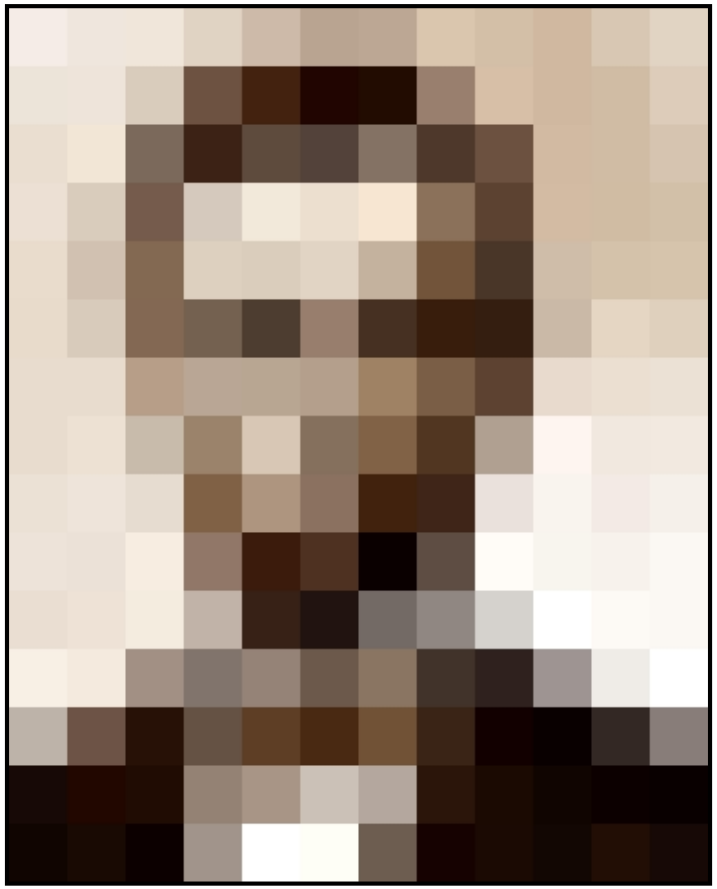
Slide Credit: http://karpathy.github.io/2012/10/22/state-of-computer-vision/

What humans see



Slide Credit: Larry Zitnick

What computers see



Slide Credit: Larry Zitnick

“I saw her duck”

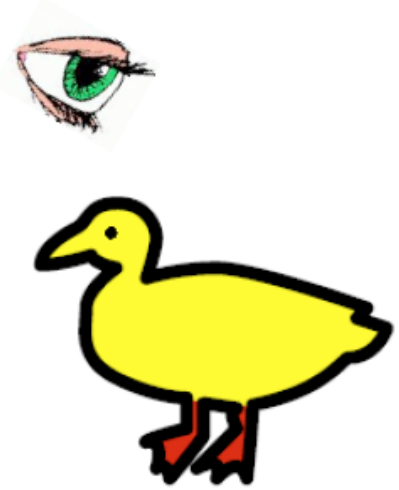


Image Credit: Liang Huang

“I saw her duck”

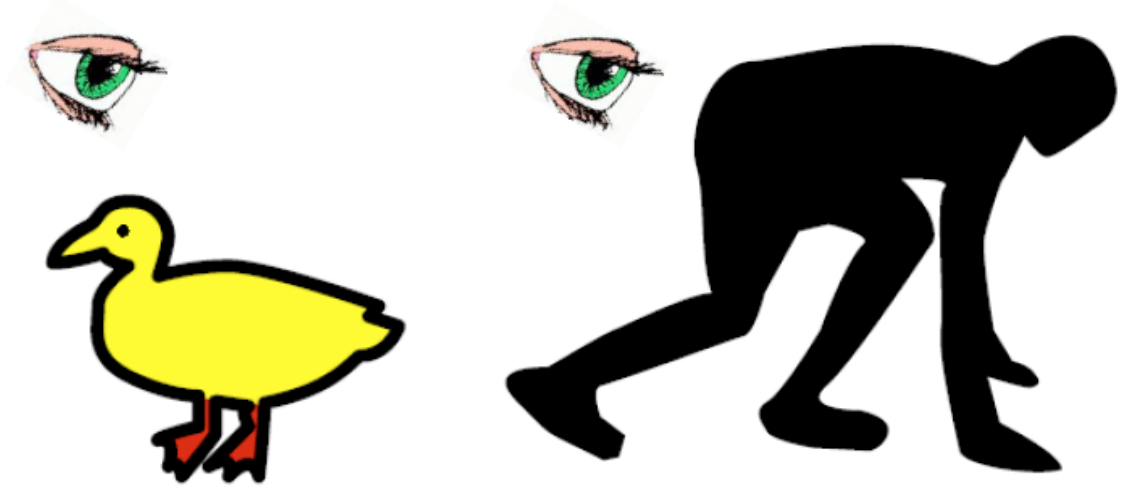


Image Credit: Liang Huang

“I saw her duck”

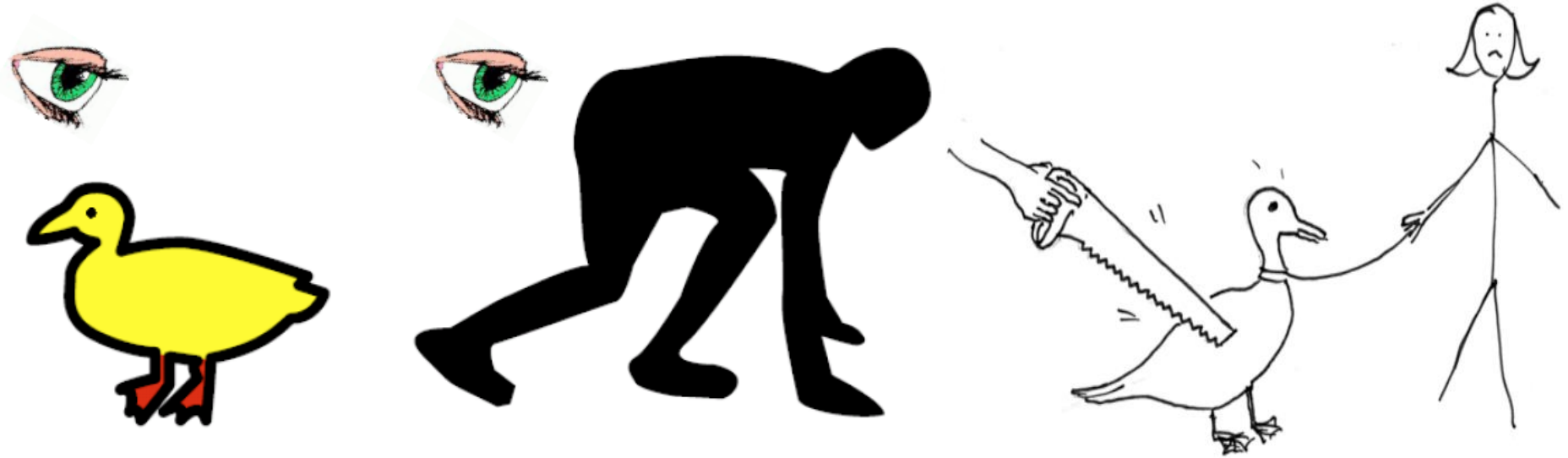


Image Credit: Liang Huang

“I saw her duck with a telescope…”

