10-701: Introduction to

~~Deep Neural Networks~~

Machine Learning

http://www.cs.cmu.edu/~pradeepr/701

Organizational info

- All up-to-date info is on the course web page (follow links from my page).
- Instructors
  - Pradeeep Ravikmur
  - Ziv Bar-Joseph

- TAs: See info on website for recitations, office hours etc.
- See web page for contact info, office hours, etc.
- Piazza would be used for questions / comments and for class quizzes. Make sure you are subscribed.
- We will also use piazza for detemining class participation

Pradeep Ravikumar (pradeepr@cs.cmu.edu)

- Office: GHC 8111
- Office hours: TBD
- Course Instructor

- Research Interests: Statistical Machine Learning

Daniel Bird (dpbird@andrew)

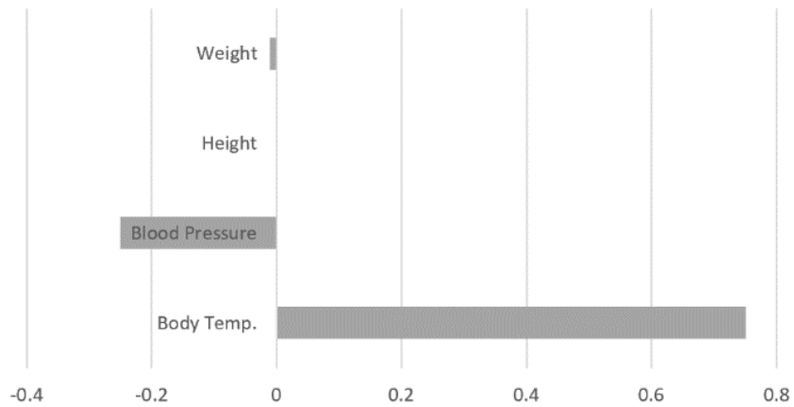
- Office: GHC 8110
- Office hours: Tuesday 9:00am-10:00am
- Education Associate for 10-701 and 10-601A

Shubhranshu Shekhar (shubhras@andrew)

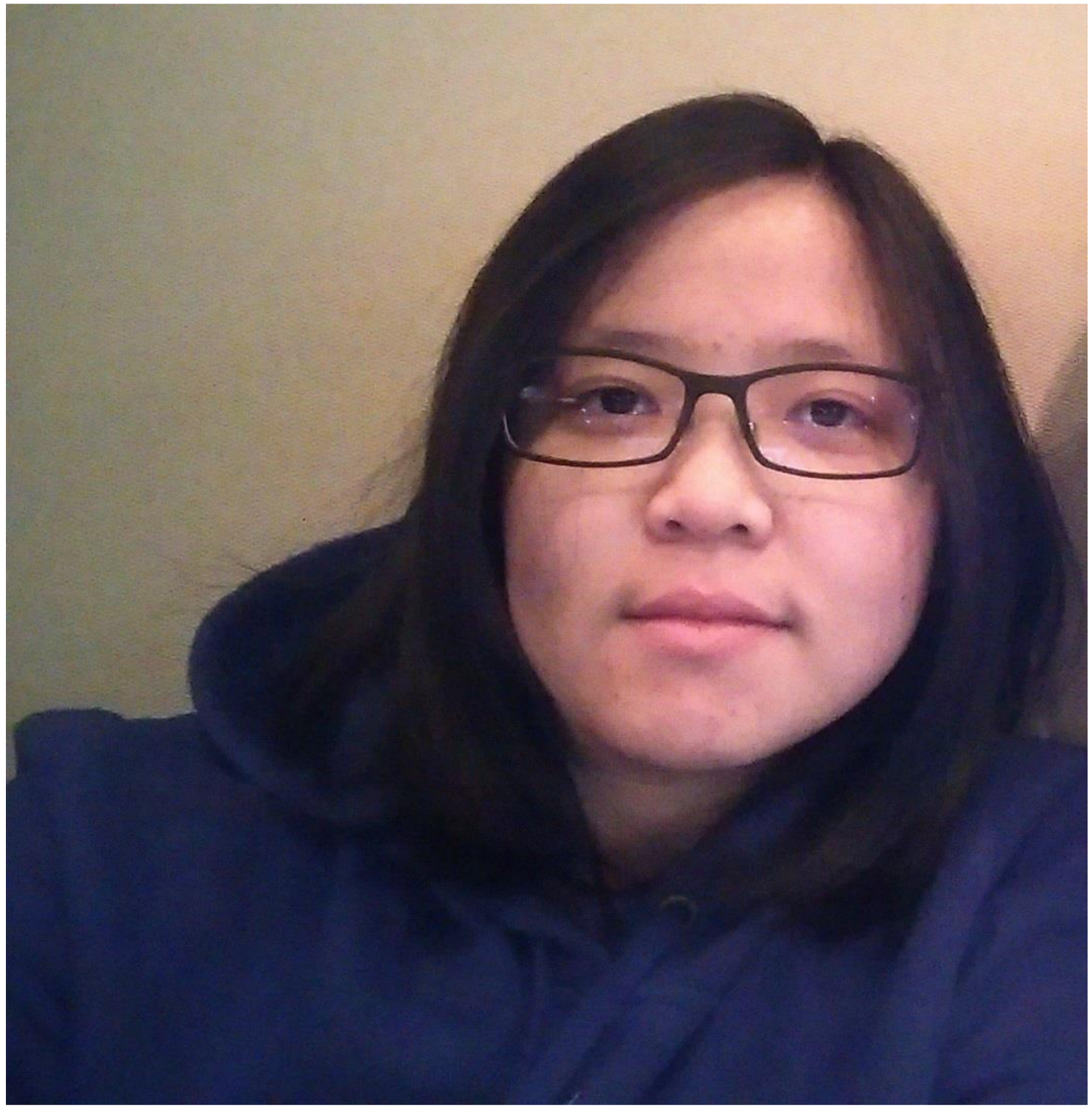
- Office space: HBH 3037
- Office hours: Thursday 10:30am-11:30am
- Research topic: Data Mining
  - Online Social System
  - Anomaly Detection
  - Graph Mining

Umang Bhatt (usb@andrew)

- Office hours: Monday 11:00am-12:00pm
- Research topic: model interpretability
  - Feature attribution
  - Influence functions



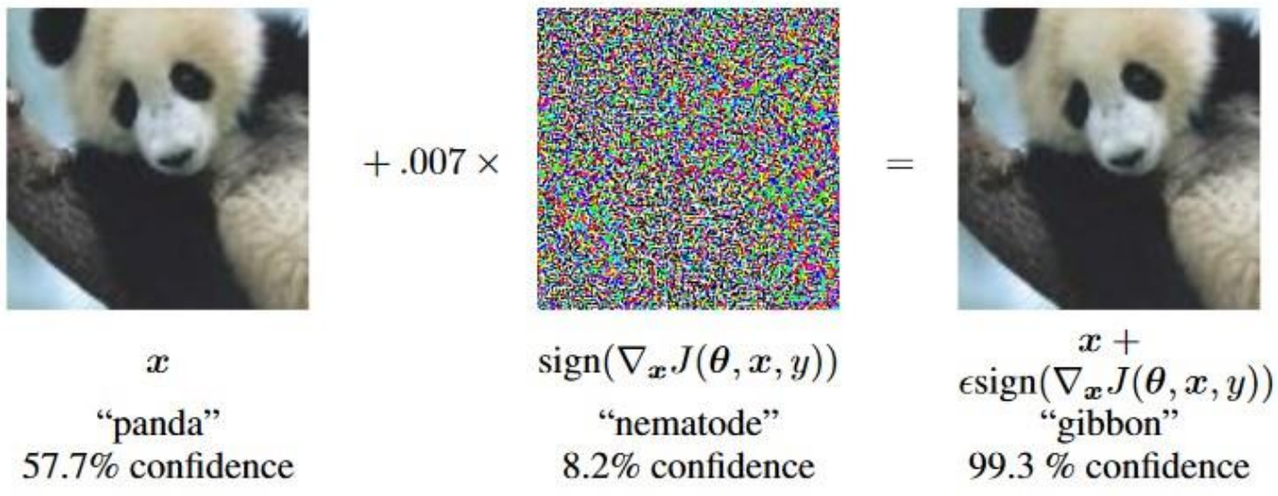
Chieh Lin (chiehl1@cs.cmu.edu)



