

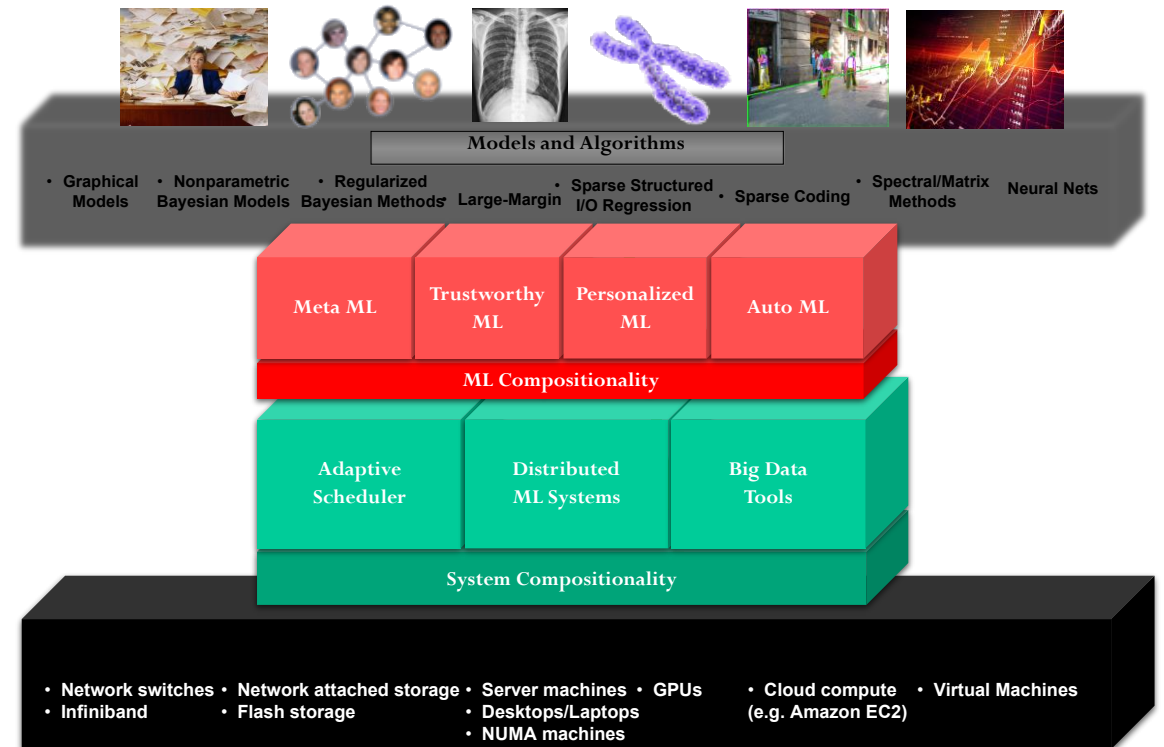
10-701: Introduction to
~~Deep Neural Networks~~
Machine Learning

Organizational info

- All up-to-date info is on piazza.
- Instructors
 - Ziv Bar-Joseph
 - Eric Xing
- TAs: See info on piazza for recitations, office hours etc.
- See also piazza 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 determining class participation

Eric Xing

(epxing@cs.cmu.edu)



- Research Interests
 - Machine Learning: Theory & System
 - Healthcare and other Applications
 - Way to Learn: Auto, Trustworthy, Personalizable, and Transferable ML

Daniel Bird (dpbird@andrew.cmu.edu)



Education Associate for 10-701

Please email me if you have any issues in the course!

Roger Iyengar (raiyinga@andrew.cmu.edu)



PhD in Computer Science Department

Interests: Edge Computing, Wearable Cognitive Assistance, Distributed Systems

Research Interests in ML: Computer Vision, Natural Language Processing

Abhi Adduri (aadduri@andrew.cmu.edu)

PhD in Computational Biology



Clay Yoo (hyungony@andrew.cmu.edu)

Masters in Language Technology Institute

Area of interest: Natural Language Processing, Data Visualization, Model Interpretability



John Grace (jmgrace@andrew.cmu.edu)



Masters in Computer Science Department

Area of interest: Parallel Computing and Automated Program Synthesis.

Chandreyee Bhaumik

(cbhaumik@andrew.cmu.edu)



Masters in the Robotics Institute

Area of Interest: Reinforcement Learning, Computer Vision

Bhuvan Agrawal (bhuvana@andrew.cmu.edu)



Masters in Computer Science Department

Jie Jiao (jiejiao@andrew.cmu.edu)



Undergraduate in Computer Science Department

Area of interest: Reinforcement Learning, Natural Language Processing

8/31 Intro, Three Axes of ML: Data, Algorithms, Tasks, Intro to probability

9/2 Bayesian Estimation, MAP, MLE

9/7 – no class, labor day

9/9 – Decision Theory, Risk Minimization, K nearest neighbors

9/14 – Naive Bayes, Generative vs Discriminative

9/16 – Decision Trees

9/21 - Linear regression

9/23 - Logistic Regression

9/28 – No class, Yom Kippur

9/30 – Support Vector Machines 1

10/05 – SVM2

10/07 – Neural Networks and Deep Learning

10/12 – Neural Networks and Deep Learning II

10/14 – Boosting, Surrogate Losses, Ensemble Methods

10/19 - Clustering, Kmeans

10/21 - Clustering: Mixture of Gaussians, Expectation Maximization

10/26– Representation Learning: Feature Transformation, Random Features, PCA

10/28 – Representation Learning: PCA Contd, ICA/ project proposals due

11/02 - Graphical Models (Bayesian Networks)

11/04 - Graphical Models (BN2) 11/09 - Sequence Models: HMMs

11/11 - Sequence Models: State Space Models, other time series models

11/16– Learning Theory: Statistical Guarantees for Empirical Risk Minimization

11/18 – Generalization, Model Selection

11/23 - Exam **11/23 (Wednesday): Exam**

11/25 – No class, Thanksgiving break

11/30– Industry lecture

12/02 – Reinforcement Learning

12/07– Reinforcement Learning 2

12/09 - Project **12/09 (Wednesday) Poster presentations**

Foundations and Non-Parametric Methods

Prediction, Parametric Methods

Unsupervised Learning

Graphical and sequence models

Theoretical considerations

Actions

Grading

- **5 Problem sets - 40%**
- **Exam - 30%**
- **Project - 30%**

Class assignments

- 5 Problem sets
 - Both theoretical and programming assignments
- Project
 - Select from a small list of suggested topics
 - We expect that multiple groups would work on a similar project
 - Groups of 3
 - Poster session (recorded) and a short writeup
- Exams
 - A single exams covering all topics taught in class up to the date
 - During class dates but likely in the afternoon (5-7pmt)
- Recitations
 - Every Friday
 - Expand on material learned in class, go over problems from previous classes etc.
- Office hours based on your section

