Machine Learning - Intro

Aarti Singh

Machine Learning 10-315
Aug 26, 2019

Teaching team

Instructor:

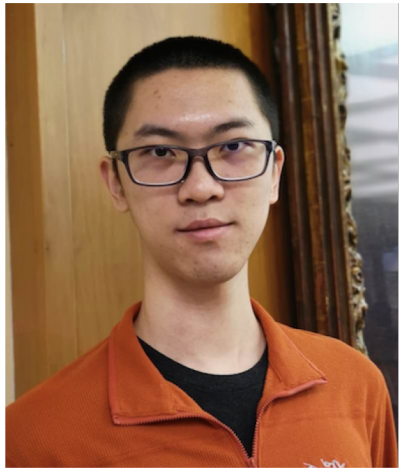
Aarti Singh

Education Associate:

Fabricio Flores

TAs:

Siddharth Ancha



Yue Wu

Other relevant people

- Admin – Mary Stech (mstech@andrew.cmu.edu)

- Qatar instructor and TA:
  - Gianni Di Caro (gdicaro@cmu.edu)
  - Aliaa Essameldin (aeahmed@andrew.cmu.edu)

Logistics

Lectures:	Mon, Wed	10:30-11:50 am HOA 160
Recitations:	Fri	10:30-11:50 am HOA 160
Office hours:	M, T, W, R – times and locations on course webpage	

Webpage: Syllabus, schedule of lectures, slides, policies, office hours, HWs, projects, exams, …
https://www.cs.cmu.edu/~aarti/Class/10315_Fall19

Piazza: discussion forum for students, announcements
piazza.com/cmu/fall2019/10315

Communication channel

- Ask questions in class!
- Office hours (M, T, W, R)
- Recitations (F)
- Note to instructors on Piazza
- DO NOT expect moderation on every question by instructors or TAs on Piazza discussion forum – discuss with class!

Grading

- Grading
  - 4 homework assignments (4 x 10% = 40%)
  - 4 QnAs (15%)
  - 1 midterm, 1 final (both in class): (10+15 = 25%)
  - Kaggle-style project (20%)

- Late days
  - total 4 across homeworks and QnAs
  - no late days for project

Homeworks & QnAs

- 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
    - list resources used (references, discussants) on top of submitted homework
- Homeworks are hard, start early ☺
- Due on gradescope at 11:59 pm

Waitlist + Audits + Pass/Fail

- Waitlist

  we'll let everyone in as long as there is space in room

  wait to see how many students drop

  keep attending lectures and doing HW

- Audits and Pass/Fail

  Audits NOT allowed

  Pass/Fail allowed

About the course

- Machine Learning **Algorithms and Principles**
  - Classification: Naïve Bayes, Logistic Regression, Neural Networks, Support Vector Machines, k-NN, Decision Trees, Boosting
  - Regression: Linear regression, Kernel regression, Nonparametric regression
  - Unsupervised methods: Kernel density estimation, k-means and hierarchical clustering, PCA, nonlinear dimensionality reduction
  - Core concepts: Probability, Optimization, Theory, Model selection, overfitting, bias-variance tradeoffs …

- See **tentative** lecture schedule on webpage – MAY CHANGE
- Material: Class slides + Reading material

Recommended textbooks

- Textbooks (Recommended, not required):

  Pattern Recognition and Machine Learning, Christopher 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

Related courses

- Related courses – Intro to ML algorithms and principles

  10-301 – Undergrad version for non-SCS majors

  10-601 – Masters version

  10-701 – PhD version

  10-715 – PhD students doing research in machine learning (hardest, most mathematical)

  Other related courses:

  10-606, 10-607 – Math background for ML

  10-605, 10-805 – Machine Learning with Large Datasets

  11-663 – Machine Learning in Practice (ML software)

  10-702, 10-704, 10-707, 10-708, 10-709, 15-859(B) – related advanced topics

Pre-requisites

- **Assume mathematical maturity**
  - Multivariate Calculus

    Derivatives, integrals of multi-variate functions
  - Linear Algebra

    Matrix inversions, eigendecomposition, …
  - Basic Probability and Statistics

    Probability distributions – discrete and continuous, Mean, Variance, Conditional probabilities, Bayes rule, Central limit theorem…
  - Programming

- **Tutorial videos**
  - Probability, Calculus, Functional Analysis, SVD

    https://www.youtube.com/channel/UC7gOYDYEgXG1yIH_rc2LgOw/playlists
  - Linear Algebra

    http://www.cs.cmu.edu/~zkolter/course/linalg/index.html

Recitations

- Strongly recommended
  - Brush up pre-requisites
  - Hands-on exercises
  - Review material (difficult topics, clear misunderstandings, extra new topics, HW and exam solutions)
  - Ask questions

