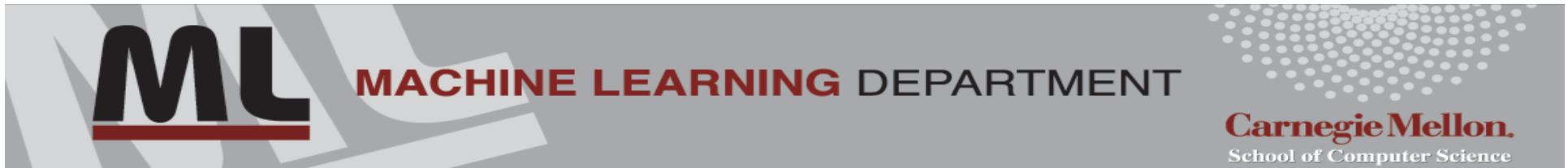


Machine Learning - Intro

Manuela Veloso

Co-instructor: Pradeep Ravikumar

Machine Learning 10-701
Jan 17, 2018



Instructors



Manuela Veloso



Pradeep Ravikumar

TAs

- Shaojie Bai shaojieb@andrew.cmu.edu
- Adarsh Prasad adarshp@andrew.cmu.edu
- Otilia Stretcu, ostretcu@andrew.cmu.edu
- Dimitris Konomis, dkonomis@andrew.cmu.edu
- Satyapriya Krishna satyaprk@andrew.cmu.edu
- Sreena Nallamothu snallamo@andrew.cmu.edu
- Lam Wing Chan, lamwingc@andrew.cmu.edu
- Wenhao Qin, wqin@andrew.cmu.edu
- George Stoica, gis@andrew.cmu.edu

SOME LOGISTICS

Class Webpage

- <http://www.cs.cmu.edu/~pradeepr/701>
- The class schedule, logistics, and lecture materials will be posted there.

Piazza Discussion Board

- We will use Piazza for announcements, as well as the discussion board for the class.
- Enrollment:
 - Those who have been admitted will be auto-added onto Piazza board for the course
 - In case you are not / those on waitlist: please request to be added

Auditing

- To satisfy the auditing requirement, you must:
 - Get a passing grade in the course.

Prerequisites

- Probabilities
 - Distributions, densities, marginalization...
- Basic statistics
 - Moments, typical distributions, regression...
- Algorithms
 - Dynamic programming, basic data structures, complexity...
- Programming
 - Mostly your choice of language, but Matlab will be very useful
- We provide some background, but the class will be fast paced
- Ability to deal with “abstract mathematical concepts”

Recitations

- Strongly recommended
 - Brush up pre-requisites
 - Review material (difficult topics, clear misunderstandings, extra new topics)
 - Ask questions
 - Schedule TBD

Reading Materials

- Lectures are intended to be self-contained. For supplementary readings, with each lecture, we will also have pointers to either online reference materials, or chapters from the following books:
 - Pattern Recognition and Machine Learning; Chris Bishop
 - Machine Learning: A probabilistic perspective; Kevin Murphy.
 - Machine Learning, Tom Mitchell
 - The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Trevor Hastie, Robert Tibshirani, Jerome Friedman

Grading

- 5 Homeworks (50%)
 - Enjoy, Learn a lot, Delve into ML, and... Start early, Start early...
- Final project (25%)
 - Stay tuned for further details
- Midterm (25%)
 - March 21, in class

Homeworks

- Homeworks are hard and designed to help you learn the most, start early 😊
- Homeworks will typically have three components
 - Short problems
 - Programming assignment
 - Using off-the-shelf packages
 - Explore the consequences of ML algorithms
 - Multiple choice
- Homeworks are due at the beginning of class
- 10 late days, which you can distribute among the five homeworks – not more than 3 late days per homework.

Homeworks

- Collaboration
 - You may **discuss** the questions
 - Each student writes their own answers
 - Each student must write their own code for the programming part
 - **Please don't search for answers on the web, Google, previous years' homeworks, etc.**
 - please ask us if you are not sure if you can use a particular reference

First Point of Contact for HWs

- To facilitate interaction, TA(s) will be assigned to each homework question
- These will be your “first point of contact” for this question
 - But, you can always ask any of us

Communication Channel

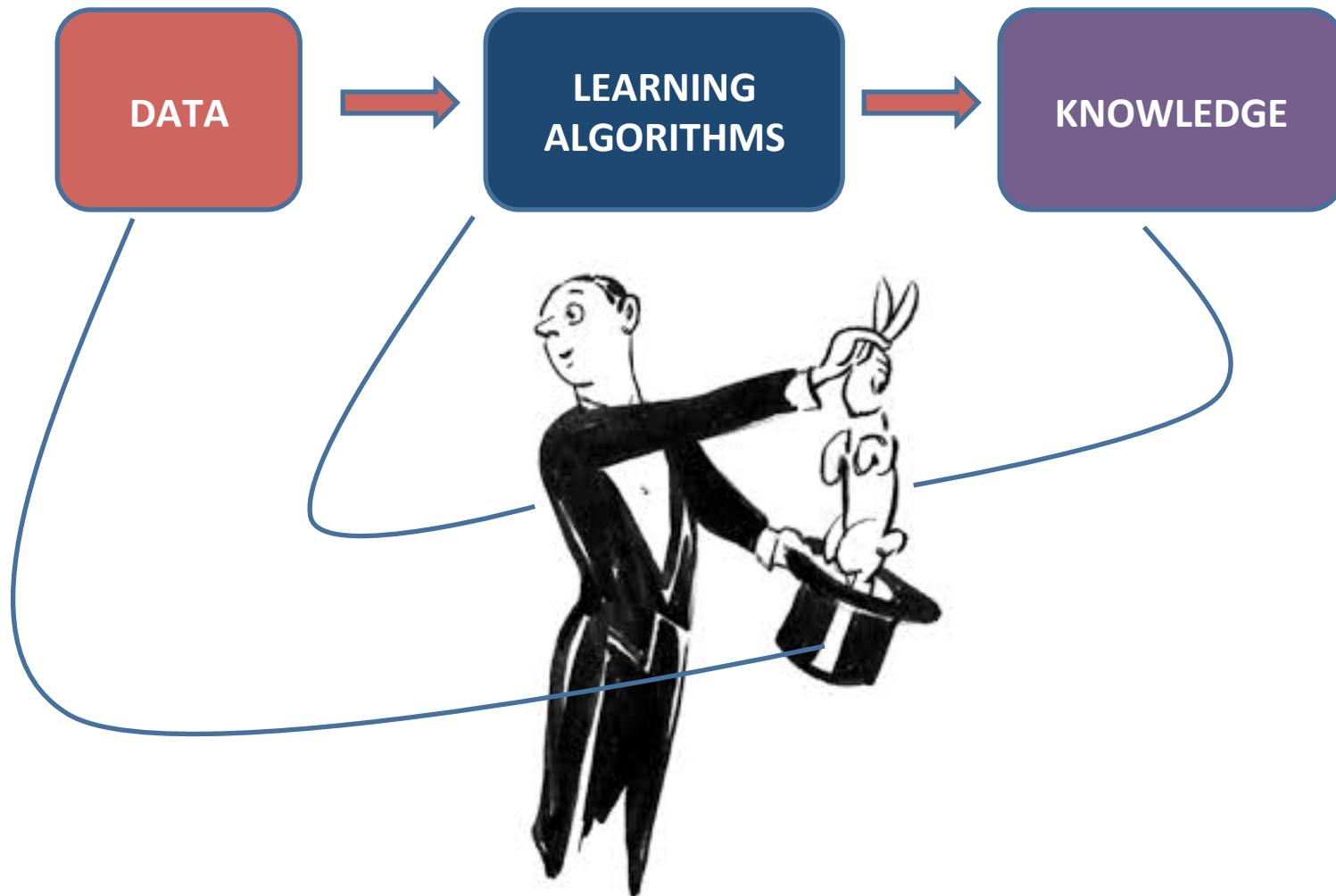
- For e-mailing instructors, always use:
 - 10701-instructors@cs.cmu.edu
- Please use Piazza to ask questions
- For announcements, we will use Piazza

What is Machine Learning?

What is Machine Learning?



What is Machine Learning?



Machine Learning



- Algorithms that improve their knowledge towards some task with data
- Machine learning as Artificial Intelligence

Herbert A. Simon and Allen Newell



Learning

- Gathering more knowledge
 - “Knowing *more* than was known before learning”
- Processing, understanding data
- Experience, feedback, refinement
- Learning modifies decision making to improve performance

Scientists Use Data

- Statistics: understanding of given data
- Data Mining: extracting patterns from large data
- Data Science: collection, analysis, and interpretation of data
- Signal Processing: analyze, transform, model data

Machine Learning

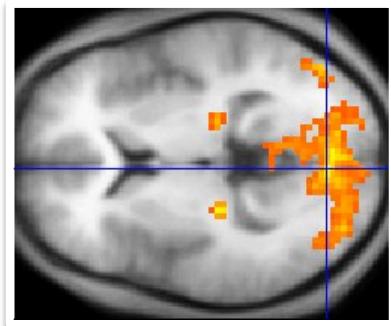
- The goal of machine learning is the underlying mechanisms and algorithms that allow improving our knowledge with more data
 - Data construed broadly, e.g., “experiences”
 - Knowledge construed broadly, e.g., possible actions

From Data to Understanding ...

Machine Learning in Action

Machine Learning in Action

- Decoding thoughts from brain scans



Rob a bank ...