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

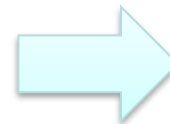
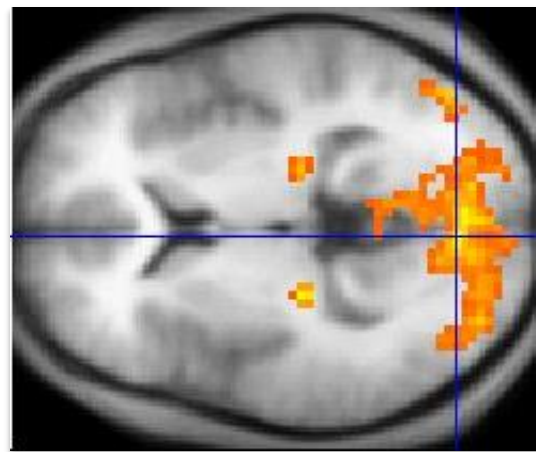
From Data to Understanding ...

Machine Learning in Action

Machine Learning in Action

Supervised learning

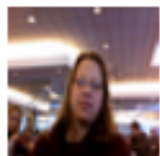
- Decoding thoughts from brain scans



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](#)

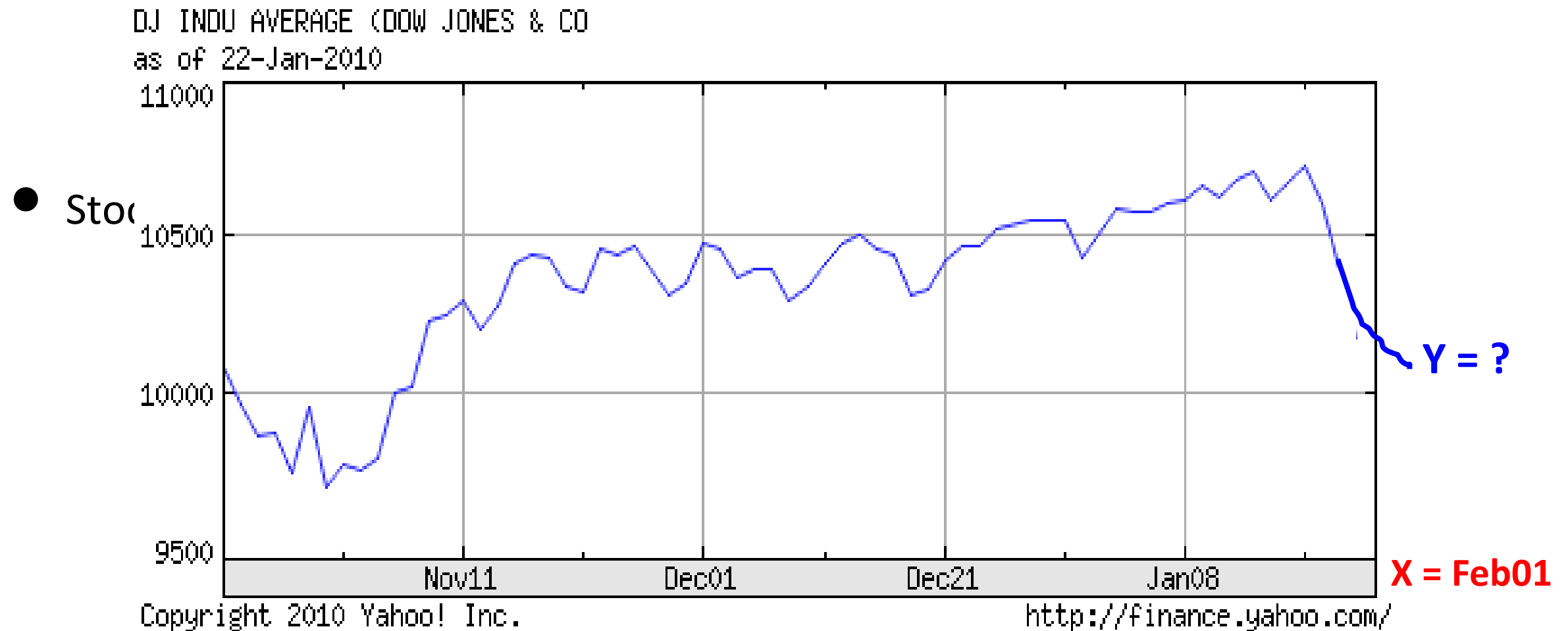
MRI Scans as Courtroom Evidence

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



Machine Learning in Action

Supervised and
unsupervised learning

● [



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.

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

=== 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

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.





















So far, NELL has accumulated over 50 million candidate beliefs by reading the web, and it is considering these with 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](#).



semi supervised
learning

Recently-Learned Facts

[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)

Z

Google

Translate

机器学习研究“我们如何构建能够通过经验自动改善其性能的计算机程序？”的问题。这包括学习基于许多类型的经验执行许多类型的任务。例如，它包括根据漫游环境获得的经验学习机器人学习更好的导航，学习基于历史健康记录数据挖掘预测哪些疗法最适合哪些疾病的医学决策辅助工具，以及学习更好的语音识别系统根据听取您的经验了解您的演讲。

本课程旨在为博士生提供在机器学习中进行研究和应用所需的方法，数学和算法的全面基础。具有概率，统计学和算法的预先存在的工作知识进入班级的学生将处于优势，但课程设计使得具有强大计算背景的任何人都可以赶上并充分参与。