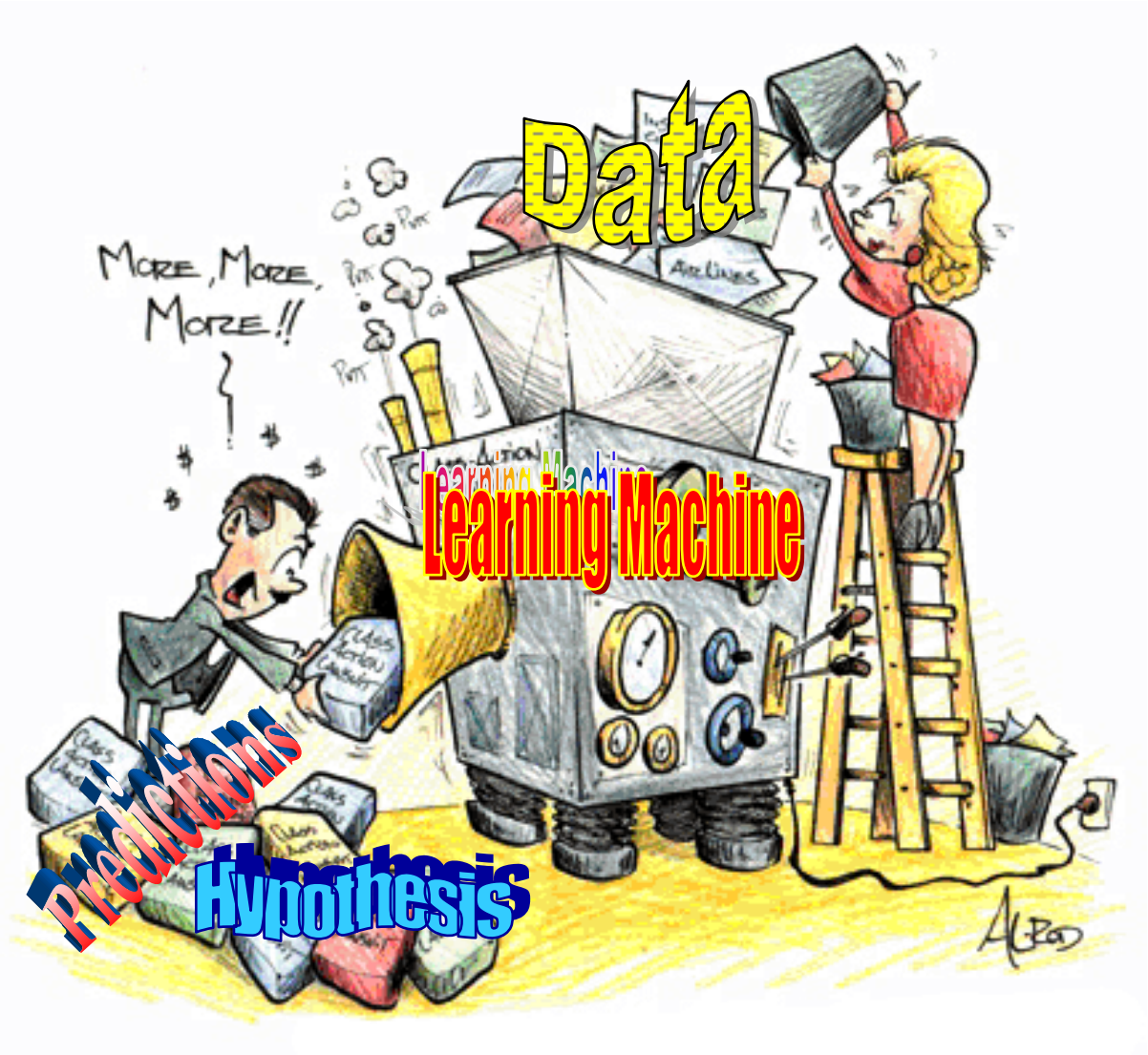
- 1st Probability Review - **FRIDAY**

  by Siddharth Ancha

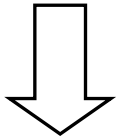
  Fri Aug 30 10:30-11:50 am HOA 160

What is Machine Learning?

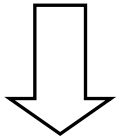
What is Machine Learning?



