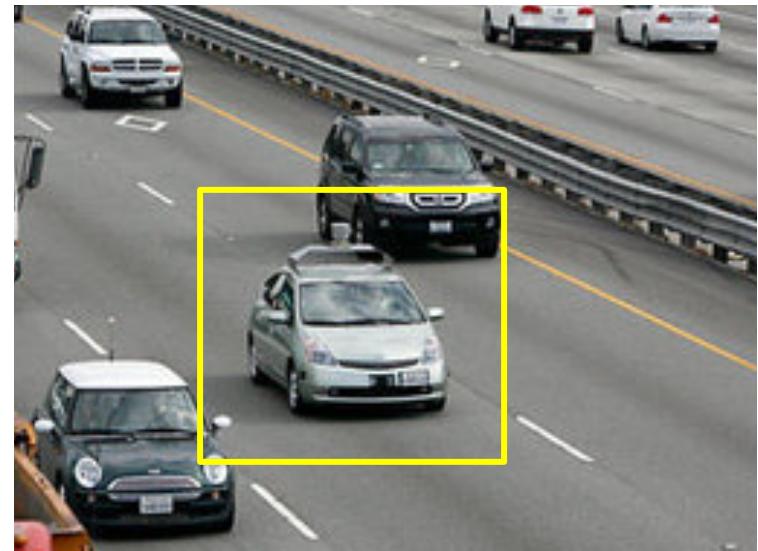
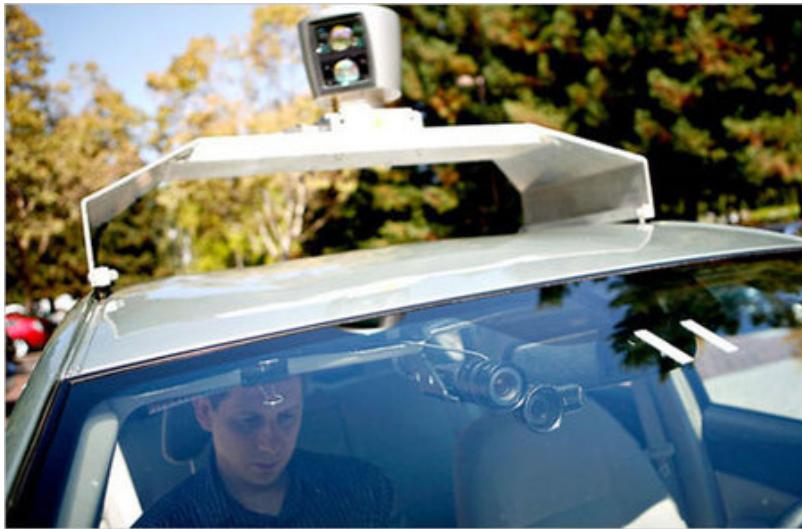


What is Machine Learning ? (by examples)

