

Machine Learning - Intro

Aarti Singh

Machine Learning 10-315
Jan 19, 2022



MACHINE LEARNING DEPARTMENT

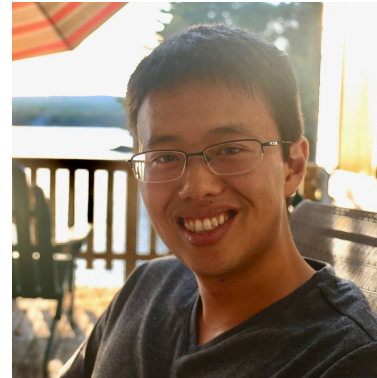


Teaching team

Instructor:

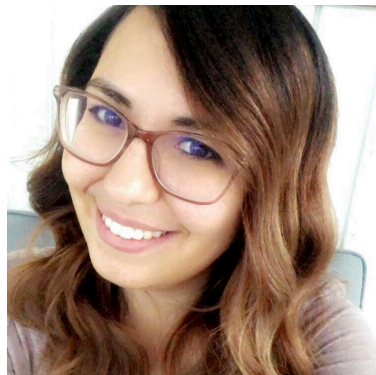


Aarti



Henry

Educational
Associate:



Fatima

Admin:



Mary

Teaching team

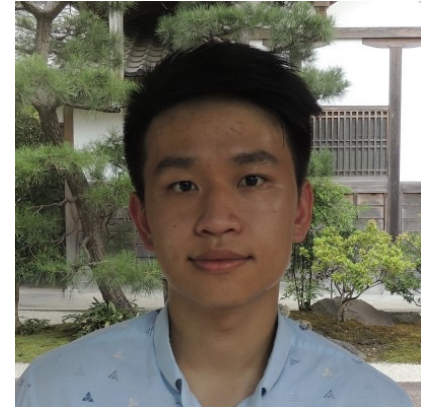
TAs:



Devanshi



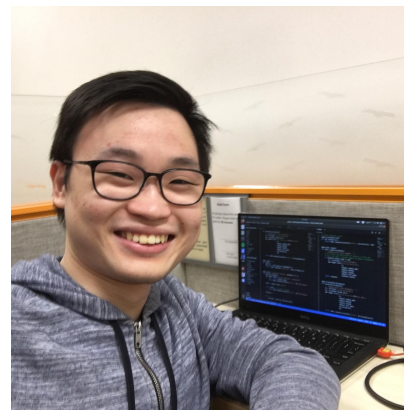
Julia



Owen



Roochi



Sedrick



Tarun

Logistics

Lectures: Mon, Wed 11:50 am-1:10 pm

Recitations: Fri 11:50 am-1:10 pm

In person: HOA 160 Zoom links: [Canvas](#)

Office hours: TBA

Lectures and recitations will be recorded (**Panopto** link on Canvas).
Strictly for your use only.

Breakout rooms and Office hours will NOT be recorded.

Logistics

Webpage: https://www.cs.cmu.edu/~aarti/Class/10315_Spring22/

Syllabus, policies, schedule of lectures, recitations, office hours, slides, reading material, homeworks, ...

Piazza: <https://piazza.com/class/ky21u0jv2op2xc>

announcements, questions for Teaching team, discussion forum for students

Homework submission: [Gradescope](#)

Grades: [Canvas](#)

Polls: Zoom (log in **without audio**) even if you are in class

Expectations

- Remote students
 - Keep yourself muted unless asking or responding to questions
- Interact!
 - Ask questions in class by raising hand or via Zoom chat
 - Respond to questions in class by raising hand or via Zoom chat
 - Anonymous polls on Zoom
- In-person conduct (lectures, recitations, office hours)
 - Follow CMU guidance on masks, distancing and physical space, etc.
 - Do not come to class if you have symptoms of ANY infectious illness

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 Devanshi and Owen
 - Fri Jan 21 11:50 am-1:10 pm