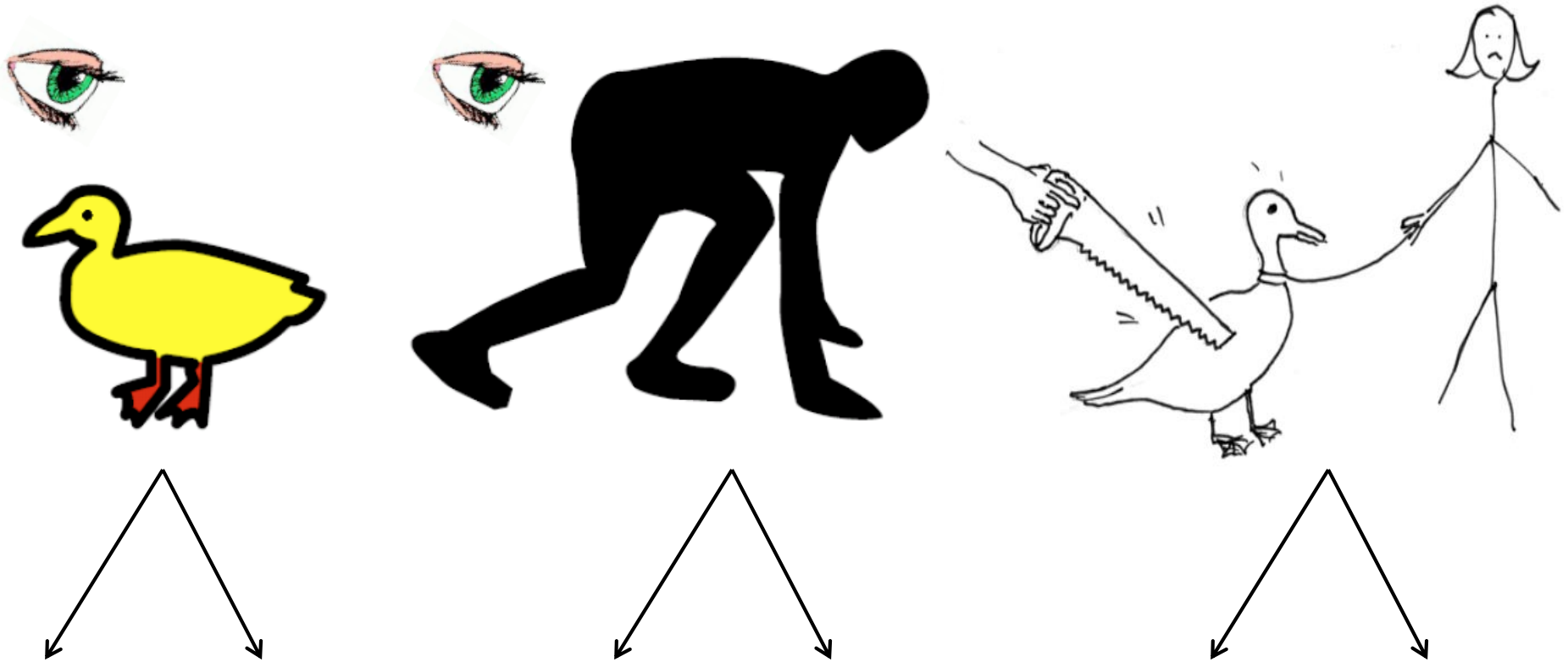


Image Credit: Liang Huang

We've come a long way…

What is Jeopardy?

- http://youtu.be/Xqb66bdsQlw?t=53s

Challenge:

- http://youtu.be/_429UIzN1JM

- Watson Demo:
  - http://youtu.be/WFR3lOm_xhE?t=22s
- Explanation
  - http://youtu.be/d_yXV22O6n4?t=4s

- Future: Automated operator, doctor assistant, finance

AlphaGo vs Lee Sedol

Slide Credit: Aykut Erdem

NVIDIA BB8 AI Car

End to End Learning for Self-Driving Cars

Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao
NVIDIA Corporation
Holmdel, NJ 07735

Karol Zieba
NVIDIA Corporation
Holmdel, NJ 07735

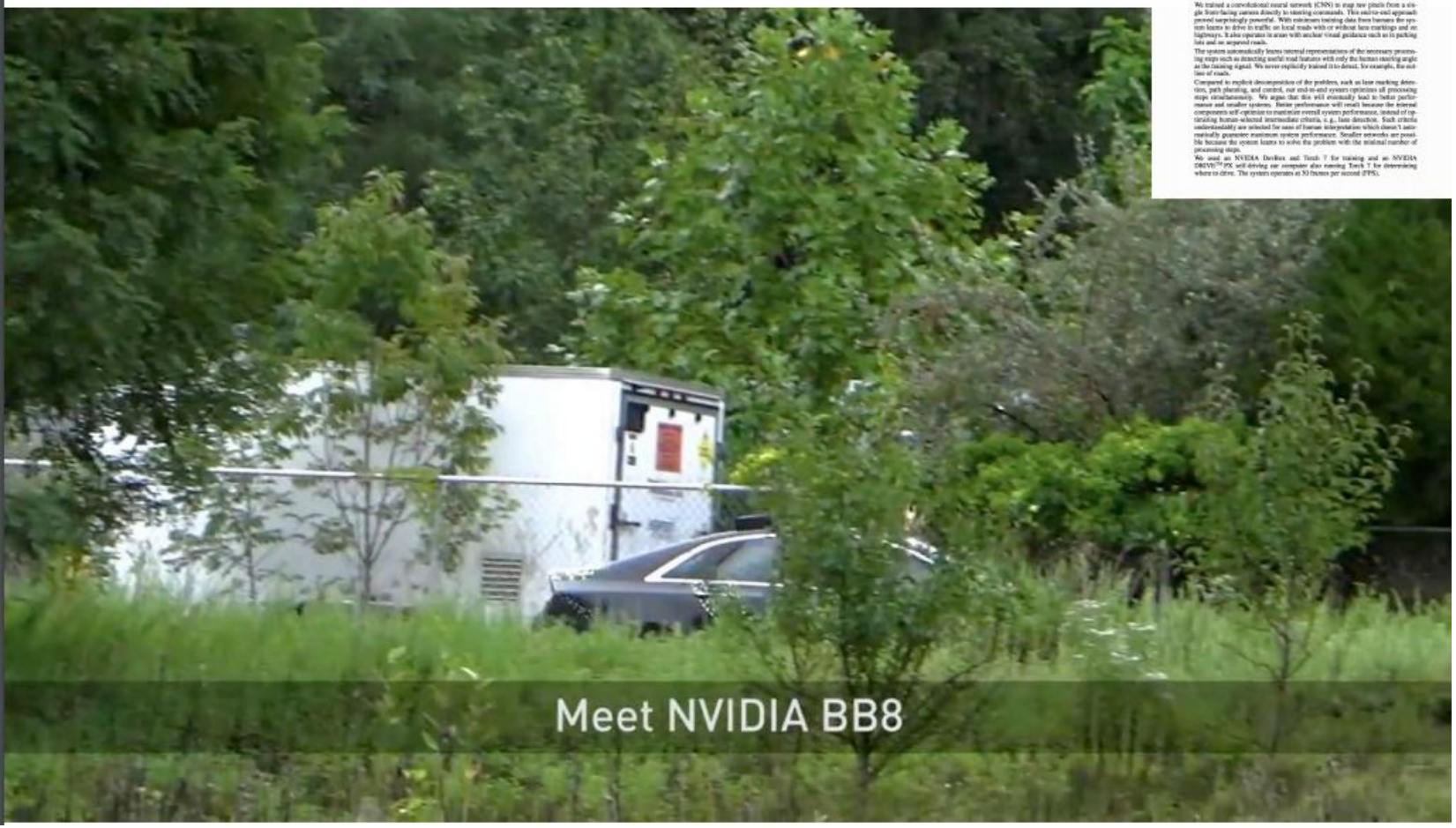
Abstract

We trained a convolutional neural network (CNN) to map raw pixels from a single front-facing camera directly to steering commands. This end-to-end approach proved surprisingly powerful. With minimum training data from humans the system learns to drive in traffic on local roads with or without lane markings and on highways. It also operates in areas with unclear visual guidance such as in parking lots and on unpaved roads.

The system automatically learns internal representations of the necessary processing steps such as detecting useful road features with only the human steering angle as the training signal. We never explicitly trained it to detect, for example, the outlines of roads.

Compared to explicit decomposition of the problem, such as lane marking detection, path planning, and control, our end-to-end system optimizes all processing steps simultaneously. We argue that this will eventually lead to better performance and smaller systems. Better performance will result because the internal components self-optimize to maximize overall system performance, instead of optimizing human-selected intermediate criteria, e.g., lane detection. Such criteria understandably are selected for ease of human interpretation which doesn't automatically guarantee maximum system performance. Smaller networks are possible because the system learns to solve the problem with the minimal number of processing steps.

We used an NVIDIA DevBox and Torch 7 for training and an NVIDIA DRIVE™ PX self-driving car computer also running Torch 7 for determining where to drive. The system operates at 30 frames per second (FPS).



https://www.youtube.com/watch?v=mCmO_5ZxdvE
https://www.youtube.com/watch?v=LVBBKppAaV4

Slide Credit: Aykut Erdem

ML in a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year

- Decades of ML research oversimplified:
  - All of Machine Learning:
  - Learn a mapping from input to output f: X → Y
  - X: emails, Y: {spam, notspam}

Slide Credit: Pedro Domingos

ML in a Nutshell

- Input: x (images, text, emails…)
- Output: y (spam or non-spam…)
- (Unknown) Target Function
  - f: X → Y (the “true” mapping / reality)
- Data
  - $(x_1,y_1), (x_2,y_2), …, (x_N,y_N)$
- Model / Hypothesis Class
  - g: X → Y
  - $y = g(x) = \text{sign}(w^T x)$

ML in a Nutshell

- Every machine learning algorithm has three components:
  - Representation / Model Class
  - Evaluation / Objective Function
  - Optimization

Slide Credit: Pedro Domingos

Representation / Model Class

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Slide Credit: Pedro Domingos

Evaluation / Objective Function

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Slide Credit: Pedro Domingos

Optimization

- Discrete/Combinatorial optimization
  - greedy search
  - Graph algorithms (cuts, flows, etc)
- Continuous optimization
  - Convex/Non-convex optimization
  - Linear programming

Types of Learning

- Supervised learning
  - Training data includes desired outputs
- Unsupervised learning
  - Training data does not include desired outputs
- Weakly or Semi-supervised learning
  - Training data includes a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

Spam vs Regular Email

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

January 21, 2015 2:53 PM
Hide Details

CUSP has approved ECE 4424 with the
copy of the proposal with these items ad
Thanks!!!
Tracy

vs

nadia bamba
To: undisclosed recipients: ;
Reply-To: nadia bamba
From Miss Nadia BamBa,

January 19, 2015 5:57 AM
Hide Details

From Miss Nadia BamBa,

Greeting, Permit me to inform you of my desire of going into business relationship with you. I am Nadia BamBa the only Daughter of late Mr and Mrs James BamBa, My father was a director of cocoa merchant in Abidjan, the economic capital of Ivory Coast before he was poisoned to death by his business associates on one of their outing to discus on a business deal. When my mother died on the 21st October 2002, my father took me very special because i am motherless.