Autonomous Cars

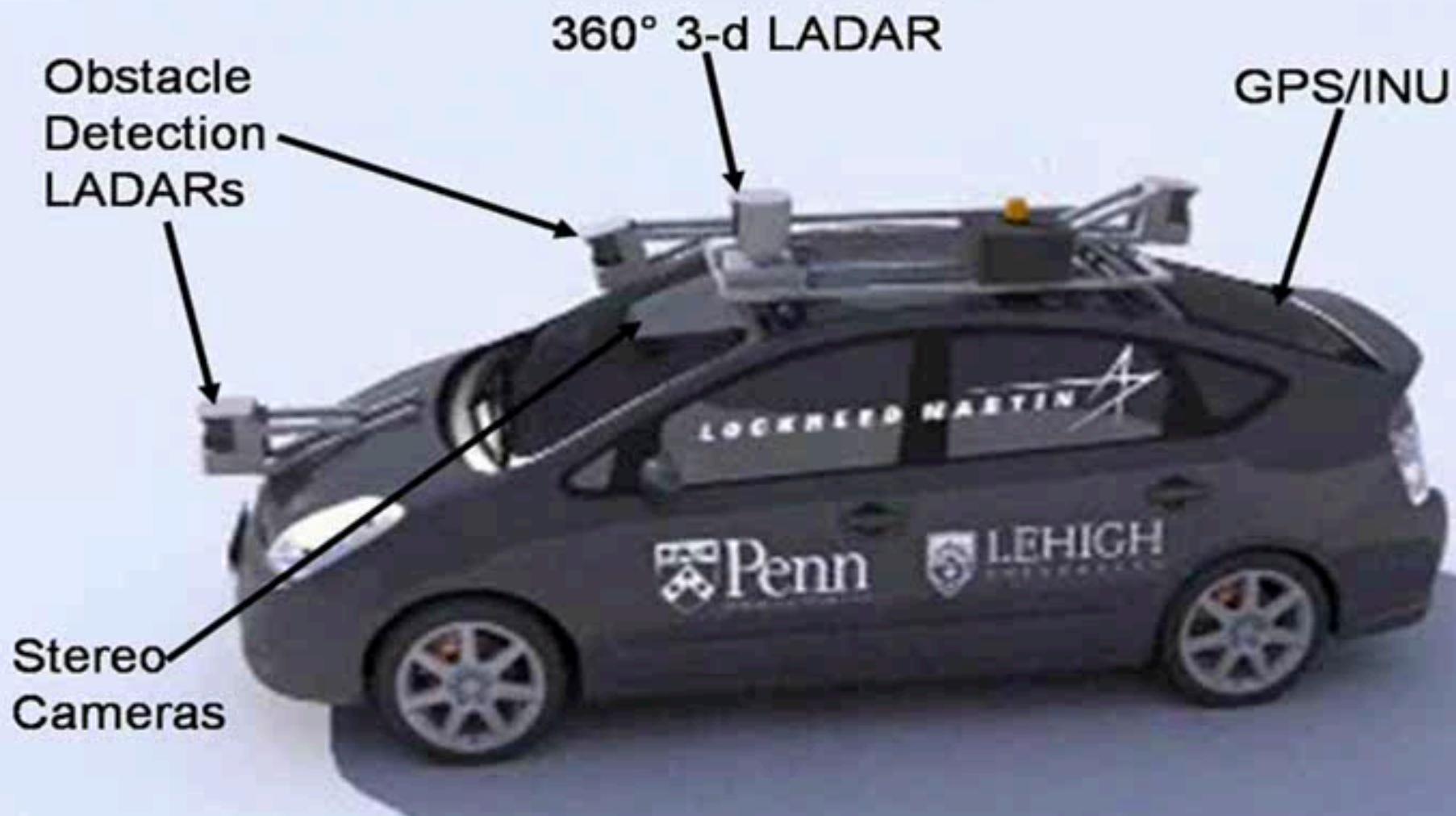


- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

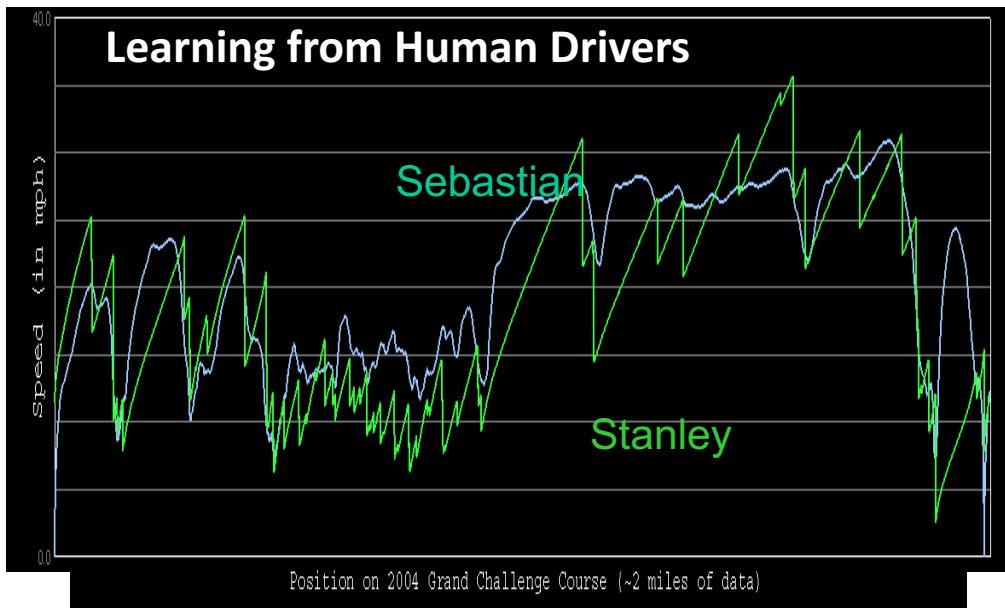
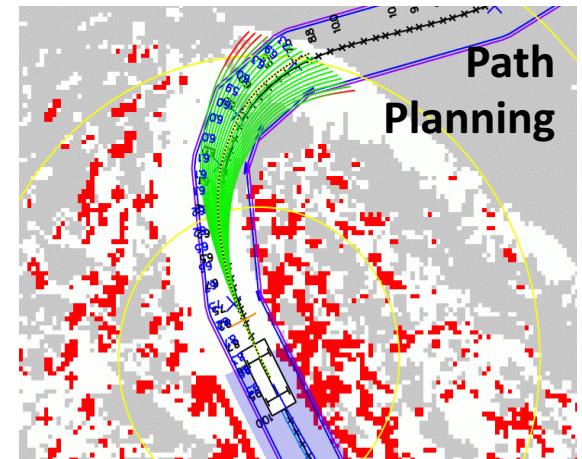
Penn's Autonomous Car →
(Ben Franklin Racing Team)



Autonomous Car Sensors



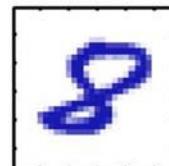
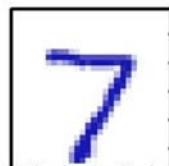
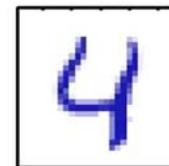
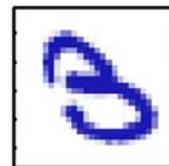
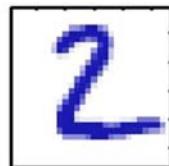
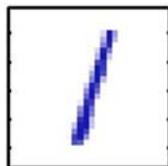
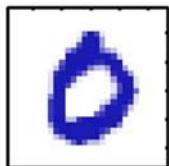
Autonomous Car Technology



Classification

from data to discrete classes

Example 1: hand-written digit recognition



Images are 28 x 28 pixels

Represent input image as a vector $\mathbf{x} \in \mathbb{R}^{784}$

Learn a classifier $f(\mathbf{x})$ such that,

$$f : \mathbf{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

How to proceed

As a supervised classification problem

Start with training data, e.g. 6000 examples of each digit

0 0 0 1 1 1 1 1 2

2 2 2 2 2 2 3 3 3

3 4 4 4 4 5 5 5 5

6 6 7 7 7 7 8 8 8

9 9 9 9 9 9 9 9 9

- Can achieve testing error of 0.4%
- One of first commercial and widely used ML systems (for zip codes & checks)

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, Michel show details 3:15 PM (8 hours ago) [Reply](#)

Hi everyone,

Welcome to New Media Installation:Art that Learns

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

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

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle,
pay only \$5.95 for shipping mfw rlk [Spam](#)

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

==== Natural WeightLOSS Solution ===

Vital Acai is a natural WeightLOSS 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 WeightLOSS
- * 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

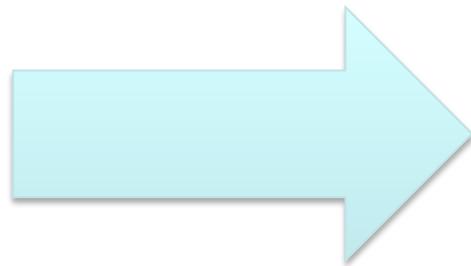
Face recognition



Example training images
for each orientation



Weather prediction



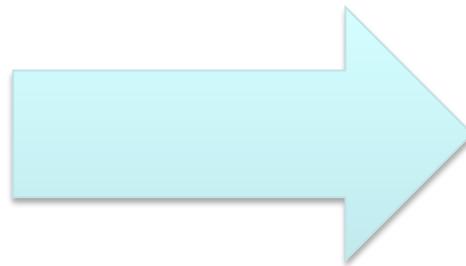
Regression

predicting a numeric value

Stock market



Weather prediction revisited



Temperature

72° F

Ranking

comparing items

Web search

The screenshot shows a Google search interface. The search bar at the top contains the query "learning to rank". Below the search bar, a dropdown menu lists several suggestions: "learning to rank", "learning to rank for information retrieval", "learning to rank using gradient descent", and "learning to rank tutorial". To the right of these suggestions is a blue button labeled "I'm Feeling Lucky »".

On the left side of the page, there is a sidebar with navigation links: "Search", "Web", "Images", "Maps", "Videos", "News", "Shopping", and "More". Below these links, there is a location entry: "Manhattan, NY 10012" followed by a "Change location" link. At the bottom of the sidebar is a "Show search tools" link.

The main content area displays search results for "learning to rank". The first result is a Wikipedia page titled "Learning to rank - Wikipedia, the free encyclopedia" with the URL en.wikipedia.org/wiki/Learning_to_rank. The snippet describes learning to rank as a machine-learned ranking (MLR) problem. Below this are links to other resources: "Applications", "Feature vectors", "Evaluation measures", and "Approaches".