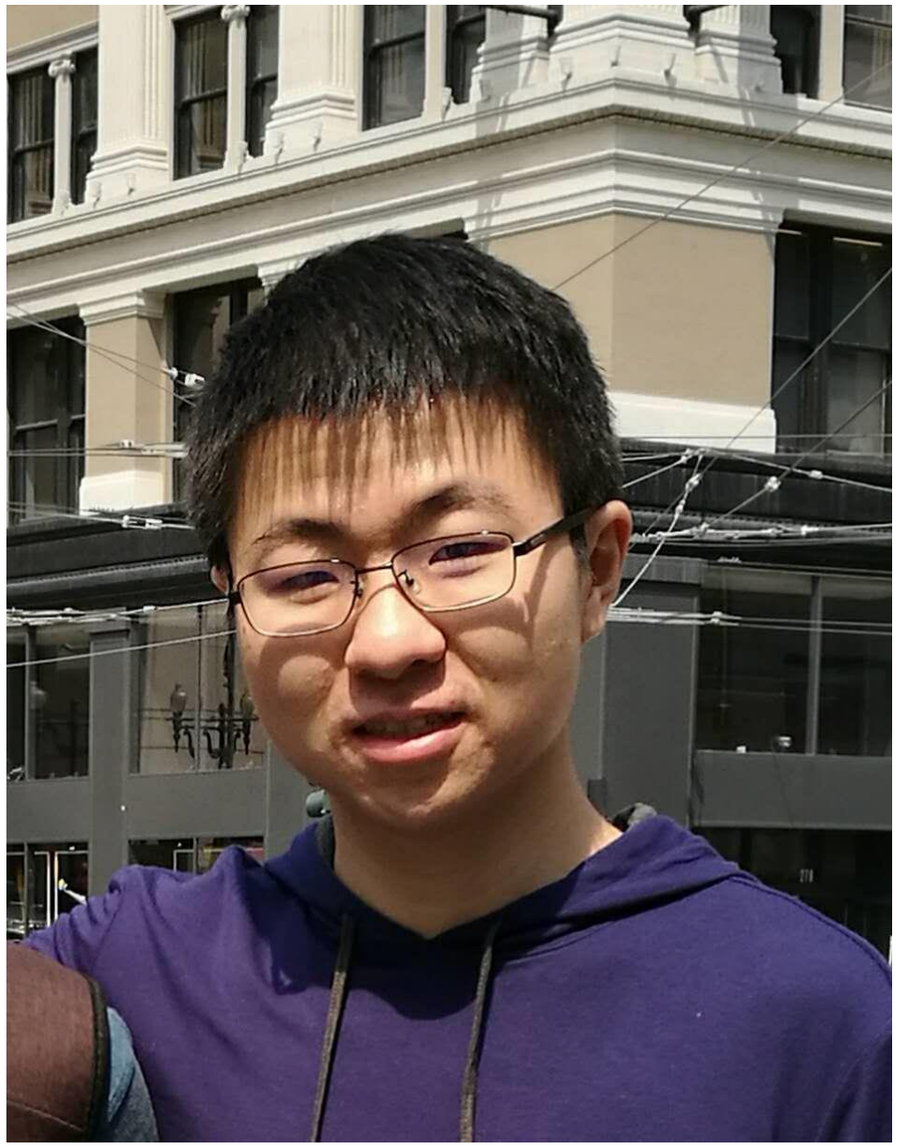
- Office: GHC 8127
- Office Hours: Tuesday 2:00-3:00
- Machine Learning PhD student
- Research in computational biology and graphical models

Sarah Mallepalle (smallepa@andrew)

- Statistics and Machine Learning
- Office Hours: Friday 11:00am-12:00pm
- Research Interests: GANs, VAEs, Adversarial Attacks, Natural Language Processing



Zirui Wang (ziruiw@andrew)

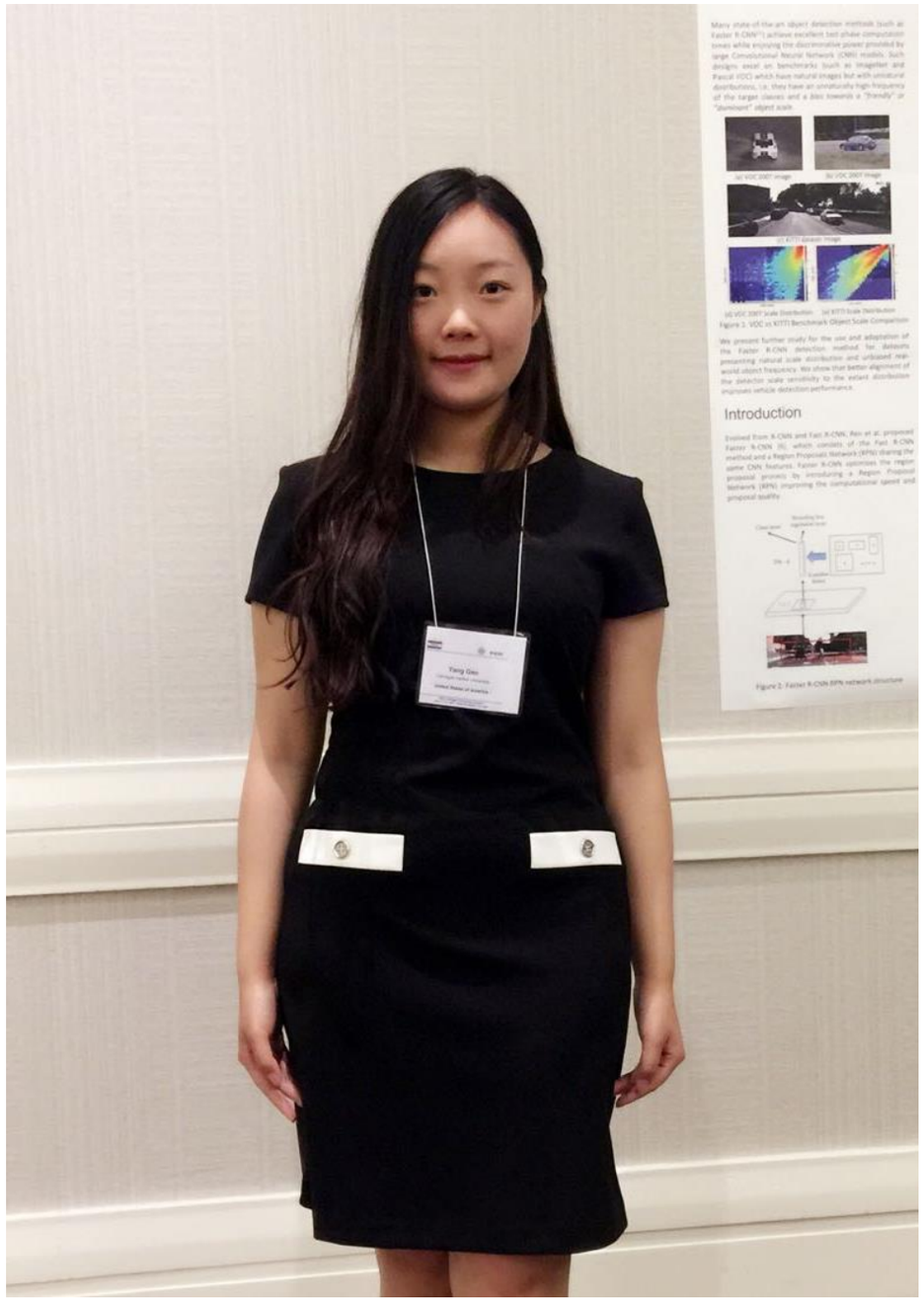


- Office hours: Thursday 3:00pm-4:00pm
- Research topics:
  - Transfer Learning
  - Active Learning
  - Adversarial Learning
  - Deep Learning

Adithya Raghuraman (araghura@andrew)

- Office hours: Wednesday 12:00-1:00
- Research interests:
  - Multimodal machine learning
  - Deep learning

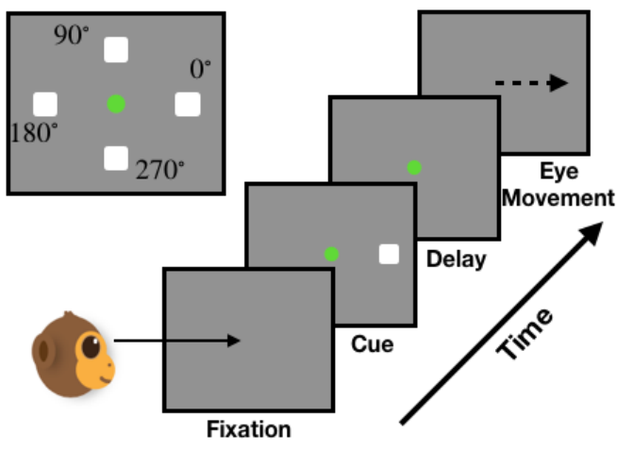
Yang Gao (yanggao@andrew)



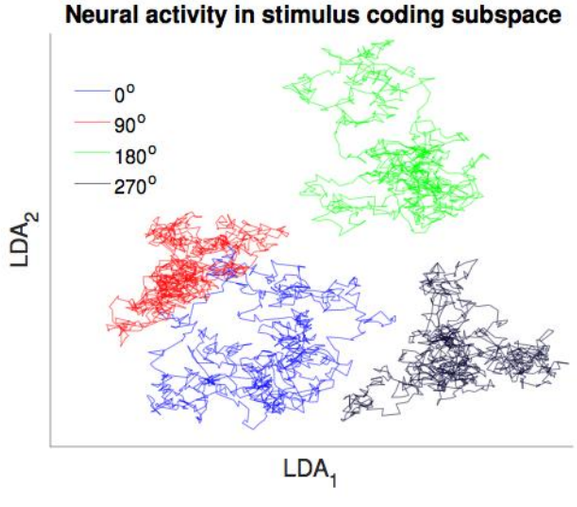
- Office: REH 244
- Office hours: Friday 4:00-5:00