Grading

- Grading
 - 4 homework assignments ($4 \times 12\% = 48\%$)
 - 4 QnAs (20%)
 - 1 midterm, 1 final (both in class): ($15+15 = 30\%$)
 - Participation (2%) Zoom polls
- Late days
 - total 6 across ALL homeworks and QnAs
 - no more than 1 for QnA, 2 for HW
 - No partial credit after late days
 - late days are for unforeseen situations (interviews, conference, etc.), do NOT include them in your plan

Homeworks & QnAs

- Collaboration
 - You may **discuss** the questions
 - Each student **MUST** write 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

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, recitations and office hours
virtually and doing HW

- Audits and Pass/Fail

Audits NOT allowed

Pass/Fail allowed – talk to your academic advisor

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
 - 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
(available online)
 - Machine Learning: A probabilistic perspective, Kevin Murphy
(available online)
 - Machine Learning, Tom Mitchell
 - The elements of statistical learning: Data mining, inference
and prediction, Trevor Hastie, Robert Tibshirani, Jerome
Friedman

Pre-requisites

- **Assume mathematical maturity**

- Basic Probability and Statistics

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

- Programming (python) and principles of computing

- Multivariate Calculus

- Derivatives, integrals of multi-variate functions

- Linear Algebra

- Matrix inversions, eigendecomposition, ...

- **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>

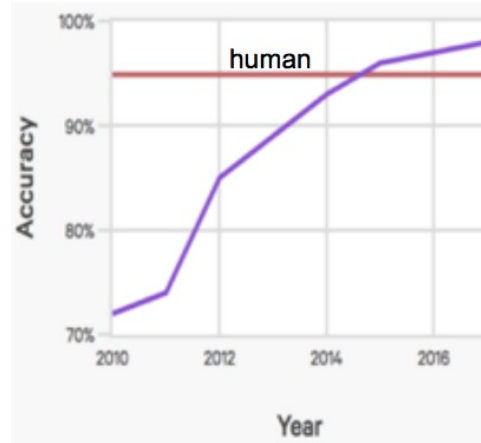
Related courses

- 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(A/B) – related advanced topics

Machine Learning in Action

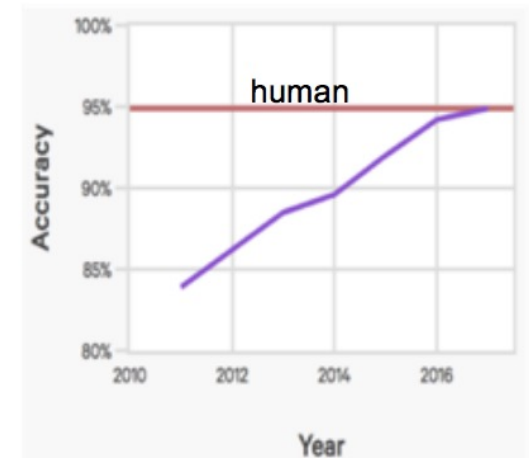
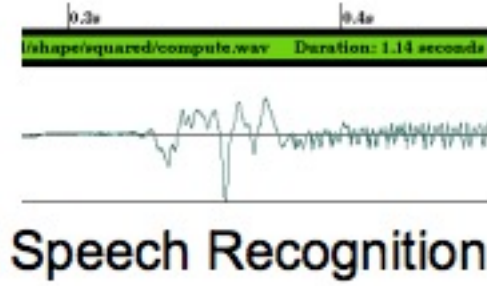