Home » Health & Wellness

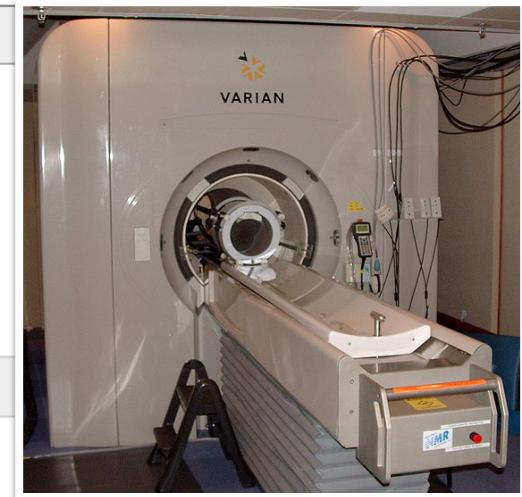
Brain Scans: Are You a Criminal?

 Published February 07, 2007 by:
Andrea Okrentowich
[View Profile](#) | [Follow](#) | [Add to Favorites](#)

More: [Brain Scans](#) | [Brain Scan](#) | [Disposition](#) | [Defendant](#) | [Criminal Behavior](#)

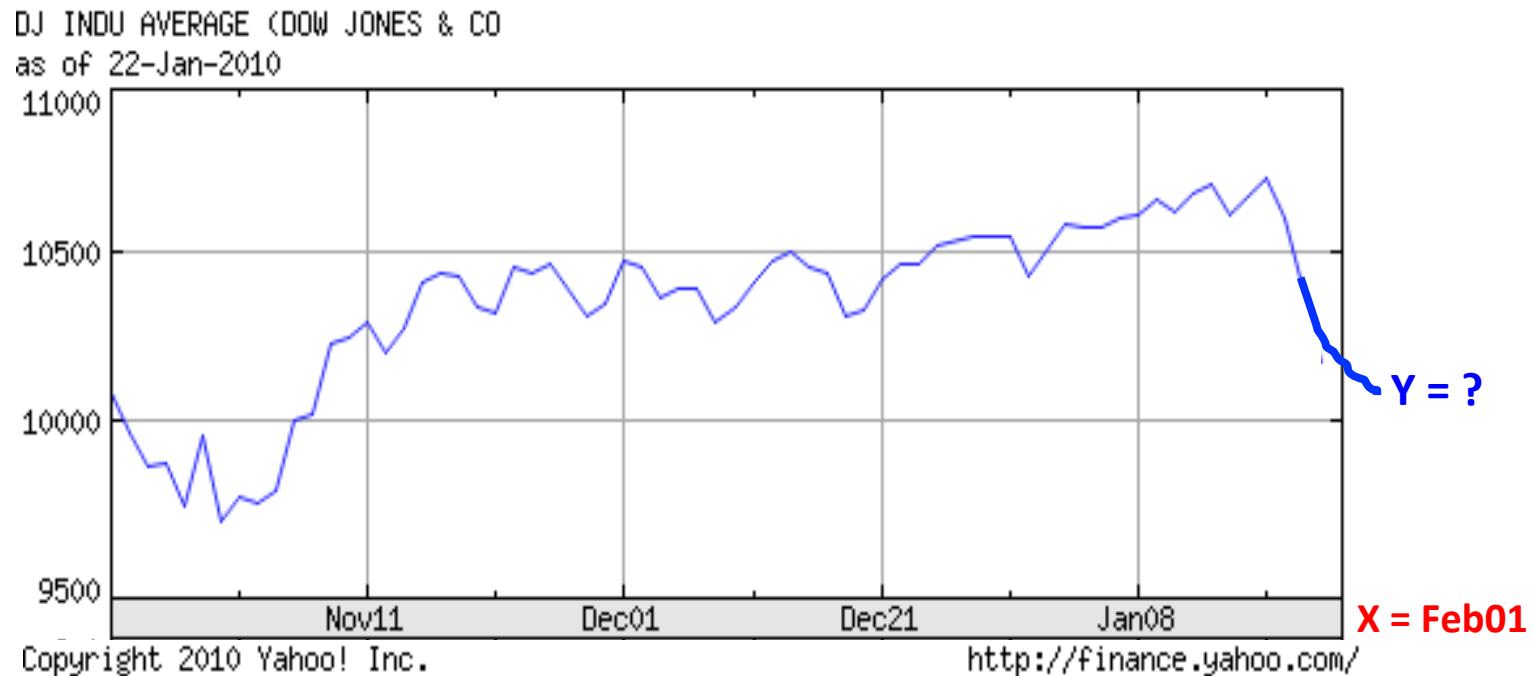
MRI Scans as Courtroom Evidence

The average Joe's MRI scan can show a brain abnormality, do we proceed to check him into the nearest mental institution or prison? That would make about as much sense as trying to prove a defendant innocent of a violent



Machine Learning in Action

- Stock Market Prediction



Machine Learning in Action

- Document classification



Sports
Science
News

Machine Learning in Action

- Spam filtering

Welcome to New Media Installation: Art that Learns

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.

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



Spam/
Not spam

==== Natural WeightLOSS Solution ===

Vital Acai is a natural WeightLOSS product that Enables people to lose weight 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

Machine Learning in Action

- Cars navigating on their own



Boss, the self-driving SUV
1st place in the DARPA Urban
Challenge.

Photo courtesy of Tartan Racing.



Machine Learning in Action

- The **best** helicopter pilot is now a computer!
 - it runs a program that learns how to fly and make acrobatic maneuvers by itself!
 - no taped instructions, joysticks, or things like that ...



Machine Learning in Action

*Learning Object Groundings for a Mobile Service Robot
from Web Access and Dialog with Users*

Tom Kollar, Vittorio Perera, Mehdi Samadi, Robin Soetens,
Joydeep Biswas, Brian Coltin, Daniele Nardi,
Manuela Veloso

School of Computer Science

Carnegie Mellon University

January 2013

Machine Learning in Action

- Many, many more...
 - Speech recognition, Natural language processing
 - Computer vision
 - Web forensics
 - Medical outcomes analysis
 - Computational biology
 - Sensor networks
 - Social networks
 - ...

ML is trending!

- Wide applicability
- Very large-scale complex systems
 - Internet (billions of nodes), sensor network (new multi-modal sensing devices), genetics (human genome)
- Huge multi-dimensional data sets
 - 30,000 genes x 10,000 drugs x 100 species x ...
- Improved machine learning algorithms
- Improved data capture (Terabytes, Petabytes of data), networking, faster computers
- New York Times is regularly talking about machine learning

ML has a long way to go ...

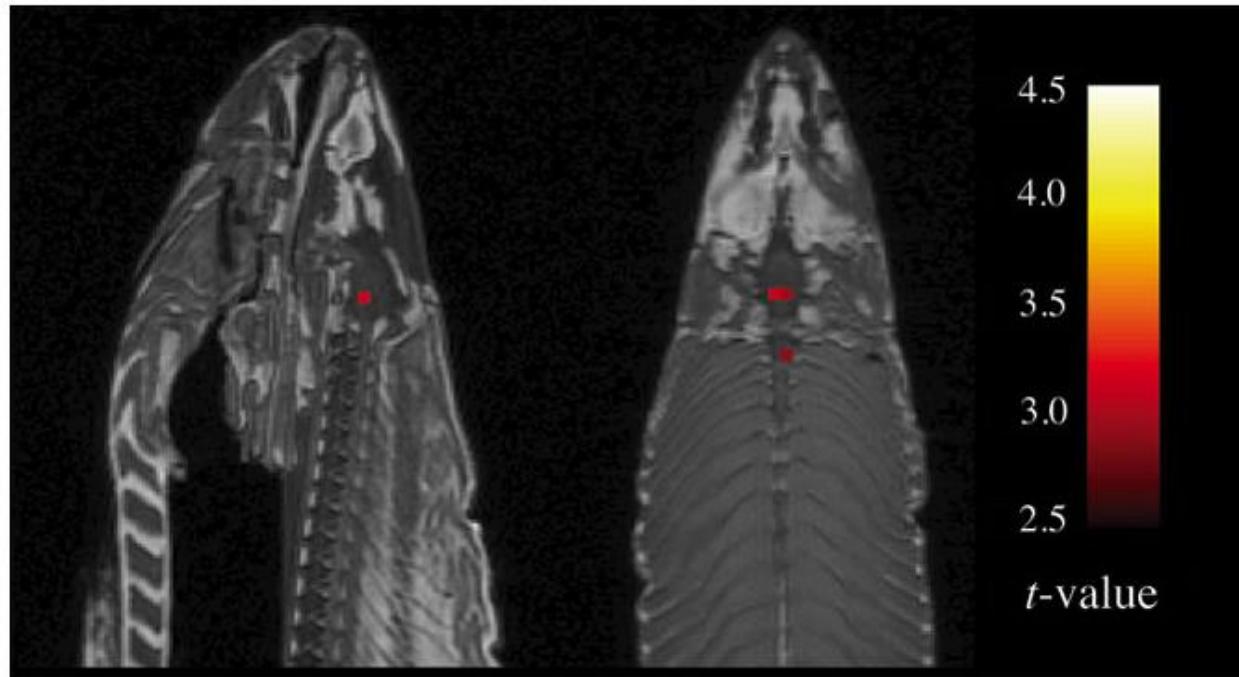
WIRED SCIENCE

NEWS FOR YOUR NEURONS



Scanning Dead Salmon in fMRI Machine Highlights Risk of Red Herrings

By Alexis Madrigal September 18, 2009 | 5:37 pm | Categories: Brains and Behavior



Unique Challenges with increased use of Machine Learning

Google researchers develop a test for machine learning bias



BY MIKE WHEATLEY
(HTTP://SILICONANGLE.COM/BLOG/AUTHOR/MIKEWHEATLEY/)

UPDATED 19:05 EST . 23 DECEMBER 2016

2 MIN READ



A team of researchers at Google Inc. has developed a method for testing whether or not machine learning algorithms inject bias, such as gender or racial bias, into their decision-making processes.