Akash Umakantha (aumakant@andrew)

- **Office: Hamerschlag B200**
- **Office hours: Monday 5:00pm-6:00pm**
- **Research topic: Computational neuroscience**
  - **Visual attention and working memory**
  - **Brain computer interfaces**



Working memory task



Jing Mao (jingmao@andrew)

- Office hours: Wednesday 4:30-5:30
- TA for 10-701
- Research interest: Machine Learning, Natural Language Processing

- 8/27 Intro: Data, Algorithms, Tasks, MLE
- 8/29 Bayesian Estimation, MAP, Decision Theory,
- Model-free, Risk Minimization
- 9/3 – no class, labor day
- 9/5 – Non-parametric Models: K nearest neighbors,
- Kernel regression / problem set 1 out
- 9/10 – no class, jewish new year
- 9/12 – Linear regression
- 9/17– Regularized, Polynomial, Logistic Regression
- 9/19 – Decision Trees / PS1 due, PS2 out
- 9/24 – Naive Bayes, Generative vs Discriminative
- 9/26 – Neural Networks and Deep Learning
- 10/1 – Neural Networks and Deep Learning, I, II
- 10/3 – Support Vector Machines 1/ PS2 due, PS3 out
- 10/8 – SVM2
- 10/10 – Boosting, Surrogate Losses, Ensemble Methods
- 10/15 - Clustering, Kmeans
- 10/17 - Clustering: Mixture of Gaussians, Expectation Maximization / PS3 due
- 10/22 (Monday): exam 1
- 10/24 – Generalization, Model Selection / Project topic selection, PS4 out
- 10/29– Learning Theory: Statistical Guarantees for Empirical Risk Minimization
- 10/31 – Representation Learning: Feature Transformation, Random Features, PCA / PS4 due, PS5 out
- 11/5 – Representation Learning: PCA Contd, ICA / PS4 due, PS5 out
- 11/7 Industry lecture
- 11/12 – Undirected Graphical models (MRF, CRF)
- 11/14 – Directed Graphical Models (incl. latent variable models), I, II
- 11/19 – Directed Graphical Models (incl. latent variable models), I, II / PS 5 due
- 12/05 (Monday): exam 2
- 11/26 – Sequence Models: HMMs
- 11/28 – Sequence Models: State Space Models, other time series models
- 12/10 (Monday): poster presentations

Foundations and Non-Parametric Methods

Prediction, Parametric Methods

Unsupervised Learning

Theoretical considerations

Representation Learning

Graphical and sequence models

Actions

Grading

- **5 Problem sets (5th has a higher weight) - 40%**
- **Exams - 15% each**
- **Project - 25%**
- **Class participation - 5%**

Class assignments

- 5 Problem sets

  - Most containing both theoretical and programming assignments

  - Last problems set: likely shorter

- Project

  - Select from a small list of suggested topics

  - We expect that multiple groups would work on a similar project

  - Groups of 2 to 3

  - Poster session and a short writeup

- Exams

  - 2 exams covering all topics taught in class

  - No final, both exams are during class dates but in the afternoon (5-7pmt)

Recitations

- Twice a week (same content in both)

- Expand on material learned in class, go over problems from previous classes etc.

What is Machine Learning?

Easy part: Machine

Hard part: Learning

- Short answer: Methods that can help generalize information from the observed data so that it can be used to make better decisions in the future

What is Machine Learning?

Machine Learning

- Algorithms that improve their knowledge towards some task with data
- How is it different from Statistics?
  - Same, but with better PR?
  - Statistics + Computation?
- What is its relationship with AI, Data Science, Data Mining?

Machine Learning

- It is useful to differentiate these different fields by their goals
- 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

While there is overlap, there are also differences

- Statistics: the goal is the understanding of the data at hand
- Artificial Intelligence: the goal is to build an intelligent agent
- Data Mining: the goal is to extract patterns from large-scale data
- Data Science: the science encompassing collection, analysis, and interpretation of data

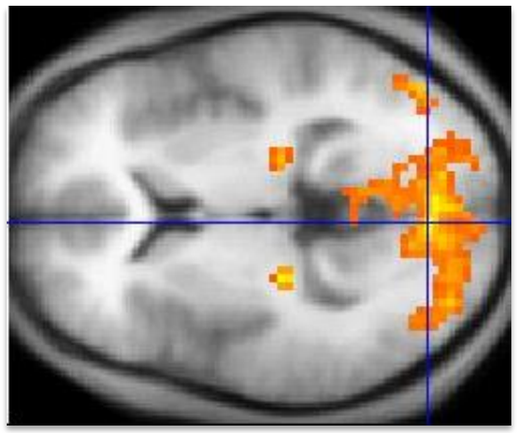
From Data to Understanding …

Machine Learning in Action

Machine Learning in Action

Supervised learning

- 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

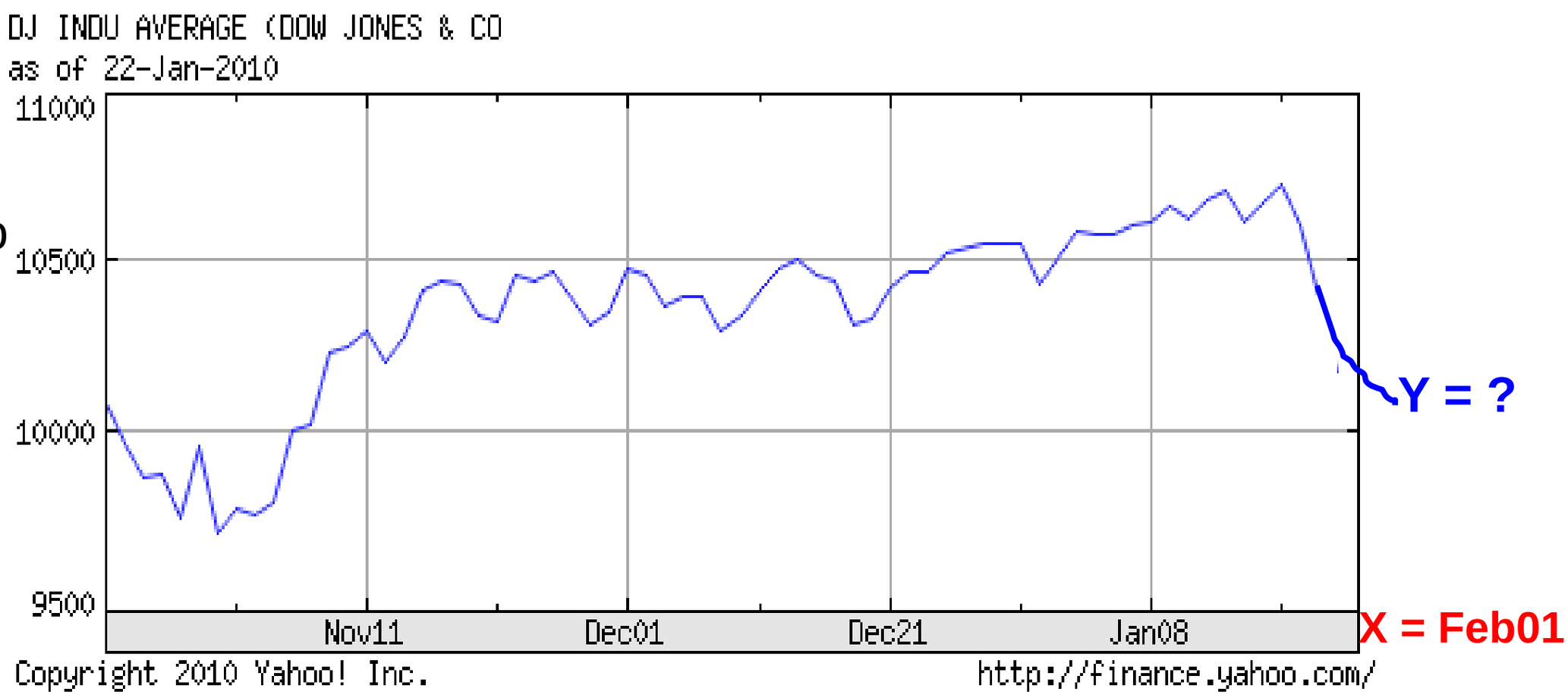
More: Brain Scans | Brain Scan | Disposition | Defendant | Criminal Behavior

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

Supervised and unsupervised learning



- Sto

Machine Learning in Action

Supervised and unsupervised learning

- E
- D

Sports
Science
News

Machine Learning in Action

Supervised learning

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.