Data



Learning algorithm



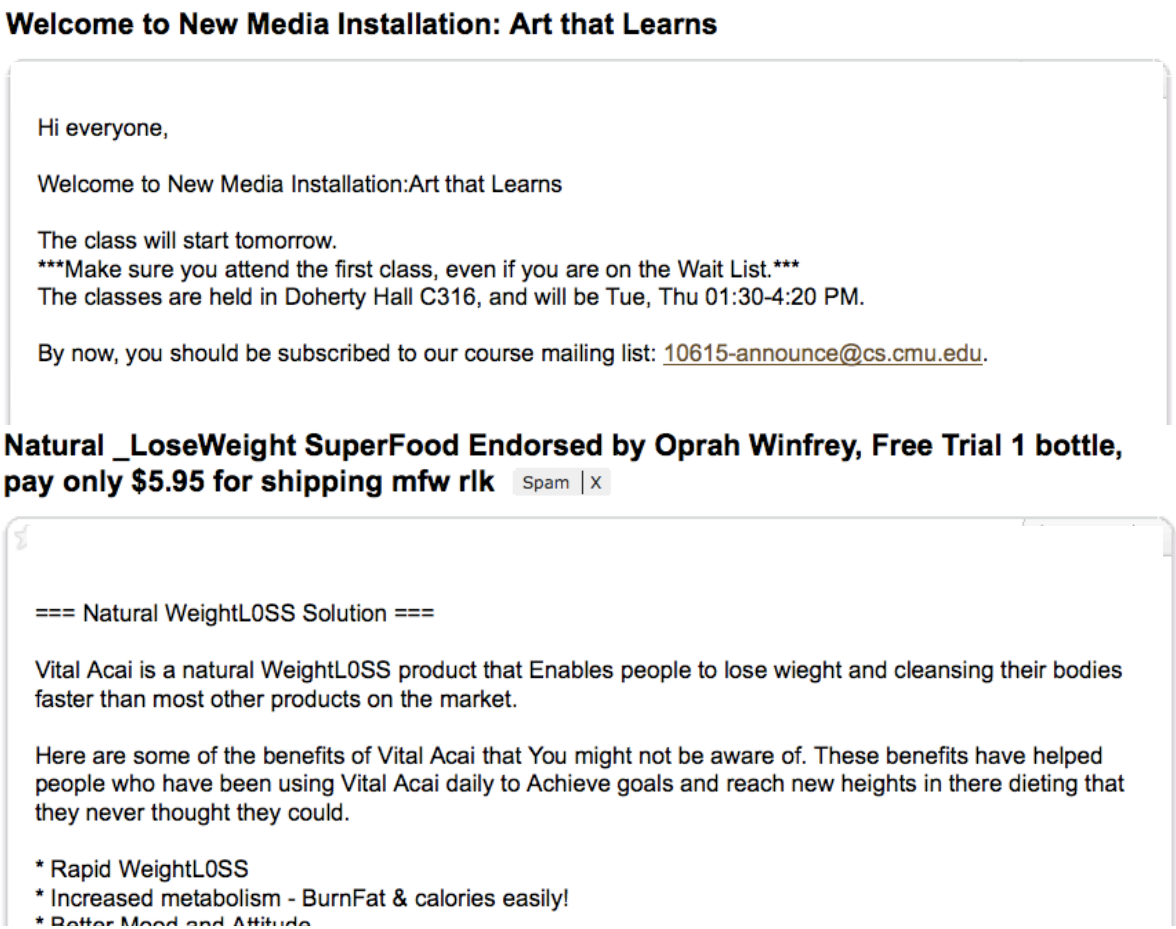
Knowledge

From Data to Knowledge …

Machine Learning in Action

Machine Learning in Action

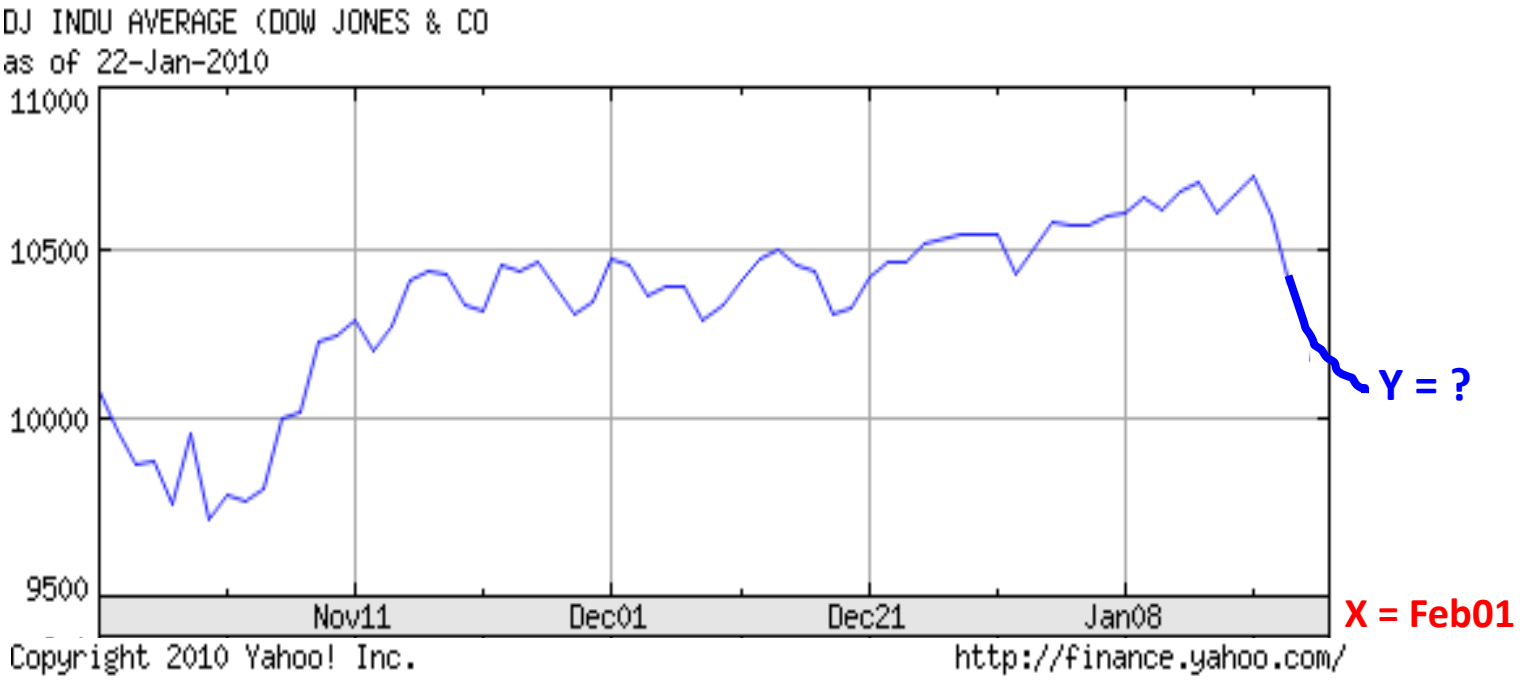
- Spam filtering



Spam/
Not spam

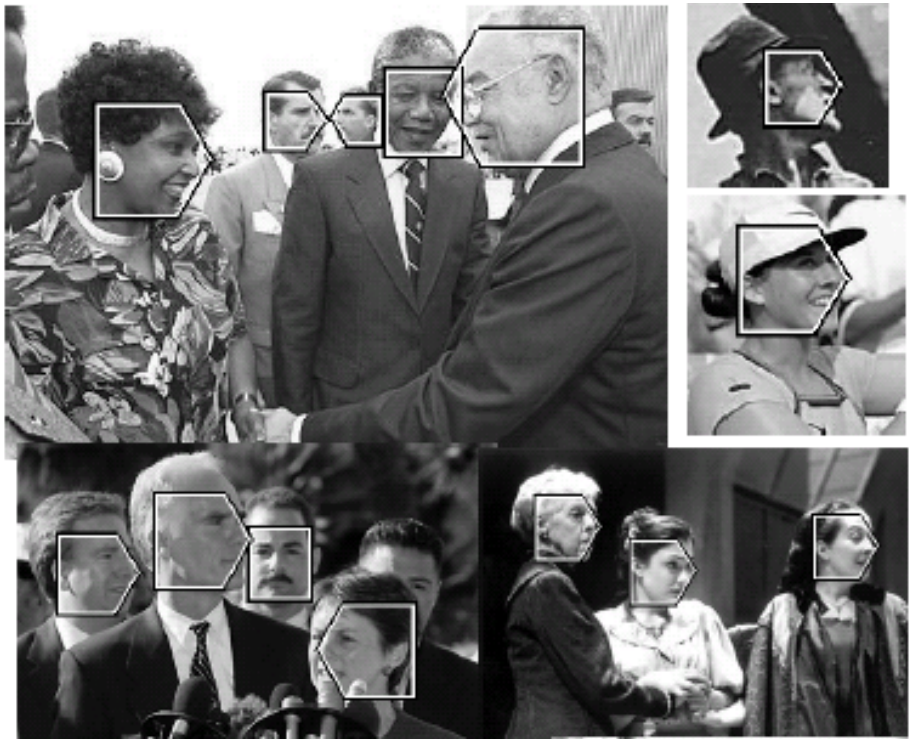
Machine Learning in Action

- Stock Market Prediction



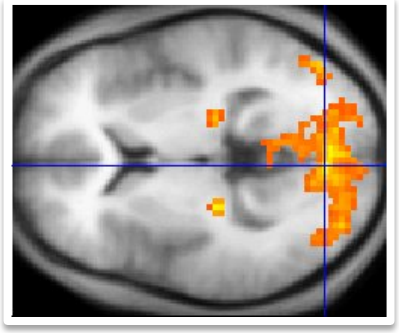
Machine Learning in Action

- Face detection



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

- 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