Computer vision

Machine Learning in Action



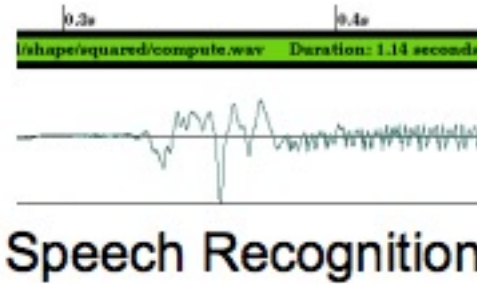
Computer vision



Machine Learning in Action



Computer vision



Robotic control

Text analysis

Peter H. van Oppen, Chairman of the Board & Chief Executive Officer
Mr. van Oppen has served as Chairman of the Board and Chief Executive Officer of ADIC since its acquisition by Interpoint in 1994 and a director of ADIC since 1986. Until its acquisition by Crane Co. in October 1996, Mr. van Oppen served as President of the Board of Directors, Chairman and Chief Executive Officer of Interpoint. Prior to 1985, Mr. van Oppen worked as a consultant at Price Waterhouse LLP and at Bain & Company in Boston and London. He has additional experience in medical electronics and venture capital. Mr. van Oppen also serves as a director of Cytosine Technologies and Spacelabs Medical, Inc.. He holds a B.A. from Whitman College and an M.B.A. from Harvard Business School, where he was a Baker Scholar.



Games & Reasoning



Protein folding

Machine Learning in Action

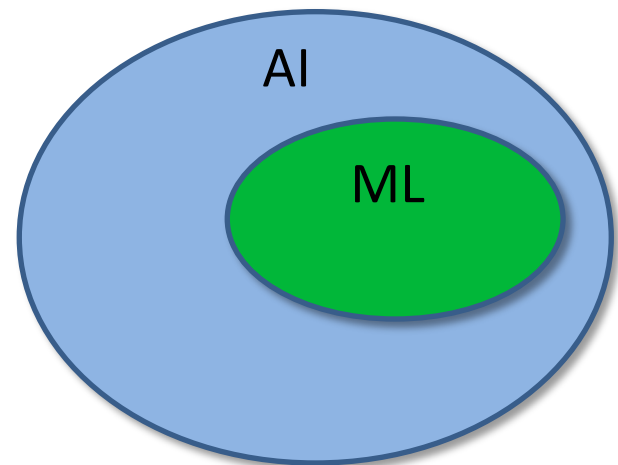
- How have you interacted with ML in your daily life so far?

ML is ubiquitous

- Wide applicability
- Software too complex to write by hand
- Improved machine learning algorithms
- Improved data capture, networking, faster computers
- Demand for self-customization to user, environment