For some time, concerns have been raised about the possibility that machine learning algorithms are injecting bias into applications such as advertising, credit, education, employment and justice. Recent examples include a crime prediction algorithm that targeted black neighborhoods (<http://uk.businessinsider.com/predictive-policing-discriminatory-police-crime-2016-10?r=US&IR=T>) and an online advertising platform that was found to show highly paid executive jobs to men more often than women (<http://www.independent.co.uk/life-style/gadgets-and-tech/news/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-10372166.html>) .

"Decisions based on machine learning can be both incredibly useful and have a profound impact on our lives," said Moritz Hardt, a senior research scientist at Google, who co-authored the paper, "Equality of Opportunity in Supervised Learning." "Despite the demand, a vetted methodology for avoiding discrimination against protected attributes in machine learning is lacking."

Challenges

- Bias
- Ethics
- Interpretability
- Explanations
- Instruction
- Correction
- Training
-

What this course is about

- Covers a wide range of Machine Learning techniques
 - from basic to state-of-the-art
- You will learn about the methods you heard about:
 - Naïve Bayes, logistic regression, nearest-neighbor, decision trees, boosting, neural nets, overfitting, regularization, dimensionality reduction, PCA, error bounds, VC dimension, SVMs, kernels, margin bounds, K-means, EM, mixture models, semi-supervised learning, HMMs, graphical models, active learning, reinforcement learning...
- Covers algorithms, theory and applications
- **It's going to be fun and hard work ☺**

Machine Learning

- The goal of machine learning is identifying the underlying mechanisms and algorithms that allow improving our knowledge with more data
- Algorithms that improve their knowledge towards some task with data

Three axes of ML

- Data
- Tasks, i.e., what is the type of knowledge that we seek from data
- Algorithms

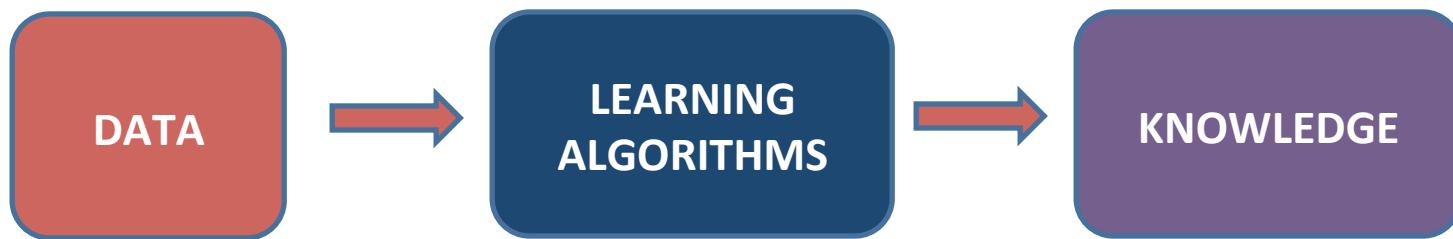
First Axis: Data

- Fully observed
- Partially observed
 - Some variables systematically not observed
 - e.g., “topic” of a document
 - Some variables missing some of the time
 - “missing data”
- Actively collect/sense data

Second Axis: Algorithms

- Model-based Methods
 - Probabilistic Model of the data
 - Parametric Models
 - Nonparametric Models
- Model-free Methods

Model-based ML



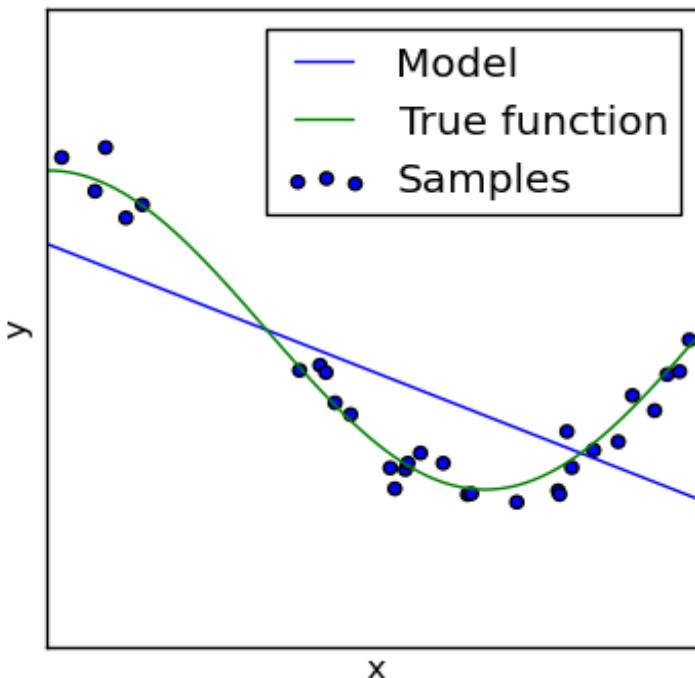
Model-based ML



- Learning: From data to model
 - A model thus is a summary of the data
 - But also a story of how the data was generated
 - Could thus be used to describe how future data can be generated
 - **E.g. given (symptoms, diseases) data, a model explains how symptoms and diseases are related**
- Inference: From model to knowledge
 - Given the model, how can we answer questions relevant to us
 - **E.g. given (symptom, disease) model, given some symptoms, what is the disease?**

Parametric Models

- “Fixed-size” models that do not “grow” with the data
- More data just means you learn/fit the model better

