

Machine Learning - Intro

Aarti Singh

Co-instructor: Barnabas Poczos

Machine Learning 10-401

Jan 12, 2016



MACHINE LEARNING DEPARTMENT



Teaching team

Instructors:



Aarti Singh



Barnabas Poczoz

TAs:



Yining Wang



Kirstin Early

Logistics

- Lectures: Tues, Thurs 12:00-1:20 pm GHC 4303
- Recitations: Wed 6:30-7:30 pm GHC 4303
- Office hours: Barnabas (Tues), Aarti (Thurs) after class outside classroom
Yining (Mon 2:30-3:30 pm outside GHC 8021)
Kirstin (Fri 4-5 pm outside GHC 8011)
- Piazza: course webpage, slides, videos, announcements, policies, office hours, HWs, projects, solutions, ...
discussion forum for students
- <https://piazza.com/cmu/spring2016/10401/home>

Communication channel

- Ask questions in class!
- Office hours (M, T, Th, F)
- Recitations (W)
- Note to instructors on Piazza (limited)
- DO NOT expect moderation by instructors or TAs on Piazza discussion forum

Grading

- Grading
 - 3 homework assignments ($3 \times 15\% = 45\%$)
 - 1 midterm, 1 in-class final: ($2 \times 10 = 20\%$)
 - paper implementation and presentation (15%)
 - project (20%)
- Late days
 - total 4 for homeworks and project, use any way you like
 - no late days for paper implementation and presentation

Homework

- 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
- Homeworks are hard, start early 😊
- Due in the beginning of class

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
 - NOT allowed

About the course

- Machine Learning **Algorithms and Principles**
 - Classification: Naïve Bayes, Logistic Regression, Perceptron, Support Vector Machines, k-NN
 - Regression: Linear regression, Gaussian Process regression, Kernel regression
 - Kernel density estimation, Hidden Markov Models, Graphical Models, k-means clustering, dimensionality reduction, neural networks, deep belief networks, Boosting, Decision Trees, etc.
 - Optimization, Theory, Model selection, overfitting, bias-variance tradeoffs ...
- See **tentative** lecture outline on Piazza – MAY CHANGE
- Material: Class slides + Reading material

About the course

- 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 – Intro to ML algorithms and principles
 - 10-601 – Masters version
 - 10-701 – PhD version
 - 10-715 – PhD students doing research in machine learning (hardest, most mathematical)

Other related courses:

10-605, 10-805 – Machine Learning with Large Datasets
11-663 – Machine Learning in Practice (ML software to solve problems)
10-702, 10-704, 10-708, 15-859(B) – related advanced topics