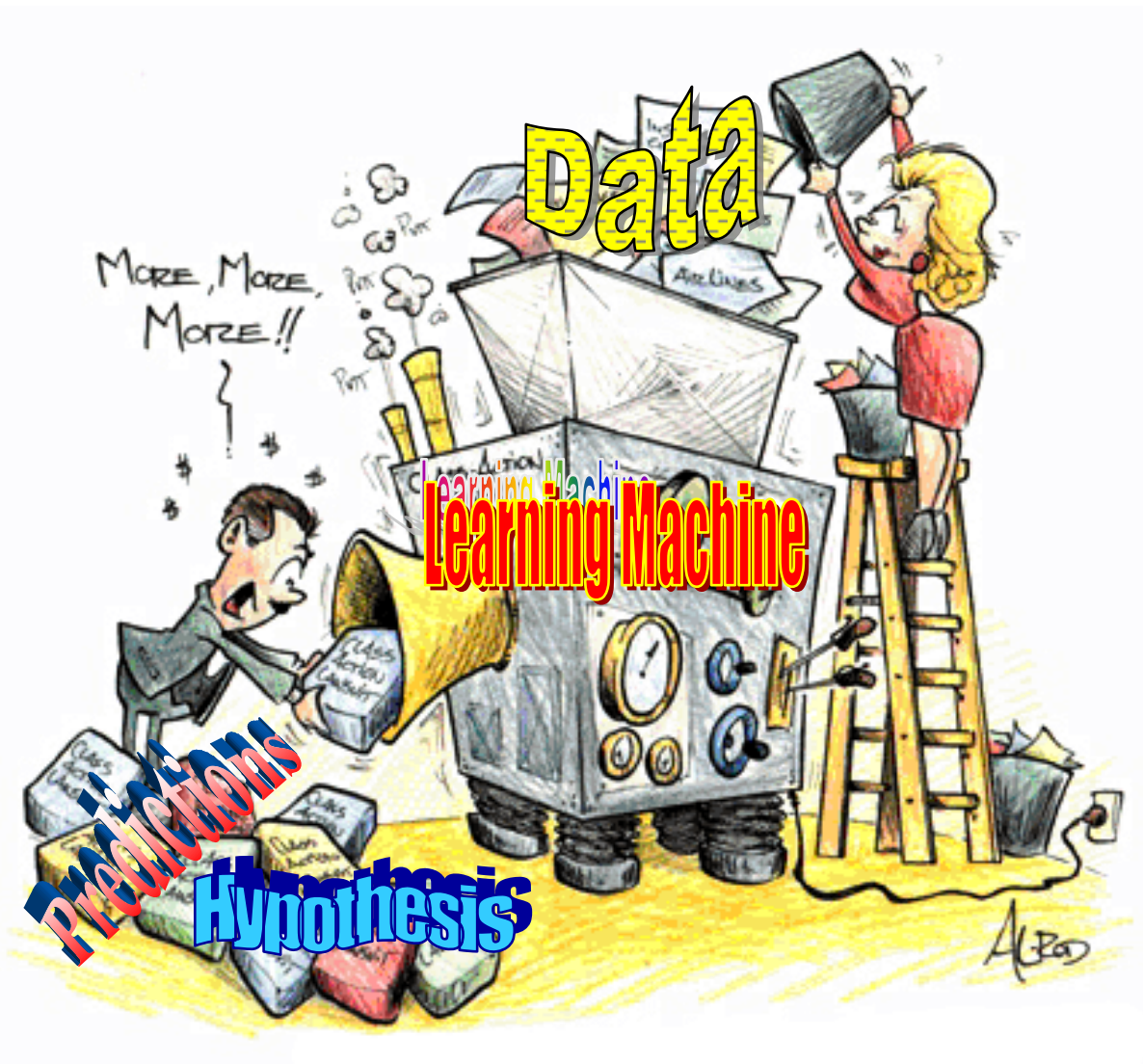
AI: develop intelligent agents

ML: learn to generalize using data

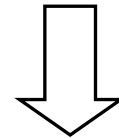


Fun begins ...

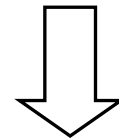
What is Machine Learning?



Data



Learning algorithm

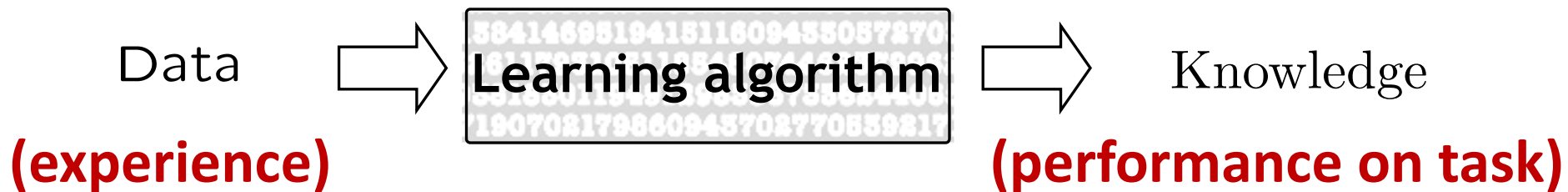


Knowledge

What is Machine Learning?