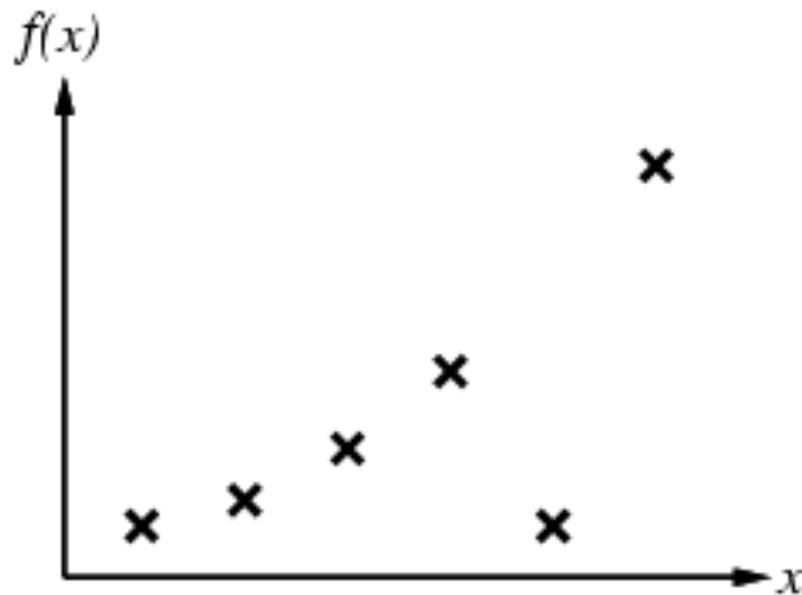


Fitting a simple line (2 params)
to a bunch of one-dim. samples

Model: data = point on line + noise

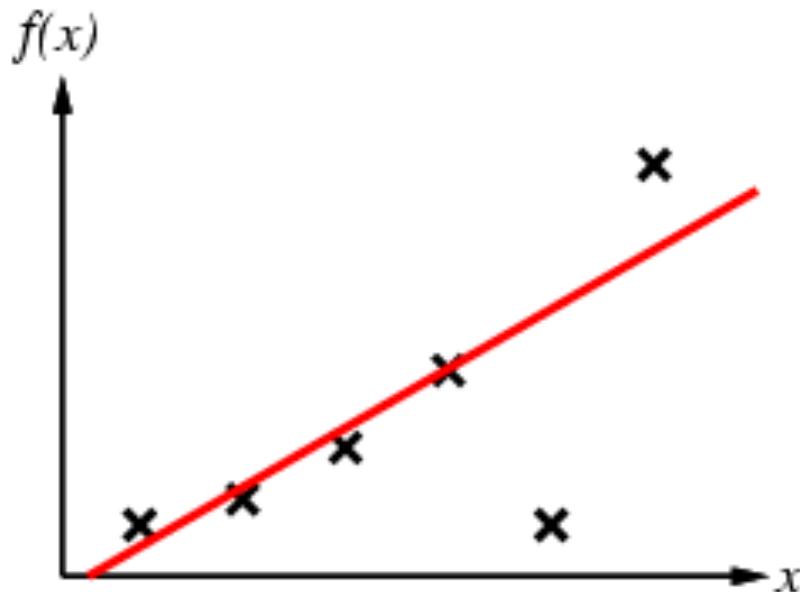
Model from Data

- Data: Examples of function
- Learning: *Curve fitting to agree with data*



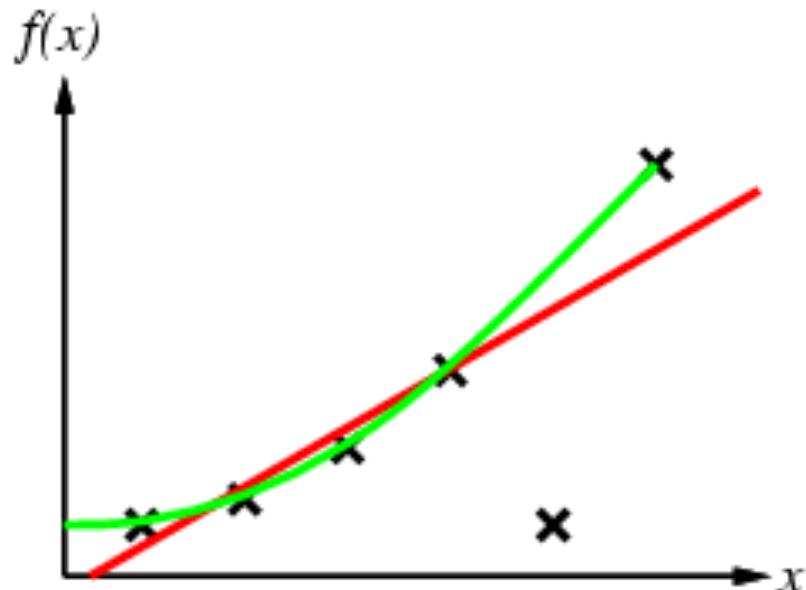
Model from Data

- Data: Examples of function
- Learning: *Curve fitting to agree with data*



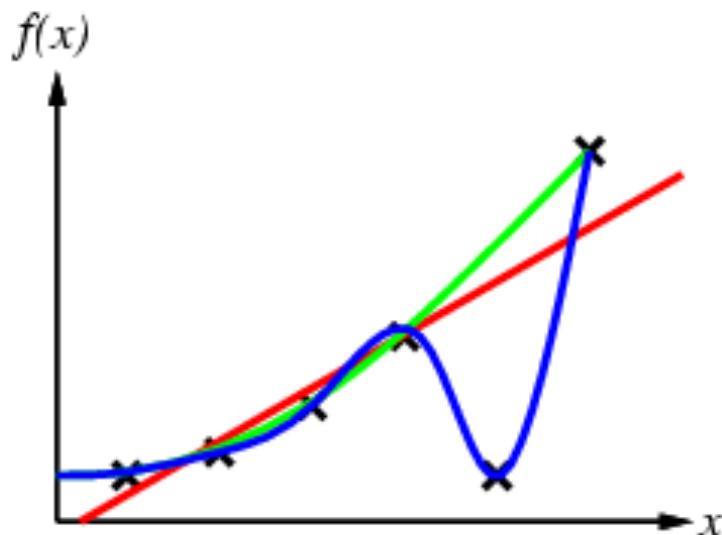
Model from Data

- Data: Examples of function
- Learning: *Curve fitting to agree with data*



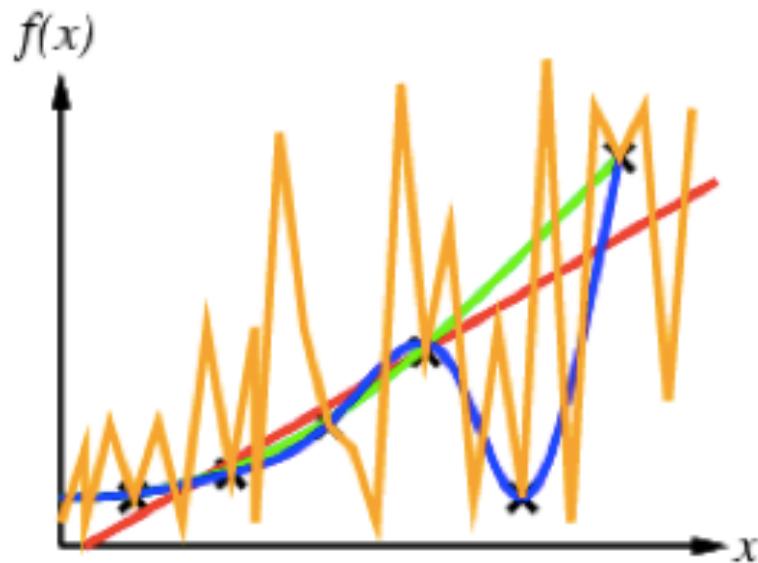
Model from Data

- Data: Examples of function
- Learning: *Curve fitting to agree with data*



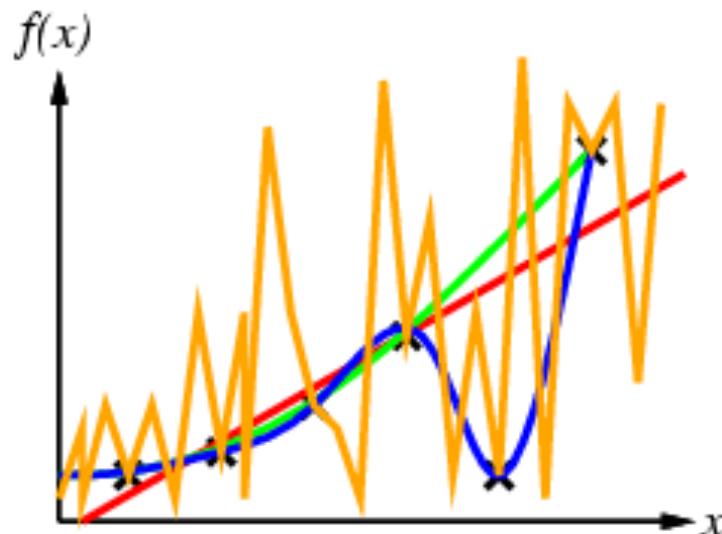
Model from Data

- Data: Examples of function
- Learning: *Curve fitting to agree with data*



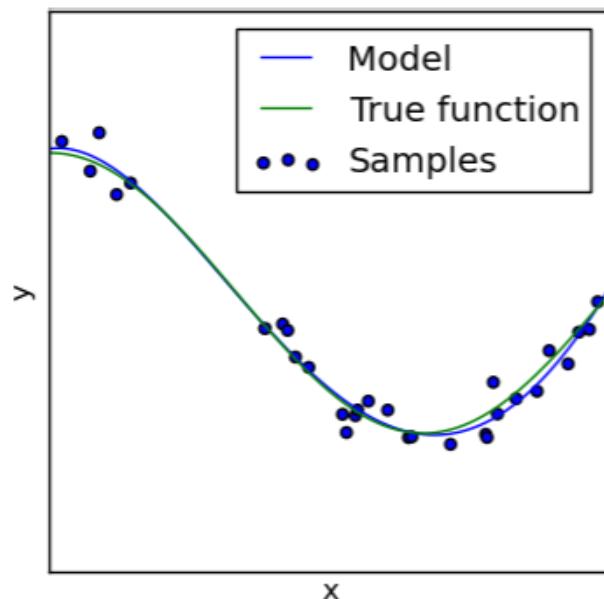
Inductive Learning Method - Bias

- Hypothesis “form” – **bias** on the learning outcome
- Ockham’s razor: prefer the simplest hypothesis consistent with data



Nonparametric Models

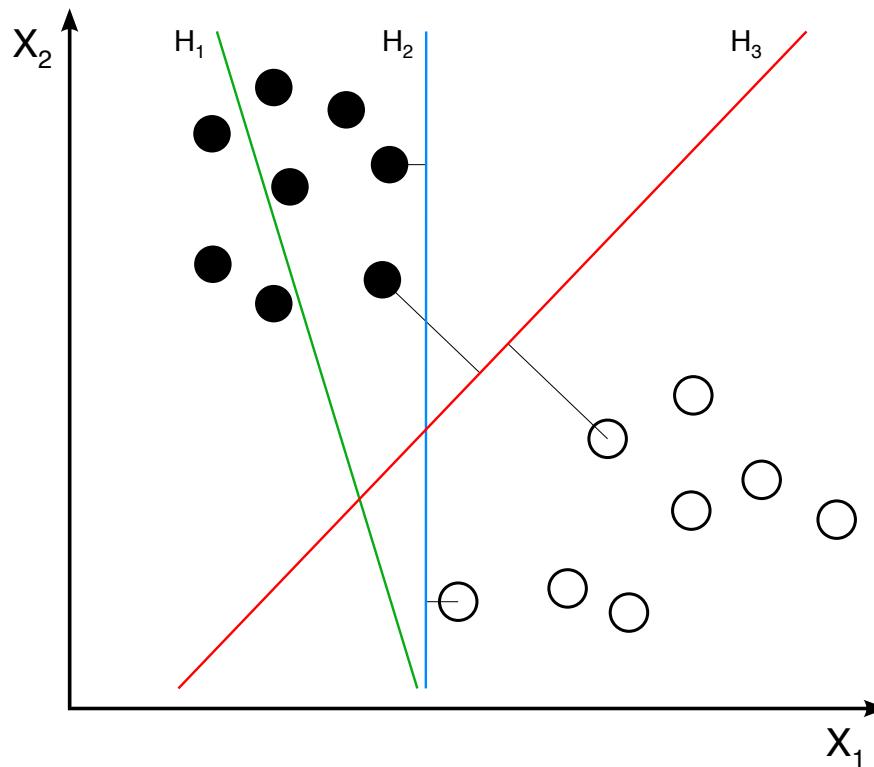
- Models that grow with the data
- More data means a more complex model