Pre-requisites

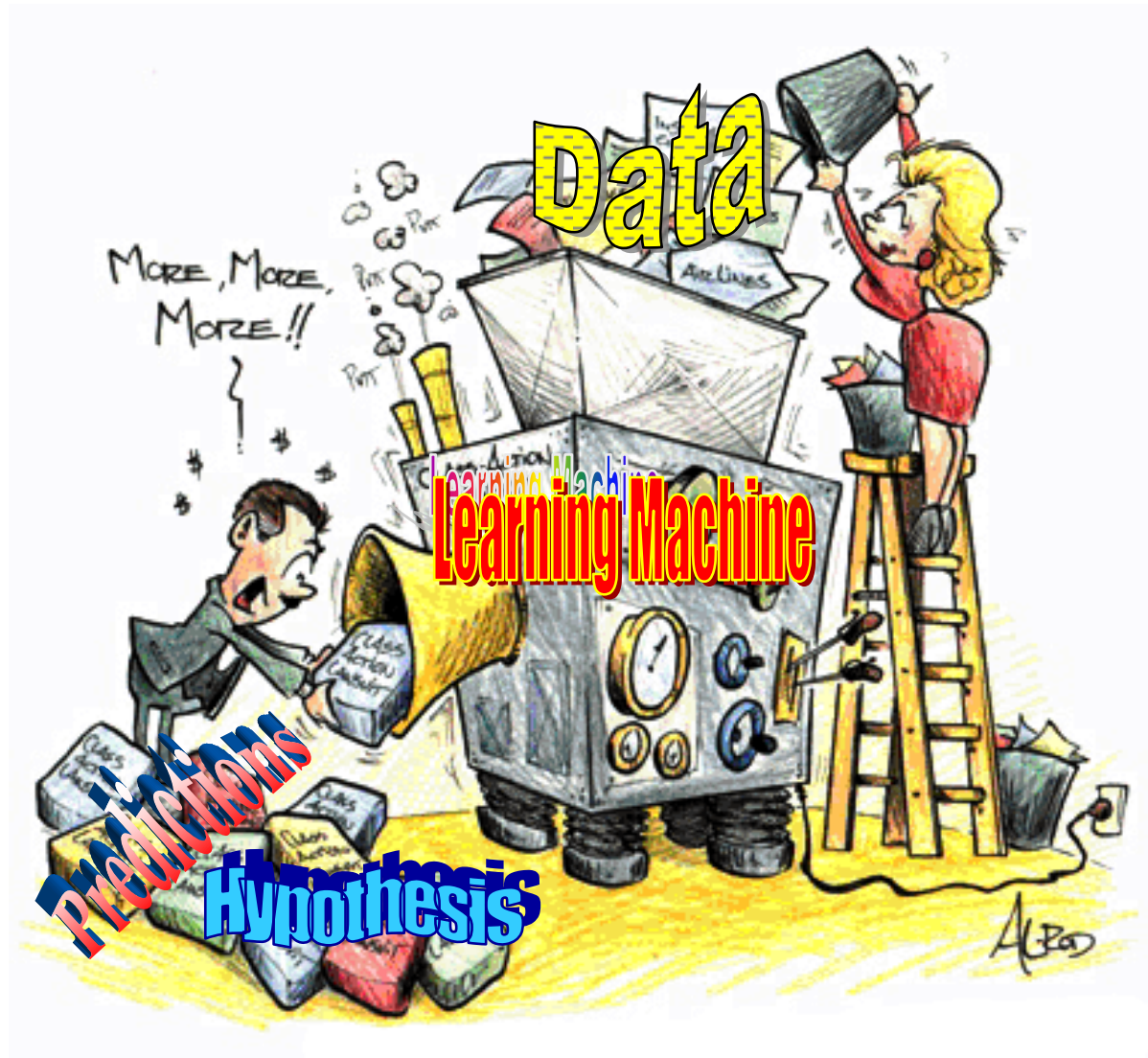
- **Assume mathematical maturity**
 - Multivariate Calculus
Derivatives, integrals of multi-variate functions
 - Linear Algebra
Matrix inversions, eigendecomposition, ...
 - Basic Probability and Statistics
Probability distributions, Mean, Variance, Conditional probabilities, Bayes rule, ...
- **Tutorial videos**
 - Probability, Calculus, Functional Analysis, SVD
<http://www.cs.cmu.edu/~aramdas/videos.html>
 - Linear Algebra
<http://www.cs.cmu.edu/~zkolter/course/linalg/index.html>

Recitations

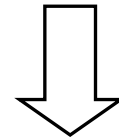
- Strongly recommended
 - Brush up pre-requisites
 - Review material (difficult topics, clear misunderstandings, extra new topics)
 - Ask questions
- 1st Probability Review - **TOMORROW**
by Kirstin Early
Wed Jan 13 6:30-7:30 pm GHC 4303

What is Machine Learning?

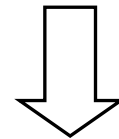
What is Machine Learning?



Data



Learning algorithm



Knowledge

From Data to Knowledge ...

Machine Learning in Action

Machine Learning in Action

- Spam filtering

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.