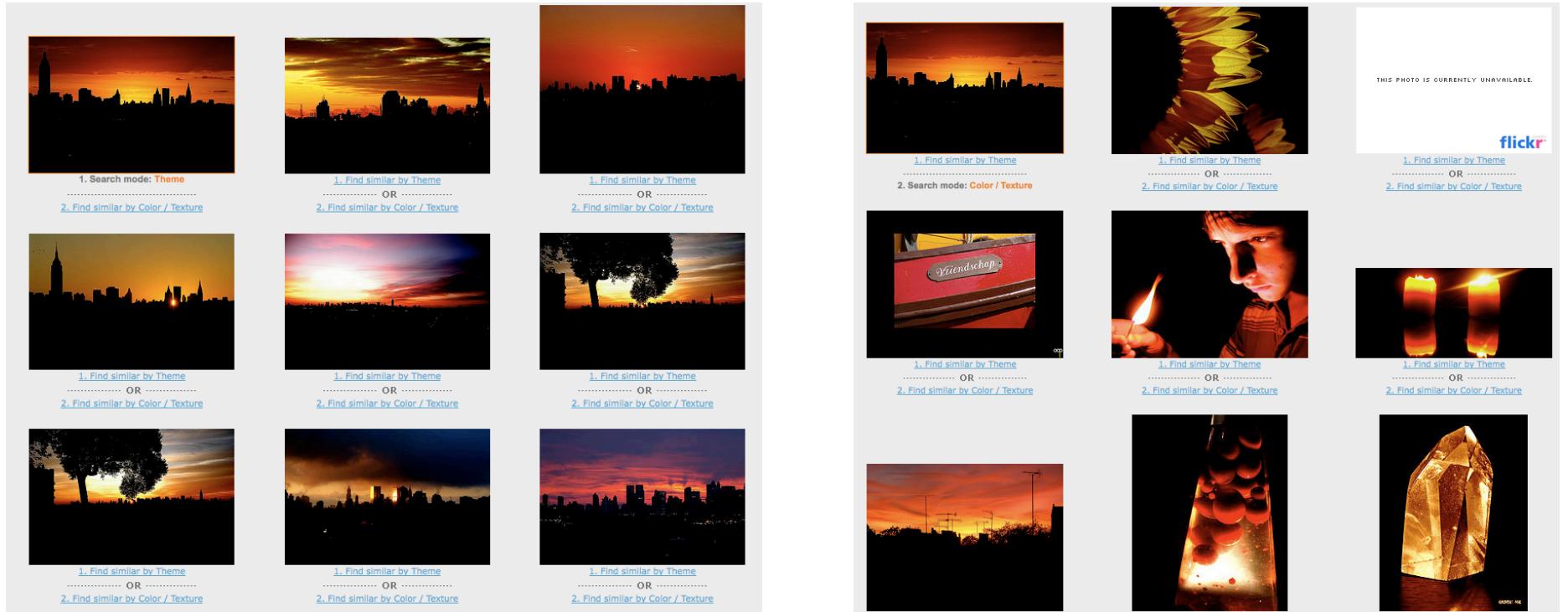
The second result is a link to the "Yahoo! Learning to Rank Challenge" with the URL learningtorankchallenge.yahoo.com/. The snippet mentions the challenge was closed and highlights competition, innovative ideas, and fierce determination.

The third result is a PDF titled "[PDF] Large Scale Learning to Rank" with the URL www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf. The snippet discusses pairwise learning to rank methods like RankSVM and their performance.

The fourth result is a Microsoft Research page titled "Microsoft Learning to Rank Datasets - Microsoft Research" with the URL research.microsoft.com/en-us/projects/mslr/. The snippet states they release two large scale datasets for research on learning to rank: L2R-WEB30k and L2R-WEB10K.

The fifth result is a Microsoft Research page titled "LETOR: A Benchmark Collection for Research on Learning to Rank ..." with the URL research.microsoft.com/~letor/. The snippet describes LETOR as a benchmark collection for research on learning to rank, noting it facilitates research in Learning TO Rank (LETOR).

Given image, find similar images



Collaborative Filtering

Recommendation systems

amazon Try Prime

David's Amazon.com | Today's Deals | Gift Cards | Sell | Help

Shop by Department ▾ Search Books ▾ Go

Hello, David Your Account ▾ Try Prime ▾ Cart ▾ Wish List ▾

Your Amazon.com Your Browsing History Recommended For You Amazon Betterizer Improve Your Recommendations Your Profile Learn More

Your Amazon.com > Recommended for You > Books > Subjects > Science & Math > History & Philosophy

Just For Today

Browse Recommended

Recommendations

History & Philosophy

History of Science

Philosophy of Biology

Philosophy of Medicine

These recommendations are based on items you own and more.

view: All | New Releases | Coming Soon

1.  **Causality: Models, Reasoning and Inference**
by Judea Pearl (September 14, 2009)
Average Customer Review: ★★★★★ (10)
In Stock
List Price: \$50.00
Price: \$32.49
61 used & new from \$28.00

Add to Cart Add to Wish List

I own it Not interested Rate this item
Recommended because you purchased Probabilistic Graphical Models and more (Fix this)

2.  **The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century**
by David Salsburg (May 1, 2002)
Average Customer Review: ★★★★★ (26)
In Stock
List Price: \$18.99
Price: \$13.88
81 used & new from \$9.00

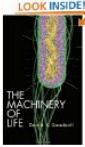
Add to Cart Add to Wish List

I own it Not interested Rate this item
Recommended because you added The Theory That Would Not Die to your Wish List (Fix this)

3.  **The Eighth Day of Creation: Makers of the Revolution in Biology, 25th Anniversary Edition**
by Horace Freeland Judson (November 1, 1996)
Average Customer Review: ★★★★★ (10)
In stock on September 4, 2013
List Price: \$56.00
Price: \$36.09
59 used & new from \$26.95

Add to Cart Add to Wish List

I own it Not interested Rate this item
Recommended because you purchased Molecular Biology of the Cell (Fix this)

4.  **The Machinery of Life**
by David S. Goodsell (April 28, 2009)
Average Customer Review: ★★★★★ (41)
In Stock
List Price: \$25.00
Price: \$17.49
92 used & new from \$12.00