- S

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

=== 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

Spam/
Not spam

NELL: Never-Ending Language Learning

Can computers learn to read? We think so. "Read the Web" is a research project that attempts to create a computer system that learns over time to read the web. Since January 2010, our computer system called NELL (Never-Ending Language Learner) has been running continuously, attempting to perform two tasks each day:

- First, it attempts to "read," or extract facts from text found in hundreds of millions of web pages (e.g., `playsInstrument(George_Harrison, guitar)`).
- Second, it attempts to improve its reading competence, so that tomorrow it can extract more facts from the web, more accurately.

semi supervised learning

So far, NELL has accumulated over 50 million candidate beliefs by reading the web, and it is considering these confidence. NELL has high confidence in 3,938,530 of these beliefs — these are displayed on this website. It is not perfect, but NELL is learning. You can track NELL's progress below or @cmunell on Twitter, browse and download its knowledge base, read more about our technical approach, or join the discussion group.

Recently-Learned Facts twitter

Refresh

instance	iteration	date learned	confidence
glass_window_restoration is a household item	1069	03-aug-2017	97.5
bracelets_curb is a kind of clothing	1069	03-aug-2017	90.9
hillsborough_lista_d_attesa_crea_un_gruppo_meetup is a visualizable thing	1069	03-aug-2017	99.1
parison_levitra_viagra_cialis is a drug	1069	03-aug-2017	97.7
the_democratic_daily is a newspaper	1069	03-aug-2017	100.0
barcelona_international_airport is an airport in the city barcelona	1073	22-aug-2017	100.0
john003 has brother james	1073	22-aug-2017	100.0
omaha_world_herald is a newspaper in the city new_york	1073	22-aug-2017	93.8
abc is a company headquartered in the city new_york	1073	22-aug-2017	100.0
arachnids001 is an arthropod as well as mites also is	1073	22-aug-2017	93.8

Machine Learning in Action

Supervised and reinforcement learning

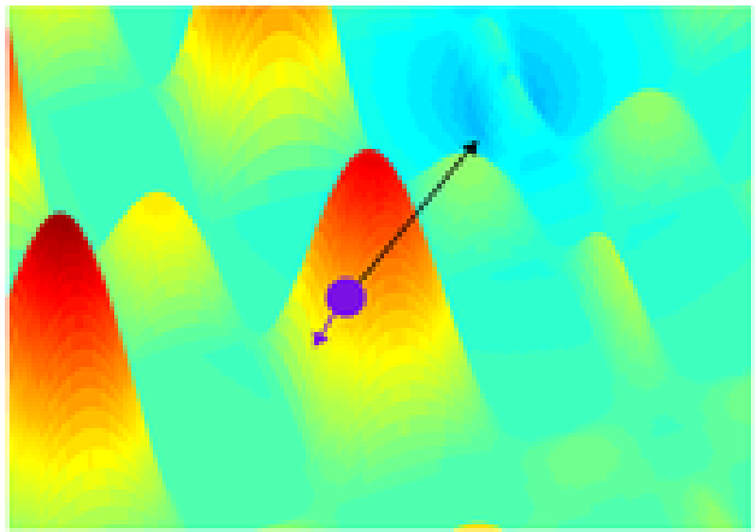
Boss, the self-driving SUV
1st place in the DARPA Urban Challenge.
Photo courtesy of Tartan Racing.

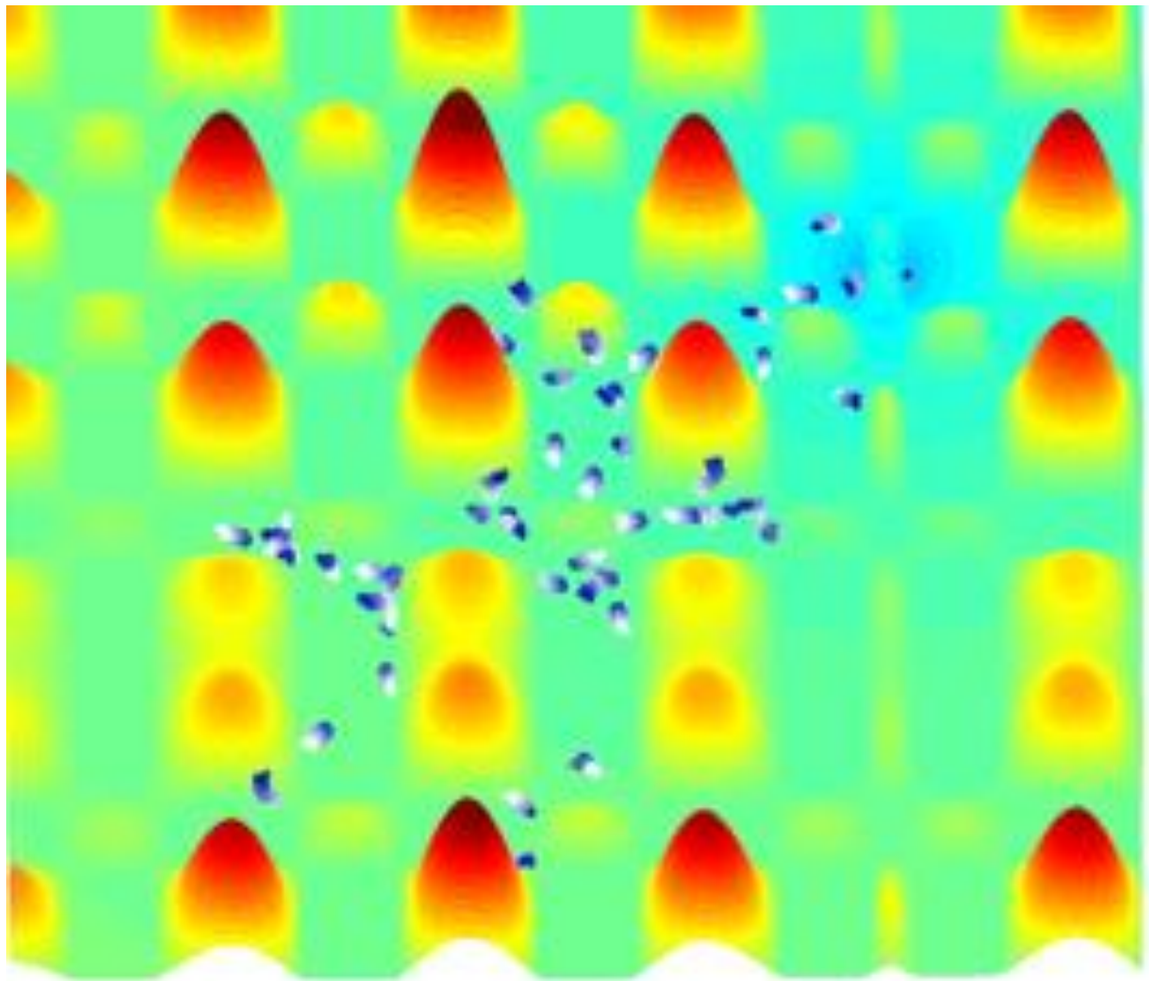
Google translate

Supervised learning (though can also be trained in an unsupervised way)

Distributed gradient descent based on bacterial movement

Reasoning under uncertainty





Biology

Which part is the gene?

Supervised and unsupervised learning (can also use active learning)

Machine Learning in Action

- Many, many more…

  Speech recognition, Natural language processing
  Computer vision
  Web forensics
  Medical outcomes analysis
  Robotics
  Sensor networks
  Social networks
  …

ML has a wide rech

- 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
  - 20,000 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