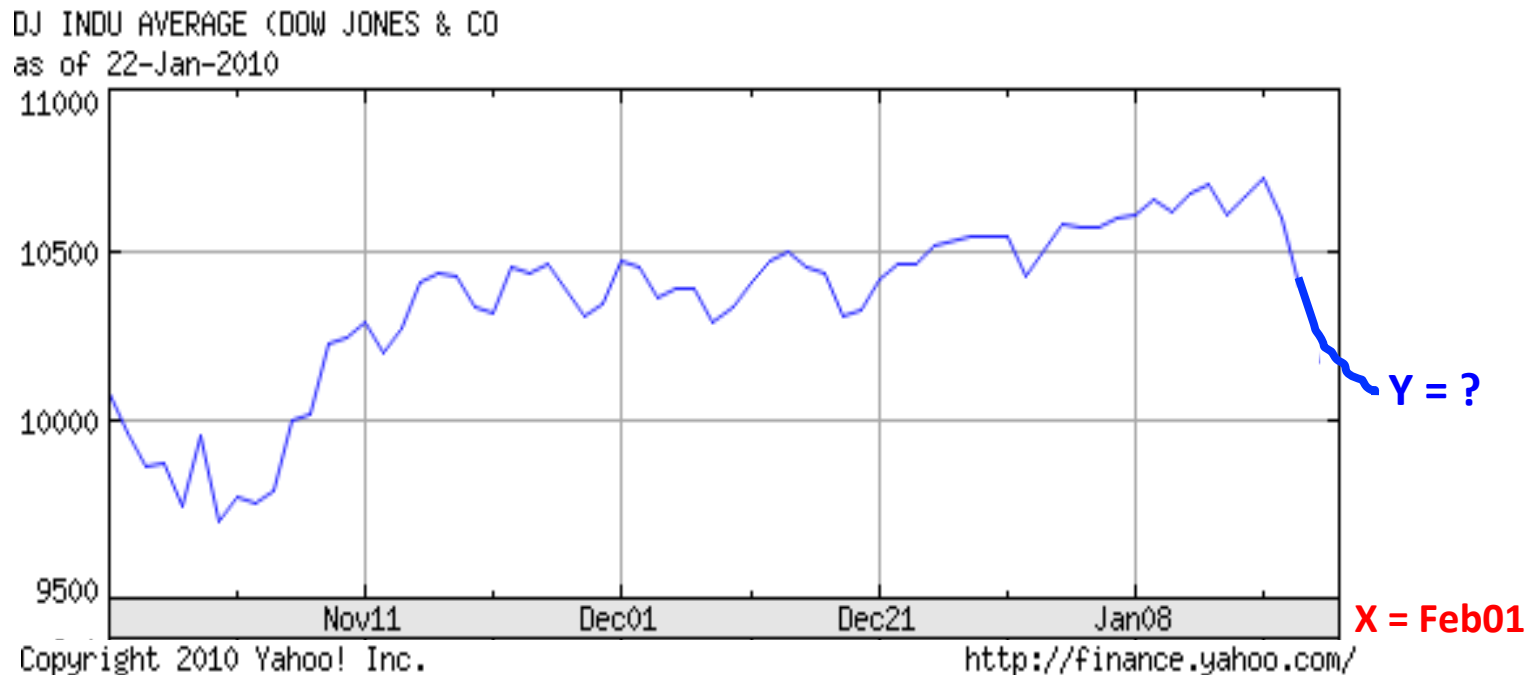
- * Rapid WeightLOSS
- * Increased metabolism - BurnFat & calories easily!
- * Better Mood and Attitude



Spam/
Not spam

Machine Learning in Action

- Stock Market Prediction



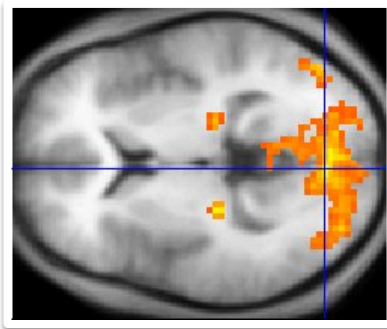
Machine Learning in Action

- 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)

