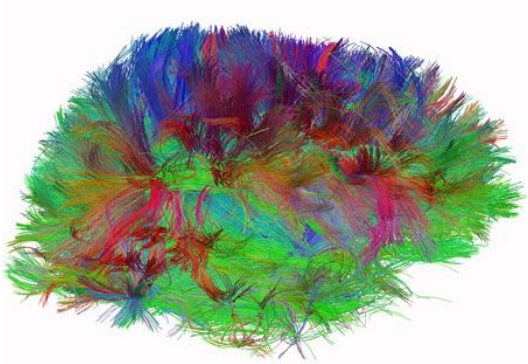
Machine Learning - Intro

Aarti Singh, Eric Xing

Machine Learning 10-701/15-781
Sept 10, 2012

The Age of Big Data

CERN Collider
320 x 10^{12} bytes/second



Personal Connectome
10^{18} bytes/human

facebook

1 billion
messages/day

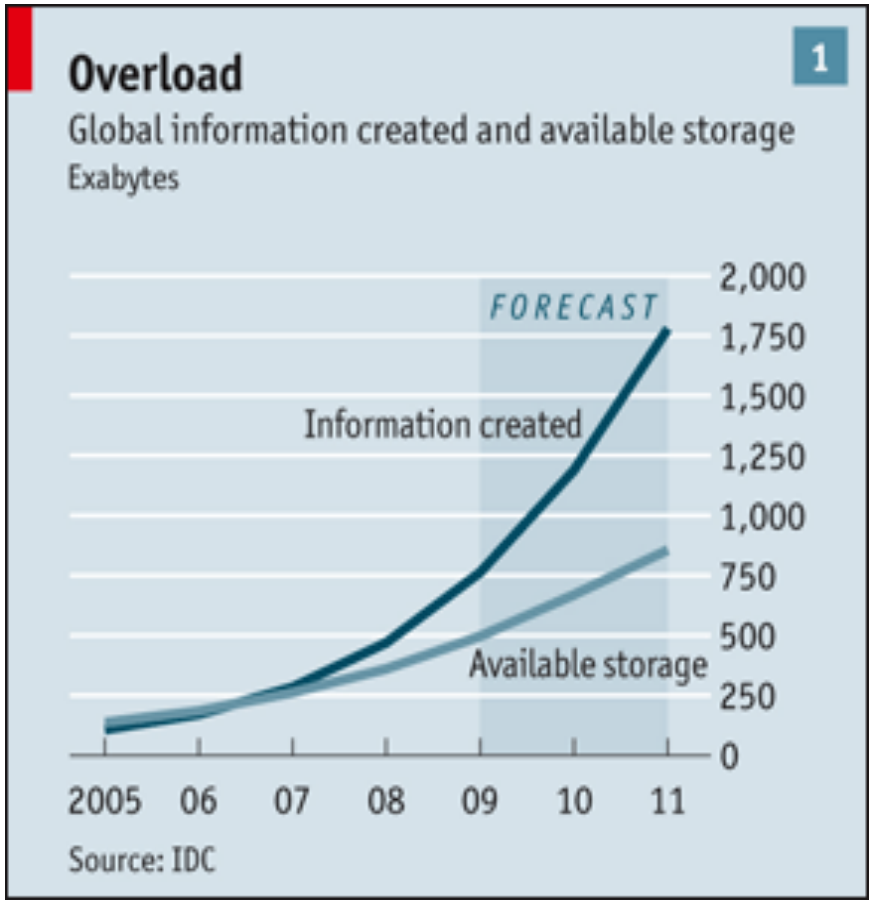
twitter

200 million
tweets/day

"Every day, people create the equivalent of 2.5 **quintillion** bytes of data from sensors, mobile devices, online transactions, and social networks:

The Huffington Post: Arnal Dayaratna: IBM Releases Big Data

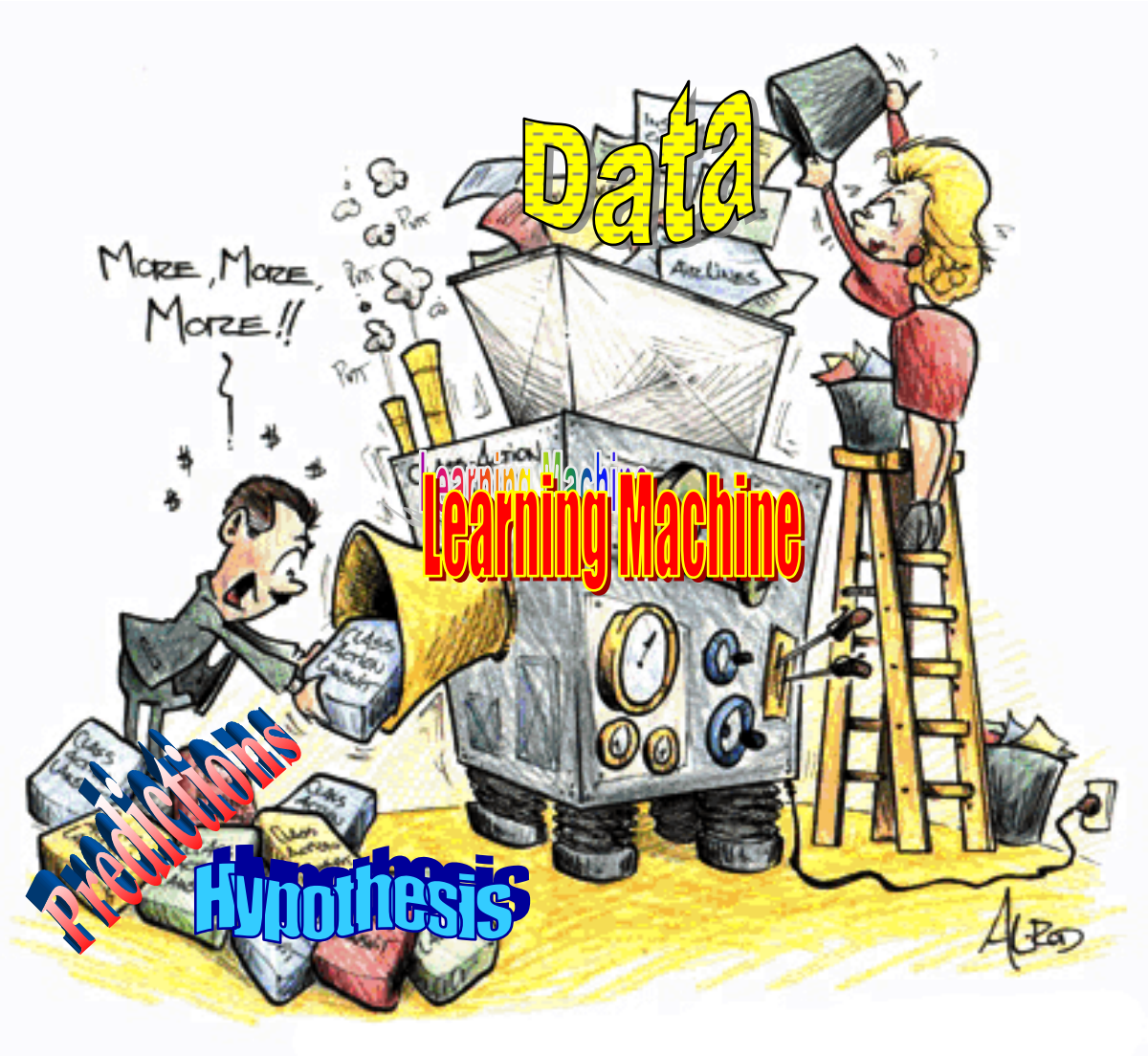
The Age of Big Data



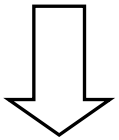
Data ⇏ Knowledge

From Data to Knowledge …

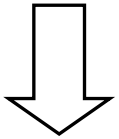
What is Machine Learning?



Data



Learning algorithm



Knowledge

What is Machine Learning?

Design and Analysis of algorithms that

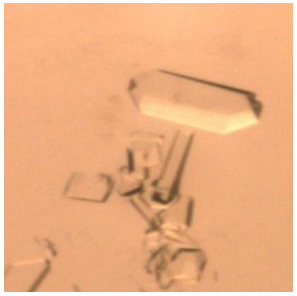
- improve their performance
- at some task
- with experience

(experience) **(performance on task)**

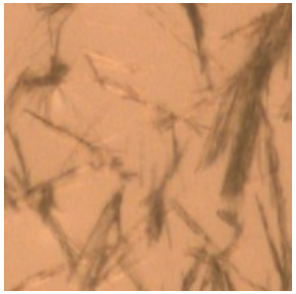
http://phillips-lab.biochem.wisc.edu/

Human learning

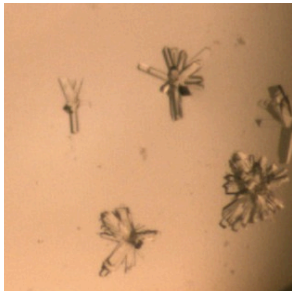
Task: Learning stage of protein crystallization



Crystal

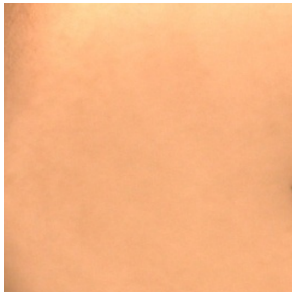


Needle

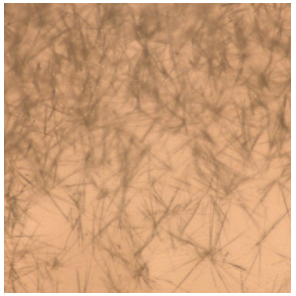


Tree

Tree

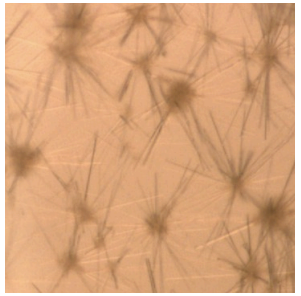


Empty



Needle

Experience



?