Three axes of ML

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

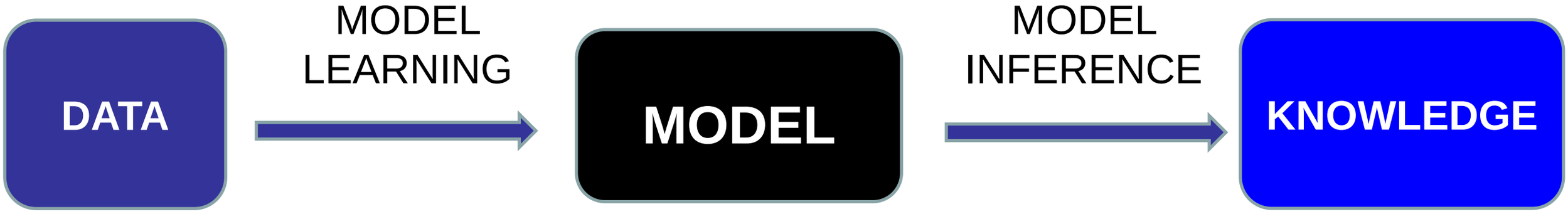
First Axis: Data

- 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

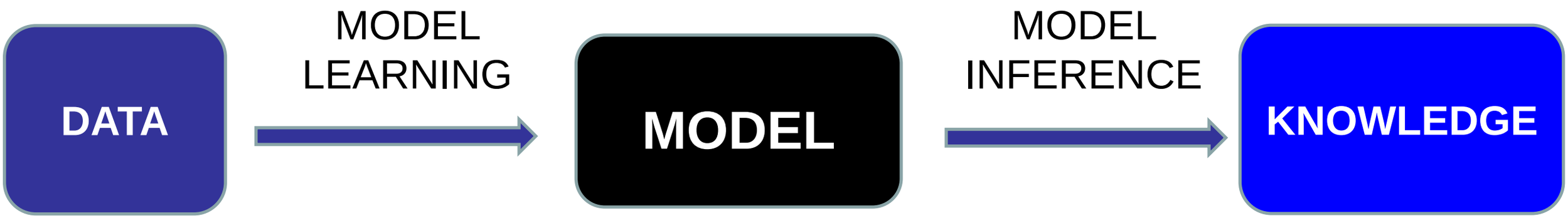
Second Axis: Algorithms

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

Model-based ML



Model-based ML

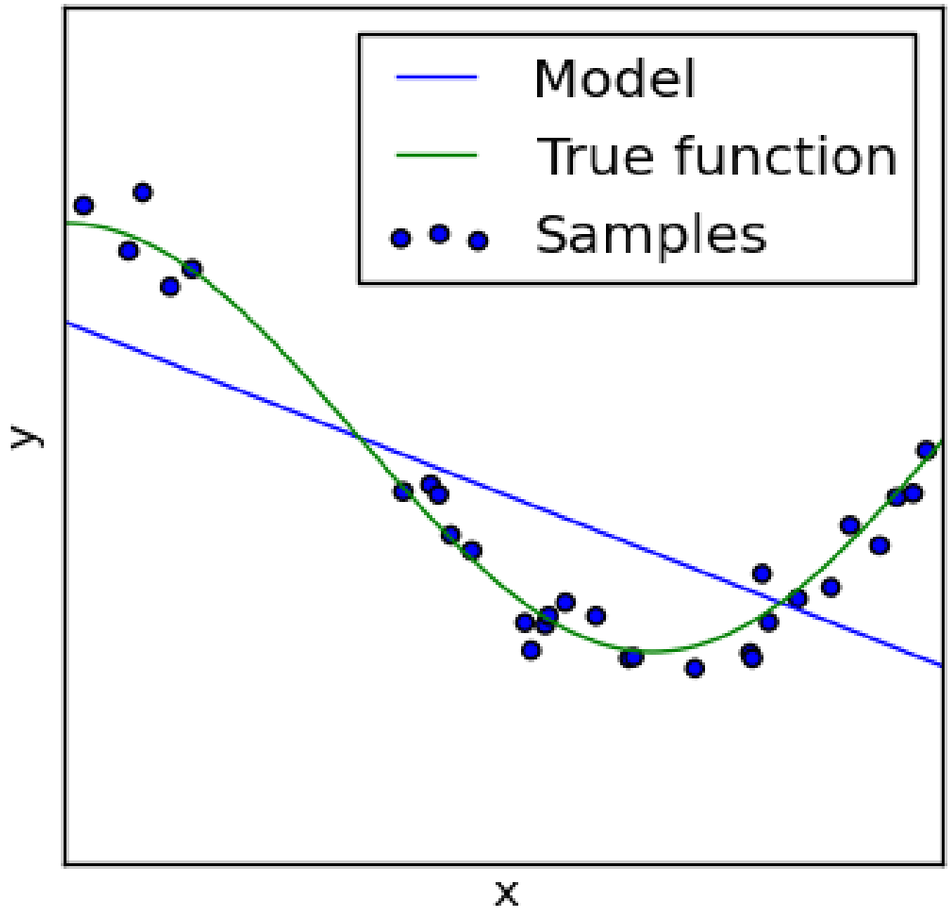


- Learning: From data to model
  - A model thus is a summary of the data
  - But can also reflect of how the data was generated
  - Could thus be used to describe how future data can be generated
  - **E.g. given (symptoms, diseases) data, a model explains how symptoms and diseases are related**

- 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

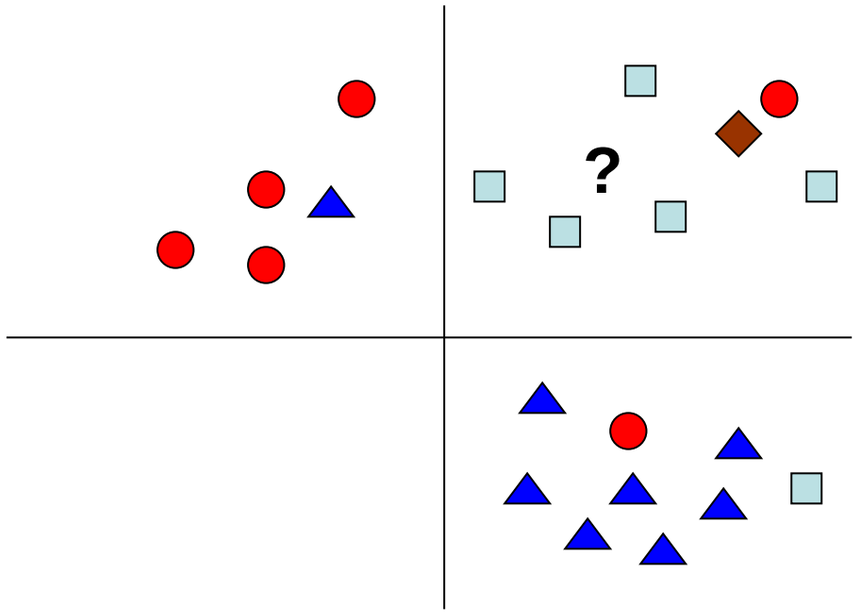


Fitting a simple line (2 params)
to a bunch of one-dim. samples

Model: data = point on line + noise

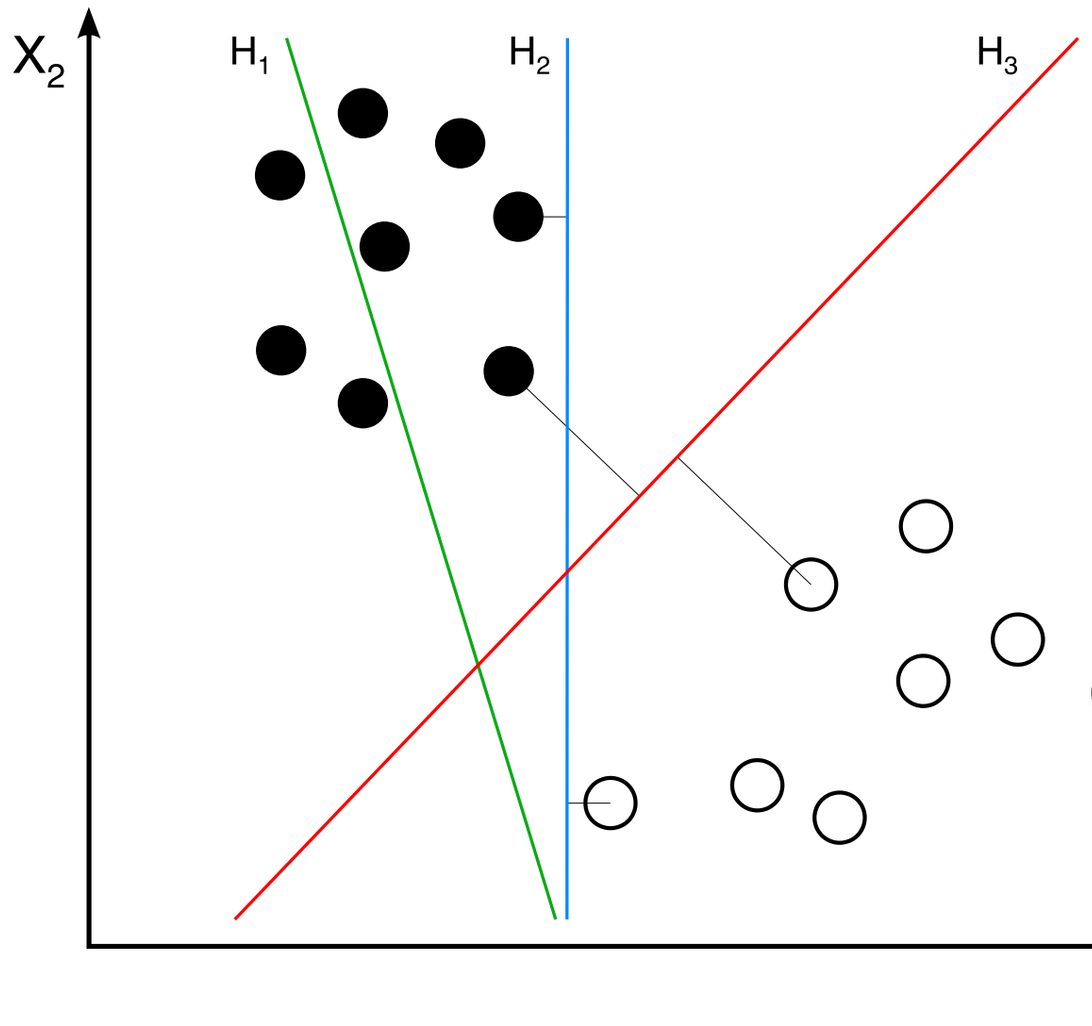
Nonparametric Models

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



- What is the class of the ? Input
- Can use the other points (k nearest neighbors) but the number of points to search scales with the input data