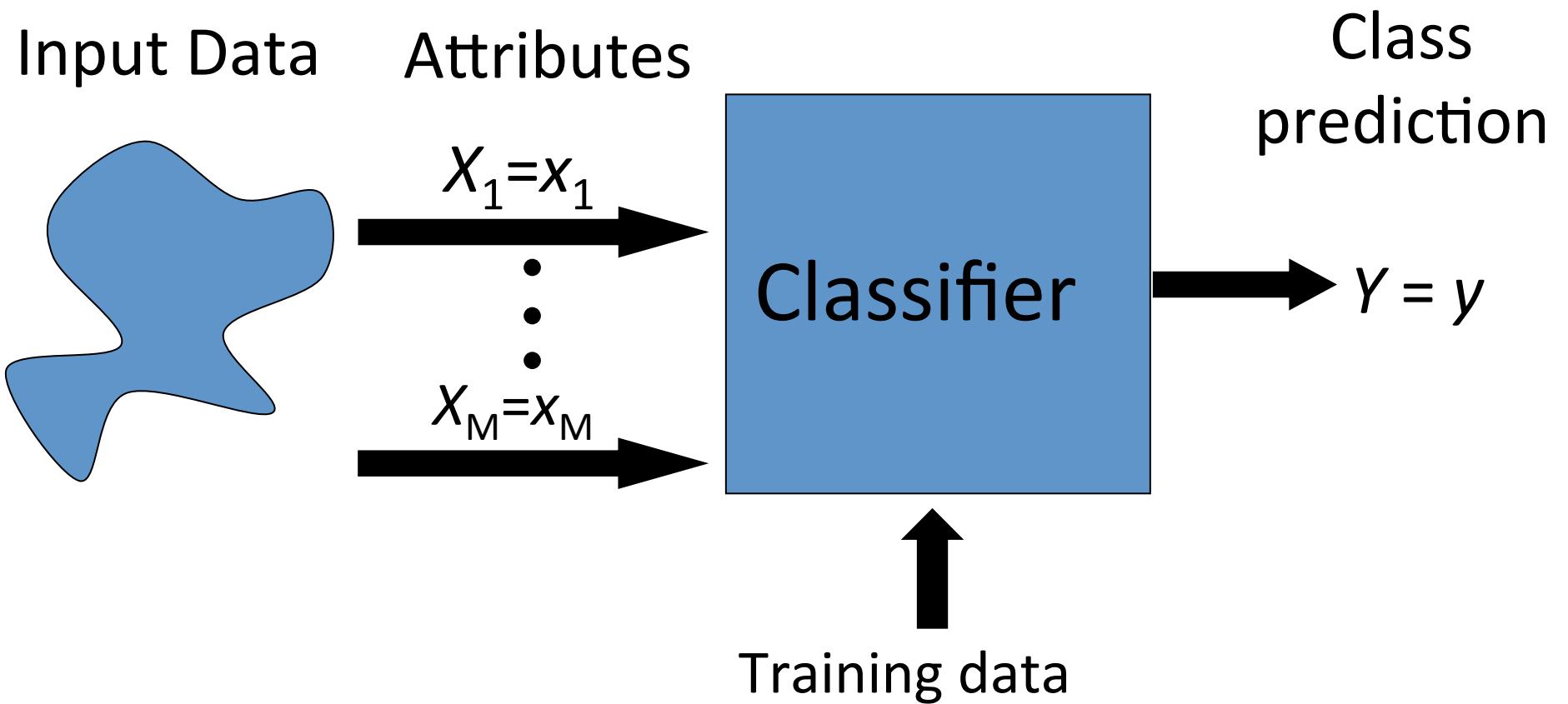
Fitting a “smooth function”
to a bunch of one-dim. samples

Model: data = point on smooth curve + noise

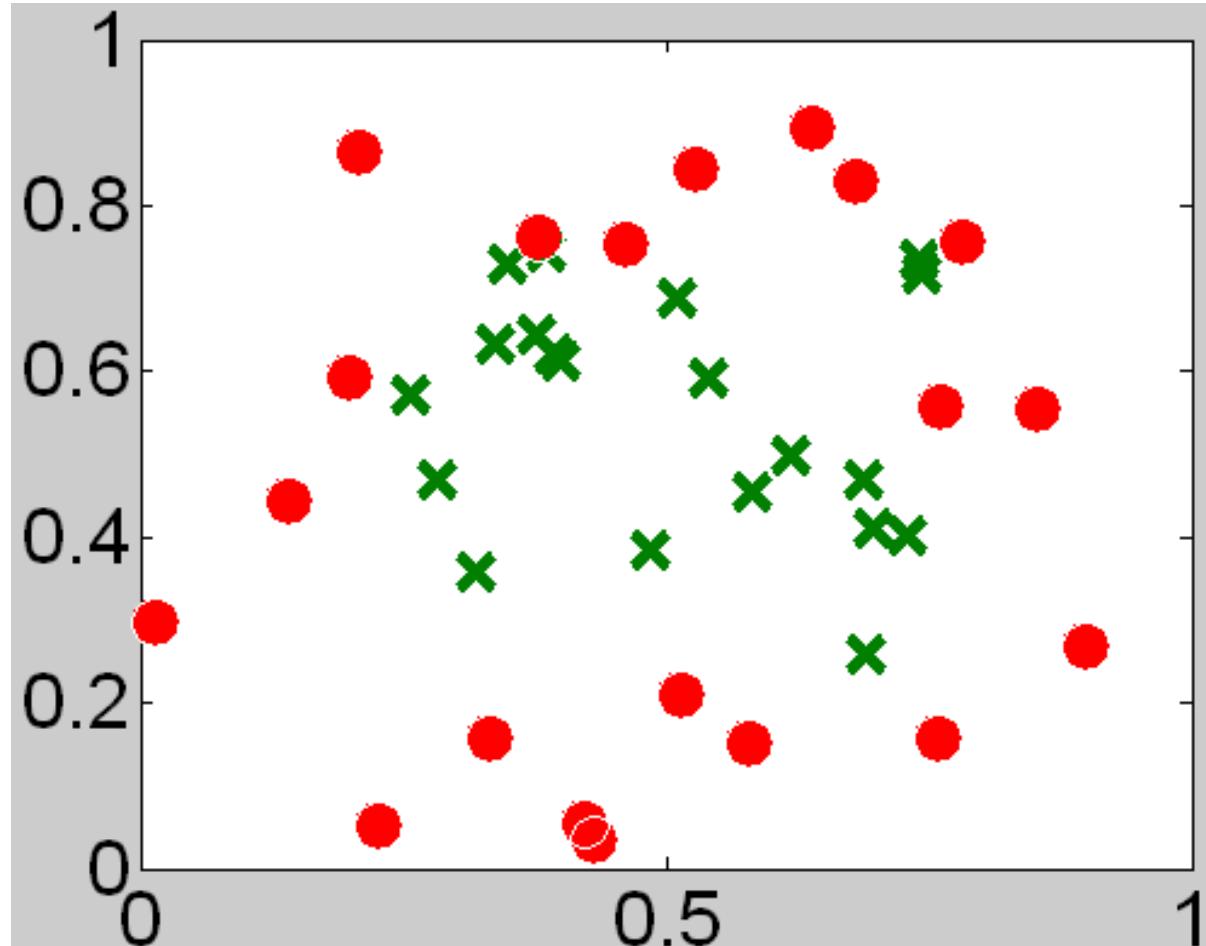
Model-free Methods



- Find best line that separates black from white points
- No modeling assumption e.g. that data generated from some point on line + noise