Performance

Machine Learning in Action

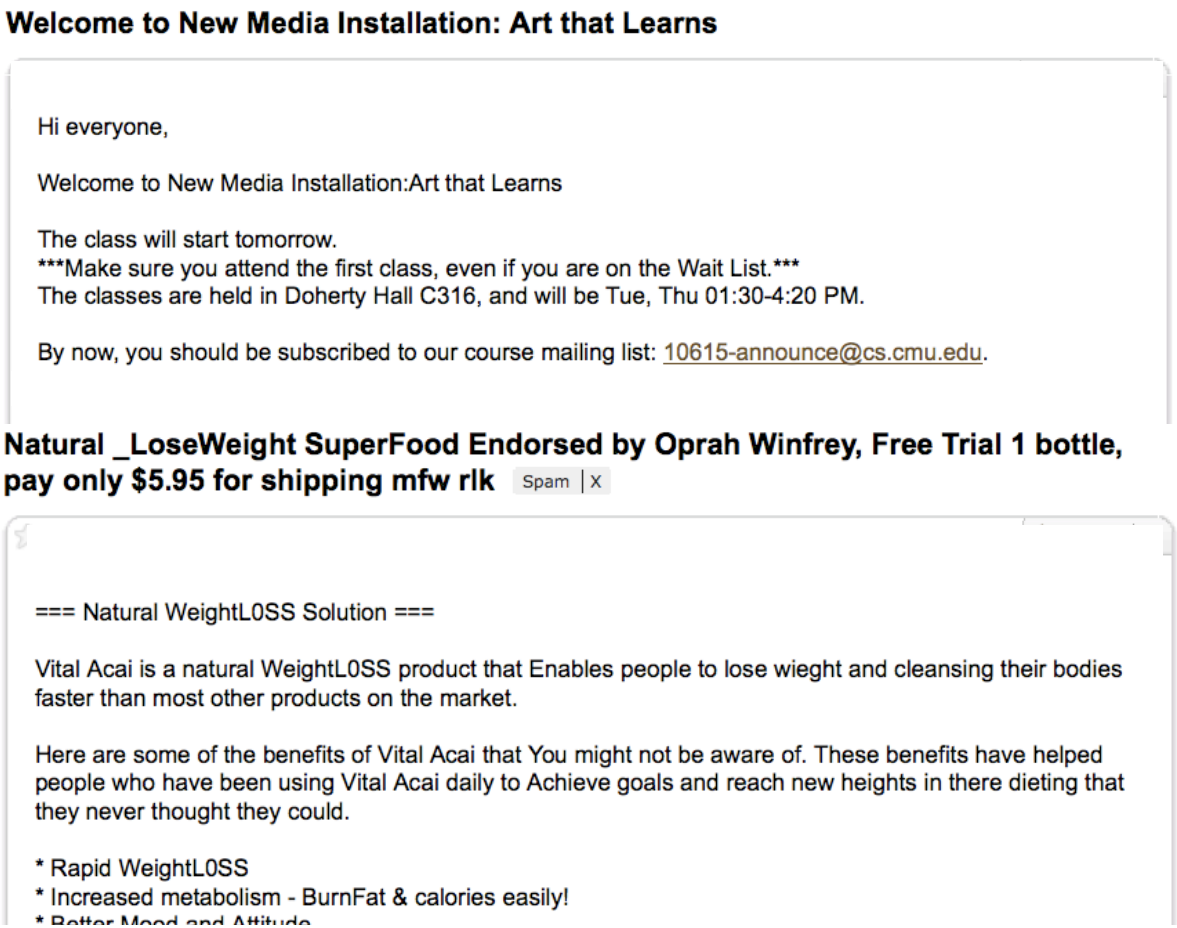
Machine Learning in Action

- Document classification

Sports
Science
News

Machine Learning in Action

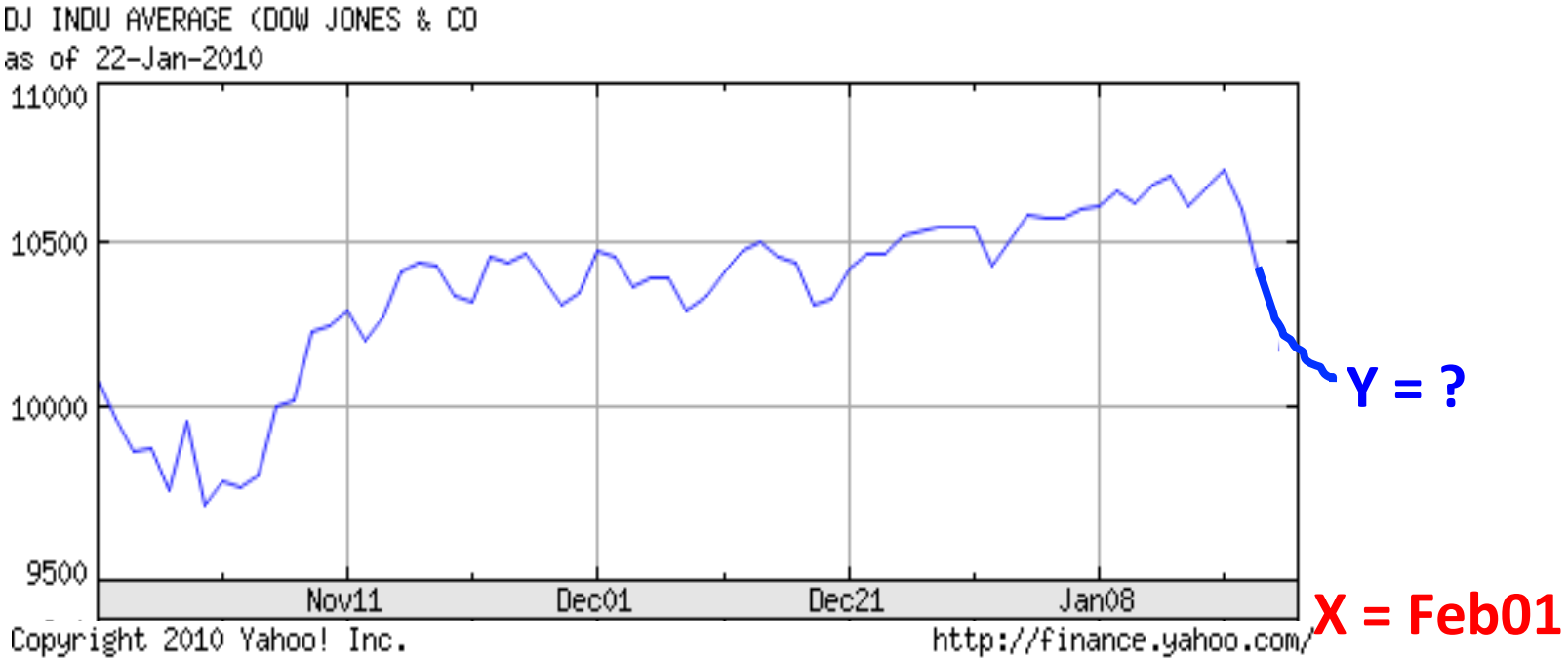
- Spam filtering



Spam/
Not spam

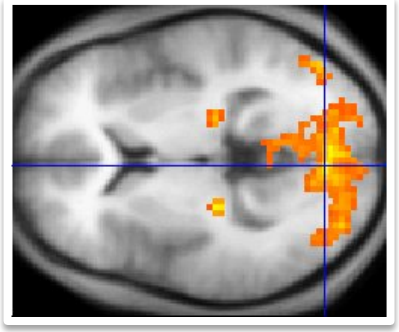
Machine Learning in Action

- Stock Market Prediction



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

- The **best** helicopter pilot is now a computer!

http://heli.stanford.edu/

Machine Learning in Action

- Many, many more…

  Speech recognition, Natural language processing
  Computer vision
  Robotics
  Web forensics
  Medical data 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 …
- Software too complex to write by hand
- Improved machine learning algorithms
- Improved data capture (Terabytes, Petabytes of data), networking, faster computers
- Demand for self-customization to user, environment

Are we there yet?