Discriminative models



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

Third Axis: Knowledge/Tasks

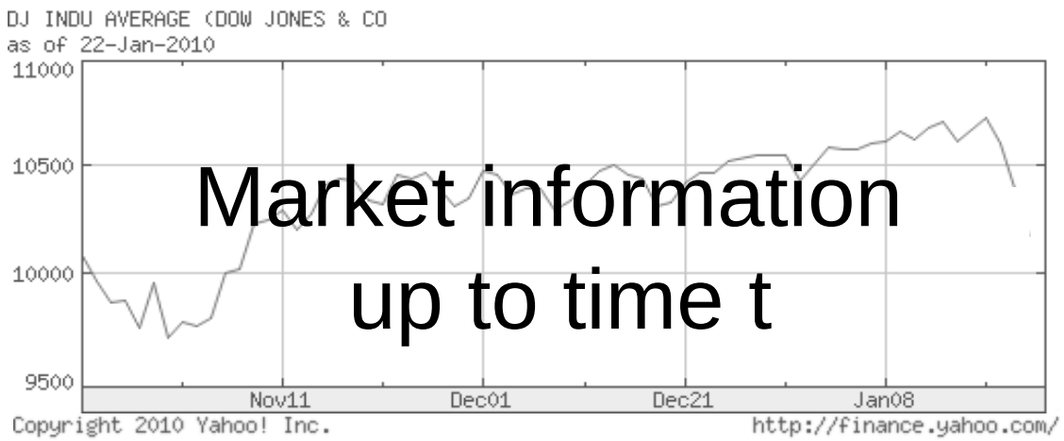
- Prediction:
  - Estimate output given input

Prediction Problems

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$

“Sports”
“News”
“Science”
…

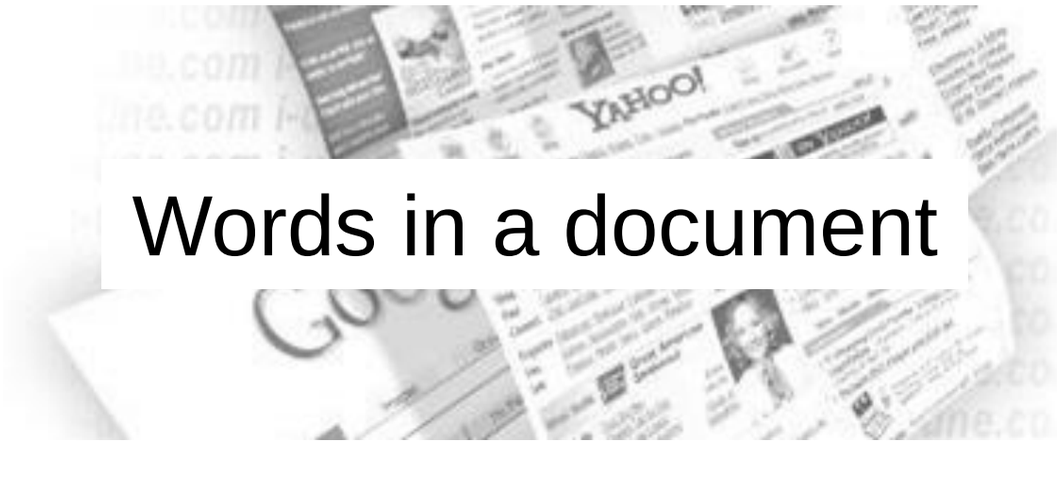


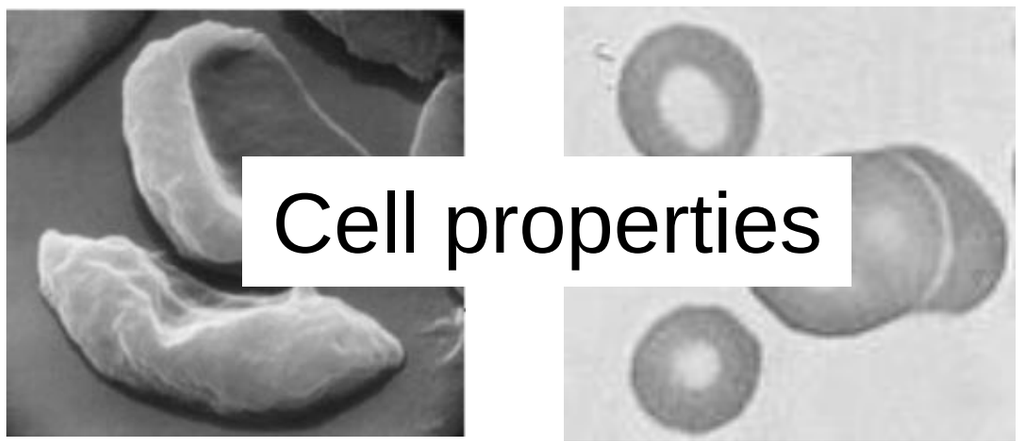
Share Price
“$ 24.50”

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

Prediction - Classification

Feature Space $\mathcal{X}$





Label Space $\mathcal{Y}$

“Sports”
“News”
“Science”
…

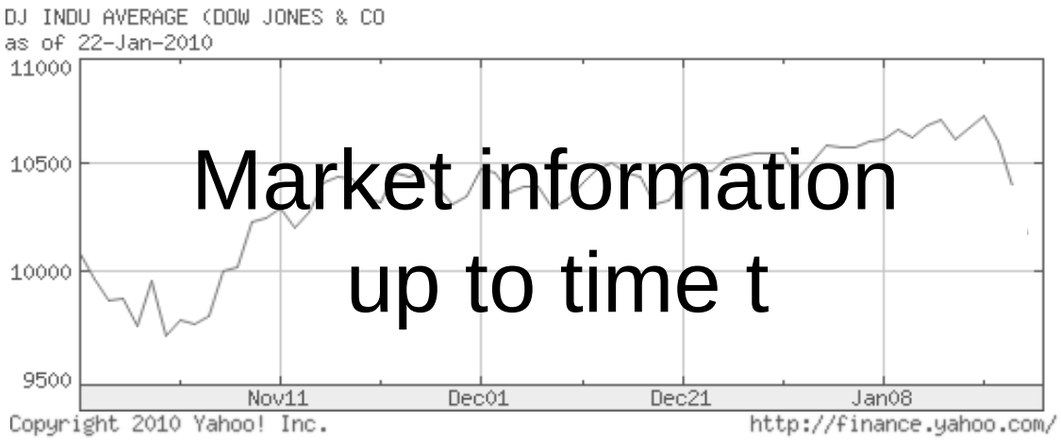
“Anemic cell”
“Healthy cell”

Discrete Labels

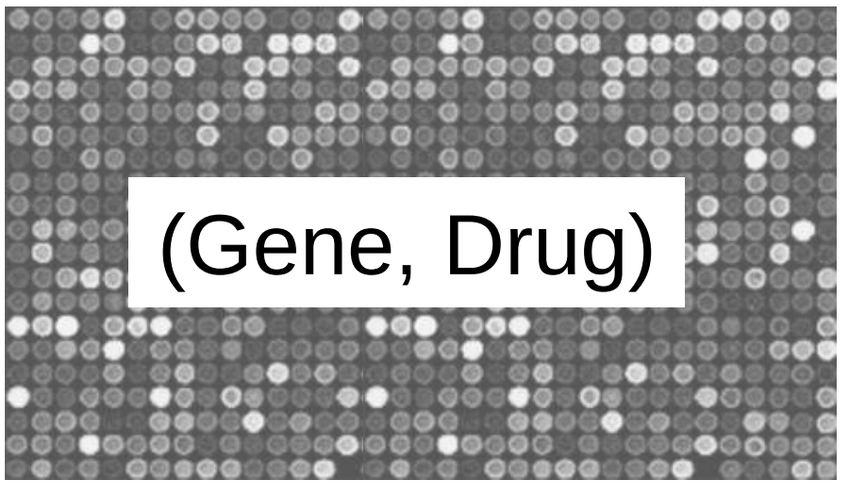
Prediction - Regression

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$



Market information up to time t



(Gene, Drug)