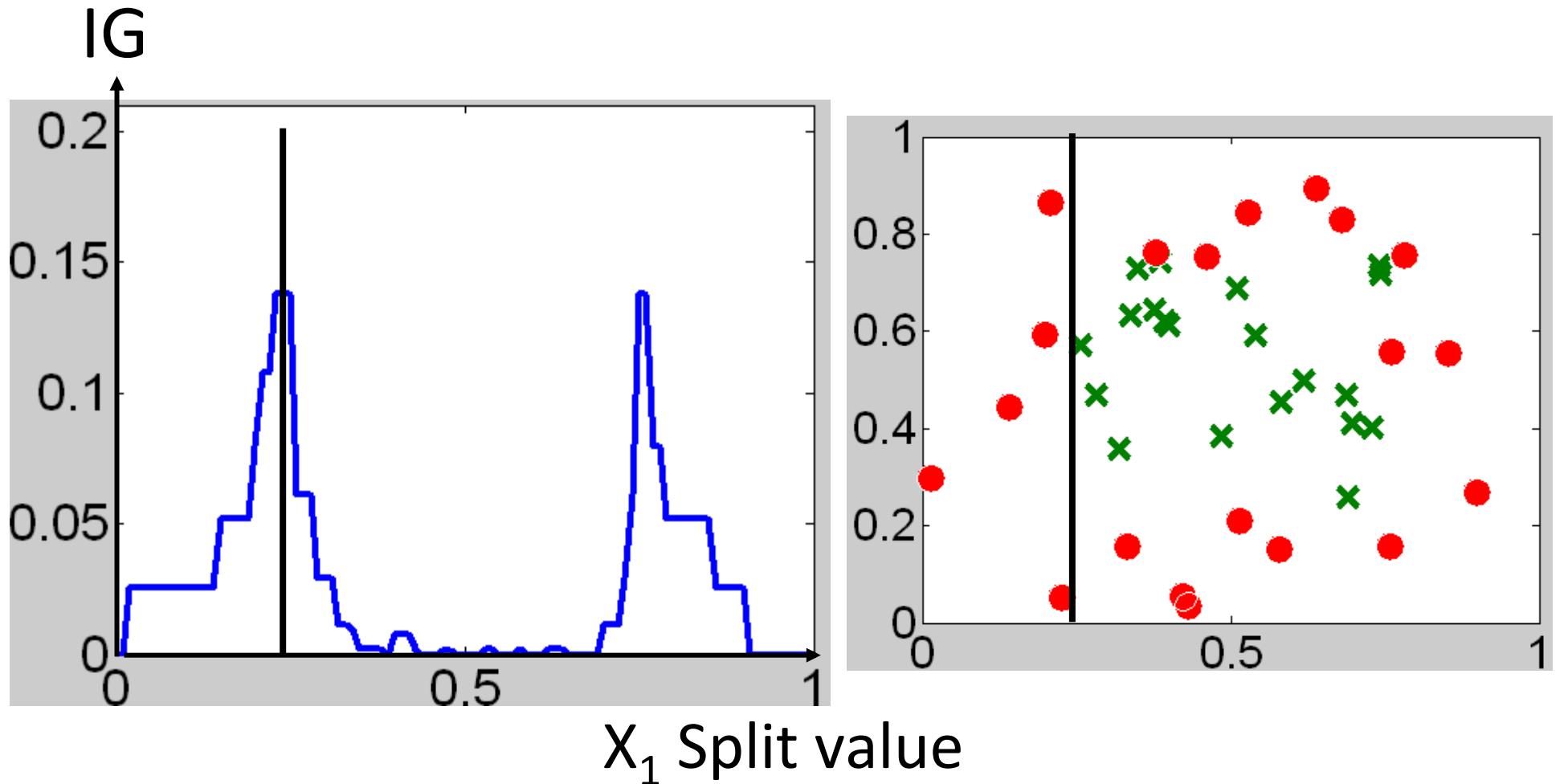


How to Split Complex Data

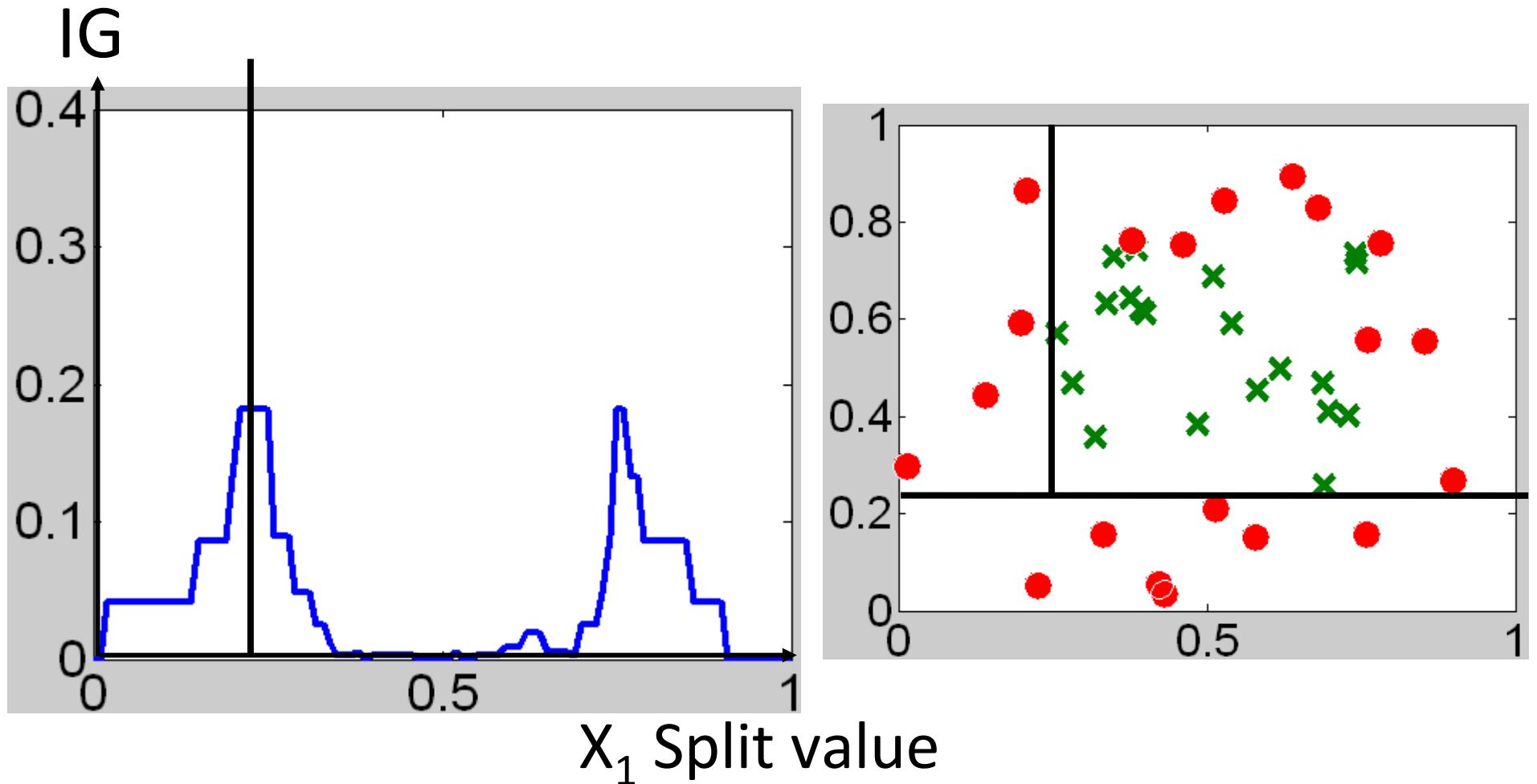


- = 20 training examples from class A
- ✗ = 20 training examples from class B

Attributes = X_1 and X_2 coordinates



Best split value (max Information Gain) for X_1
attribute: 0.24 with $IG = 0.138$



Best split value (max Information Gain) for X_1
attribute: 0.22 with $IG \sim 0.182$

Third Axis: Knowledge/Tasks

- Prediction:
 - Estimate output given input

Prediction Problems

Feature Space \mathcal{X}



Words in a document



Label Space \mathcal{Y}

“Sports”
“News”
“Science”

...

Share Price
“\$ 24.50”

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Prediction - Classification

Feature Space \mathcal{X}

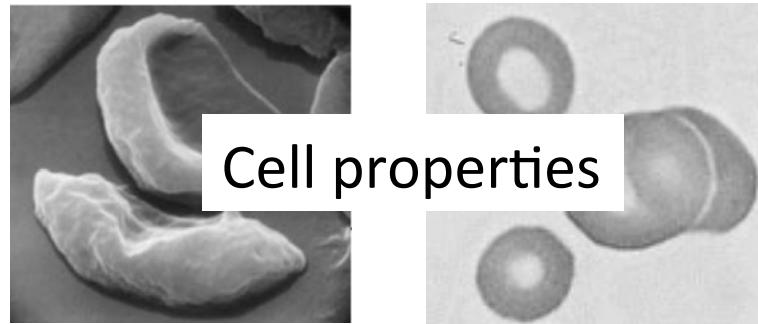


Label Space \mathcal{Y}

“Sports”
“News”
“Science”

...

Cell properties



“Anemic cell”
“Healthy cell”

Discrete Labels

Prediction - Regression

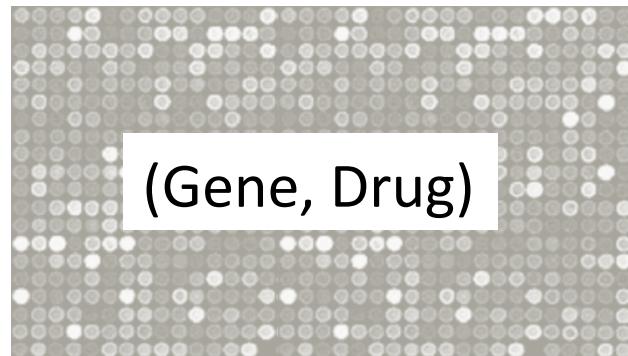
Feature Space \mathcal{X}



Label Space \mathcal{Y}



Share Price
“\$ 24.50”



Expression level
“0.01”

Continuous Labels

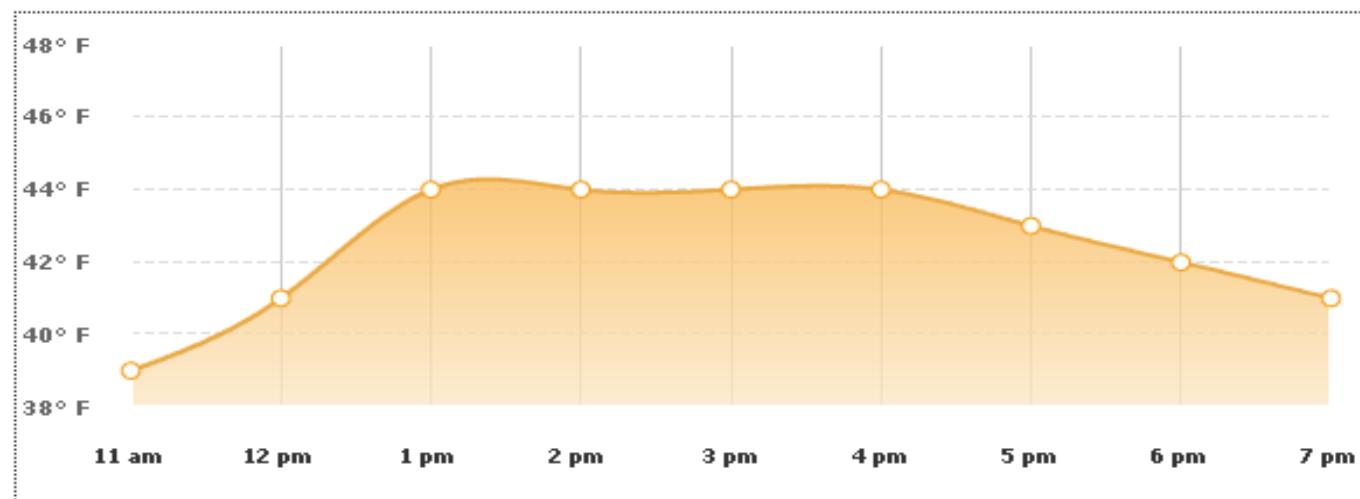
Prediction problems

Features?



Labels?

Classification/Regression?