שליחת הצעה לעריכה

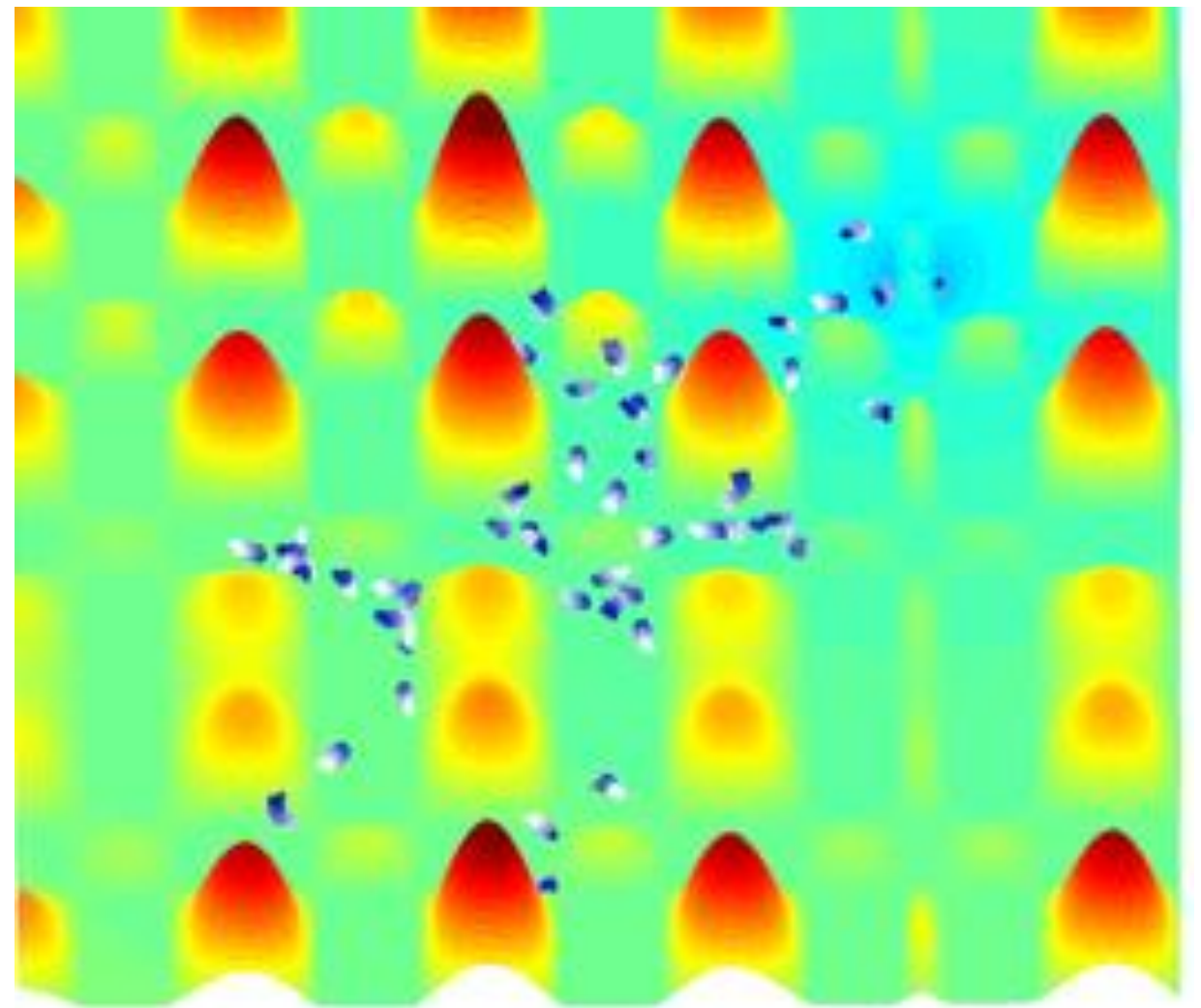
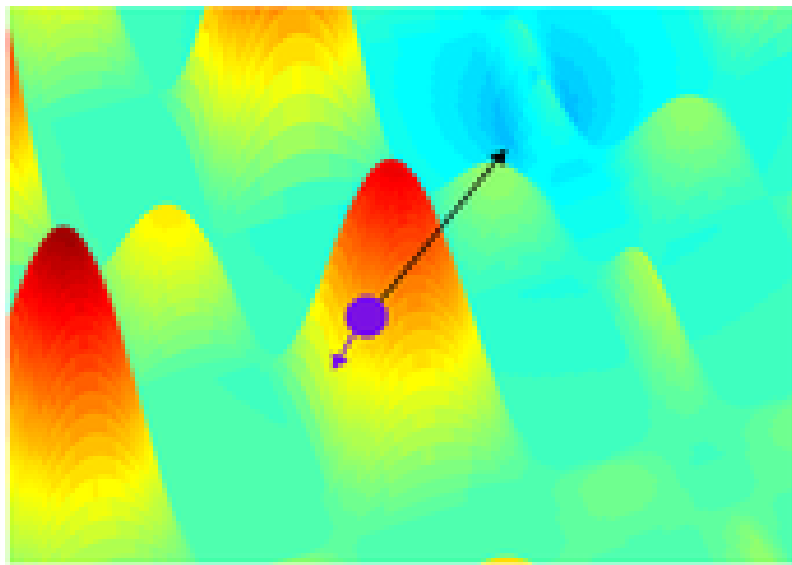
Jīqì xuéxí yánjiū “wǒmen rúhé gòujiàn nénggòu tōngguò jīngyàn zìdòng gǎishàn qí xìngnéng de jìsuànjī chéngxù?” De wèntí. Zhè bāokuò xuéxí jīyú xǔduō lèixíng de jīngyàn zhíxíng xǔduō lèixíng de rènwù. Lìrú, tā bāokuò gēnjù mànyóu huánjìng huòdé de jīngyàn xuéxí jīqìrén xuéxí gèng hǎo de dǎobǎng...

Machine learning studies the question of "How can we build computer programs that automatically improve their performance through experience?". This includes learning to perform many types of tasks based on many types of experience. For example, it includes robots learning to better navigate based on experience gained by roaming their environments, medical decision aids that learn to predict which therapies work best for which diseases based on data mining of historical health records, and speech recognition systems that learn to better understand your speech based on experience listening to you.

This course is designed to give PhD students a thorough grounding in the methods, mathematics and algorithms needed to do research and applications in machine learning. Students entering the class with a pre-existing working knowledge of probability, statistics and algorithms will be at an advantage, but the class has been designed so that anyone with a strong computational background can catch up and fully participate.

Distributed gradient descent based on bacterial movement

Reasoning under uncertainty



Biology

ACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAATTTCGATAGCAATTTC
GATAACGCTGAGCAATCGGATAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACG
CTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATCGGATATCGATAGCAATTTCGATAAATC
GGATAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAATTTCGATAGC
AATTCGATAACGCTGAGCAATCGGATATCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCA
ATTCGATAGCAATTTCGATAACGCTGAGCAATCGGATAACGCTGAGCAATTTCGATAGCATTCGAT
AACGCTGAGCAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATTCGGATAACGCTG
AGCAATTTCGATAGCAATTTCGATAACGCTGAGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGA
GCAATCGGATAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAATTTC
GATAGCAATTTCGATAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAATTTCGAT
AGCAATTTCGATAACGCTGAGCAATCGGATAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCT
GAGCAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATCGGATATCGATAGCAATT
CGATAACGCTGAGCAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATCGGATAAC
GCTGAGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATTTCGATAACGCTGAGCA
CTGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAACGCTGAGCA
ATTCGGATATCGATAGCAATTTCGATAACGCTGAGCA
ACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATCGGATAACGCTGAGCAATTTCGAT
AGCAATTTCGATAACGCTGAGCAACGCTGAGCAATTTCGATAGCAATTTCGATCGGATAACGCTGAGC
AATTCGATAGCAATTTCGATAACGCTGAGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCA
ATCGGATAACGCTGAGCAATTTCGATAGCA
AGCAATTTCGATAACGCTGAGCAATCGGAT
GCAATTTCGATAGCAATTTCGATAACGCTGA
GATAACGCTGAGCAACGCTGAGCAATTTCG
CTGAGCAATTTCGATAGCAATTTCGATAACG
TGAGCAATCGGATAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAA
TTCGATAGCAATTTCGATAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAACGCTGAGCAATTTC
GATAGCAATTTCGATAACGCTGAGCAATCGGATAACGCTGAGCAATTTCGATAGCAATTTCGATAAC
GCTGAGCAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATCGGATATCGATAGCA
ATTCGATAACGCTGAGCAACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCAATCGGAT
AACGCTGAGCAATTTCGATAGCAATTTCGATAACGCTGAGCTGAGCAATTTCGATAGCAATTTCGATA
ACGCTGAGCAATCGGA