Before the death of my father in a private hospital here in Abidjan, He secretly called me on his bedside and told me that he had a sum of $6, 8000.000(SIX Million EIGHT HUNDRED THOUSAND), Dollars) left in a suspense account in a Bank here in Abidjan, that he used my name as his first Daughter for the next of kin in deposit of the fund.

He also explained to me that it was because of this wealth and some huge amount of money That his business associates supposed to balance him from the deal they had that he was poisoned by his business associates, that I should seek for a God fearing foreign partner in a country of my choice where I will transfer this money and use it for investment purposes, (such as real estate Or Hotel management).please i am honourably seeking your assistance in the following ways.

1) To provide a Bank account where this money would be transferred to.
2) To serve as the guardian of this Money since I am a girl of 19 years old.
3)Your private phone number's and your family background' s that we can know each order more.

Moreover i am willing to offer you 15% of the total sum as compensation for effort input after the successful transfer of this fund to your designated account overseas,

Anticipating to hear from you soon.
Thanks and God Bless.
Best regards.

Intuition

- Spam Emails
  - a lot of words like
    - “money”
    - “free”
    - “bank account”
    - “viagara” ... in a single email
- Regular Emails
  - word usage pattern is more spread out

Simple Strategy: Let us count!

This is X

I know you will be very much interested, kindly provide me with the details below.

First Name.....
Surname.....
Address.....
City.....
State/Province.....
Country.....
Telephone No.....
Occupation …….
Date of Birth (date/m/yr) ……..
Copy of International Passport Or ID card…….

$$\begin{pmatrix} \text{free} & 100 \\ \text{money} & 2 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

From: Ross Girshick
Subject: **Re: hey**
Date: January 17, 2013 7:48:18 PM EST
To: Dhruv Batra

Hi Dhruv,

sorry for the high latency. I just got back from Singapore last night ar

$$\begin{pmatrix} \text{free} & 1 \\ \text{money} & 1 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

Final Procedure

$$\begin{pmatrix} \text{free} & 100 \\ \text{money} & 2 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

Why these words?

$$\begin{pmatrix} 100 \times 0.2 \\ 2 \times 0.3 \\ \vdots \\ 2 \times 0.3 \\ \vdots \end{pmatrix}$$

$$\begin{pmatrix} 100 \times 0.01 \\ 2 \times 0.02 \\ \vdots \\ 2 \times 0.01 \\ \vdots \end{pmatrix}$$

Where do the weights come from?

= 3.2

= 1.03

Why linear combination?

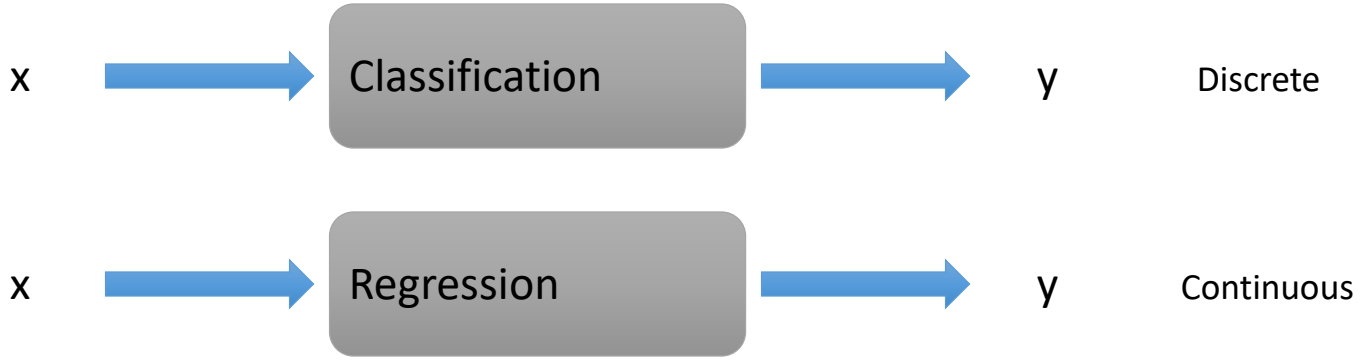
Confidence / performance guarantee?

Types of Learning

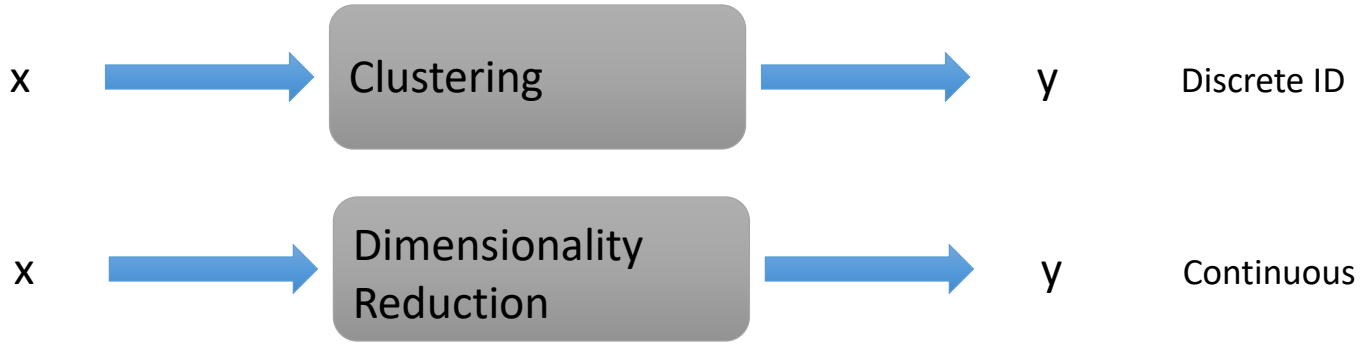
- Supervised learning
  - Training data includes desired outputs
- Unsupervised learning
  - Training data does not include desired outputs
- Weakly or Semi-supervised learning
  - Training data includes a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