Document classification

Speech recognition, Natural language processing

Computer vision

Robotics

Web forensics

Medical data analysis

Sensor networks

Social networks

Smart buildings

…

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
  - 1.6 million images, 1000 object categories
  - 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

"Data scientist: The sexiest job of the 21st century"

(Harvard Business Review)

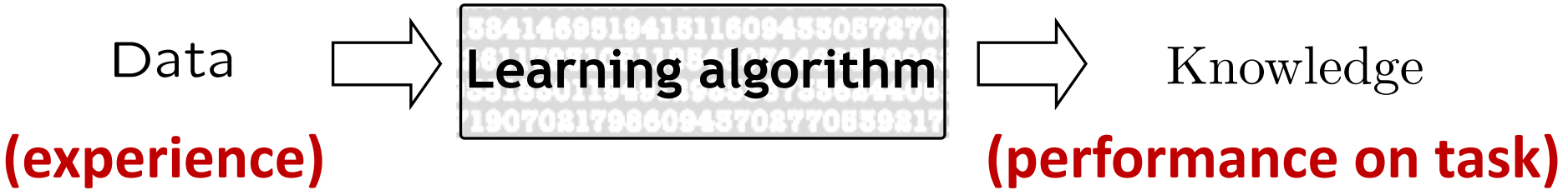
Enjoy!