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 reach

- 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

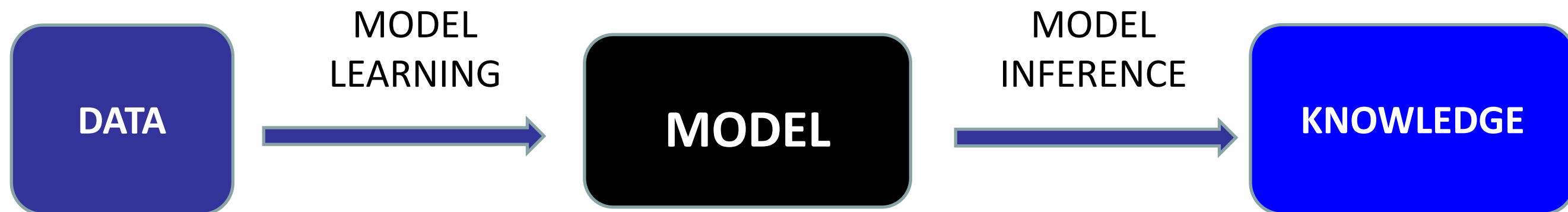
First Axis: Data

- 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

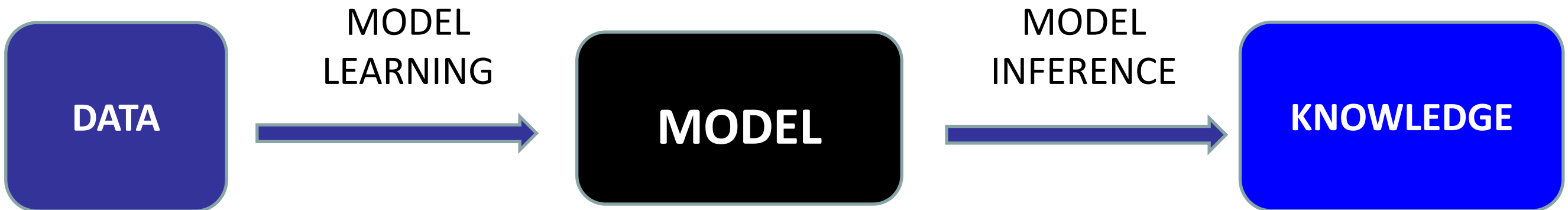
Second Axis: Algorithms

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

Model-based ML



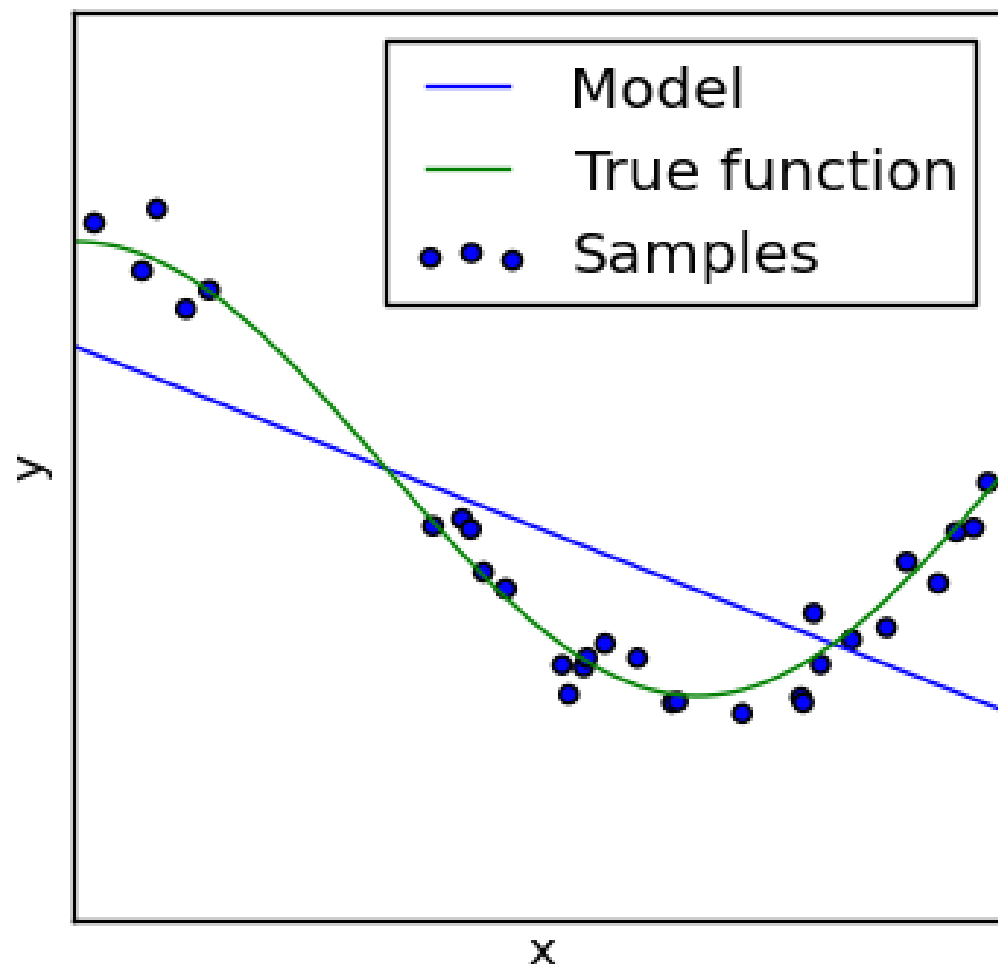
Model-based ML



- Learning: From data to model
 - A model is a summary of the data
 - But can also inform on 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