Tasks

Supervised Learning



Unsupervised Learning

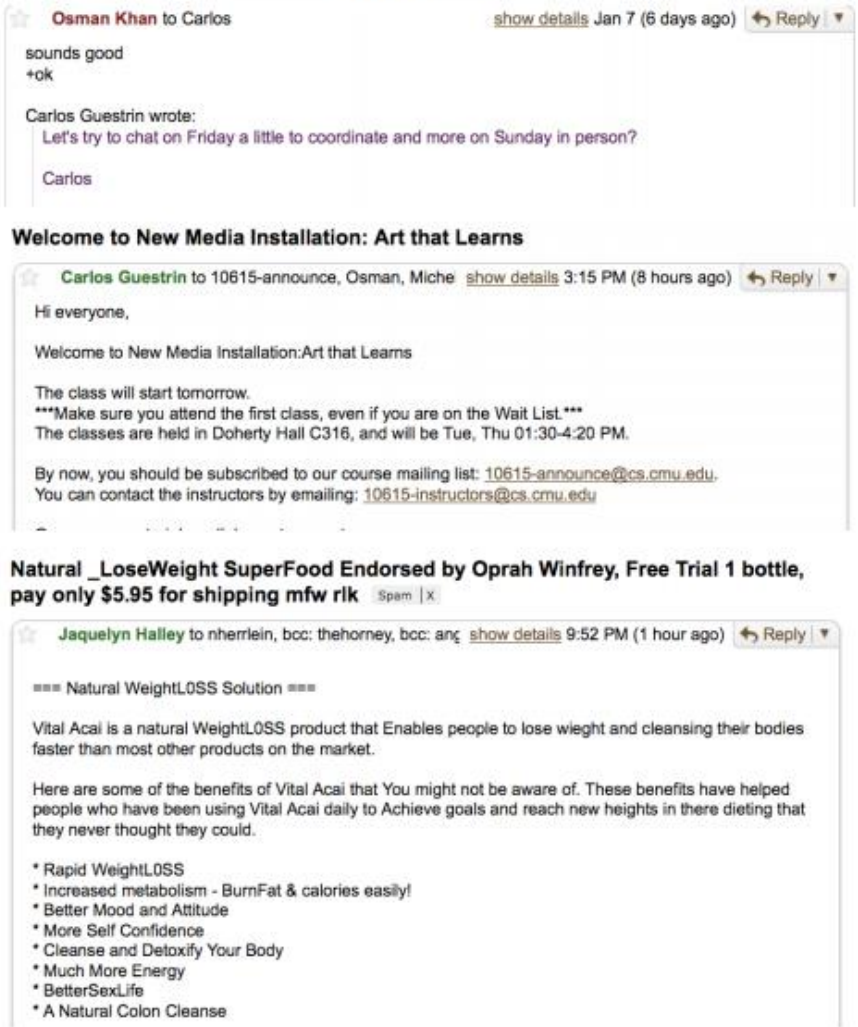


Classification

- From data to discrete classes

Spam filtering

data



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

sounds good
+ok

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

Carlos

Welcome to New Media Installation: Art that Learns

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

Hi everyone,

Welcome to New Media Installation:Art that Learns

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

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

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

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

=== Natural WeightL0SS Solution ===

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

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

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

prediction

Spam
vs.
Not Spam

Slide Credit: David Sontag

Face detection

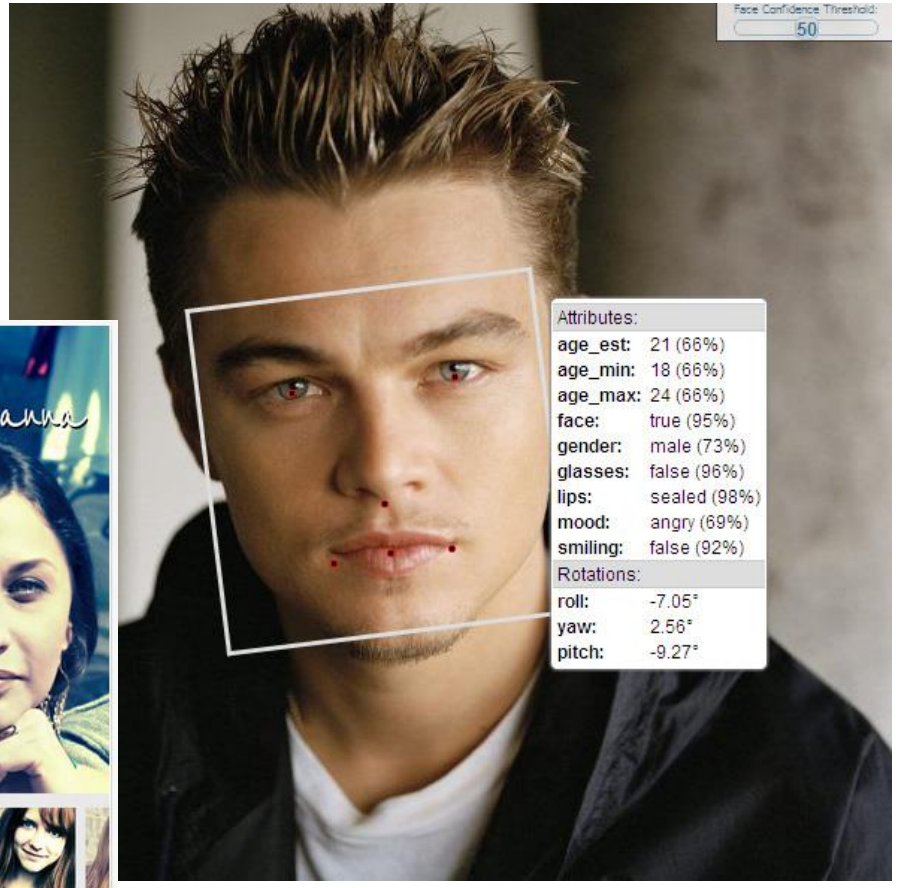
Example training images
for each orientation

Slide Credit: David Sontag

Face Recognition

face.com

http://developers.face.com/tools/



Slide Credit: Noah Snavely

Speech Recognition

Slide Credit: Carlos Guestrin

Weather prediction

Slide Credit: David Sontag

Image Classification

- Im2tags; Im2text
- http://deeplearning.cs.toronto.edu/

Pizza
Wine
Stove

Machine Translation

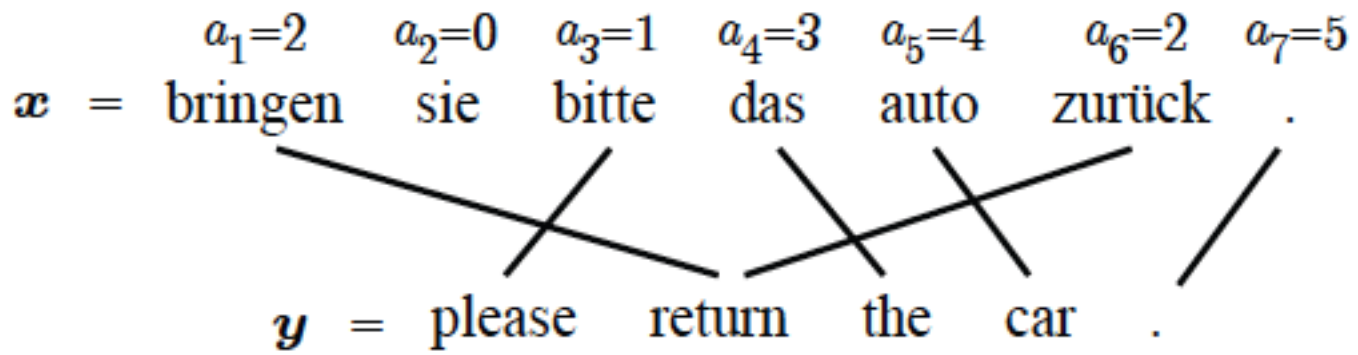


Figure Credit: Kevin Gimpel

Seeing is worse than believing

- [Barbu et al. ECCV14]

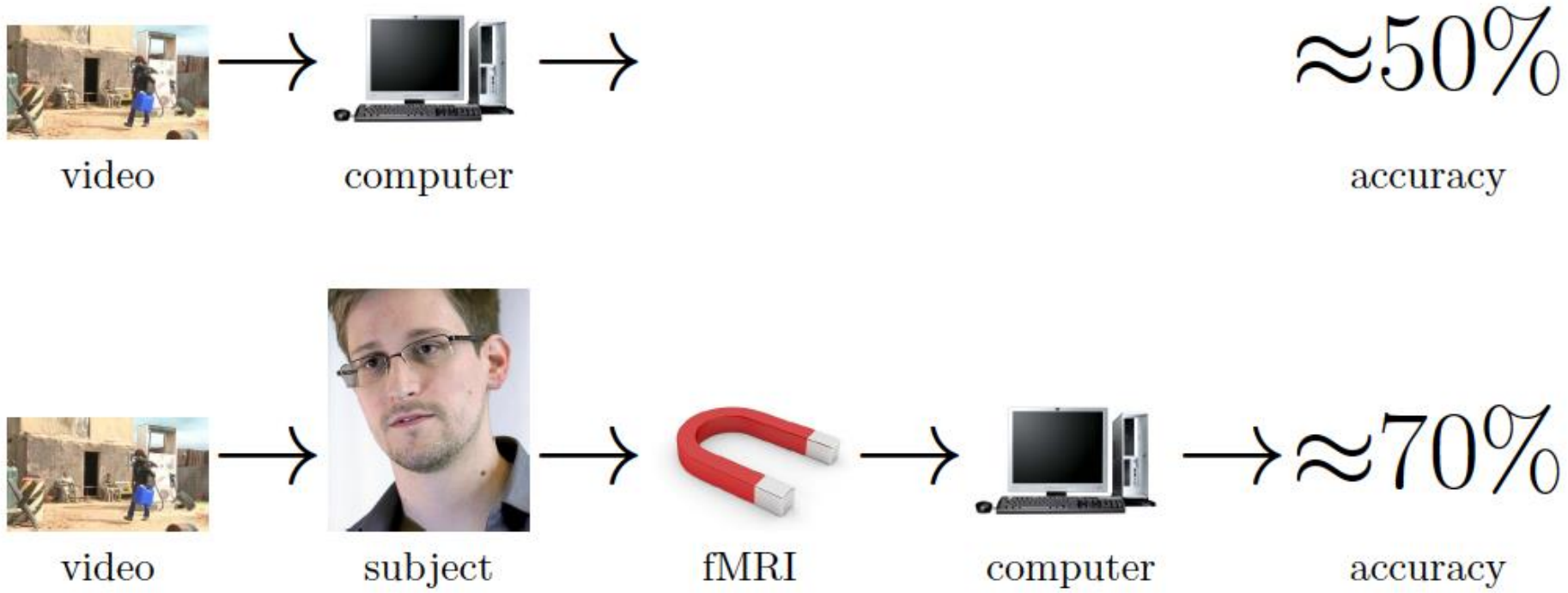


Image Credit: Barbu et al.