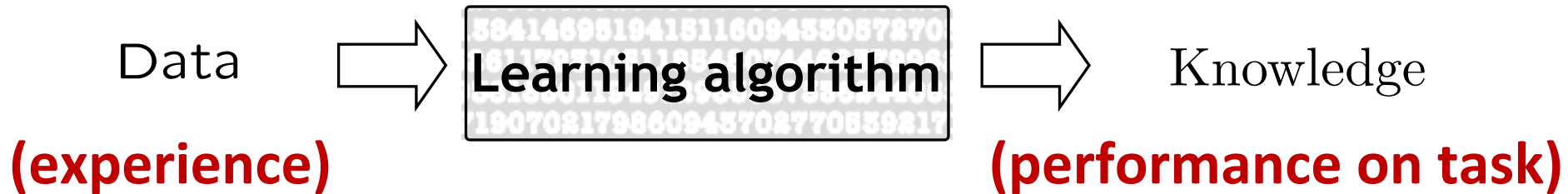
Enjoy!

- 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

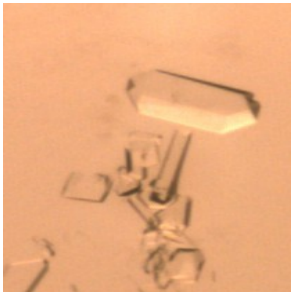
- improve their performance
- at some task
- with experience



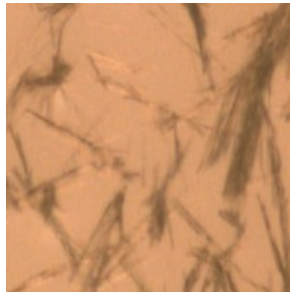
Human learning



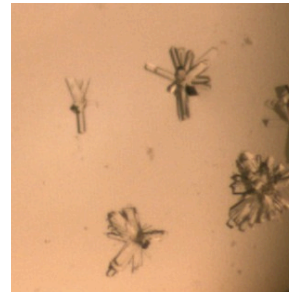
Task: Learning stage of protein crystallization



Crystal



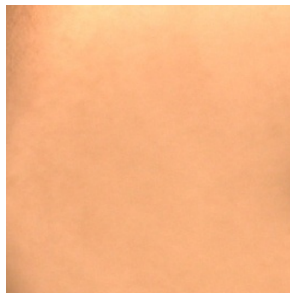
Needle



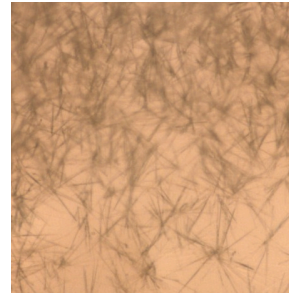
Tree



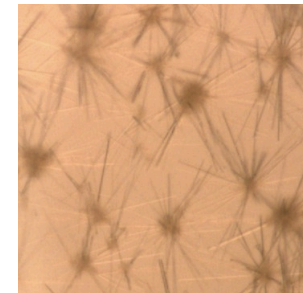
Tree



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



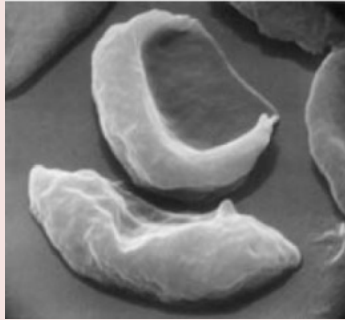
Share Price
"\$ 24.50"

Continuous Labels
Regression

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

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

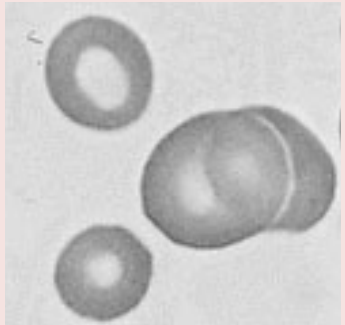
Classification or Regression?