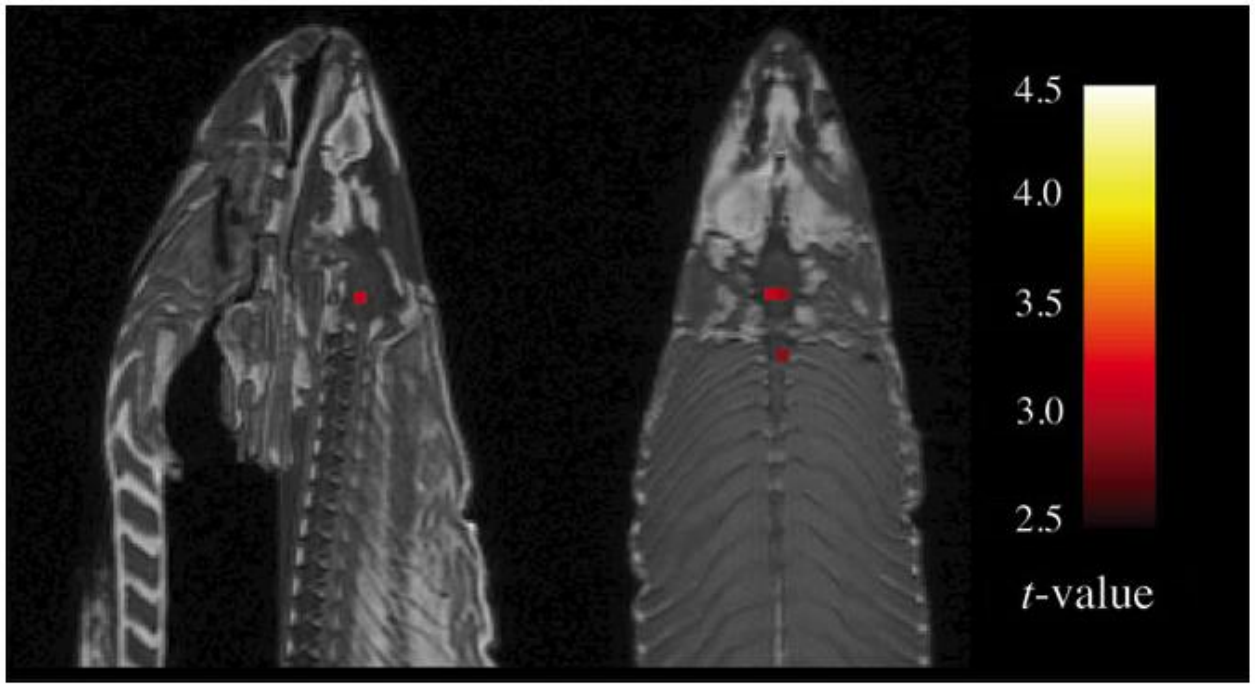
ML has a long way to go …

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



ML has a long way to go …

Speech Recognition gone Awry

http://www.google.com/url?sa=t&source=web&cd=5&ved=0CCwQtwIwBA&url=http://video.google.com/videoplay?docid=-11232212177827774 72&rct=j&q=bad%20speech%20recognition&ei=nvyGTN3kOMOAlAezu_HHDg&usg=AFQjCNHDTf0w6VudgJfbP3xAvDTFhbzWCQ&cad=rja

Machine Learning Tasks

Broad categories -

- **Supervised learning**

  Classification, Regression

- **Unsupervised learning**

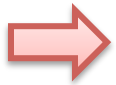
  Density estimation, Clustering, Dimensionality reduction

- Semi-supervised learning
- Active learning
- Reinforcement learning
- Many more …

Supervised Learning

Feature Space $\mathcal{X}$

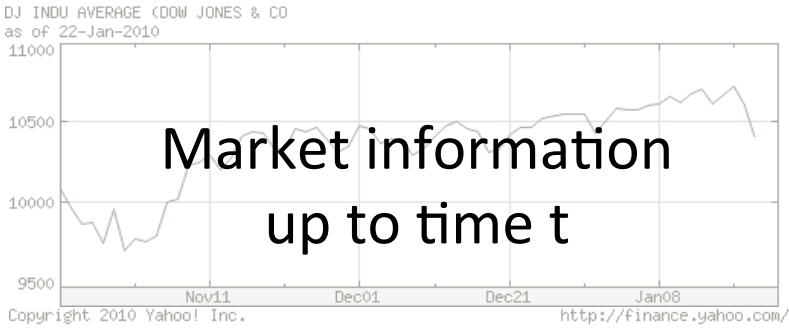
Words in a document



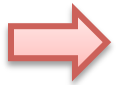
Label Space $\mathcal{Y}$

"Sports"
"News"
"Science"
…

Discrete Labels
Classification



Market information up to time t



Share Price
"$ 24.50"

Continuous Labels
Regression

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

Unsupervised Learning

Aka "learning without a teacher"