Design and Analysis of algorithms that

- improve their performance
- at some task
- with experience



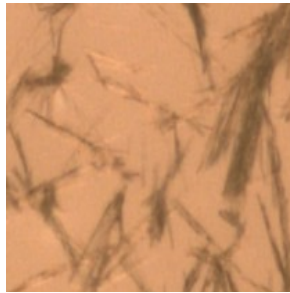
Human learning



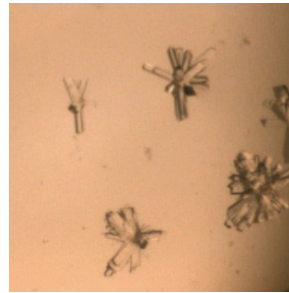
Task: Learning stage of protein crystallization



Crystal



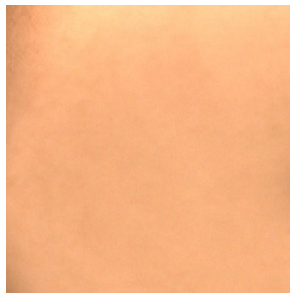
Needle



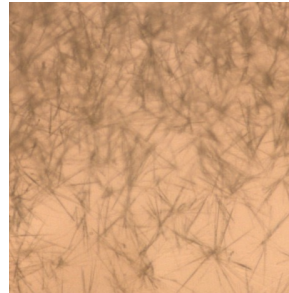
Tree



Tree

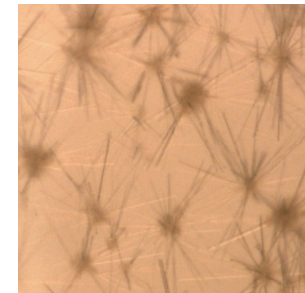


Empty



Needle

➤ Predict the label of the test image?



?

Experience

Performance

Tasks, Experience, Performance

Tasks, Experience, Performance

Machine Learning Tasks

Broad categories -

- **Supervised learning**

Classification, Regression

- **Unsupervised learning**

Density estimation, Clustering, Dimensionality reduction

- Graphical models
- Semi-supervised learning
- Active learning
- Bayesian optimization
- Reinforcement learning
- Many more ...

Supervised Learning

Input $X \in \mathcal{X}$

Document/Article

Label $Y \in \mathcal{Y}$

"Sports"
"News"
"Science"

...

Discrete Labels
Classification

DJ INDU AVERAGE (DOW JONES & CO
as of 22-Jan-2010



Copyright 2010 Yahoo! Inc.

<http://finance.yahoo.com/>

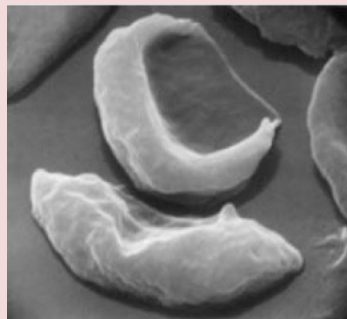
Share Price
"\$ 24.50"

Continuous Labels
Regression

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

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

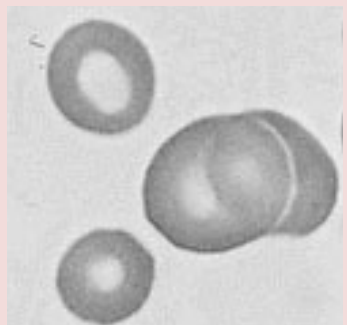
Classification or Regression?



Medical Diagnosis



“Anemic”
“Healthy”