Back-to-School Savings Daily Lightning Deals Shop now

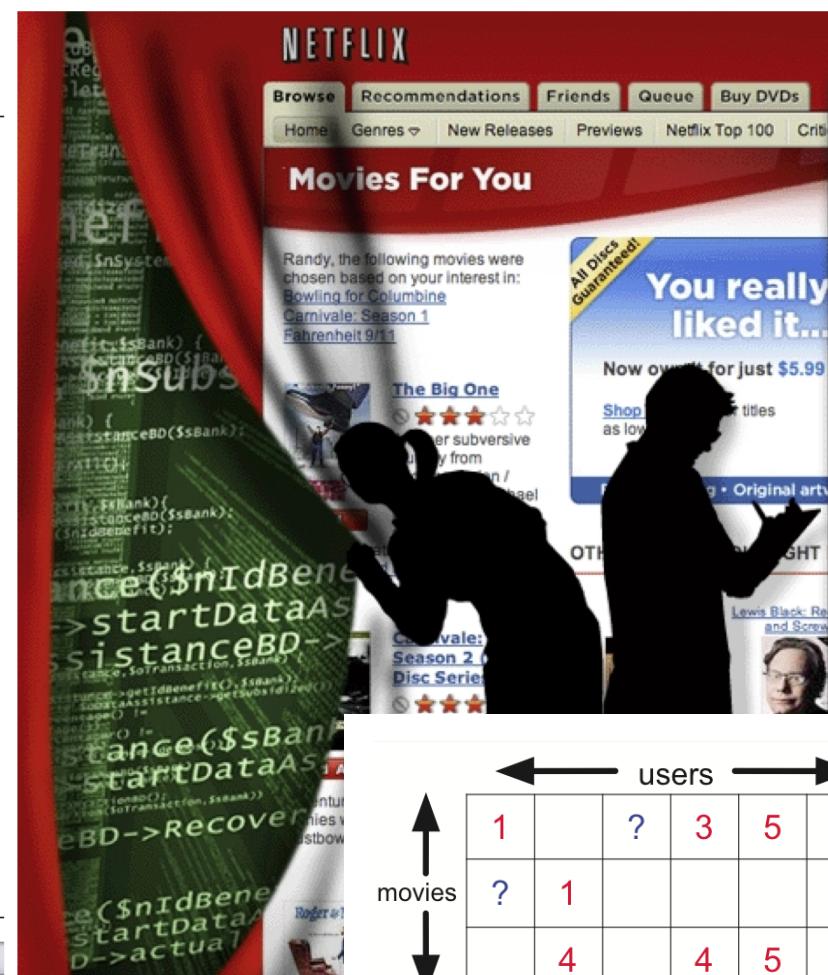
Recommendation systems

Machine learning competition with a \$1 million prize

Leaderboard

Display top 20 leaders.

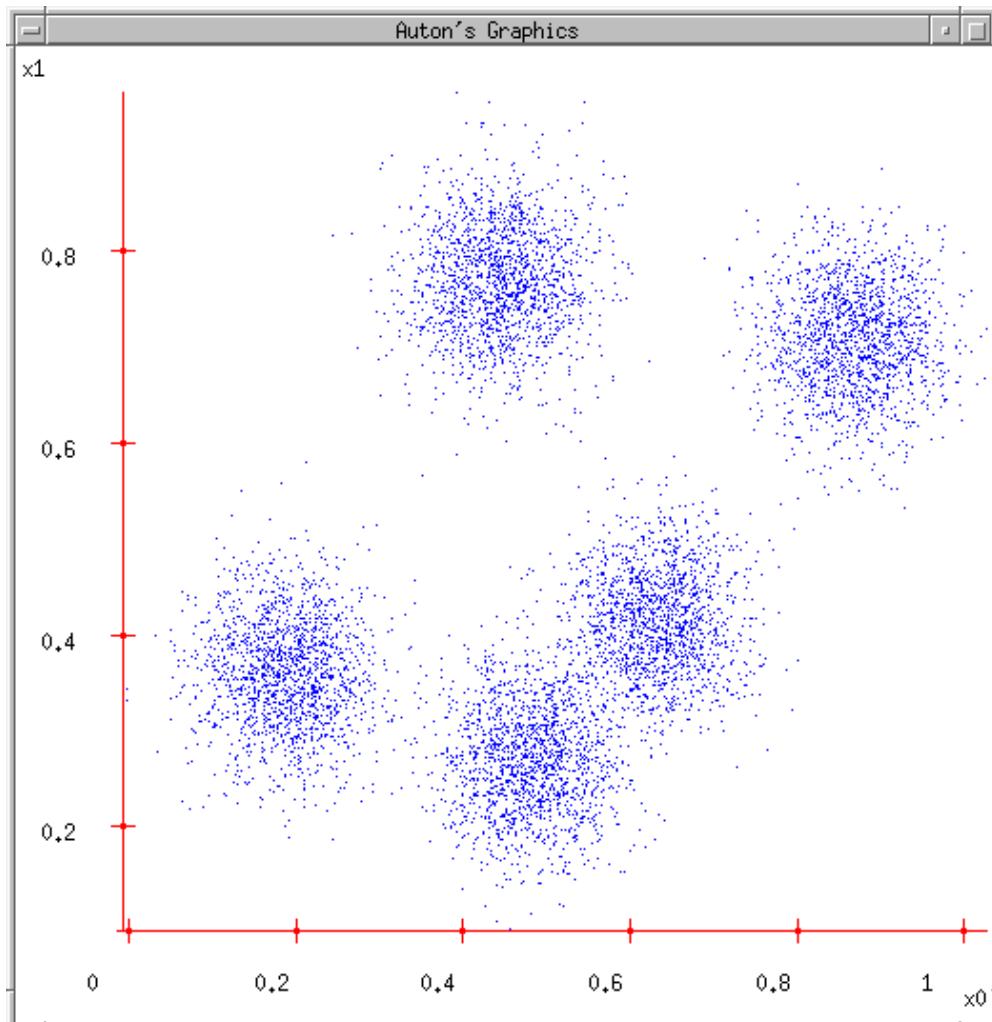
Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	The Ensemble	0.8553	10.10	2009-07-26 18:38:22
2	BellKor's Pragmatic Chaos	0.8554	10.09	2009-07-26 18:18:28
Grand Prize - RMSE <= 0.8563				
3	Grand Prize Team	0.8571	9.91	2009-07-24 13:07:49
4	Opera Solutions and Vandelay United	0.8573	9.89	2009-07-25 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				