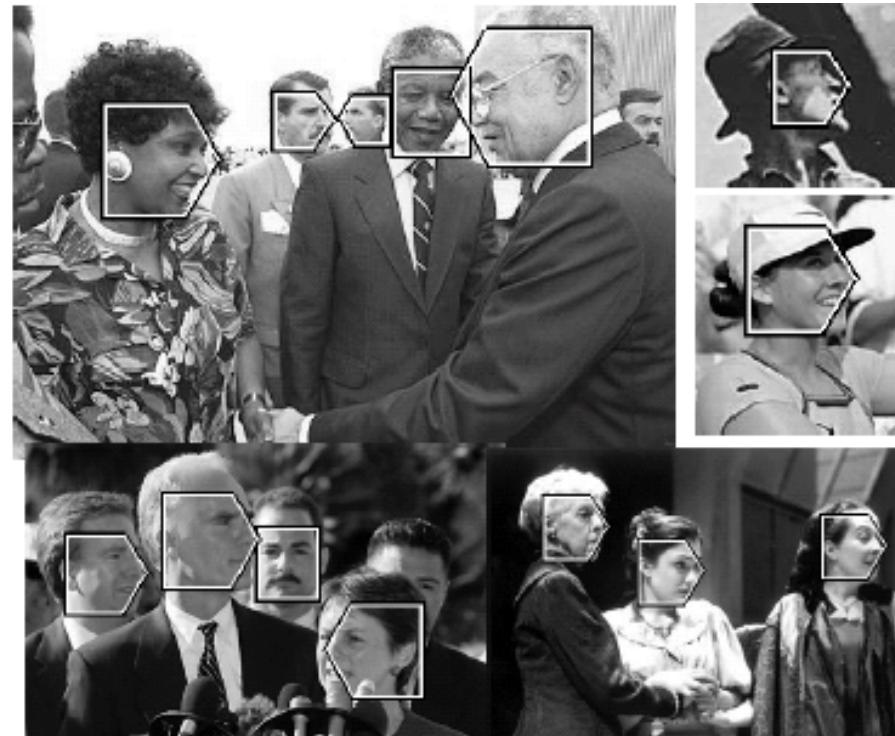
Temperature/Weather prediction

Prediction problems

Features?

Labels?

Classification/Regression?



Face Detection

Prediction problems

Features?

Labels?

Classification/Regression?



Robot Control

Third Axis: Tasks

- Other than prediction problems, another class of tasks are **description** problems
- Examples:
 - Density estimation
 - Clustering
 - Dimensionality reduction
- Also called **unsupervised learning**
 - When first axis (data) consists only of inputs
 - No “supervision” in data as to the descriptive outputs

Unsupervised Learning

Aka “learning without a teacher”

Feature Space \mathcal{X}



Words in a document

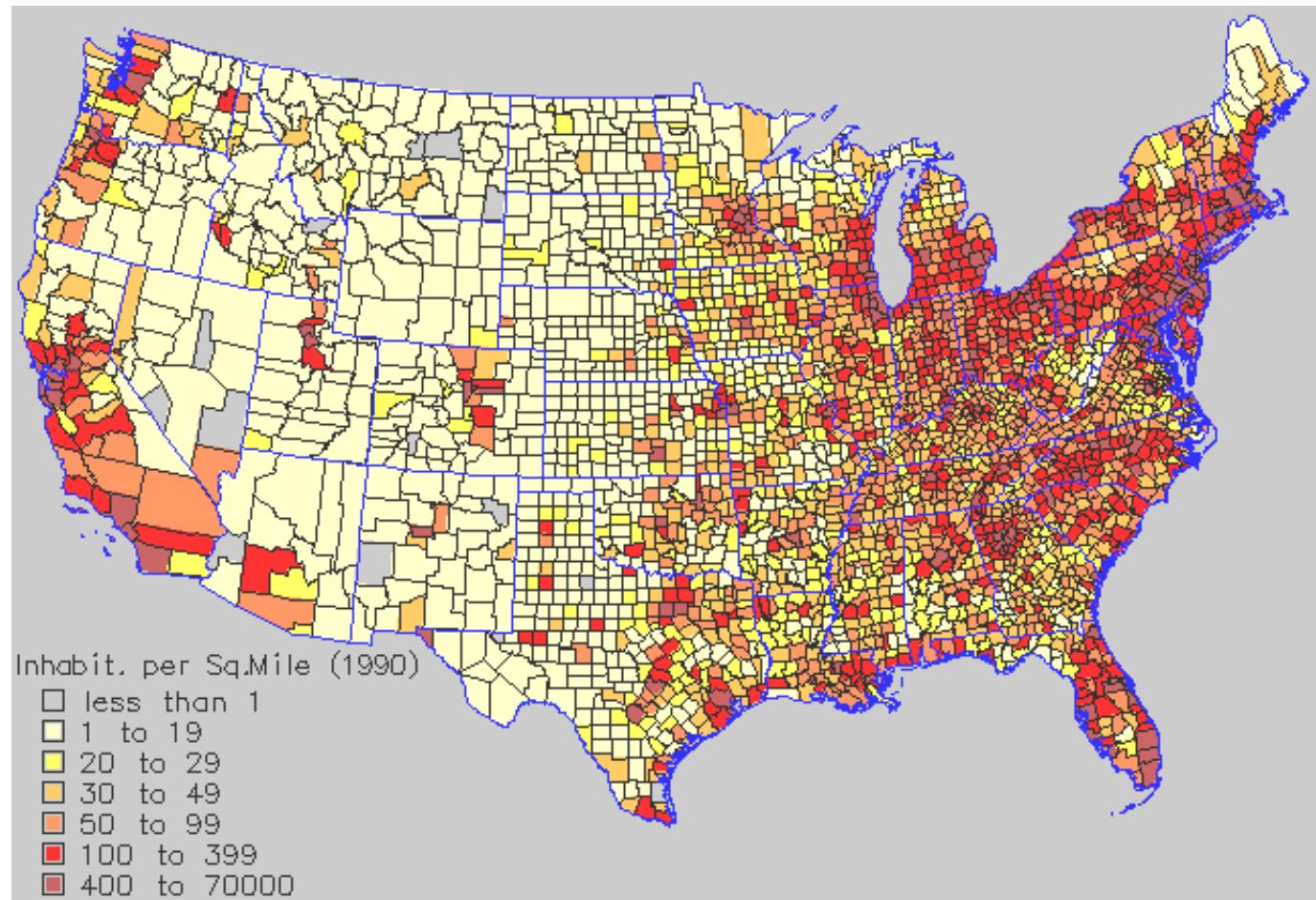


Word distribution
(Probability of a word)

Task: Given $X \in \mathcal{X}$, learn $f(X)$.

Unsupervised Learning – Density Estimation

Population density



Unsupervised Learning – Clustering

Group similar things e.g. images

[Goldberger et al.]



Unsupervised Learning – clustering web search results

The screenshot shows a search results page from Clusty. At the top, there's a navigation bar with links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. A search bar contains the query 'race'. Below the search bar are 'advanced preferences' and a 'Search' button.

The main content area starts with a header: 'Cluster Human contains 8 documents.' On the left, there's a sidebar titled 'clusters' which lists various document categories. Some of these categories are circled in red:

- Car (28)
- Race cars (1)
- 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)

The main list of results includes:

- 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 visible traits (especially skin color, cranial or facial features and hair texture), and self-identified by culture and over time, and are often controversial for scientific as well as social and political reasons. History · McGraw-Hill · 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 sail of **human beings** **Race** and ethnicity in the United States Census, official definitions of "race" used by the US Census Bureau, Genetics. Historical definitions of **race**; **Race** (bearing), the inner and outer rings of a rolling-element bearing. **RACE** · 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 ...
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 · 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. 100, issue 1, pp. 1-12, 1999. 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 ...
www.answers.com/topic/race-1 - [cache] - Live

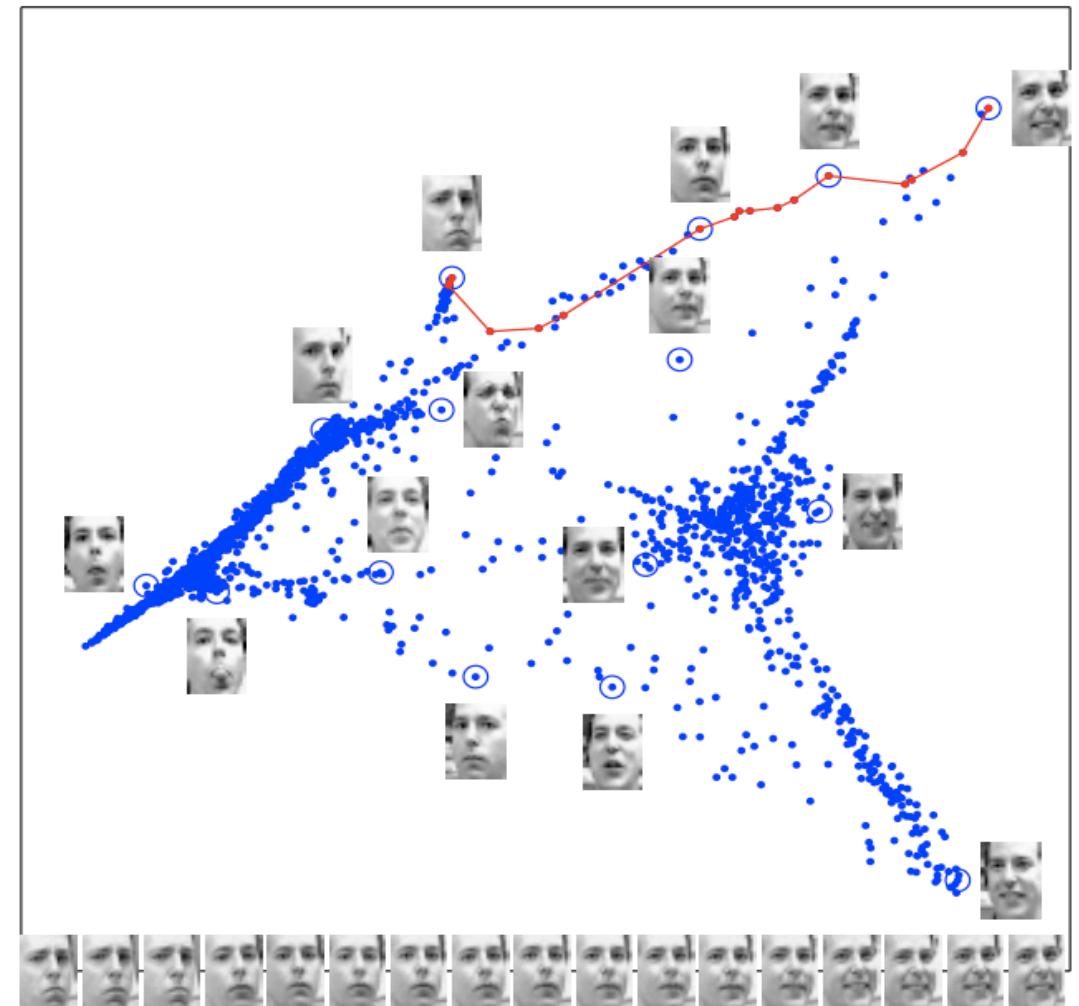
Unsupervised Learning - Embedding

Dimensionality Reduction

[Saul & Roweis '03]

Images have thousands or millions of pixels.