Estimating Environmental
Contamination



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

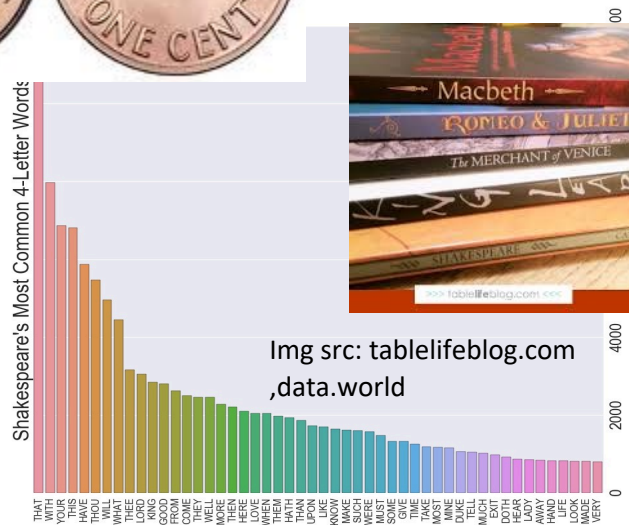
Learning a Distribution

Clustering



Bias of a coin

Distribution of words in text



➤ What other distribution would be interesting to learn?

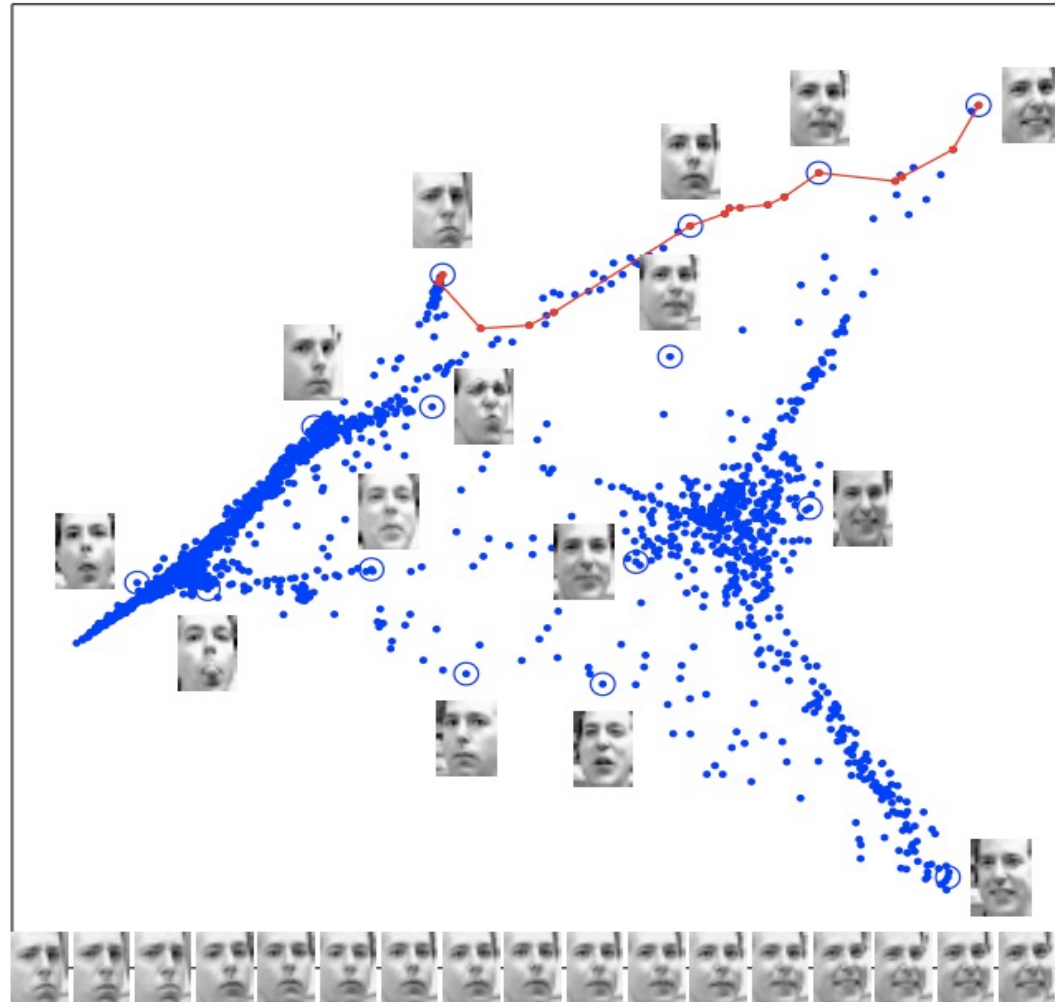
Unsupervised Learning

Dimensionality Reduction/Embedding

[Saul & Roweis '03]

Images have thousands or millions of pixels.