- 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…

What is Machine Learning?

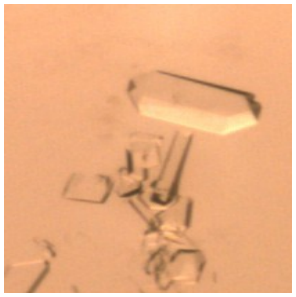
Design and Analysis of algorithms that

- improve their performance
- at some task
- with experience

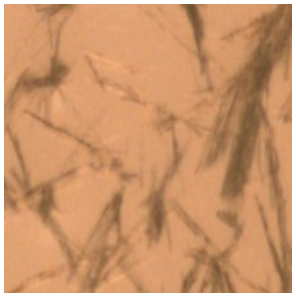


Human learning

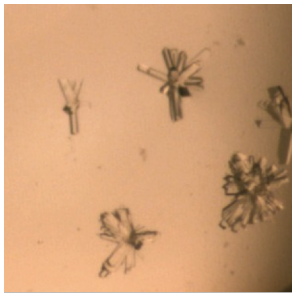
Task: Learning stage of protein crystallization



Crystal

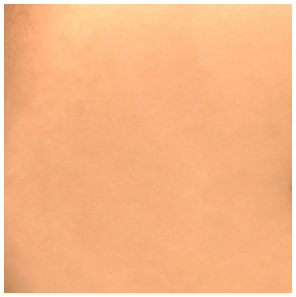


Needle

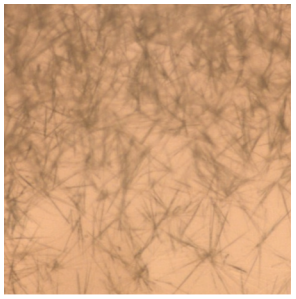


Tree

Tree

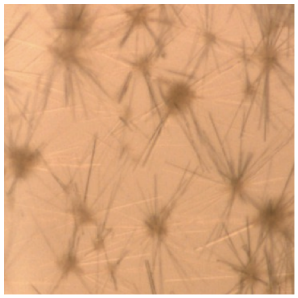


Empty



Needle

Experience



?

Performance

Tasks, Experience, Performance

Tasks, Experience, Performance

Machine Learning Tasks

Broad categories -

- **Supervised learning**

  Classification, Regression

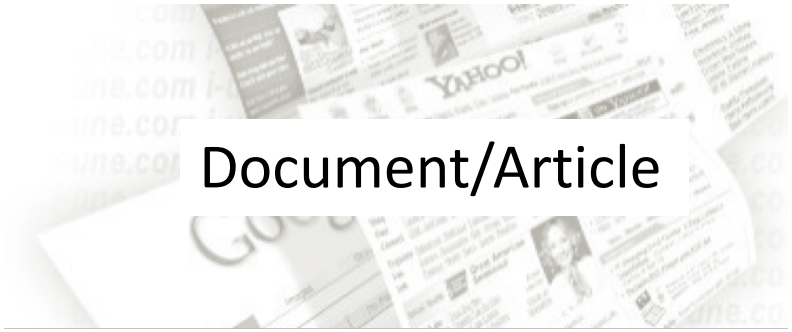
- **Unsupervised learning**

  Density estimation, Clustering, Dimensionality reduction

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

Supervised Learning

Input $X \in \mathcal{X}$



Document/Article

Label $Y \in \mathcal{Y}$

"Sports"
"News"
"Science"
…

Discrete Labels
Classification

Market information

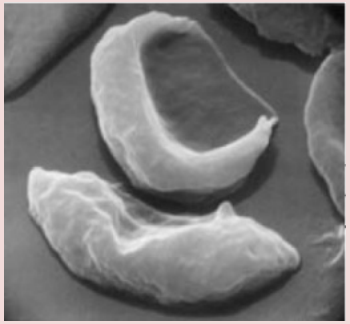
Share Price
"$ 24.50"