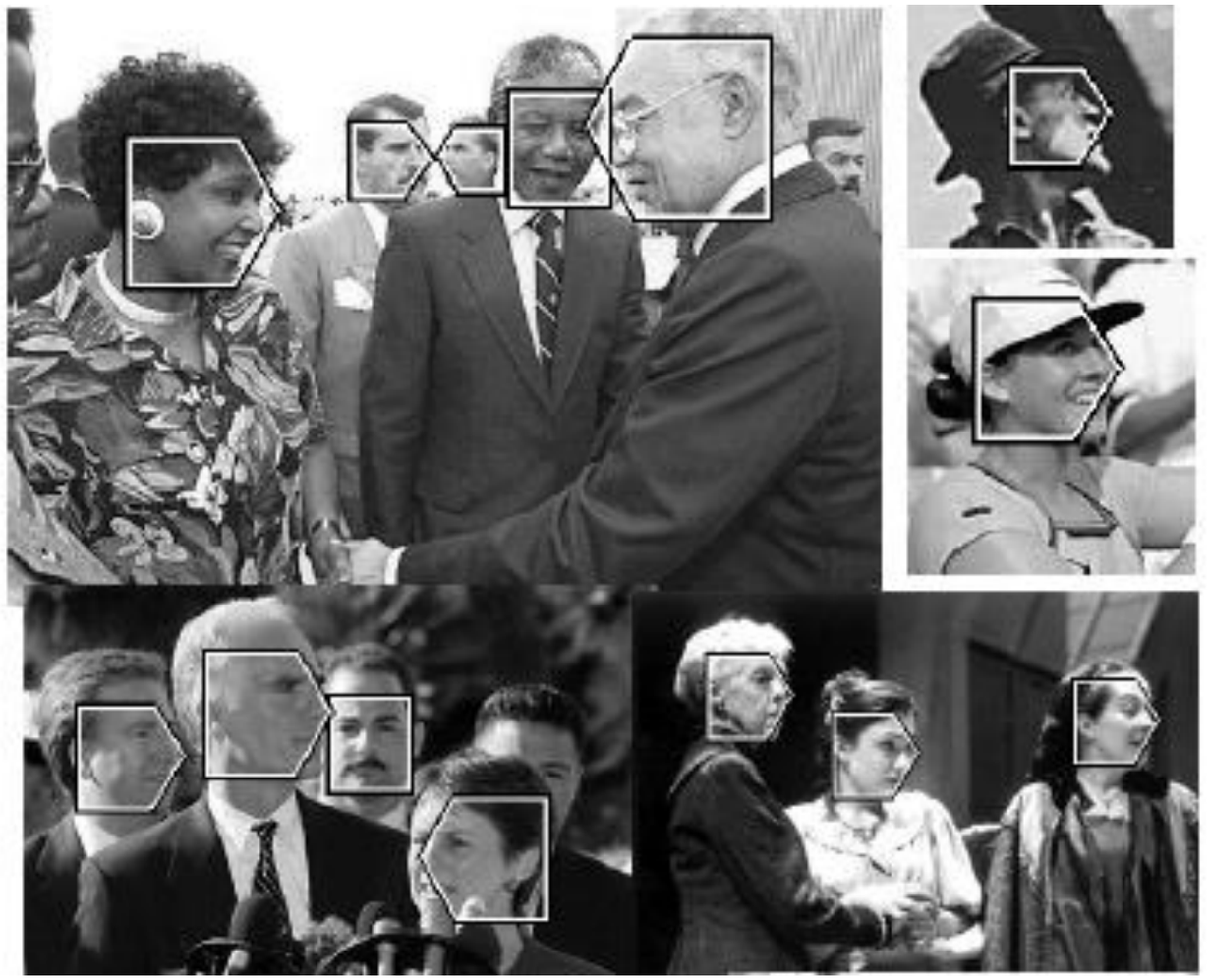
Share Price
"$ 24.577"

Expression level
"6.88"

Continuous Labels

Prediction problems

Features? Labels? Classification/Regression?



Face Detection

Prediction problems

Features? Labels? Classification/Regression?

Robotic 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"

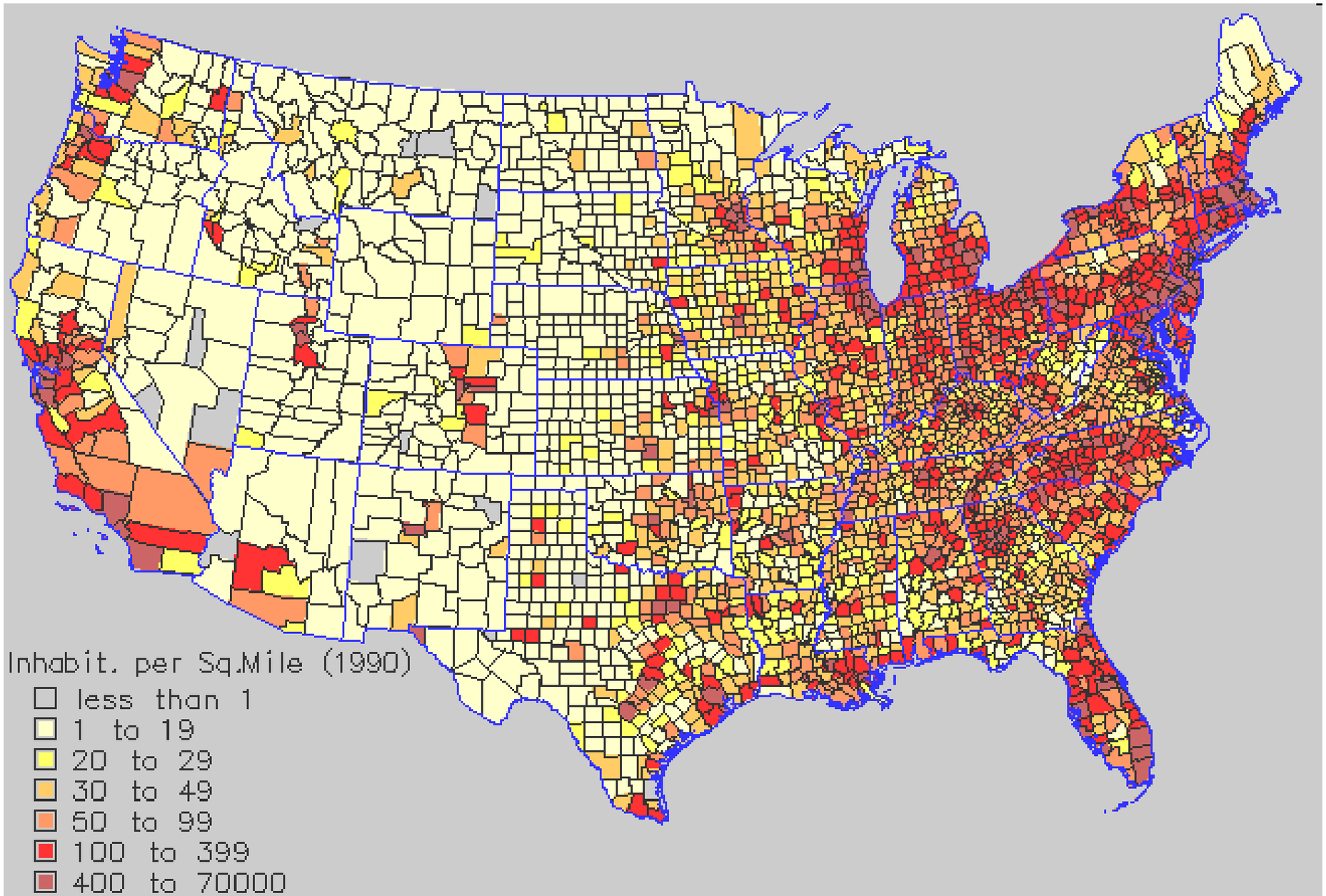
Feature Space $\mathcal{X}$

Word distribution
(Probability of a word)

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

Unsupervised Learning – Density Estimation

Population density



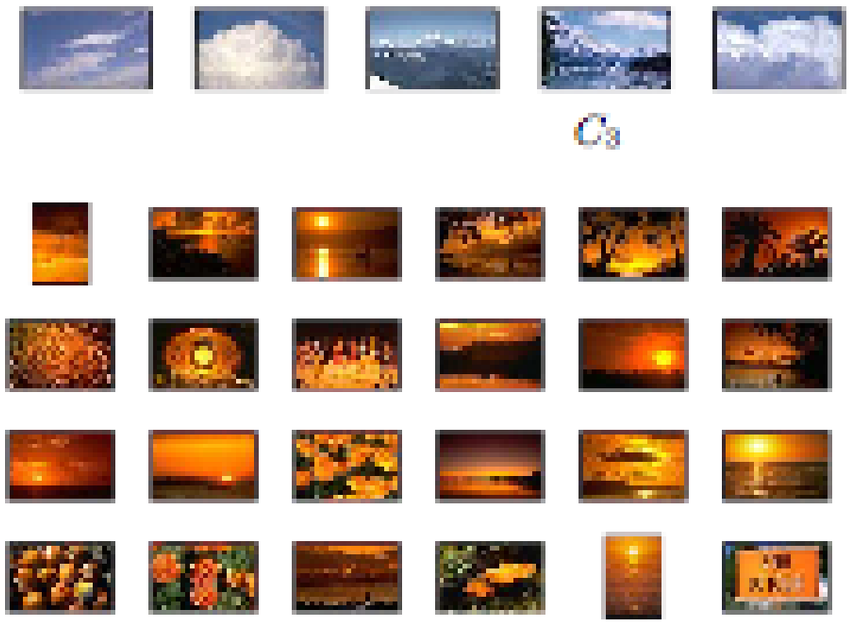
Unsupervised Learning – Clustering

Group similar things e.g. images

[Goldberger et al.]