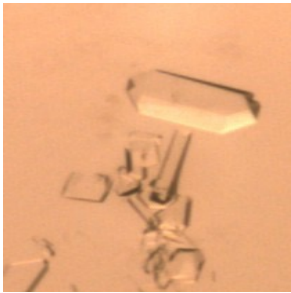
Can we give each image a small set of coordinates, such that similar images are near each other?



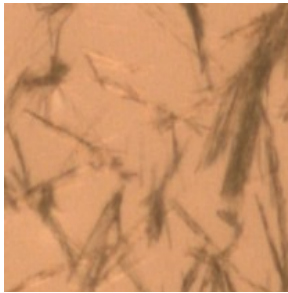
Tasks, **Experience**, Performance

Experience = Training Data

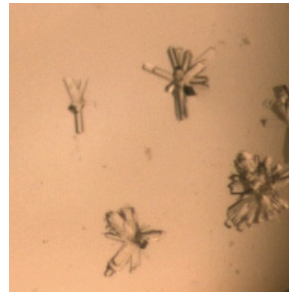
Task: Learning stage of protein crystallization



Crystal



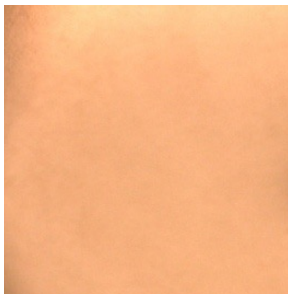
Needle



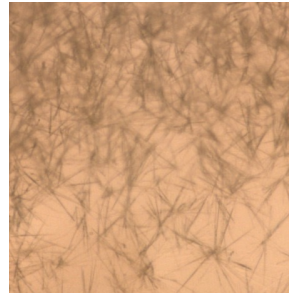
Tree



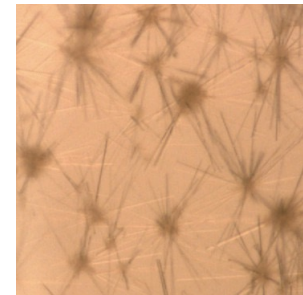
Tree



Empty



Needle



?

Experience

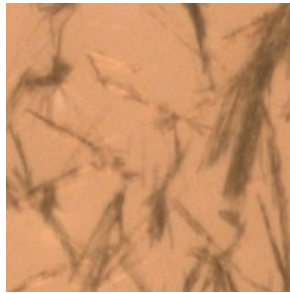
Performance

Training Data vs. Test Data

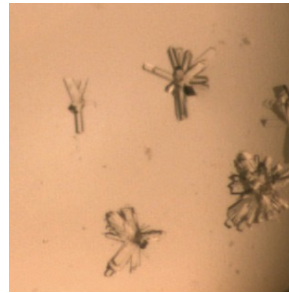
Task: Learning stage of protein crystallization



Crystal



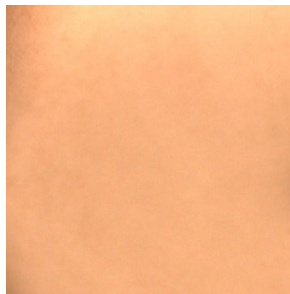
Needle



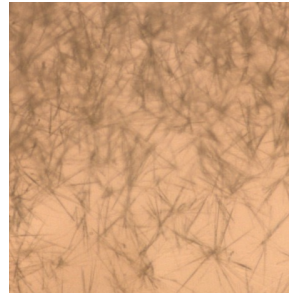
Tree



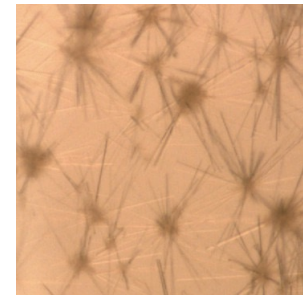
Tree



Empty



Needle



?