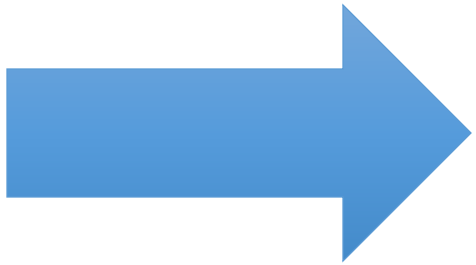
Regression

- predicting a numeric value

Stock market

Slide Credit: David Sontag

Weather prediction



Slide Credit: Carlos Guestrin

Pose Estimation

Slide Credit: Noah Snavely

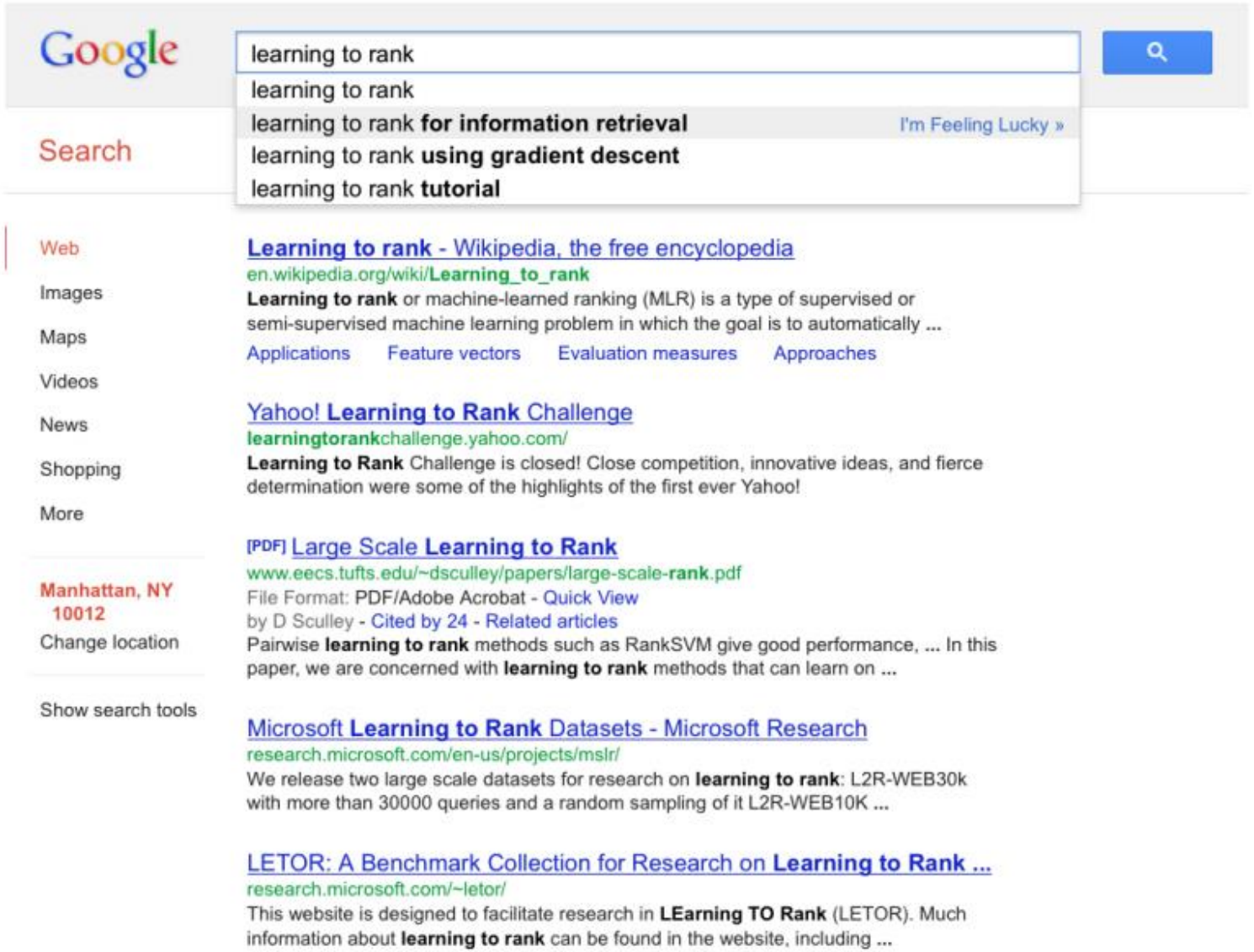
Pose Estimation

- 2010: (Project Natal) Kinect
  - http://www.youtube.com/watch?v=r5-zZDSsgFg
- 2012: Kinect One
  - http://youtu.be/Hi5kMNfgDS4?t=28s
- 2013: Leap Motion
  - http://youtu.be/gby6hGZb3ww

Ranking

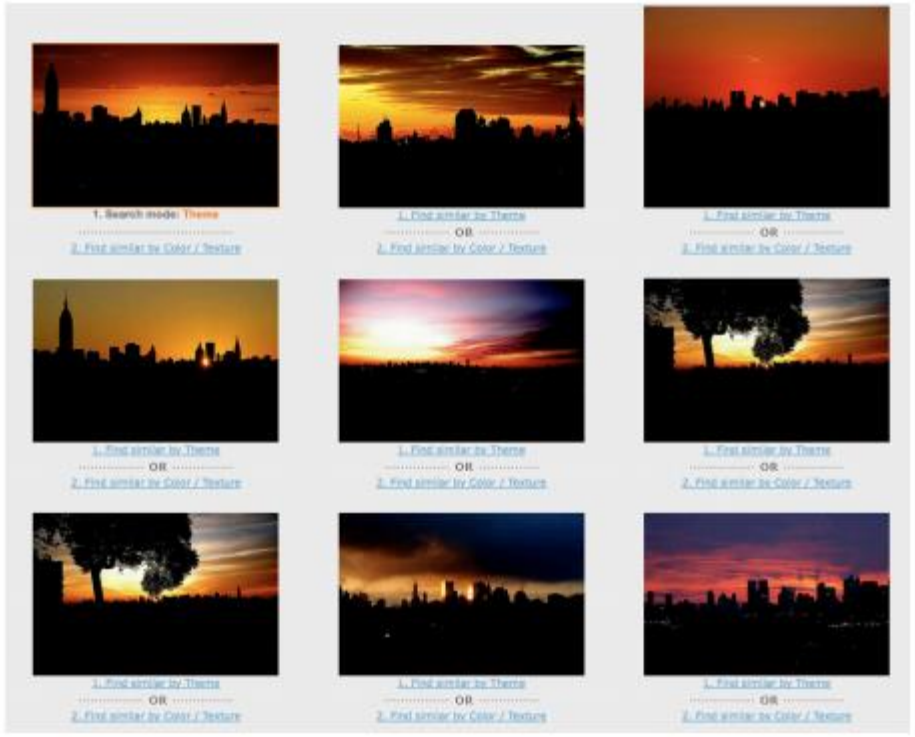
- Comparing items

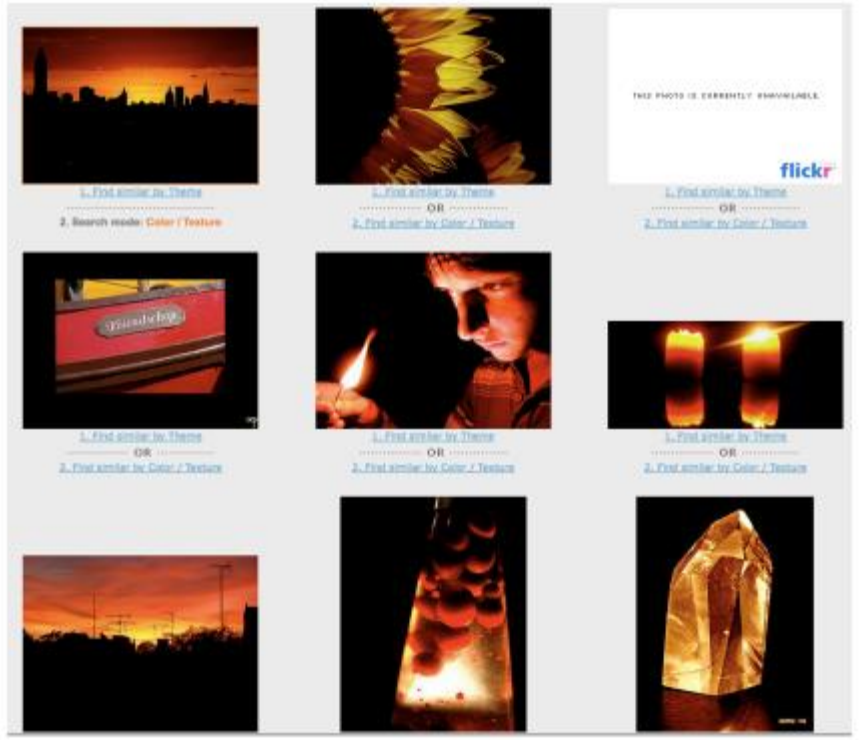
Web search



Slide Credit: David Sontag

Given image, find similar images





http://www.tiltomo.com/

Slide Credit: David Sontag

Recommendation systems

Slide Credit: David Sontag

Recommendation systems

Machine learning competition with a $1 million prize

Leaderboard

Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	The Ensemble	0.8553	10.10	2009-07-26 18:38:22
2	BellKor's Pragmatic Chaos	0.8554	10.09	2009-07-26 18:18:28
Grand Prize - RMSE <= 0.8563				
3	Grand Prize Team	0.8571	9.91	2009-07-24 13:07:49
4	Opera Solutions and Vandelay United	0.8573	9.89	2009-07-25 20:05:52
5	Vandelay Industries !	0.8579	9.83	2009-07-26 02:49:53
6	PragmaticTheory	0.8582	9.80	2009-07-12 15:09:53
7	BellKor in BigChaos	0.8590	9.71	2009-07-26 12:57:25
8	Dace	0.8603	9.58	2009-07-24 17:18:43
9	Opera Solutions	0.8611	9.49	2009-07-26 18:02:08
10	BellKor	0.8612	9.48	2009-07-26 17:19:11
11	BigChaos	0.8613	9.47	2009-06-23 23:06:52
12	Feeds2	0.8613	9.47	2009-07-24 20:06:46
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
13	xiangliang	0.8633	9.26	2009-07-21 02:04:40
14	Gravity	0.8634	9.25	2009-07-26 15:58:34
15	Ces	0.8642	9.17	2009-07-25 17:42:38
16	Invisible Ideas	0.8644	9.14	2009-07-20 03:26:12
17	Just a guy in a garage	0.8650	9.08	2009-07-22 14:10:42
18	Craig Carmichael	0.8656	9.02	2009-07-25 16:00:54
19	J Dennis Su	0.8658	9.00	2009-03-11 09:41:54
20	acmehill	0.8659	8.99	2009-04-16 06:29:35
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell				
Cinematch score on quiz subset - RMSE = 0.9514				