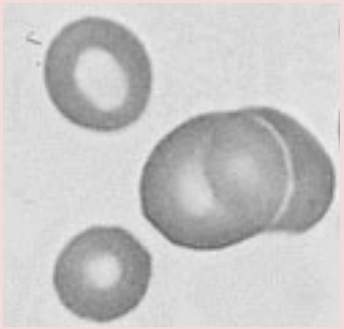
Continuous Labels
Regression

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

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

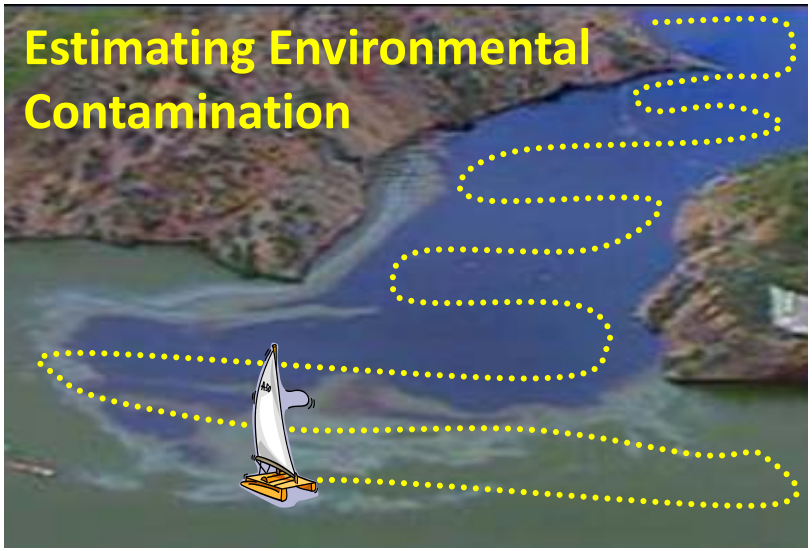
Classification or Regression?

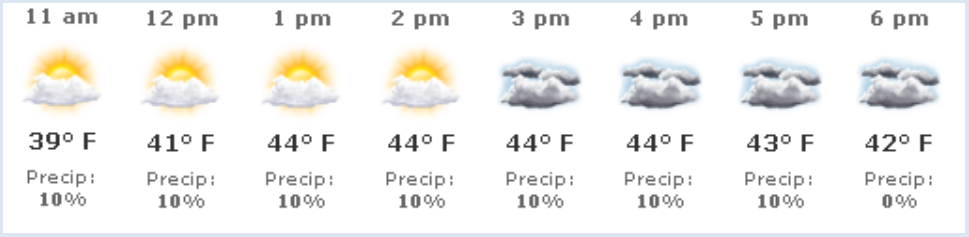




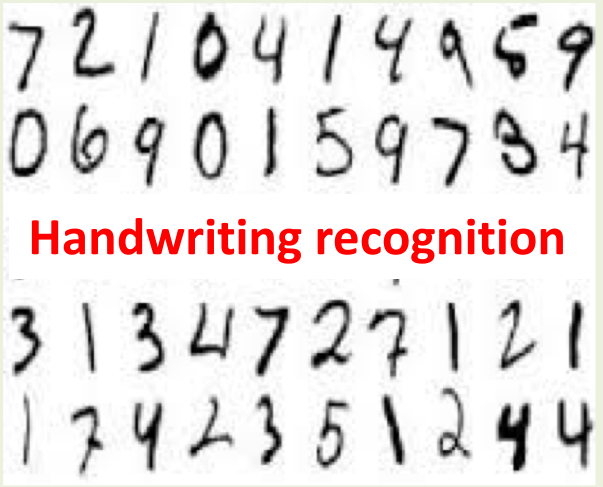
Medical Diagnosis

“Anemic”
“Healthy”





Weather prediction



Handwriting recognition

Unsupervised Learning

Aka "learning without a teacher"

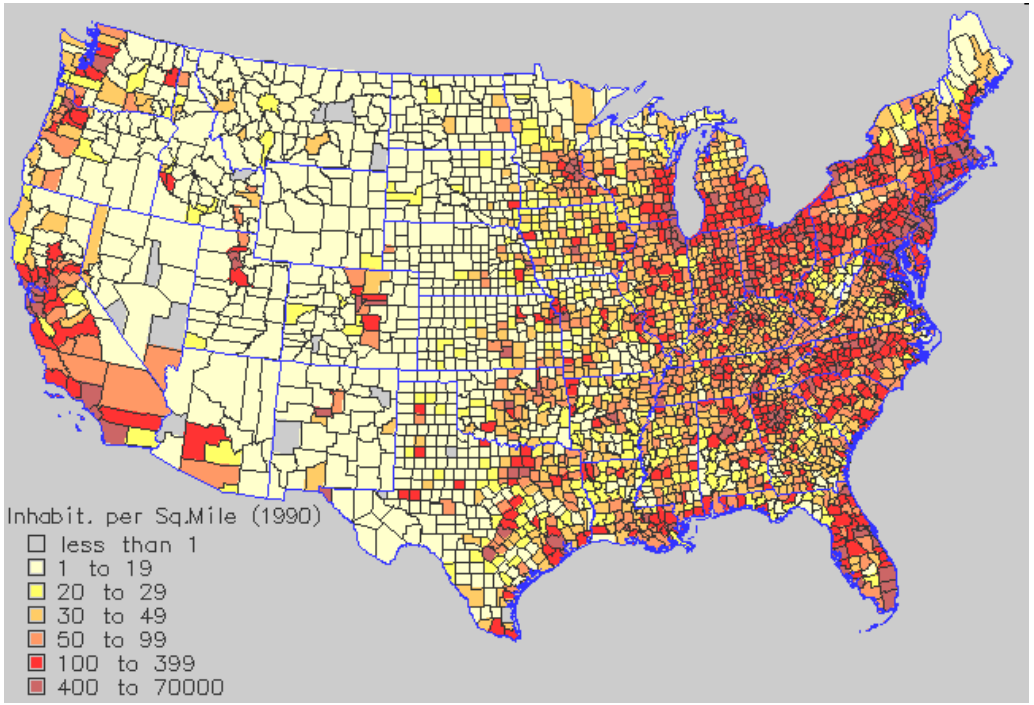
Input $X \in \mathcal{X}$

Word distribution
(Probability of a word)