Medical Diagnosis



"Anemic"
"Healthy"



11 am	12 pm	1 pm	2 pm	3 pm	4 pm	5 pm	6 pm
39° F	41° F	44° F	44° F	44° F	44° F	43° F	42° F
Precip: 10%	Precip: 10%	Precip: 10%	Precip: 10%	Precip: 10%	Precip: 10%	Precip: 10%	Precip: 0%

Weather prediction

7210414959
0690159734

Handwriting recognition

3134727121
1742351244

Unsupervised Learning

Aka “learning without a teacher”

Input $X \in \mathcal{X}$



Document/Article

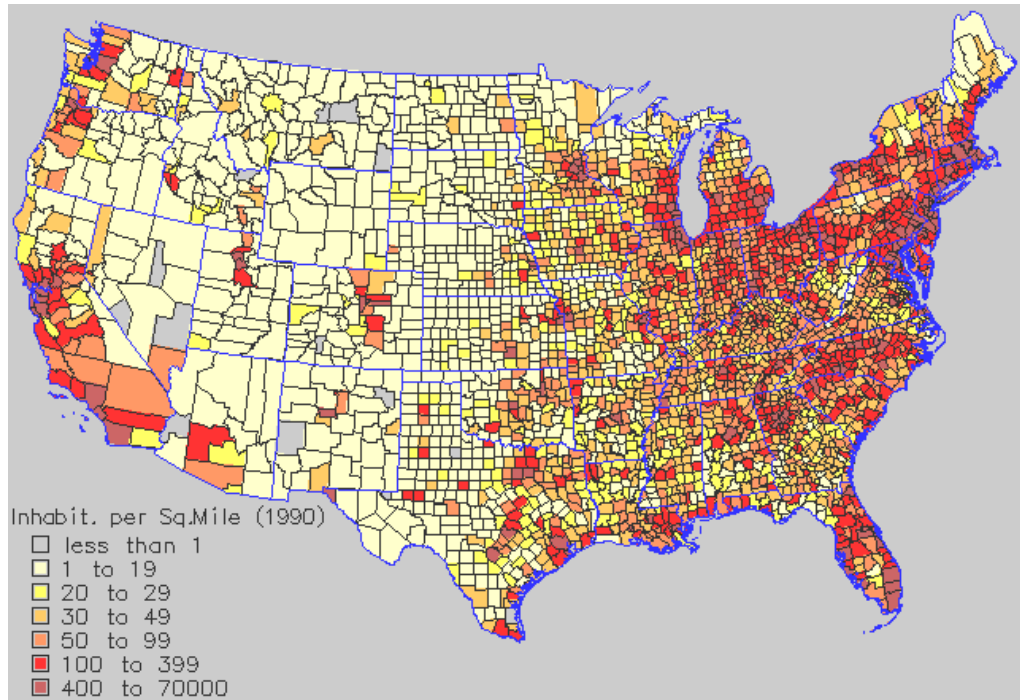


Word distribution
(Probability of a word)

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.]