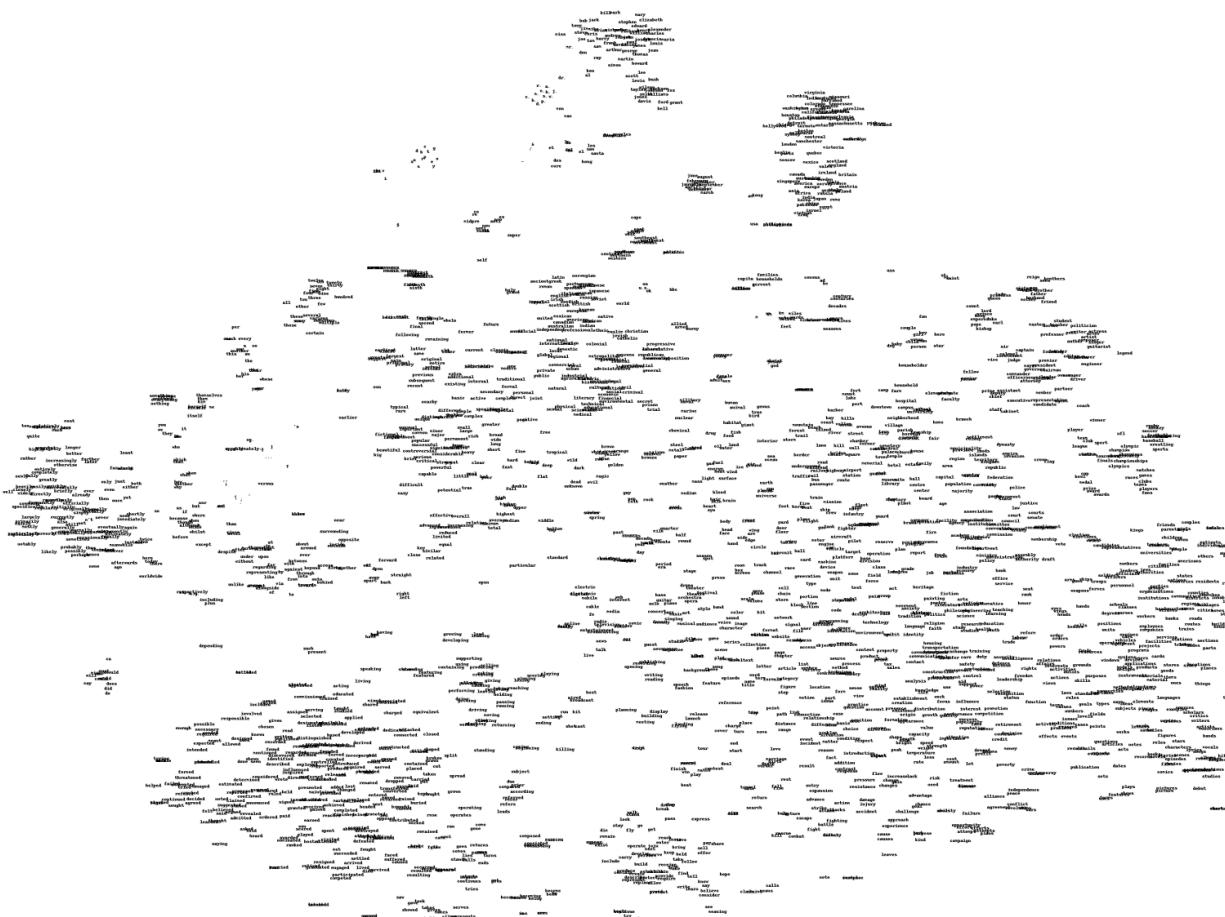
Can we give each image a coordinate,
such that similar images are near each other?



Unsupervised Learning - Embedding

Dimensionality Reduction - words

[Joseph Turian]



Unsupervised Learning - Embedding

billmark mary
bob jack stephen elizabeth
tony edward
miss jimmie richard charles
steve chris andrew william charles
joe tom harry roger john joseph francis maria
mr. sam frank paul james louis
don arthur george jean thomas
ray martin howard
simon

dr. ben lee
al scott lewis bush

r. a.
m. e. h. j.
c. s. w.
b. d. p.
von
van

diagnoses
la los
dal san
el santa
des hong
core

cape
east
south

june august
february september
january september
april september
december march

hkong

usa philippines

virginia
columbia missouri
indiana maryland
colorado tennessee
washington wisconsin
taylor johnson fox
smith williams
jones davis ford grant
bell

hollywood toronto
boston ontario
sydney melbourne
montreal manchester cambridge
london victoria
berlin quebec
moscow mexico scotland
wales england
canada ireland britain
singapore australia sweden
america norway spain
europe austria
asia germany
africa russia netherlands
india japan rome
pakistan egypt
vietnam israel
iraq

Machine Learning Tasks

Broad categories -

- **Prediction tasks**
 - Classification, Regression
- **Description tasks**
 - Typically synonymous with unsupervised learning
 - Density estimation, Clustering, Dimensionality reduction

Machine Learning Subfields

- Topics in machine learning can be categorized by these three axes
 - Data
 - Tasks i.e. type of knowledge we seek
 - Algorithms

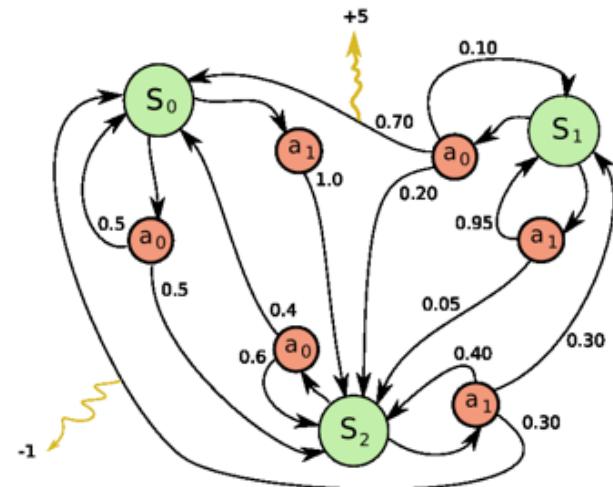
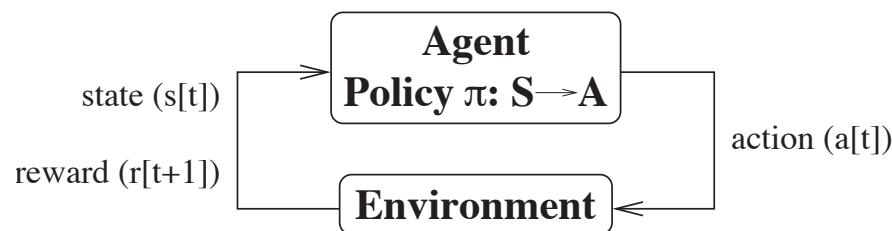
Machine Learning Subfields

- **Particular choices of the three axes**
- **Supervised learning**
 - First axis (data) consists of both inputs and outputs
 - Third axis (tasks) consists of prediction
- **Semi-supervised learning**
 - First axis (data) consists of inputs and only some of them with outputs
 - Third axis (tasks) consists of prediction
- Many more (active learning,)

Some other Machine Learning Subfields

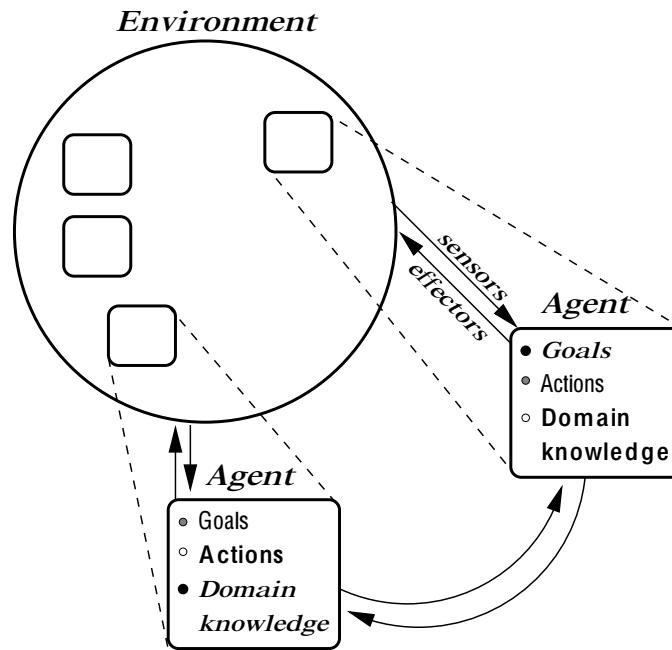
- Reinforcement learning
- Multi-agent systems
 - In addition to observations, we also have “actions” we can take

Reinforcement Learning



- Data consists of rewards that come through taking **actions** that has a feedback i.e. affects future observations
- Task: maximize reward

Multi-agent Systems



- Multiple agents
- Same setup as in reinforcement learning
- But now, the data also consists of other agents' actions

Enjoy!

- ML is becoming ubiquitous in science, engineering and beyond
- This class should give you the basic foundation for applying ML and developing new methods
- The fun and work begin!