Feature Space $\mathcal{X}$

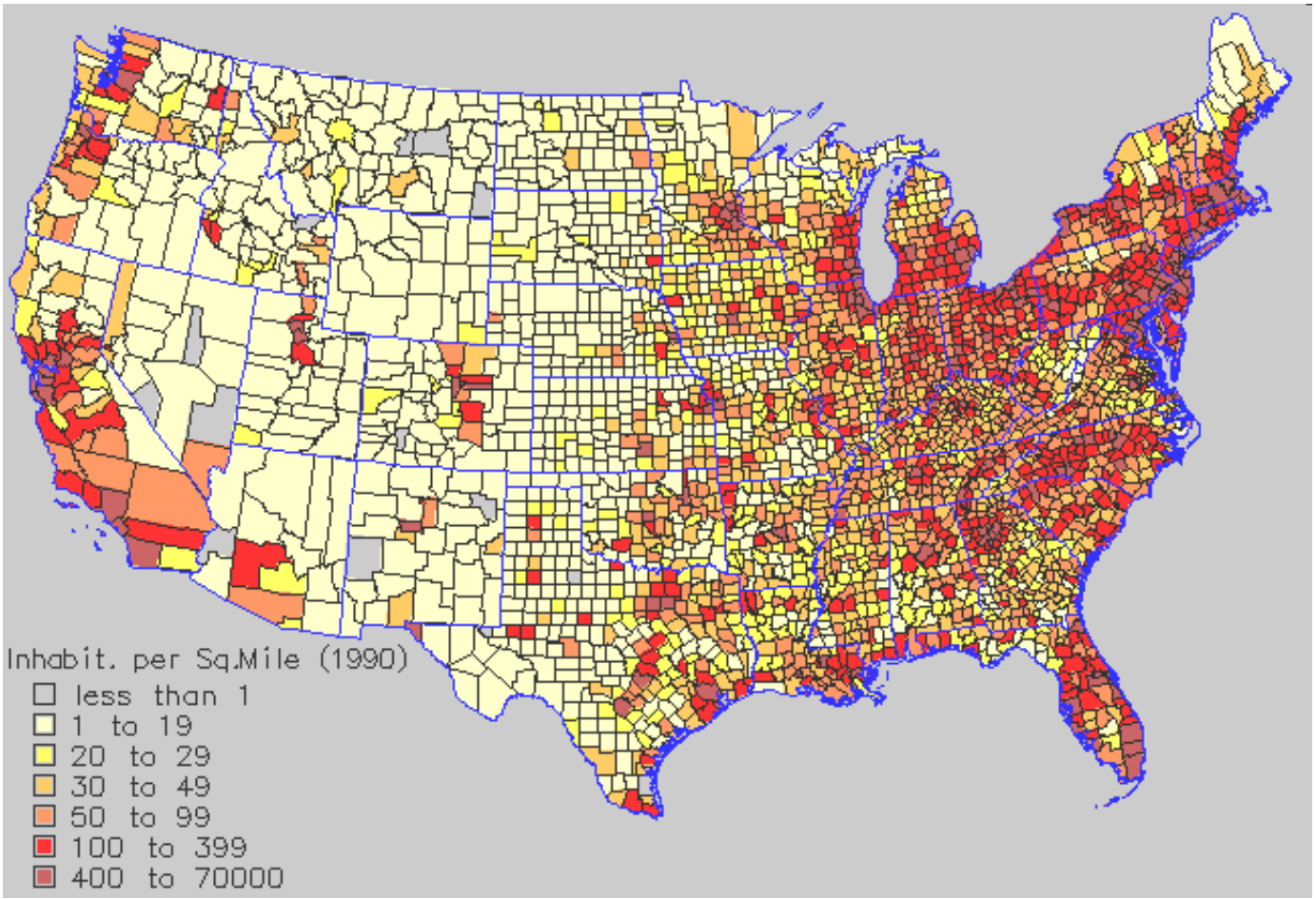
Words in a document

Word distribution
(Probability of a word)

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

Unsupervised Learning

Density/Distribution Estimation



Unsupervised Learning

Clustering - Group similar things e.g. images [Goldberger et al.]