C_1

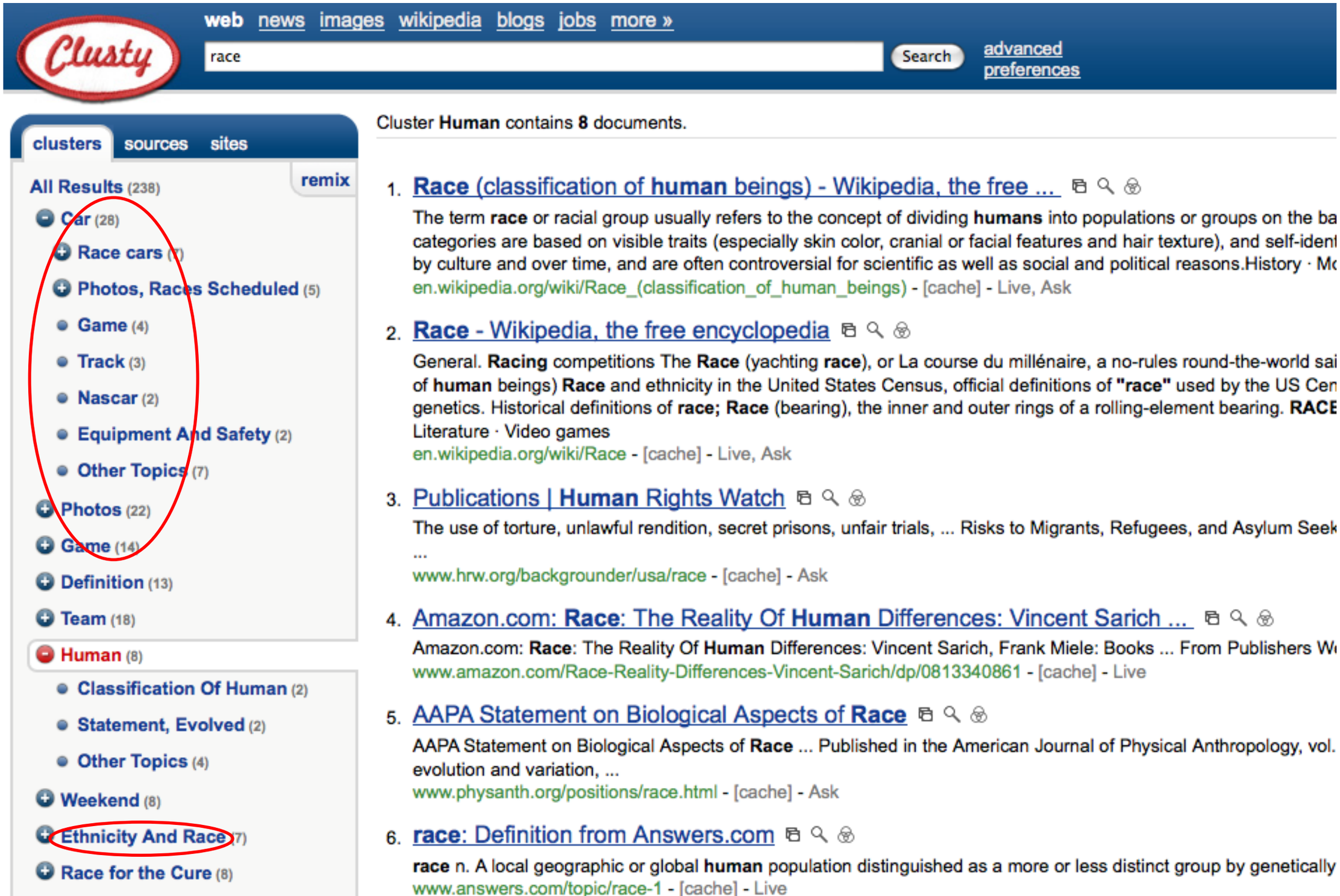
C_2



C_4

C_5

Unsupervised Learning – clustering web search results



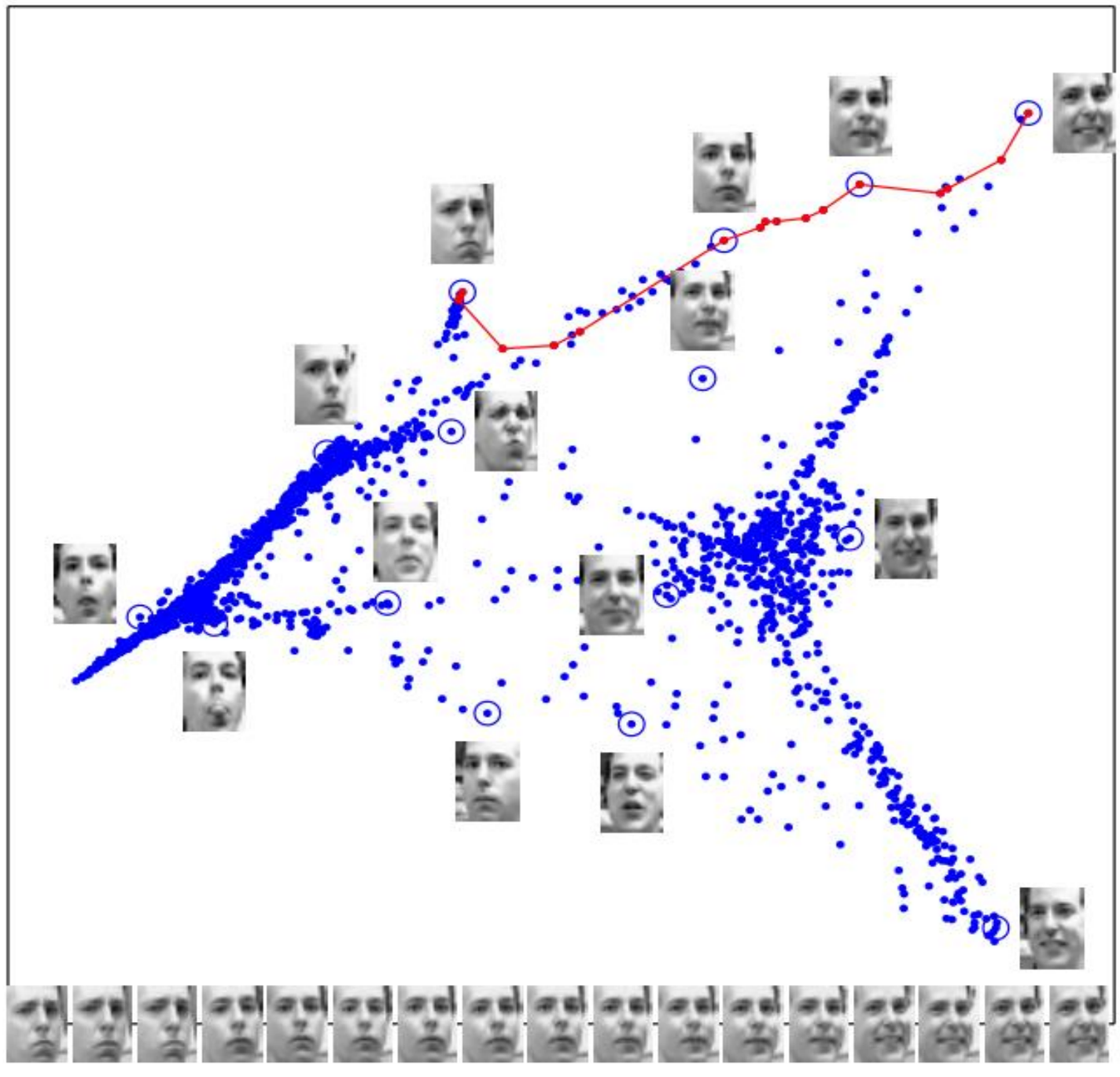
Unsupervised Learning - Embedding

Dimensionality Reduction

Images have thousands or millions of pixels.

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

[Saul & Roweis '03]



Summary: ML tasks

Supervised learning

- Given a set of features and labels learn a model that will predict a label to a new feature set

- Unsupervised learning
  - Discover patterns in data

- Reasoning under uncertainty
  - Determine a model of the world either from samples or as you go along

- Active learning
  - Select not only model but also which examples to use

A bit more formal …

- Supervised learning
  - Given $D = \{X_i, Y_i\}$ learn a model (or function) $F: X_k \rightarrow Y_k$

- Unsupervised learning
  Given $D = \{X_i\}$ group the data into Y classes using a model (or function) $F: X_i \rightarrow Y_j$

- Reinforcement learning (reasoning under uncertainty)
  Given D = {environment, actions, rewards} learn a policy and utility functions:

  policy: $F1: \{e,r\} \rightarrow a$
  utility: $F2: \{a,e\} \rightarrow R$

- Active learning
  - Given $D = \{X_i, Y_i\}$, $\{X_j\}$ learn a function $F1 : \{X_j\} \rightarrow x_k$ *to* maximize the success of the supervised learning function $F2: \{X_i , x_k\} \rightarrow Y$

Common Themes

- Mathematical framework

  - Well defined concepts based on explicit assumptions

- Representation

  - How do we encode text? Images?

- Model selection

  - Which model should we use? How complex should it be?

- Use of prior knowledge

  - How do we encode our beliefs? How much can we assume?