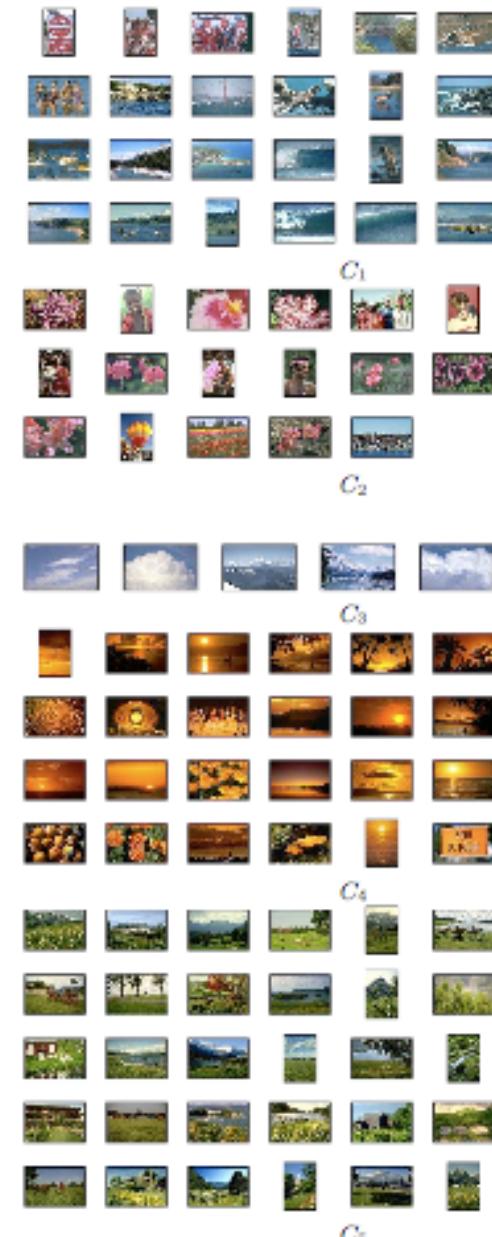
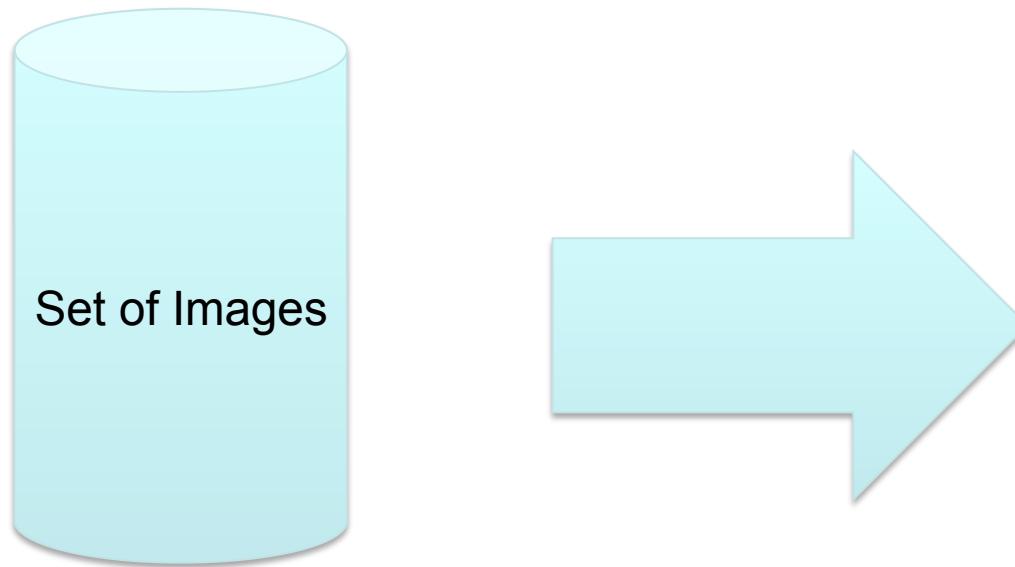
Clustering

discovering structure in data

Clustering Data: Group similar things



Clustering images



[Goldberger et al.]

Clustering web search results

Clusty

web news images wikipedia blogs jobs more »
race

Search advanced preferences

clusters sources sites

All Results (238) remix

- Car (28)
 - Race cars (7)
- Photos, Races Scheduled (5)
 - Game (4)
 - Track (3)
 - Nascar (2)
 - Equipment And Safety (2)
 - Other Topics (7)
- Photos (22)
 - Game (14)
 - Definition (13)
 - Team (18)
- Human (8)
 - Classification Of Human (2)
 - Statement, Evolved (2)
 - Other Topics (4)
- Weekend (8)
- Ethnicity And Race (7)
- Race for the Cure (8)
- Race Information (8)
- [more | all clusters](#)

Cluster Human contains 8 documents.