C_4



C_2



C_3



C_5

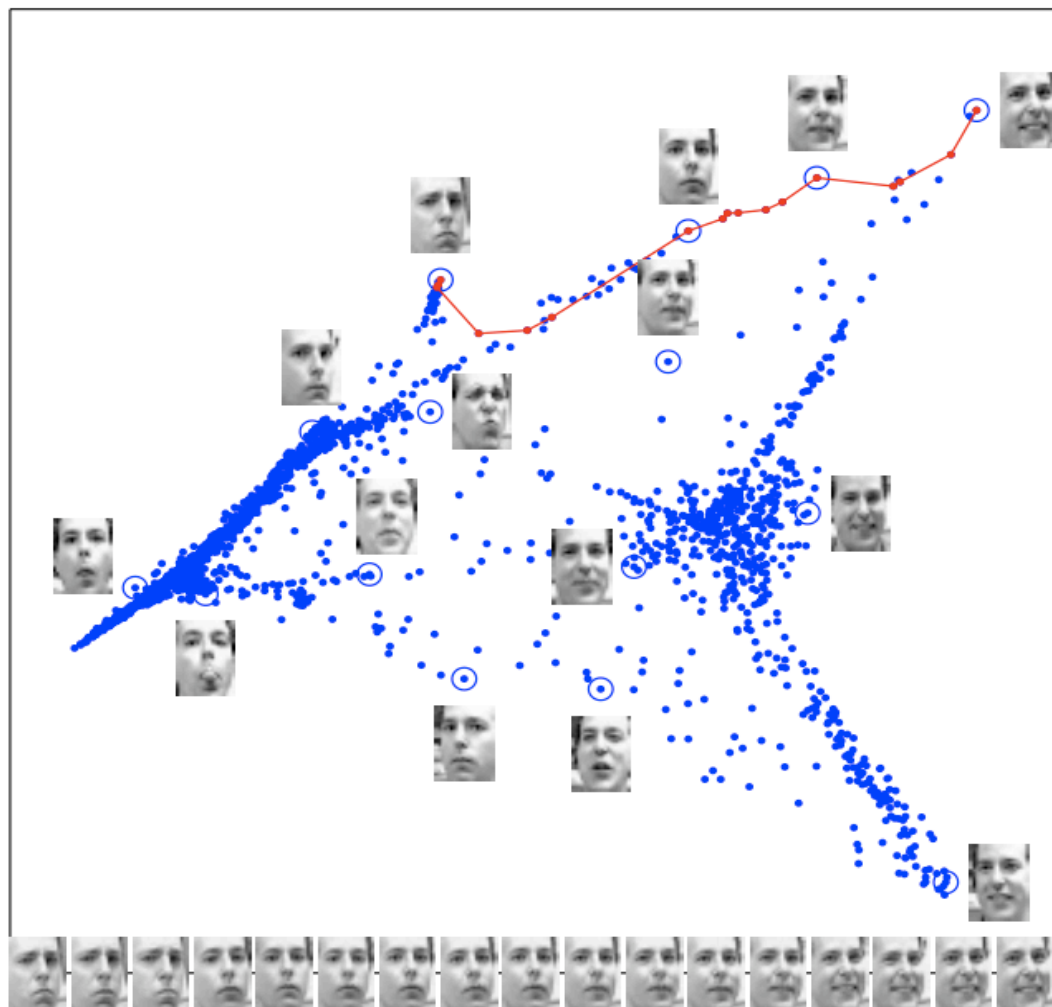
Unsupervised Learning

Dimensionality Reduction/Embedding

[Saul & Roweis '03]

Images have thousands or millions of pixels.

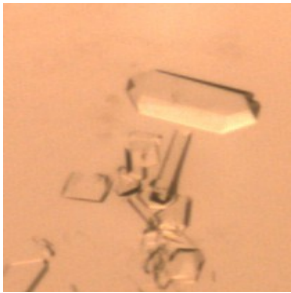
Can we give each image a small set of coordinates, such that similar images are near each other?



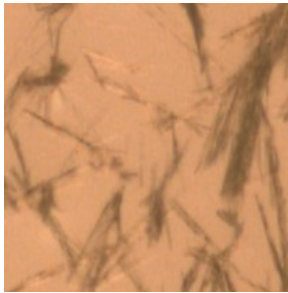
Tasks, **Experience**, Performance

Training Data vs. Test Data

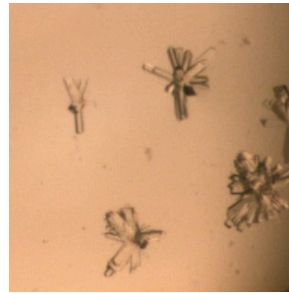
Task: Learning stage of protein crystallization