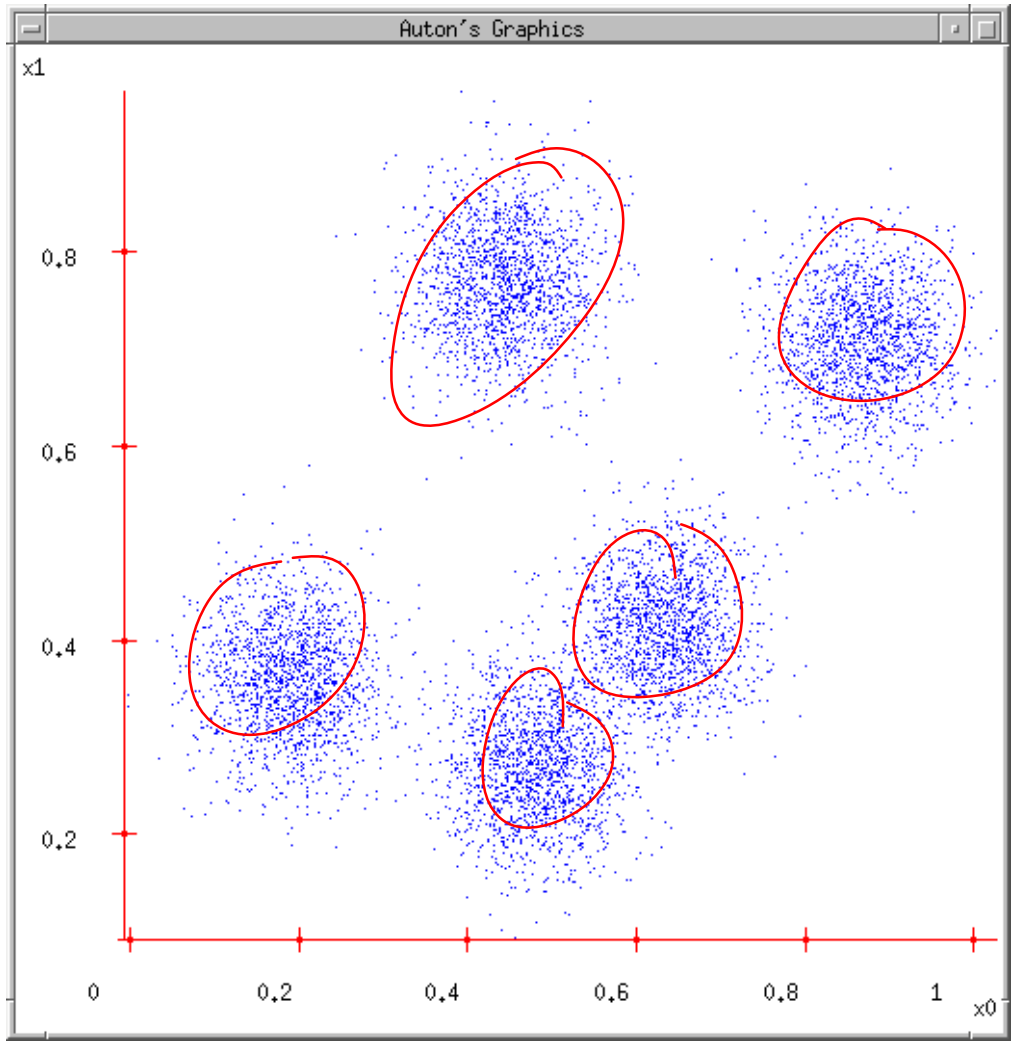
Slide Credit: David Sontag

Unsupervised Learning
Discovering structure in data

Clustering

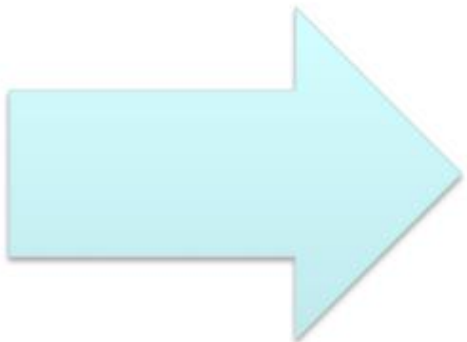
Unsupervised Learning
Y not provided

Clustering Data: Group similar things



Slide Credit: Carlos Guestrin

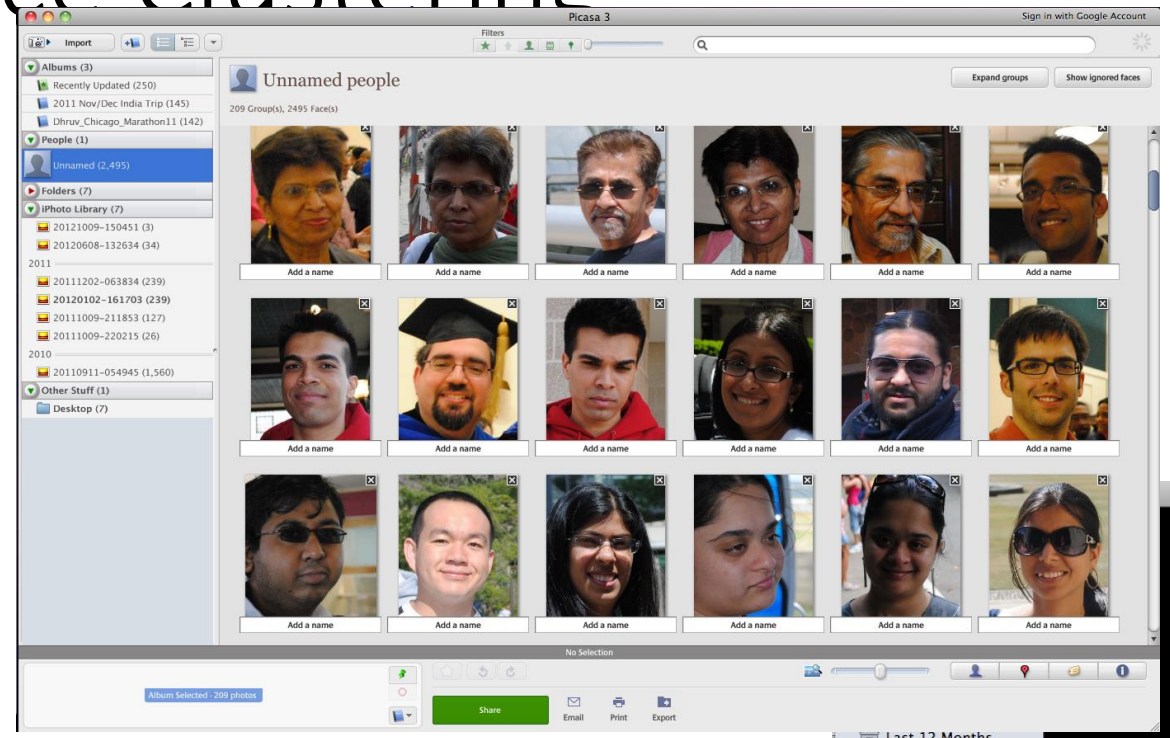
Clustering images



[Goldberger et al.]