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

Unsupervised Learning

Density/Distribution Estimation



Population density

Bias of a coin

Unsupervised Learning

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

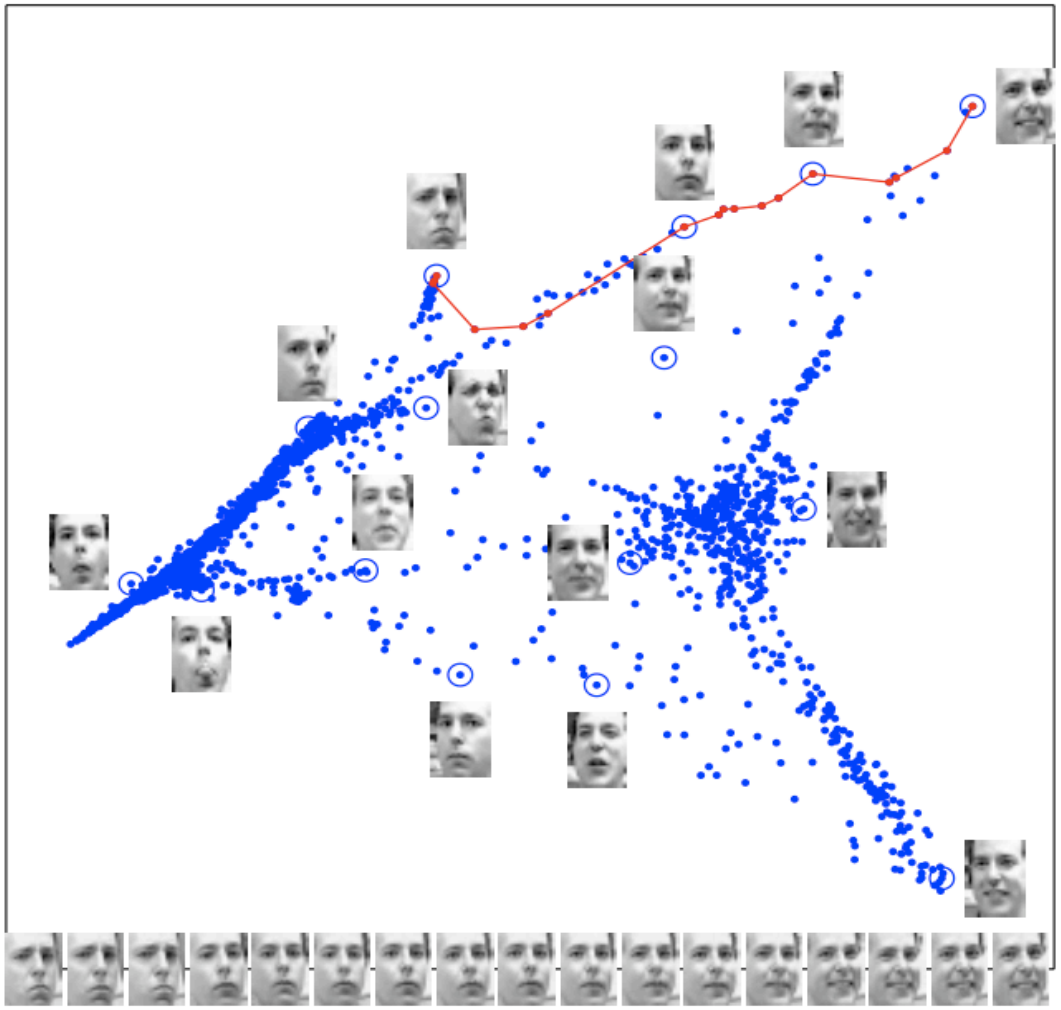
Unsupervised Learning

Dimensionality Reduction/Embedding

[Saul & Roweis '03]

Images have thousands or millions of pixels.

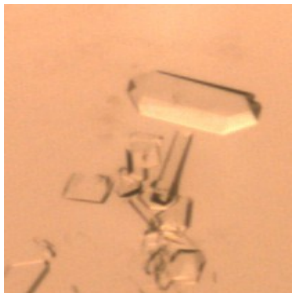
Can we give each image a small set of coordinates, such that similar images are near each other?