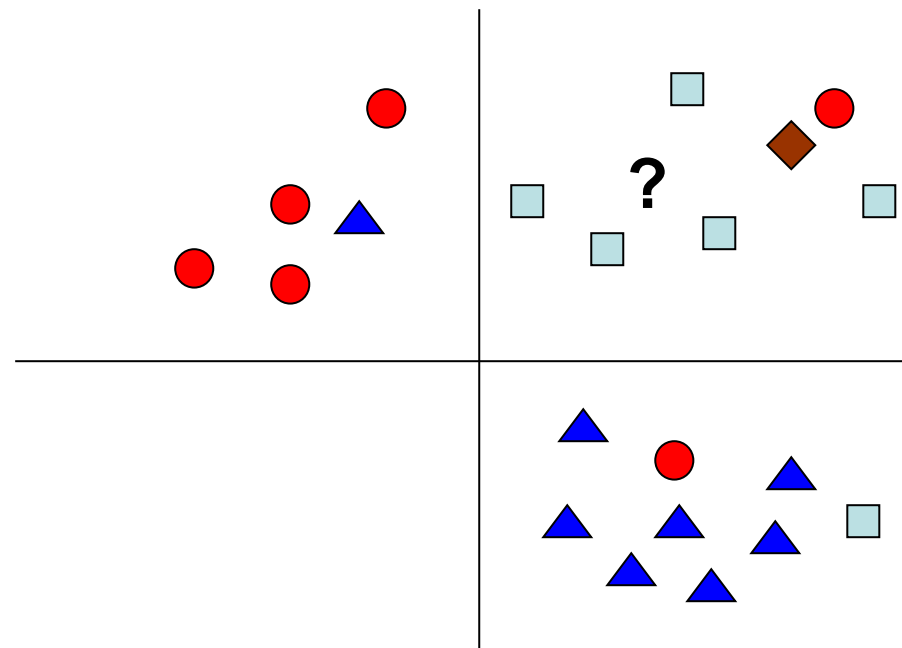


Fitting a simple line (2 params)
to a bunch of one-dim. samples

Model: data = point on line + noise

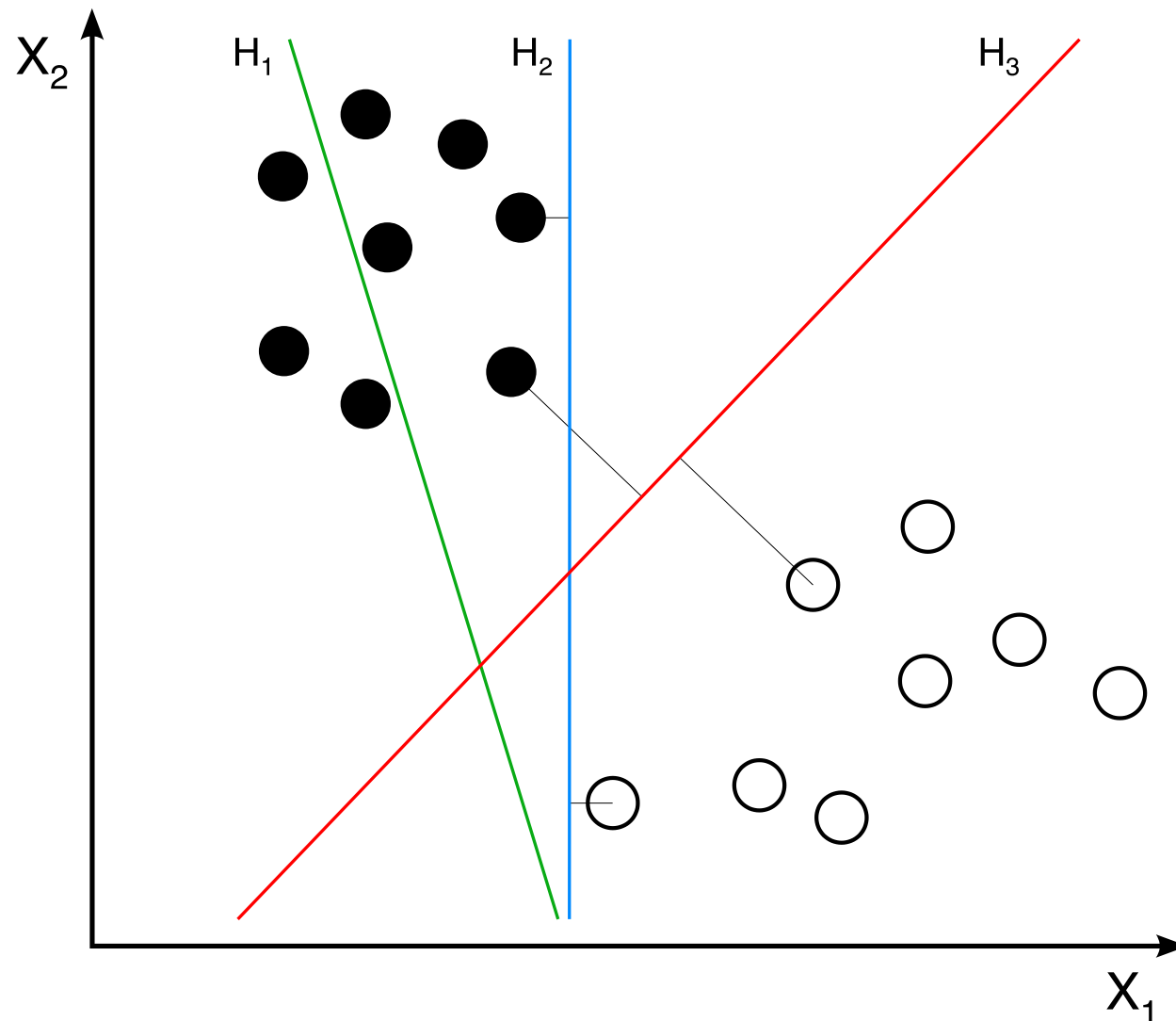
Nonparametric Models

- 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

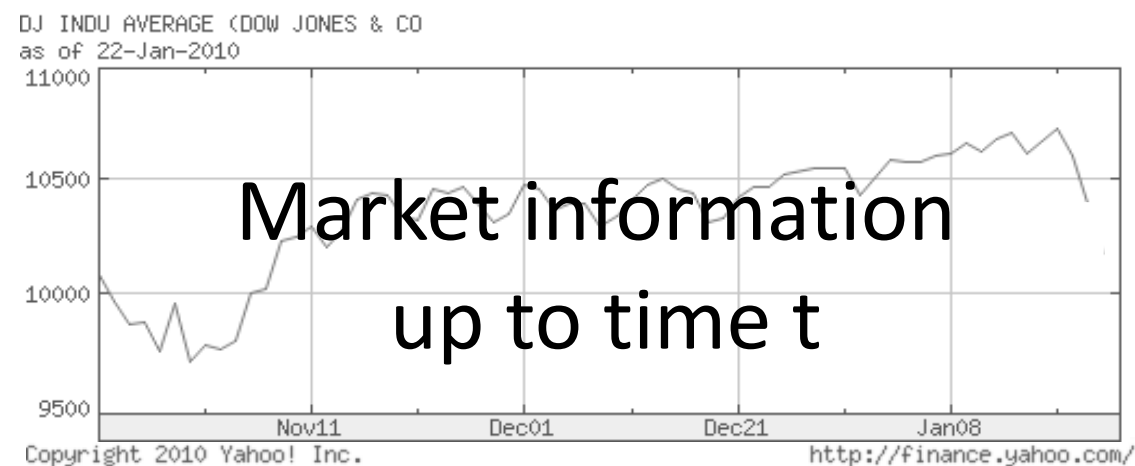
Prediction Problems

Feature Space \mathcal{X}

Words in a document

Label Space \mathcal{Y}

“Sports”
“News”
“Science”
...