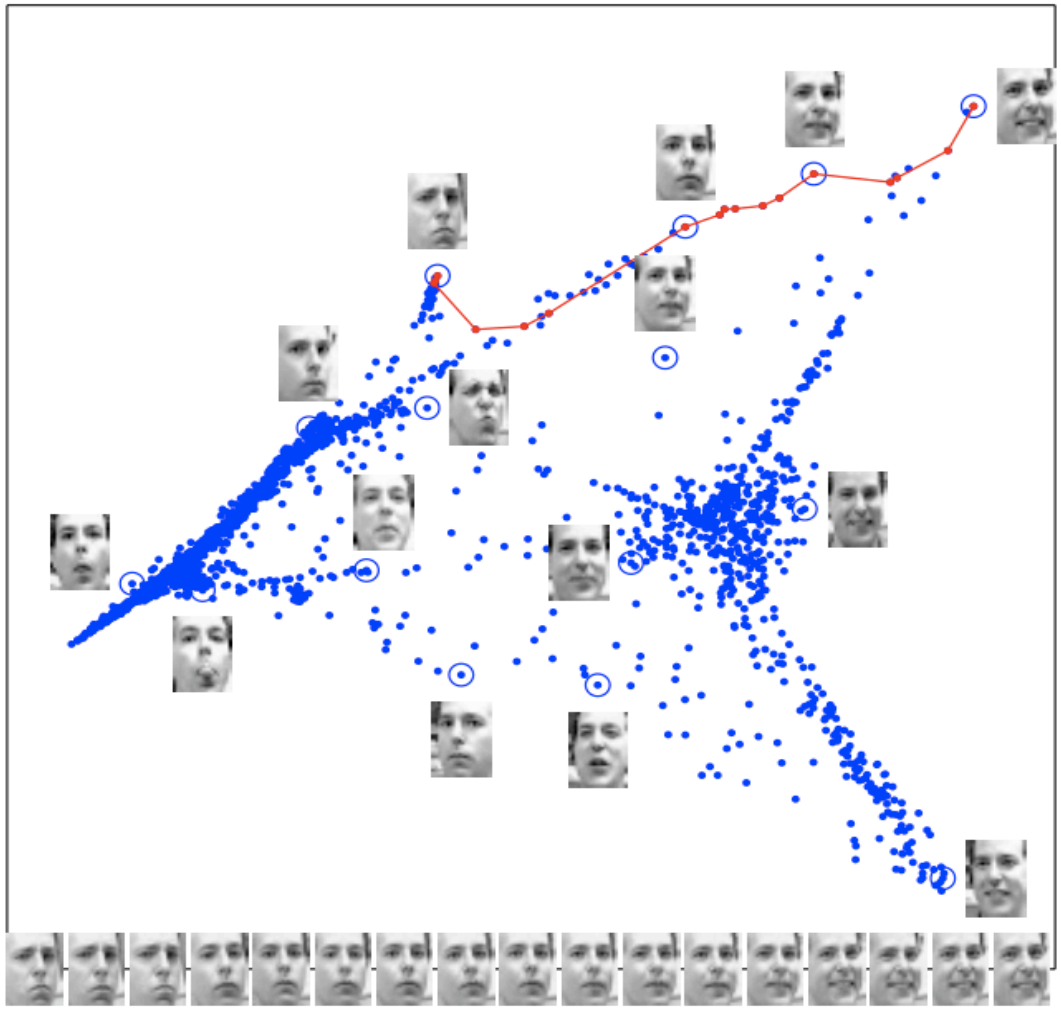
Unsupervised Learning

Dimensionality Reduction/Embedding

[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?



Machine Learning Tasks

Broad categories -

- **Supervised learning**

  Classification, Regression

- **Unsupervised learning**

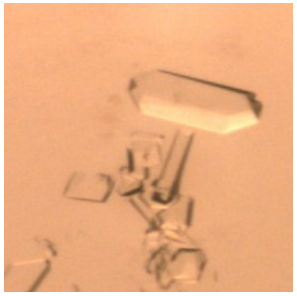
  Density estimation, Clustering, Dimensionality reduction

- Semi-supervised learning
- Active learning
- Reinforcement learning
- Many more …

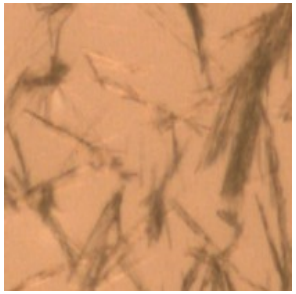
Key Issues in Machine Learning

Training Data vs. Test Data

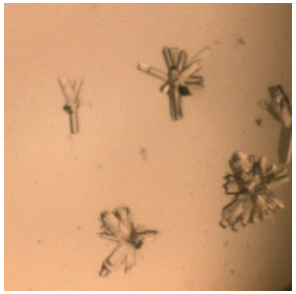
Task: Learning stage of protein crystallization