Search Results

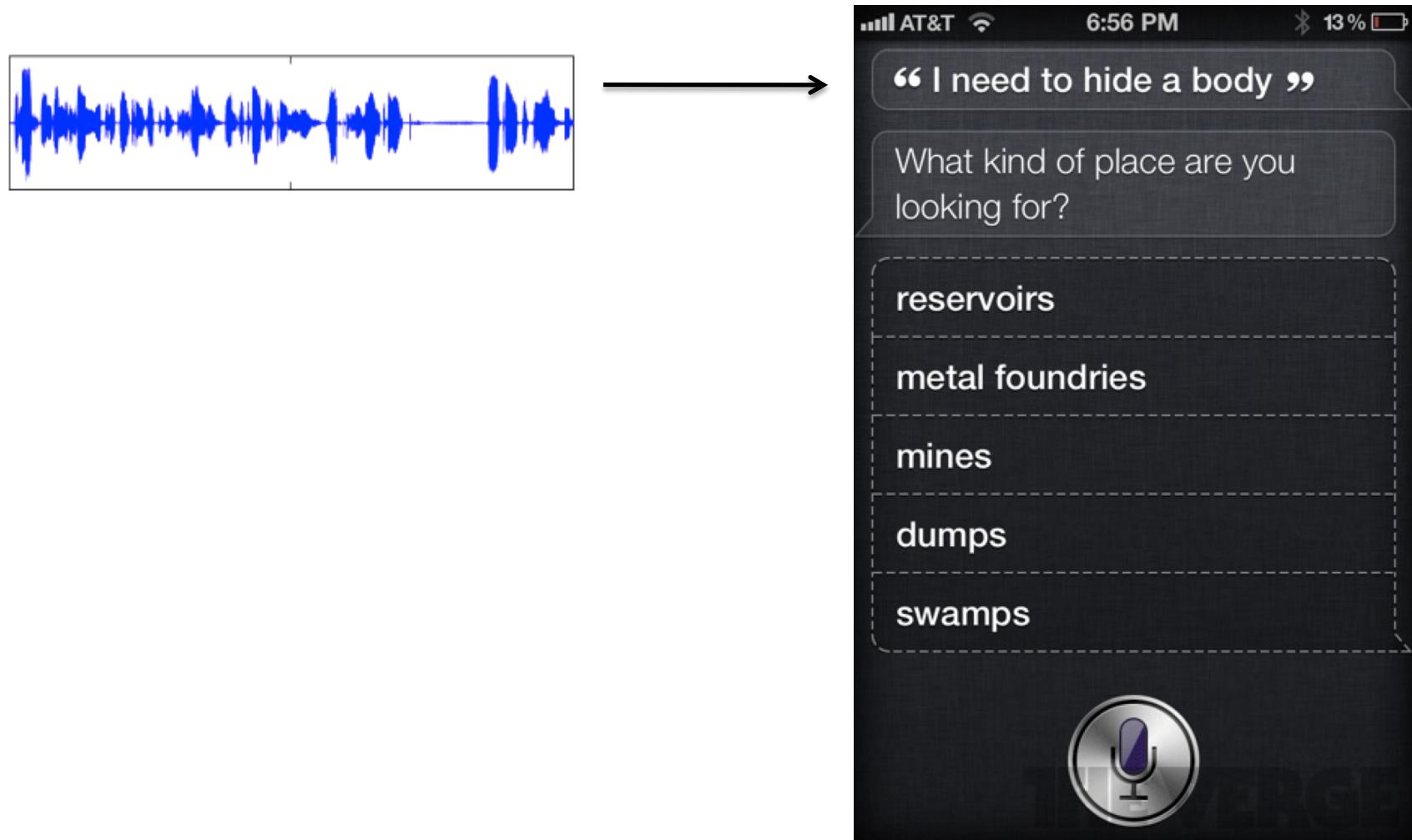
- [Race \(classification of human beings\) - Wikipedia, the free ...](#)** ⓘ 🔎 ⓘ
The term **race** or racial group usually refers to the concept of dividing **humans** into populations or groups on the basis of various sets of characteristics. The most widely used **human** racial categories are based on visible traits (especially skin color, cranial or facial features and hair texture), and self-identification. Conceptions of **race**, as well as specific ways of grouping **races**, vary by culture and over time, and are often controversial for scientific as well as social and political reasons. History · Modern debates · Political and ...
[en.wikipedia.org/wiki/Race_\(classification_of_human_beings\)](http://en.wikipedia.org/wiki/Race_(classification_of_human_beings)) - [cache] - Live, Ask
- [Race - Wikipedia, the free encyclopedia](#)** ⓘ 🔎 ⓘ
General. **Racing** competitions The **Race** (yachting **race**), or La course du millénaire, a no-rules round-the-world sailing event; **Race** (biology), classification of flora and fauna; **Race** (classification of human beings) **Race** and ethnicity in the United States Census, official definitions of "race" used by the US Census Bureau; **Race** and genetics, notion of racial classifications based on genetics. Historical definitions of **race**; **Race** (bearing), the inner and outer rings of a rolling-element bearing. **RACE** in molecular biology "Rapid ... General · Surnames · Television · Music · Literature · Video games
en.wikipedia.org/wiki/Race - [cache] - Live, Ask
- [Publications | Human Rights Watch](#)** ⓘ 🔎 ⓘ
The use of torture, unlawful rendition, secret prisons, unfair trials, ... Risks to Migrants, Refugees, and Asylum Seekers in Egypt and Israel ... In the run-up to the Beijing Olympics in August 2008, ...
www.hrw.org/backgrounder/usa/race - [cache] - Ask
- [Amazon.com: Race: The Reality Of Human Differences: Vincent Sarich ...](#)** ⓘ 🔎 ⓘ
Amazon.com: **Race**: The Reality Of Human Differences: Vincent Sarich, Frank Miele: Books ... From Publishers Weekly Sarich, a Berkeley emeritus anthropologist, and Miele, an editor ...
www.amazon.com/Race-Reality-Differences-Vincent-Sarich/dp/0813340861 - [cache] - Live
- [AAPA Statement on Biological Aspects of Race](#)** ⓘ 🔎 ⓘ
AAPA Statement on Biological Aspects of **Race** ... Published in the American Journal of Physical Anthropology, vol. 101, pp 569-570, 1996 ... PREAMBLE As scientists who study **human** evolution and variation, ...
www.physanth.org/positions/race.html - [cache] - Ask
- [race: Definition from Answers.com](#)** ⓘ 🔎 ⓘ
race n. A local geographic or global **human** population distinguished as a more or less distinct group by genetically transmitted physical
www.answers.com/topic/race-1 - [cache] - Live
- [Dopefish.com](#)** ⓘ 🔎 ⓘ
Site for newbies as well as experienced Dopefish followers, chronicling the birth of the Dopefish, its numerous appearances in several computer games, and its eventual take-over of the **human race**. Maintained by Mr. Dopefish himself, Joe Siegler of Apogee Software.
www.dopefish.com - [cache] - Open Directory

find in clusters: Find

Structured prediction

from data to discrete classes

Speech recognition



Natural language processing

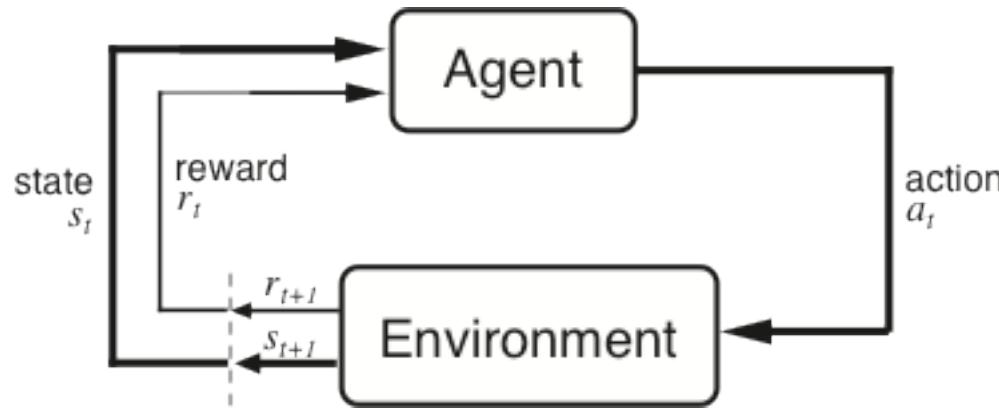
I need to hide a body
noun, verb, preposition, ...



Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
 - Policy is a mapping from states → actions that tells you what to do in a given state
- Examples:
 - Credit assignment problem
 - Game playing
 - Robot in a maze
 - Balance a pole on your hand

The Agent-Environment Interface



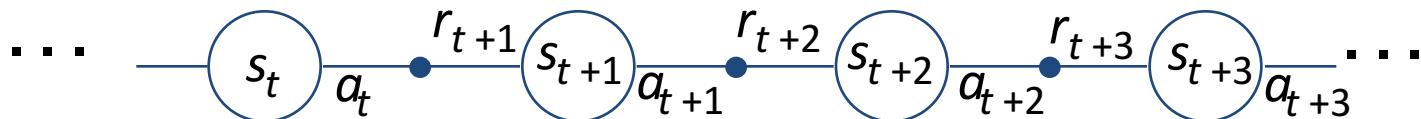
Agent and environment interact at discrete time steps : $t = 0, 1, 2, K$

Agent observes state at step t : $s_t \in S$

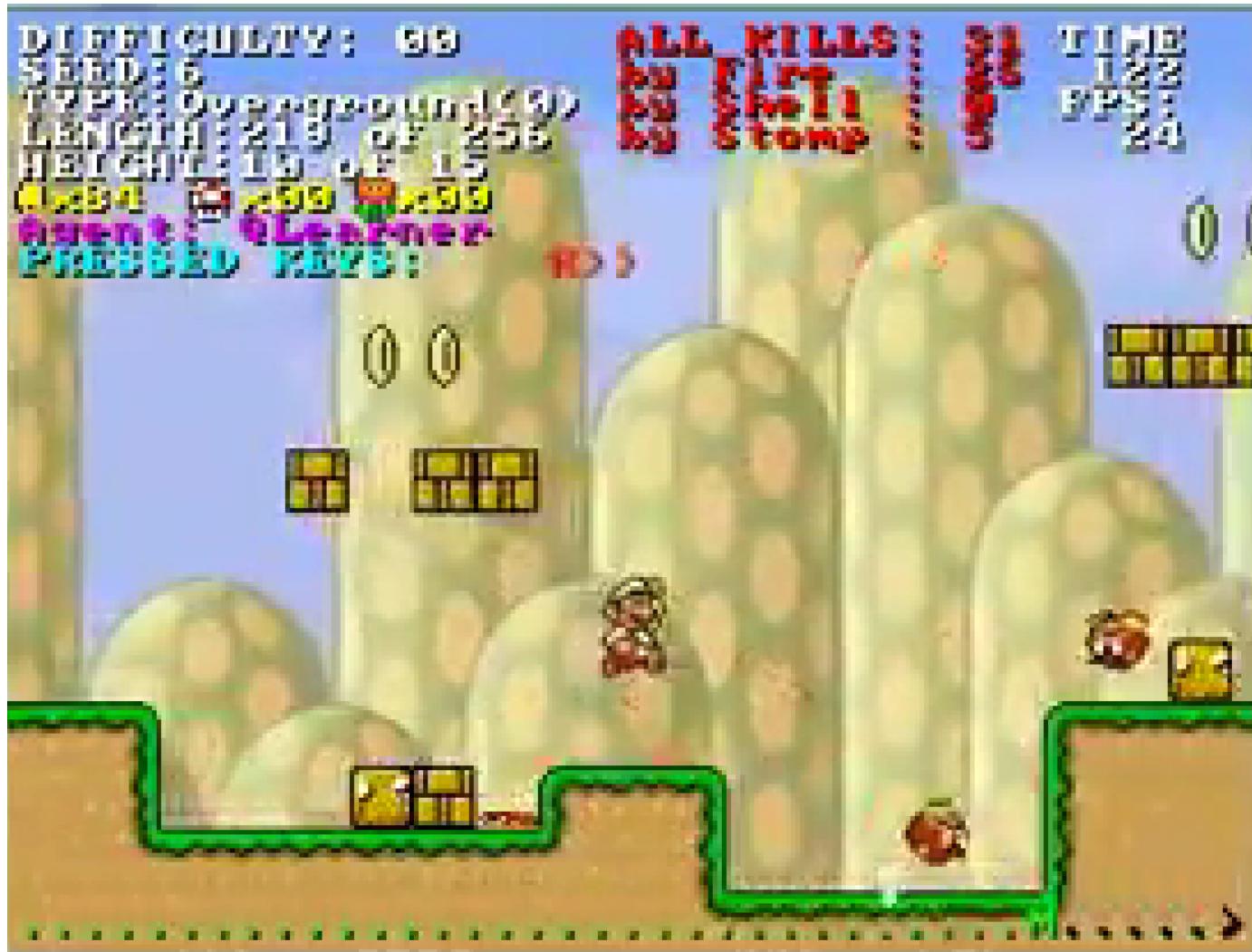
produces action at step t : $a_t \in A(s_t)$

gets resulting reward : $r_{t+1} \in \mathcal{R}$

and resulting next state : s_{t+1}



Reinforcement Learning



<https://www.youtube.com/watch?v=4cgWya-wjgY>

Inverse Reinforcement Learning

- Learn policy from user demonstrations



Stanford Autonomous Helicopter

<http://heli.stanford.edu/>

<https://www.youtube.com/watch?v=VCdxqn0fcnE>

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

Supervised Learning: find f

- Given: Training set $\{(x_i, y_i) \mid i = 1 \dots N\}$
- Find: A good approximation to $f : X \rightarrow Y$

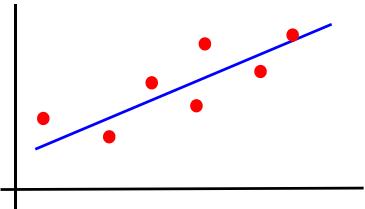
Examples: what are X and Y ?

- Spam Detection
 - Map email to {Spam, Not Spam}
- Digit recognition
 - Map pixels to {0,1,2,3,4,5,6,7,8,9}
- Stock Prediction
 - Map new, historic prices, etc. to \mathbb{R} (the real numbers)