Share Price
“\$ 24.50”

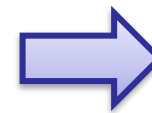
Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Prediction - Classification

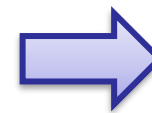
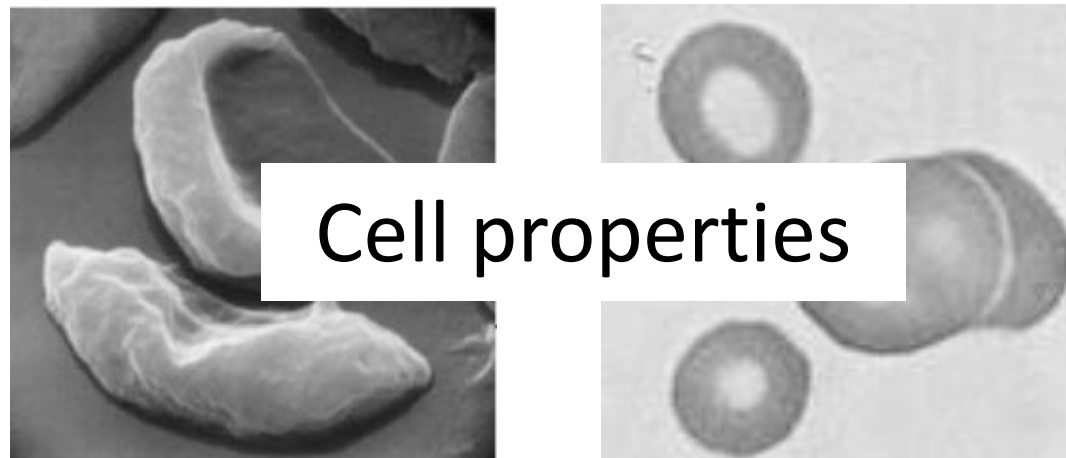
Feature Space \mathcal{X}



Label Space \mathcal{Y}



“Sports”
“News”
“Science”
...



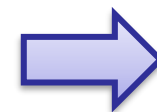
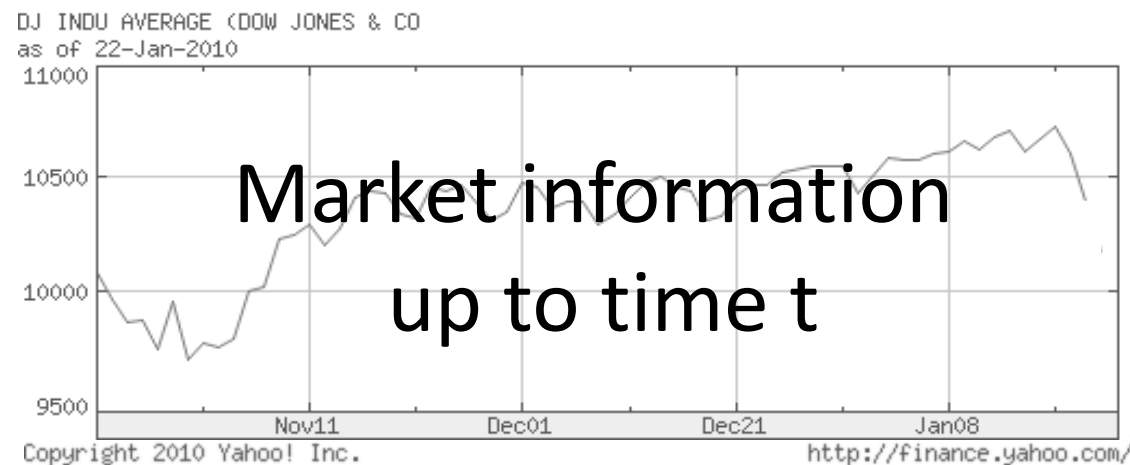
“Anemic cell”
“Healthy cell”

Discrete Labels

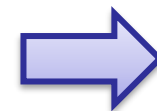
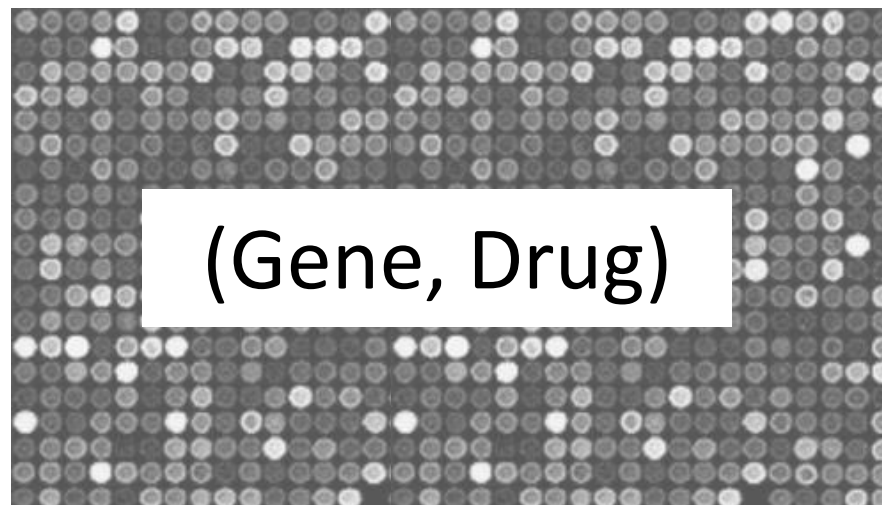
Prediction - Regression

Feature Space \mathcal{X}

Label Space \mathcal{Y}



Share Price
"\$ 24.577"



Expression level
"6.88"