Crystal



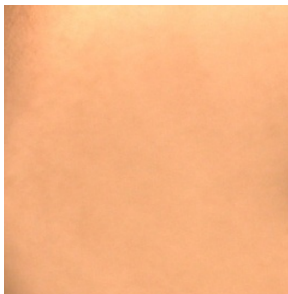
Needle



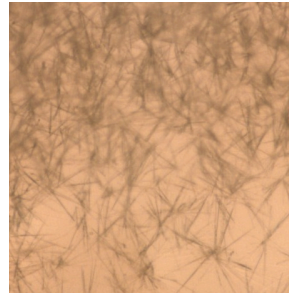
Tree



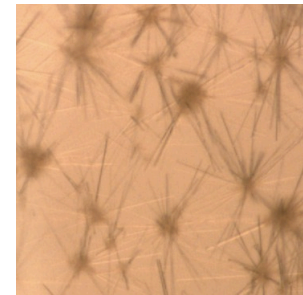
Tree



Empty



Needle

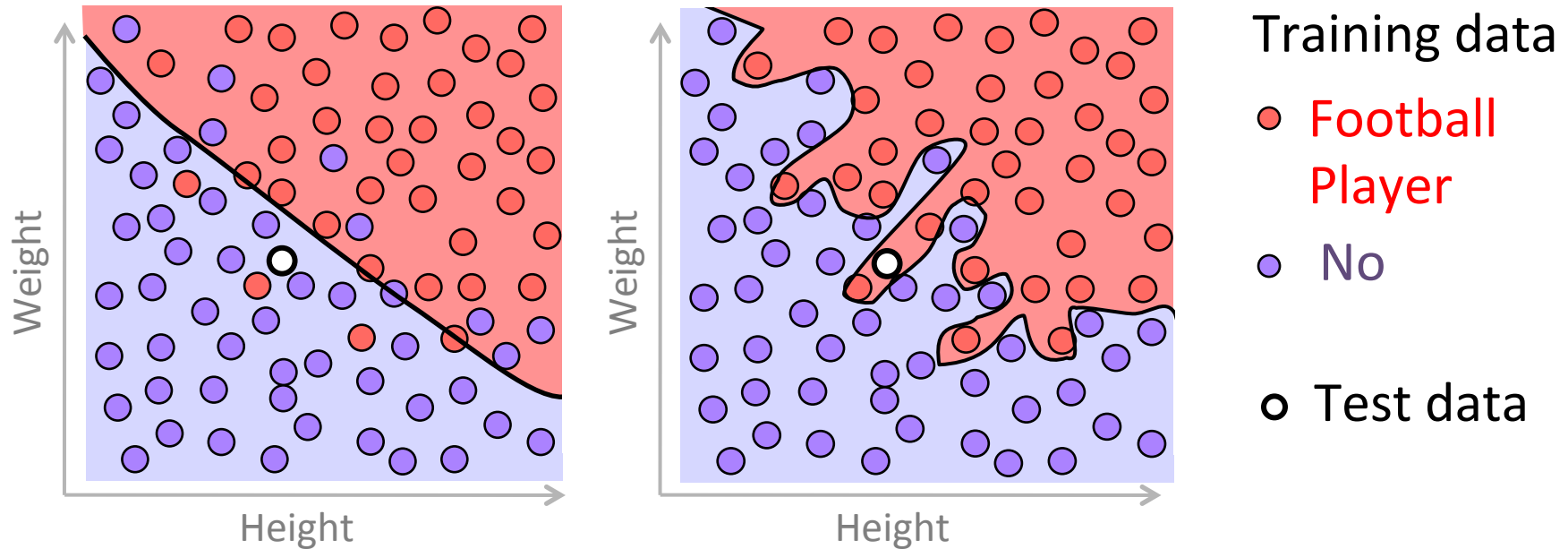


?

Experience

Performance

Training Data vs. Test Data



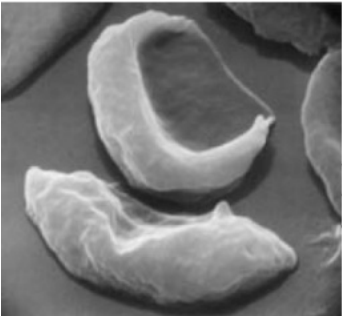
- A good machine learning algorithm
 - Does not **overfit** training data
 - **Generalizes** well to test data

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"	"Healthy cell"	1

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

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?

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

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**