Slide Credit: David Sontag

Face Clustering

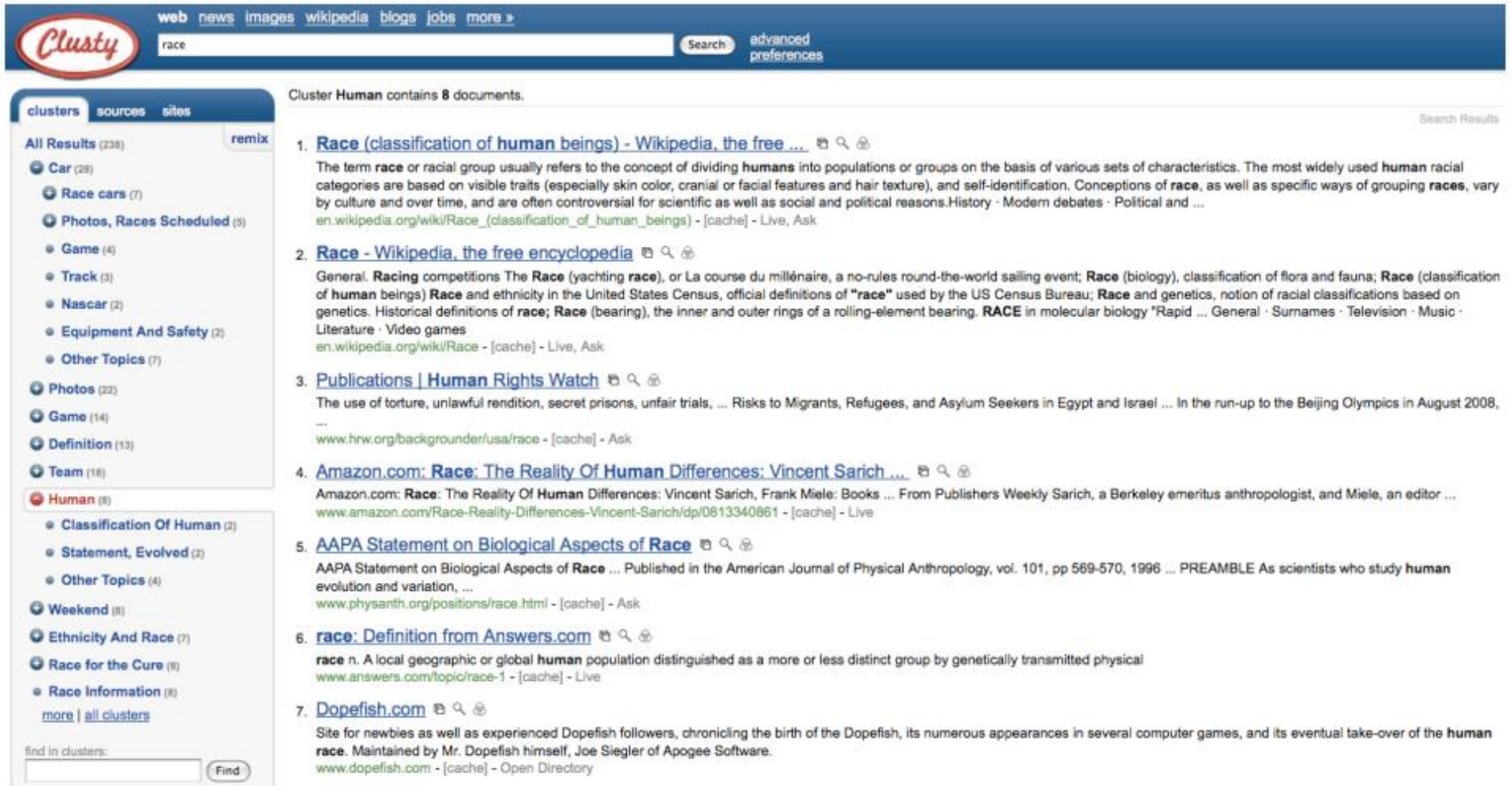


Picassa

iPhoto

Slide Credit: David Sontag

Clustering web search results



Slide Credit: David Sontag

Unsupervised Learning

Dimensionality Reduction / Embedding

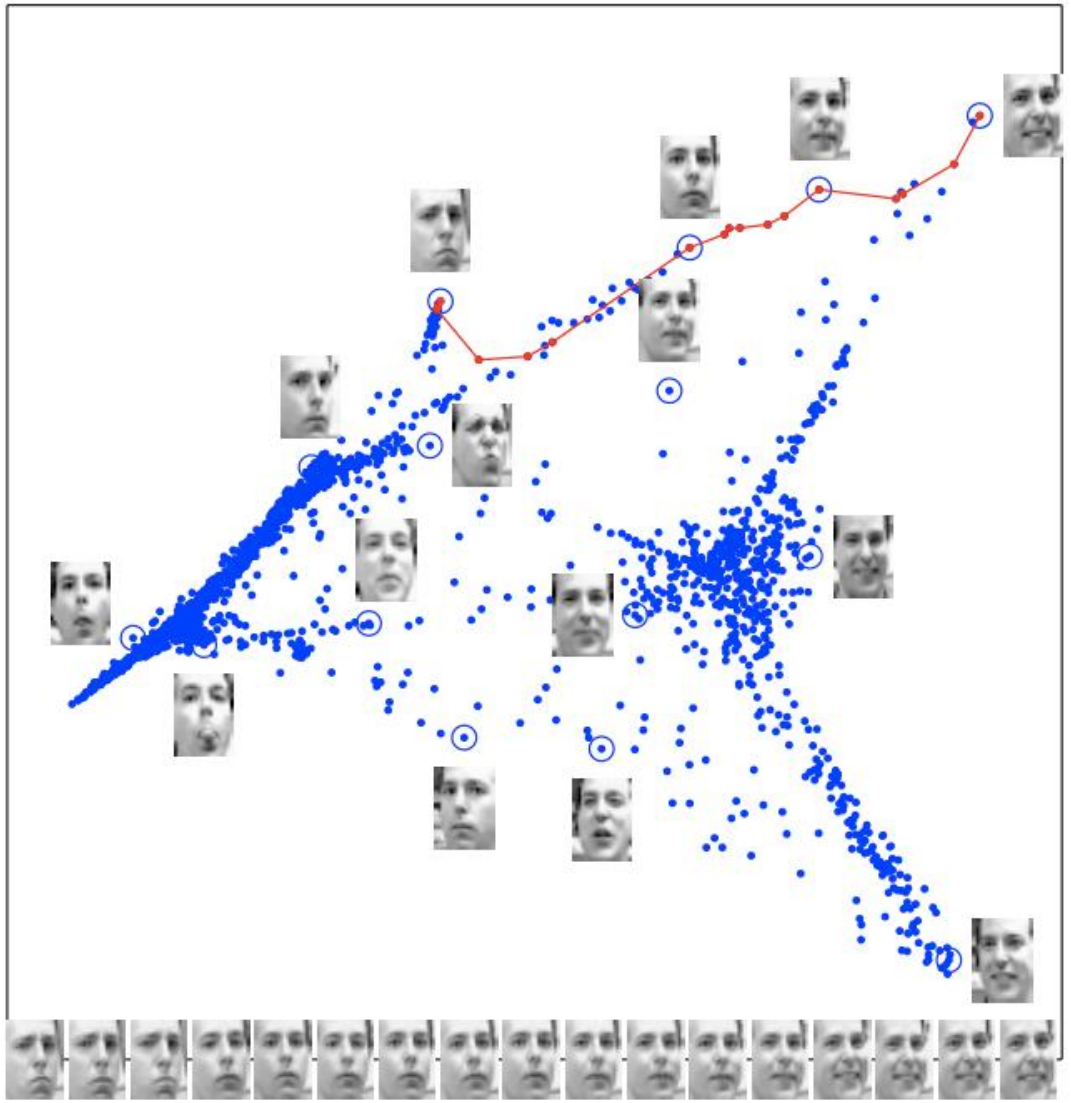
Unsupervised Learning

Y not provided

Embedding images

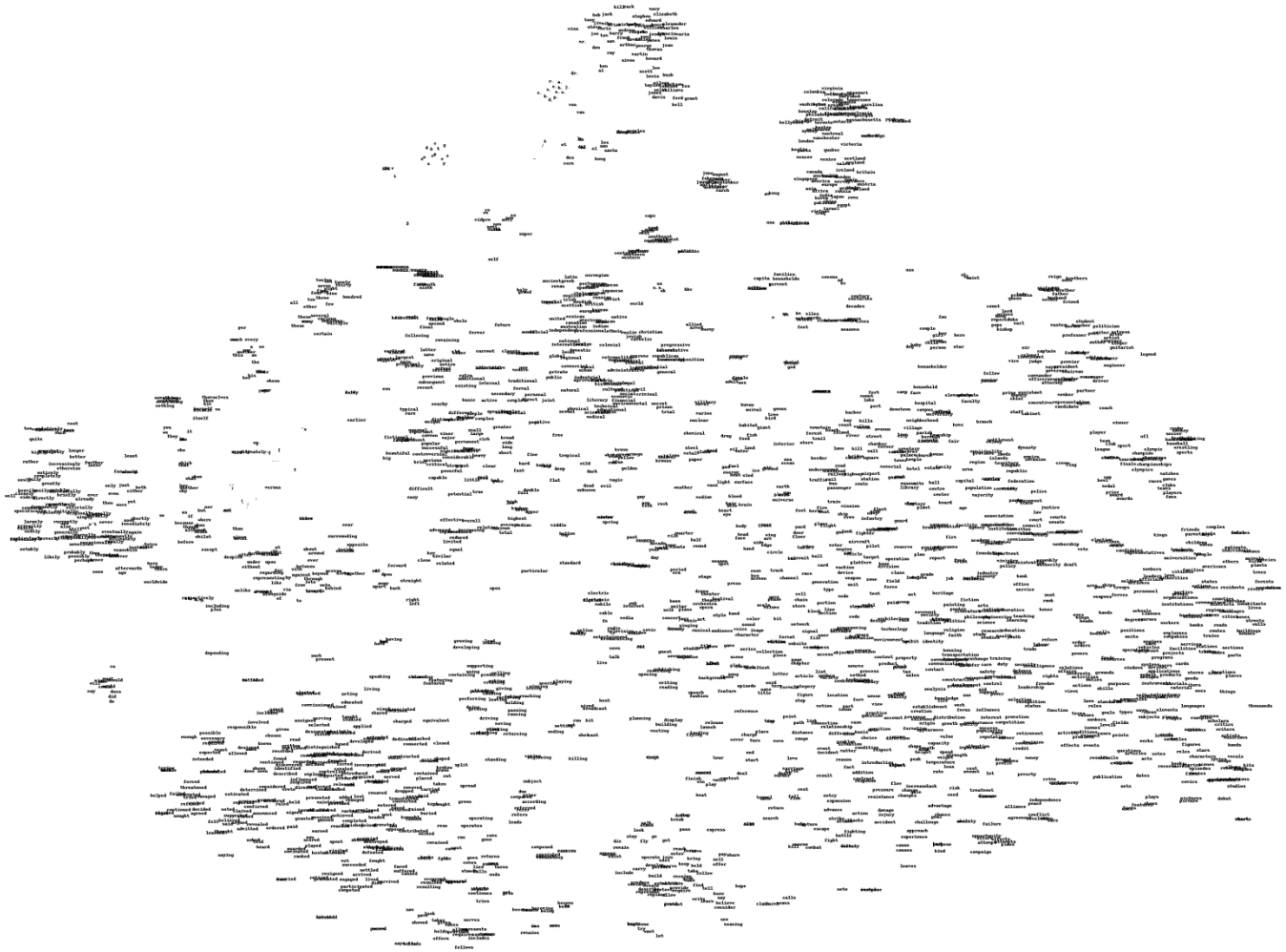
Images have thousands or millions of pixels.

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



Slide Credit: Carlos Guestrin [Saul & Roweis '03]

Embedding words



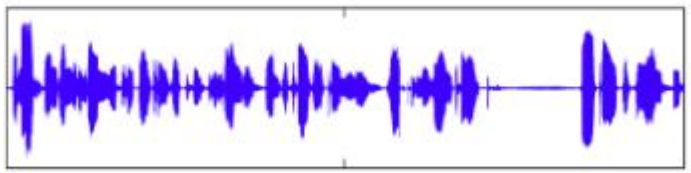
Slide Credit: Carlos Guestrin

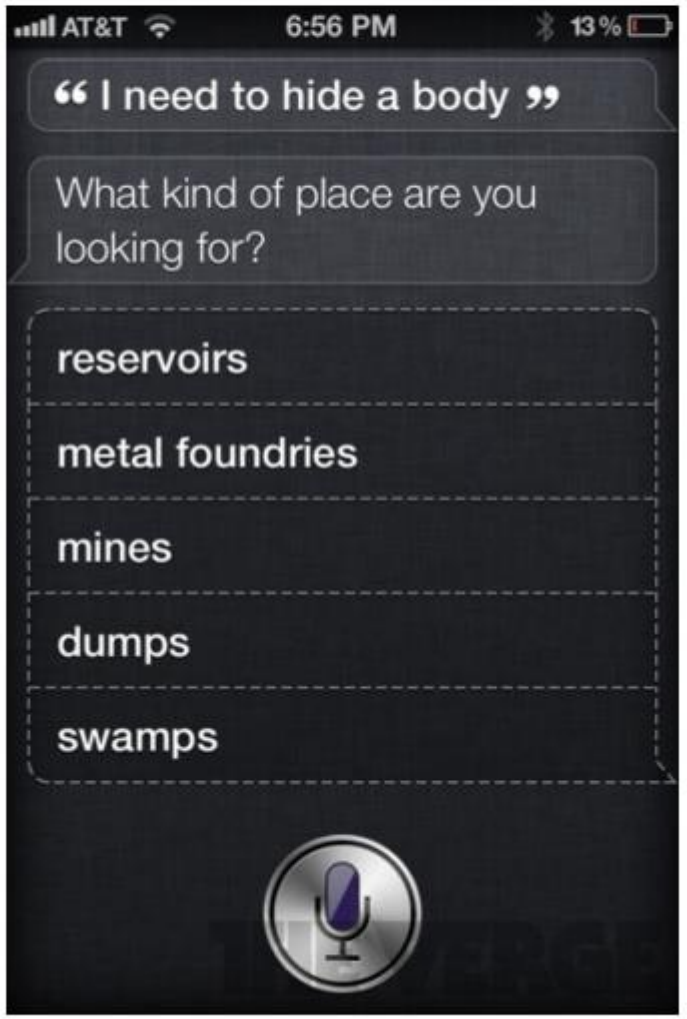
[Joseph Turian]

Structured prediction

- From data to discrete classes

Speech recognition





Slide Credit: David Sontag

Reinforcement Learning

Learning from feedback

Reinforcement Learning: Learning to act

- There is only one "supervised" signal at the end of the game.
- But you need to make a move at every step
- RL deals with "credit assignment"

Learning to act

- Reinforcement learning
- An agent
  - Makes sensor observations
  - Must select action
  - Receives rewards
    - positive for “good” states
    - negative for “bad” states
- Towel Folding
  - http://youtu.be/gy5g33S0Gzo

Growth of Machine Learning

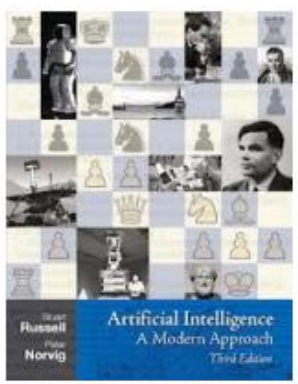
- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
  - Sensor networks
  - …
- This trend is accelerating
  - Big data
  - Improved machine learning algorithms