Three canonical learning problems

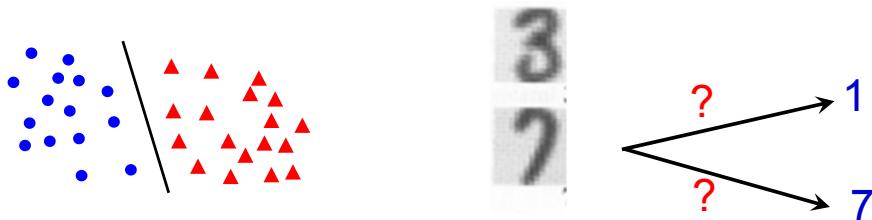
1. Regression - supervised

- estimate parameters, e.g. of weight vs height



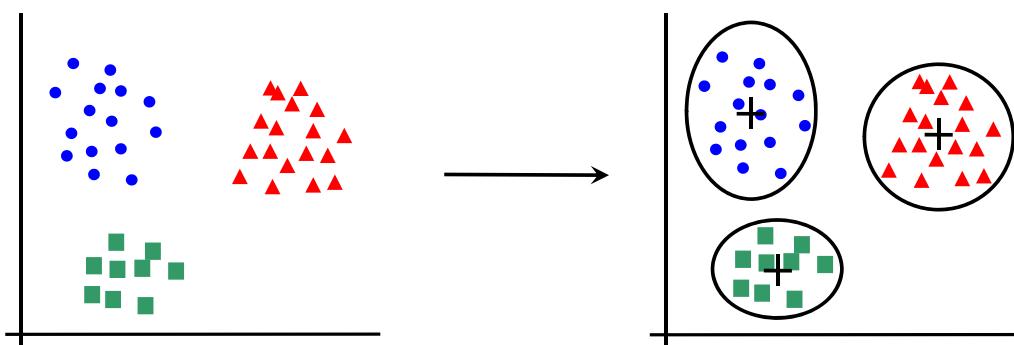
2. Classification - supervised

- estimate class, e.g. handwritten digit classification

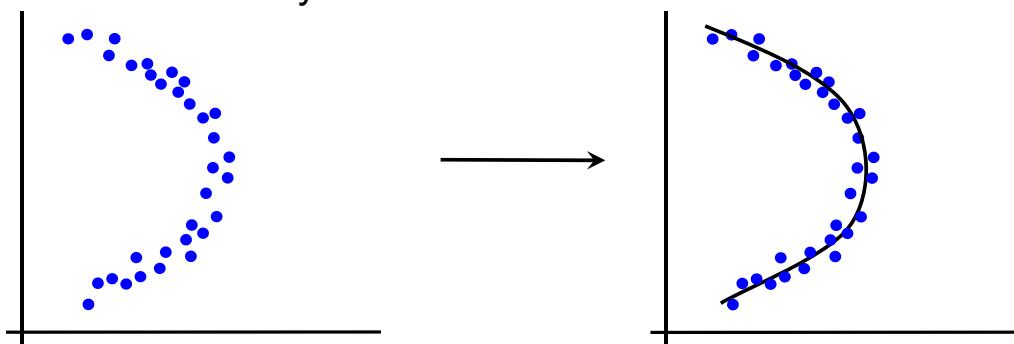


3. Unsupervised learning – model the data

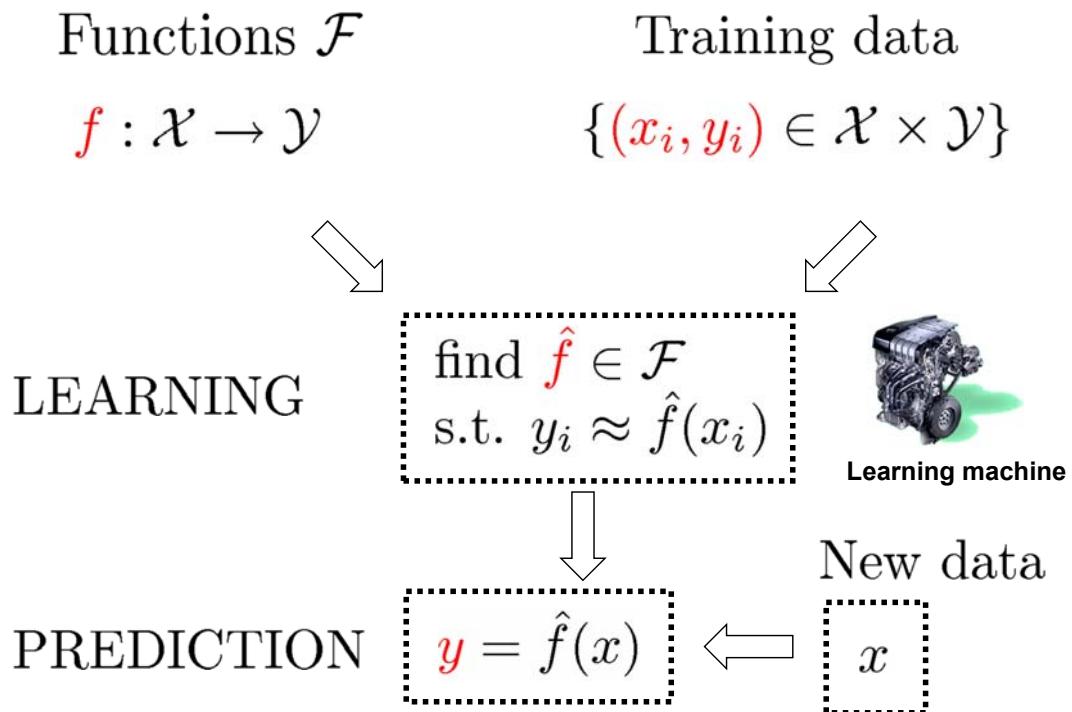
- clustering



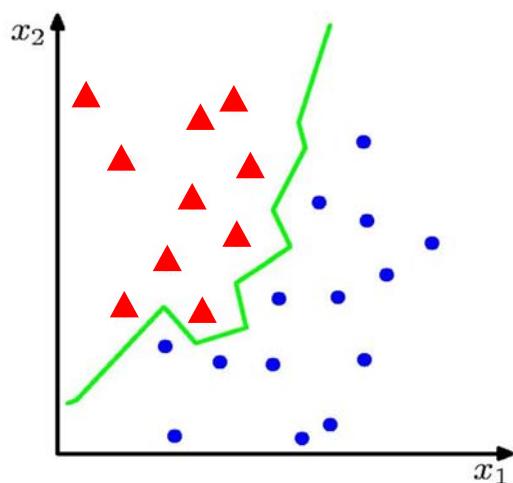
- dimensionality reduction



Supervised Learning: Overview



Classification



- Suppose we are given a training set of N observations (x_1, \dots, x_N) and (y_1, \dots, y_N) , $x_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$
- Classification problem is to estimate $f(x)$ from this data such that

$$f(x_i) = y_i$$

A Supervised Learning Problem

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

- Our goal is to find a function $f : X \rightarrow Y$
 - $X = \{0,1\}^4$
 - $Y = \{0,1\}$
- **Question 1:** How should we pick the *hypothesis space*, the set of possible functions f ?
- **Question 2:** How do we find the best f in the hypothesis space?

Occam's Razor Principle

- William of **Occam**: Monk living in the 14th century
- Principle of parsimony:

“One should not increase, beyond what is necessary, the number of entities required to explain anything”

- When **many** solutions are available for a given problem, we should select the **simplest** one
- But what do we mean by **simple**?
- We will use **prior knowledge** of the problem to solve to define what is a simple solution

Example of a prior: smoothness

[Samy Bengio]

Key Issues in Machine Learning

- How do we choose a hypothesis space?
 - Often we use **prior knowledge** to guide this choice
- How can we gauge the accuracy of a hypothesis on unseen data?
 - **Occam's razor:** use the *simplest* hypothesis consistent with data!
This will help us avoid overfitting.
 - **Learning theory** will help us quantify our ability to **generalize** as a function of the amount of training data and the hypothesis space
- How do we find the best hypothesis?
 - This is an **algorithmic** question, the main topic of computer science
- How to model applications as machine learning problems?
(engineering challenge)