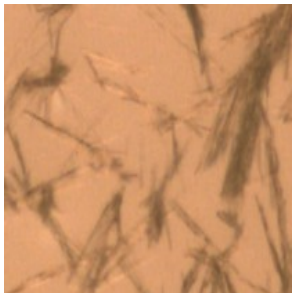
Tasks, **Experience**, Performance

Training Data vs. Test Data

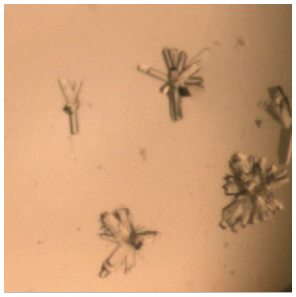
Task: Learning stage of protein crystallization



Crystal

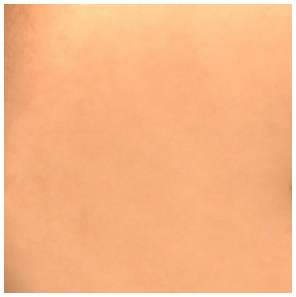


Needle

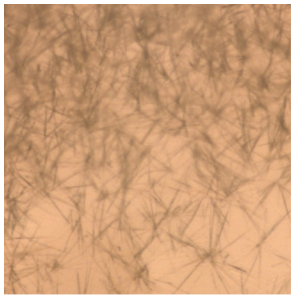


Tree

Tree

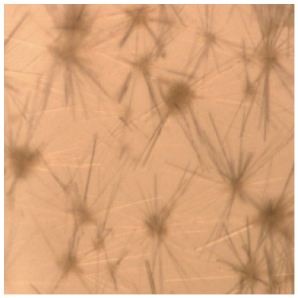


Empty



Needle

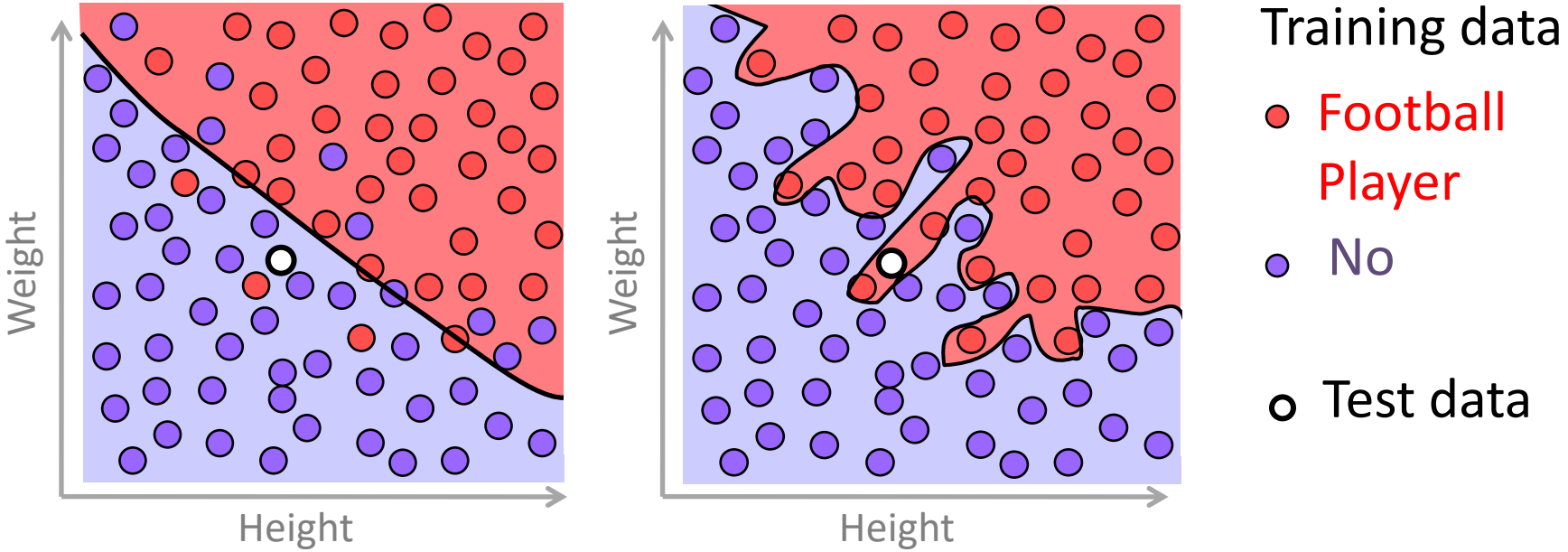
Experience



?

Performance

Training Data vs. Test Data



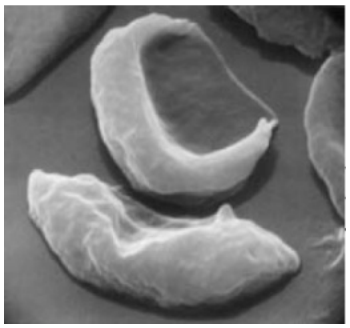
- A good machine learning algorithm
  - Does not **overfit** training data
  - **Generalizes** well to test data

Tasks, Experience, **Performance**

Performance Measure

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between true label *Y* and prediction *f*(*X*)

X	Diagnosis, *Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
	"Anemic cell"	"Anemic cell"	0
		"Healthy cell"	1

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

Performance Measure

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between true label *Y* and prediction *f*(*X*)

X	Share price, *Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
Past performance, trade volume etc. as of Sept 8, 2010	"$24.50"	"$24.50"	0
		"$26.00"	1?
		"$26.10"	2?

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

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**