 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

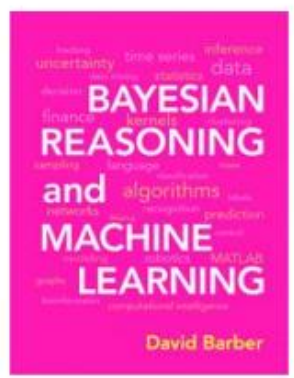
Syllabus

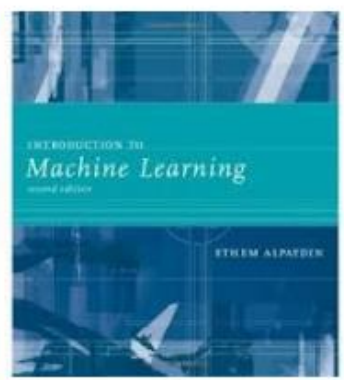
- Basics of Statistical Learning
    - Loss functions, MLE, MAP, Bayesian estimation, bias-variance tradeoff, overfitting, regularization, cross-validation
- Supervised Learning
    - Nearest Neighbour, Naïve Bayes, Logistic Regression, Support Vector Machines, Kernels, Neural Networks, Decision Trees
    - Ensemble Methods: Bagging, Boosting
- Unsupervised Learning
    - Clustering: k-means, Gaussian mixture models, EM
    - Dimensionality reduction: PCA, SVD, LDA
- Advanced Topics
    - Weakly-supervised and semi-supervised learning
    - Reinforcement learning
    - Probabilistic Graphical Models: Bayes Nets, HMM
    - Applications to Vision, Natural Language Processing

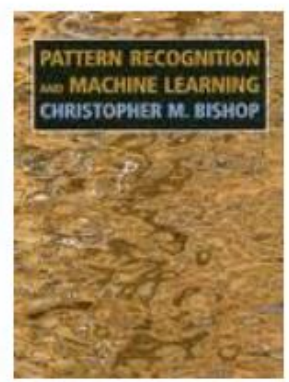
Reference Books

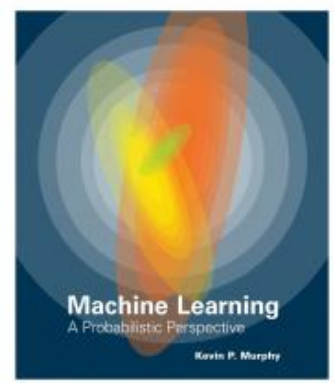
- A Course in Machine Learning, Hal Daumé III (**online version (v.0.99) available**), 2017
- Artificial Intelligence: A Modern Approach (3rd Edition), Russell and Norvig. Prentice Hall, 2009
- Bayesian Reasoning and Machine Learning, Barber, Cambridge University Press, 2012 (**online version available**)
- Introduction to Machine Learning (2nd Edition), Alpaydin, MIT Press, 2010
- Pattern Recognition and Machine Learning, Bishop, Springer, 2006
- Machine Learning: A Probabilistic Perspective, Murphy, MIT Press, 2012











Slide by Aykut Erdem