Experience

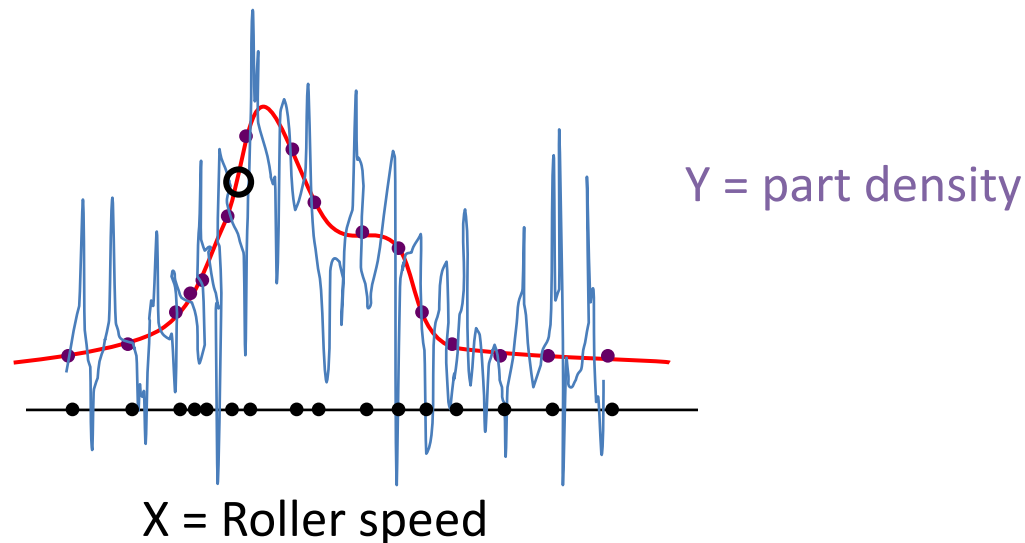
Performance

Generalization & Overfitting

A good ML algorithm

should: **generalize** aka perform well on test data

should not: **overfit** the training data



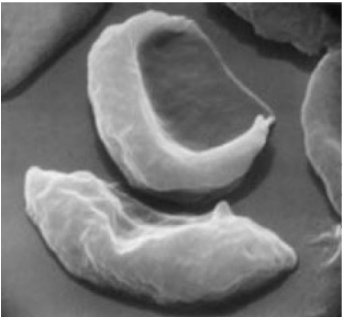
➤ Critical to report test and NOT training data accuracy

Tasks, Experience, **Performance**

Performance Measure

Performance:

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

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

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

Performance Measure

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between label Y and prediction $f(X)$ for test data 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?

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

Performance Measure

For test data X , measure of closeness between label Y and prediction $f(X)$

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

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

We will talk about more performance measures including for unsupervised learning later in course.

Poll

- A classifier with 100% accuracy on training data and 70% accuracy on test data is better than a classifier with 80% accuracy on training data and 80% accuracy on test data.
A. True B. False
- Which classifier is better, given following statistics on test accuracy?

	Mean	Best run	Std
Classifier A	92%	97%	15%
Classifier B	87%	100%	5%

Glossary of Machine Learning

- Task
- Supervised learning
 - Classification
 - Regression
- Unsupervised learning
 - Learning distribution
 - Clustering
 - Dimensionality reduction/Embedding
- Input, X
- Label, Y
- Prediction, $f(X)$
- Experience = Training data
- Test data
- Overfitting
- Generalization
- Performance measure/loss – 0/1, squared