Crystal

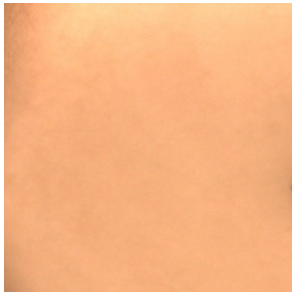


Needle

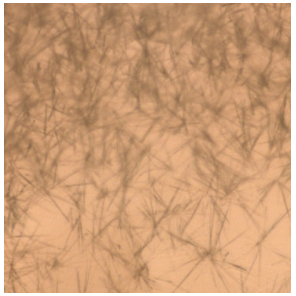


Tree

Tree

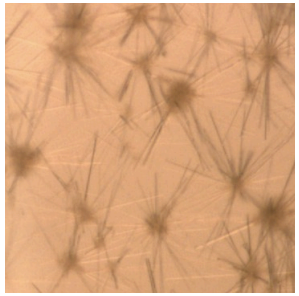


Empty



Needle

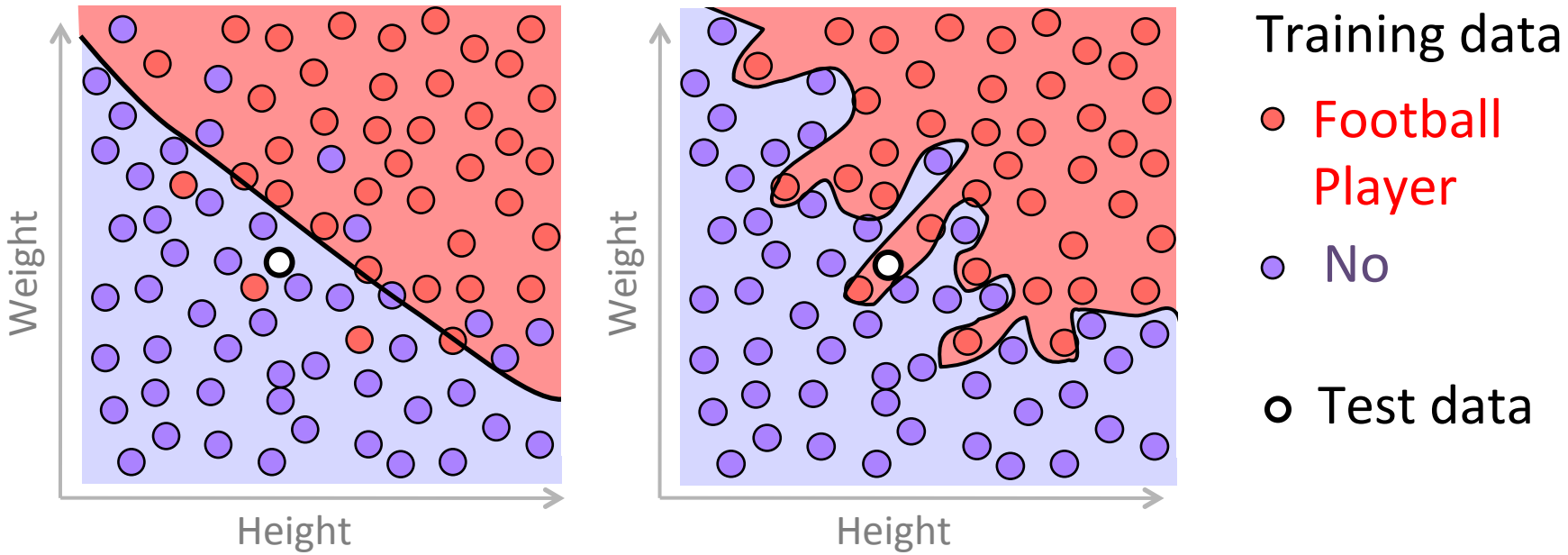
Experience



?

Performance

Training Data vs. Test Data



- A good machine learning algorithm
  - Does not **overfit** training data
  - **Generalizes** well to test data

Performance Measure

For a random test data X, measure of closeness between true label *Y* and prediction *f*(*X*)

Binary Classification $\text{loss}(Y, f(X)) = 1_{\{f(X) \neq Y\}}$ **0/1 loss**

Regression $\text{loss}(Y, f(X)) = (f(X) - Y)^2$ **square loss**

Density Estimation $\text{loss}(f(X)) = -\log(\mathbb{P}_f(X))$ **Negative log likelihood loss**

Machine Learning vs. Optimization

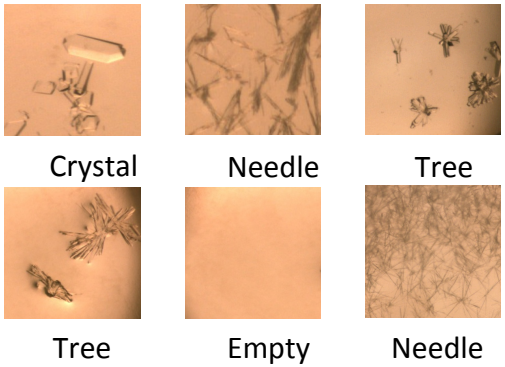
Ideal goal: Construct **prediction rule** $f : \mathcal{X} \to \mathcal{Y}$

that works well for any test data point $(X, Y) \sim P_{XY}$

Simply an optimization problem: $\min_f \mathbb{E}_{XY}\,[\text{loss}(Y, f(X))]$

BUT… Optimization depends on unknown P_{XY} !

Training data (experience) provides a glimpse of P_{XY}



$$\{(X^{(j)}, Y^{(j)})\}_{j=1}^{n} \sim P_{XY}$$

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 begins…