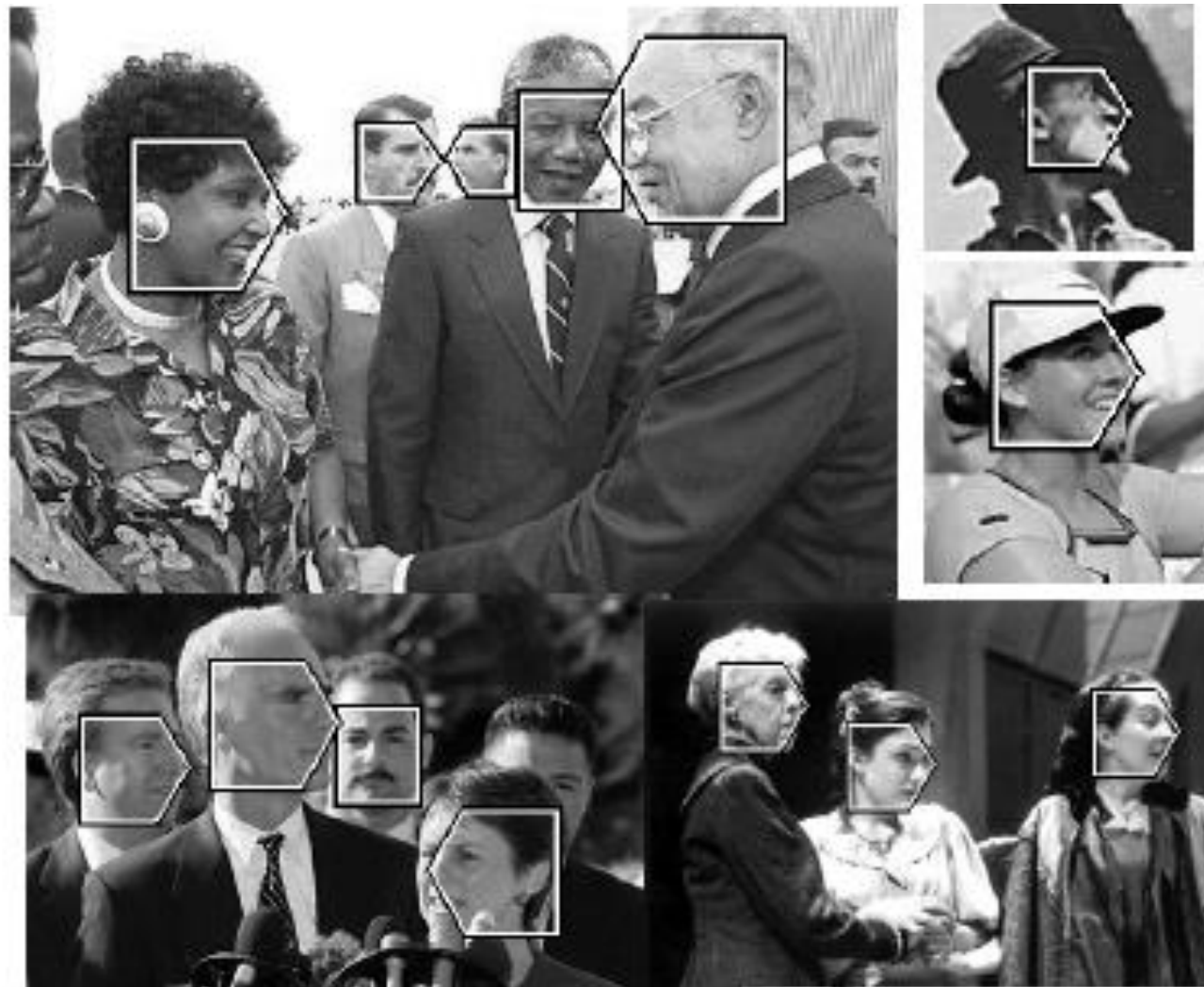
Continuous Labels

Prediction problems

Features?

Labels?

Classification/Regression?



Face Detection

Prediction problems

Features?

Labels?

Classification/Regression?



Robotic Control

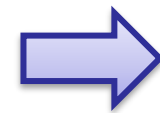
Third Axis: Tasks

- 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}

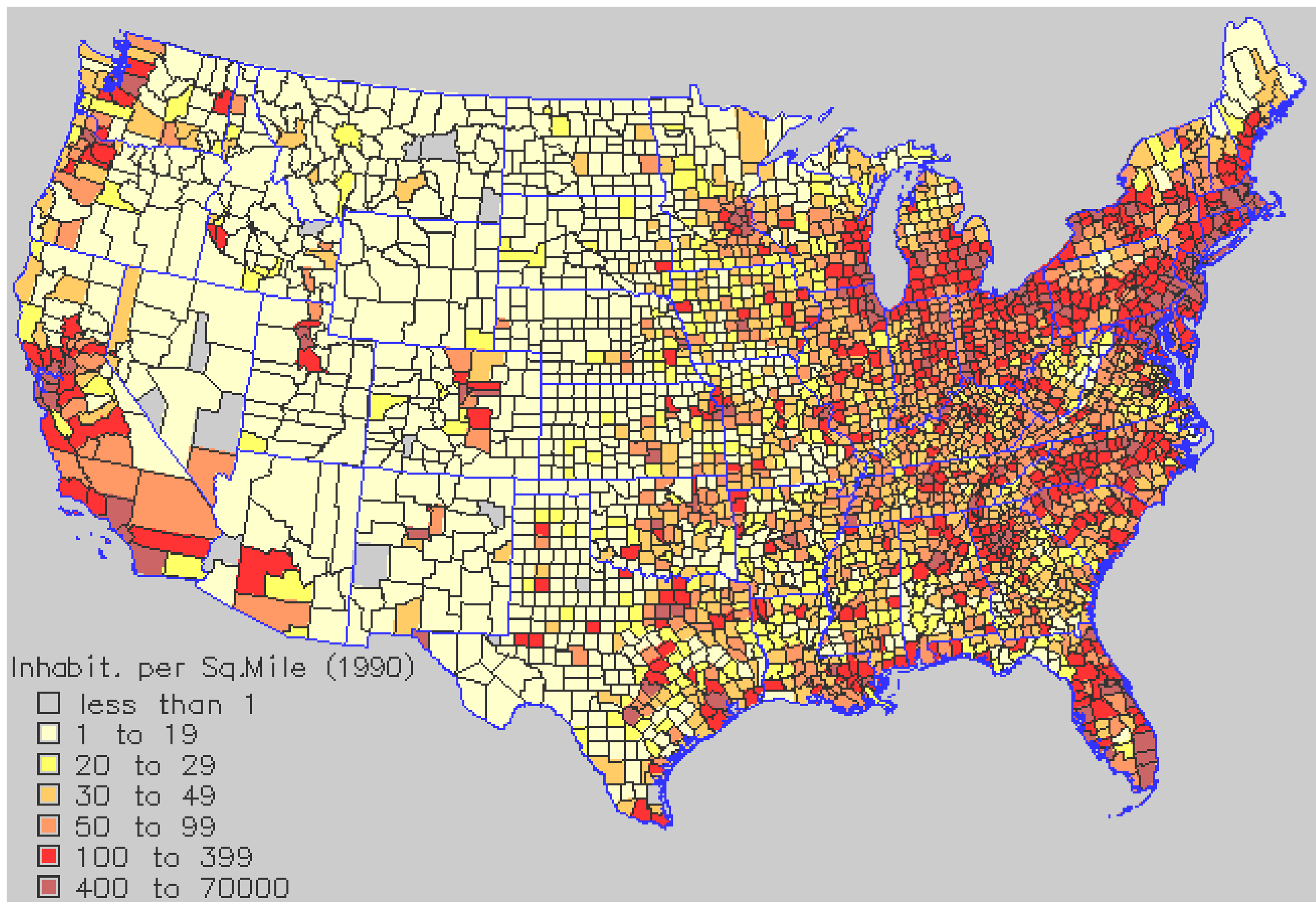


Word distribution
(Probability of a word)

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.]



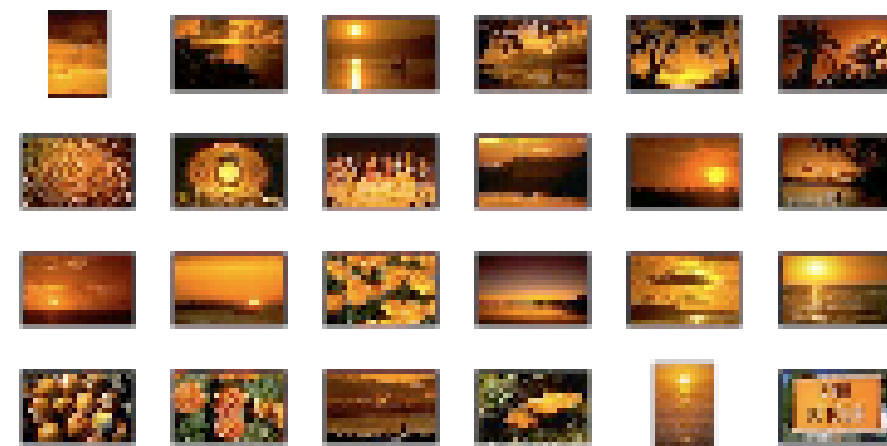
C_1



C_2



C_3




C_4



C_5

Unsupervised Learning – clustering web search results

web news images wikipedia blogs jobs more »




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)




Cluster **Human** contains 8 documents.

- [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-identification by culture and over time, and are often controversial for scientific as well as social and political reasons. History · More

[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 sailboat race. **Race** and ethnicity in the United States Census, official definitions of "**race**" used by the US Census Bureau. Historical definitions of **race**; **Race** (bearing), the inner and outer rings of a rolling-element bearing. **RACE** in 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/background/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, 1995, pp. 1-10.

[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 inherited characteristics.

[www.answers.com/topic/race-1](#) - [cache] - Live

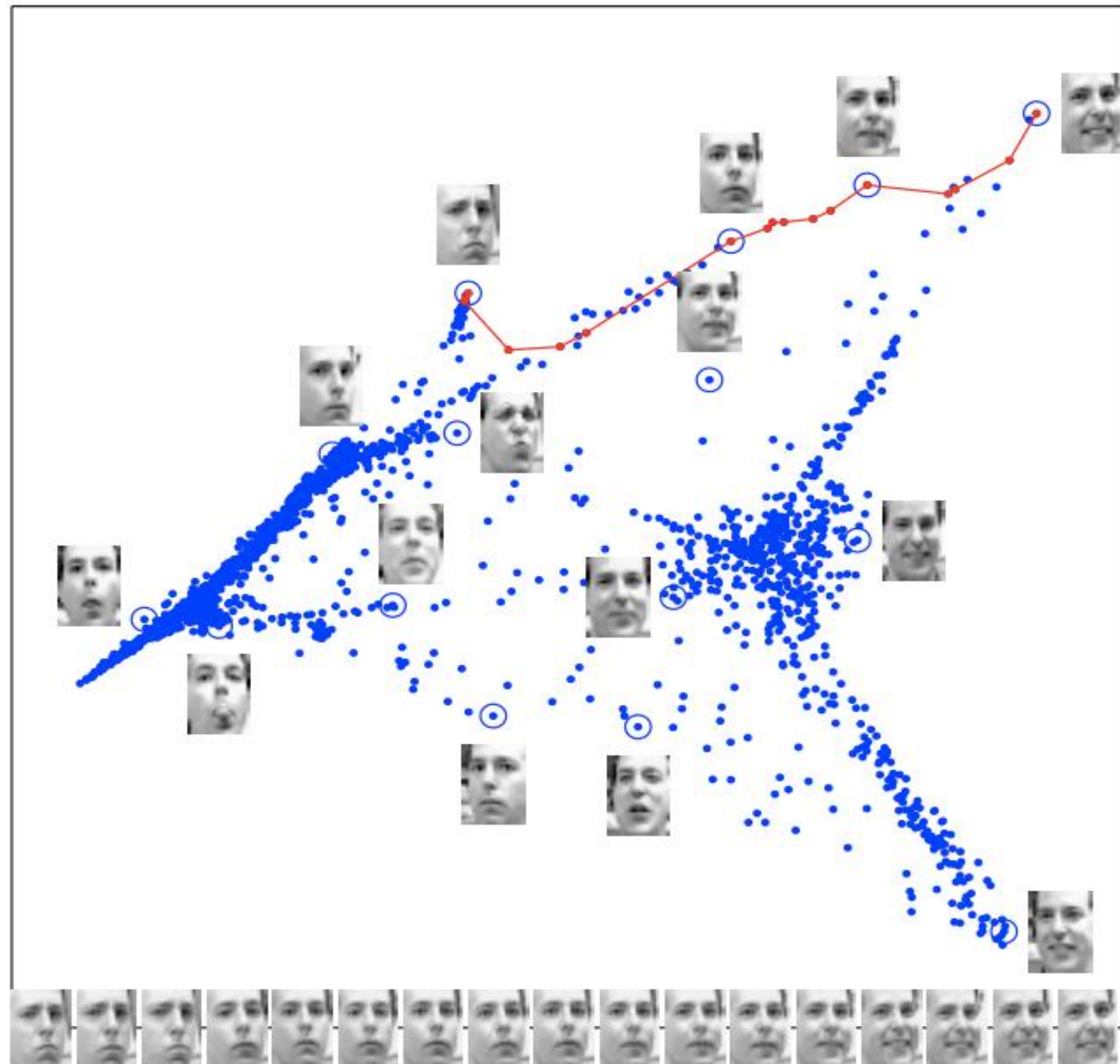
Unsupervised Learning - Embedding

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?



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 $D = \{\text{environment, actions, rewards}\}$ 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 $F2: \{X_i, x_k\} \rightarrow Y$