Machine Learning - Intro

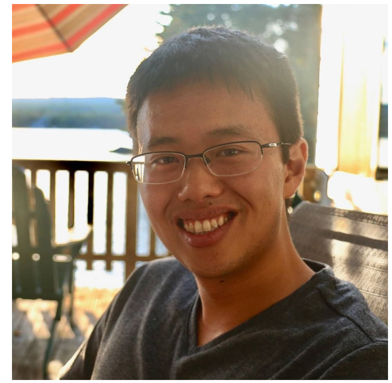
Aarti Singh

Machine Learning 10-315
Jan 19, 2022

Teaching team

Instructor:

Aarti



Henry

Educational Associate:

Fatima

Admin:

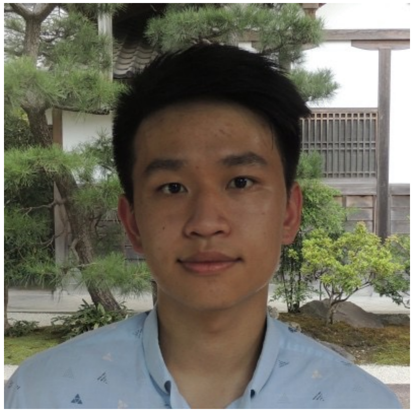
Mary

Teaching team

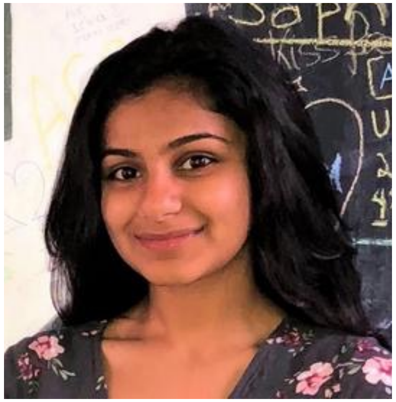
TAs:

Devanshi

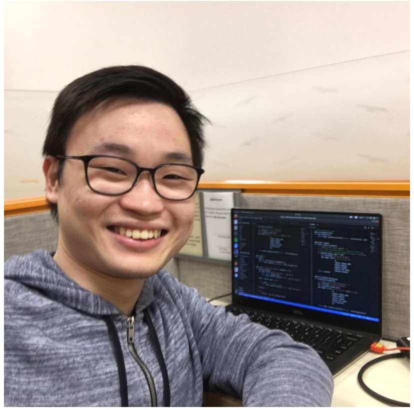
Julia



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

- <u>Strongly recommended</u>
  - 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 x 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:

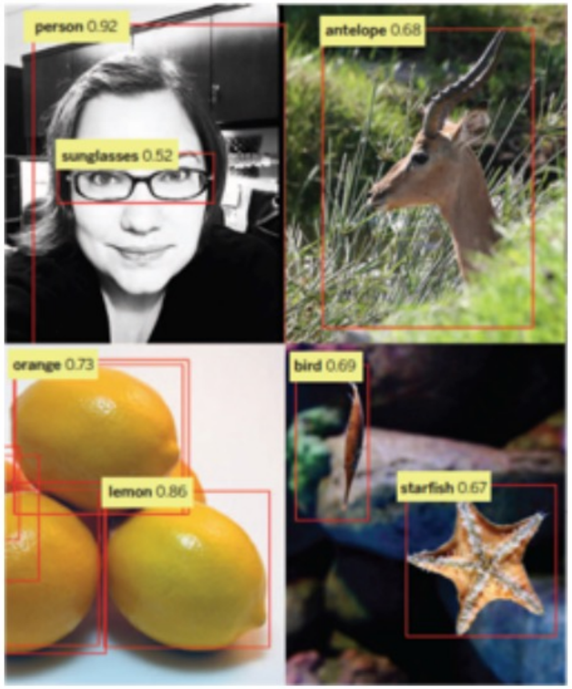
  10-606, 10-607 – Math background for ML

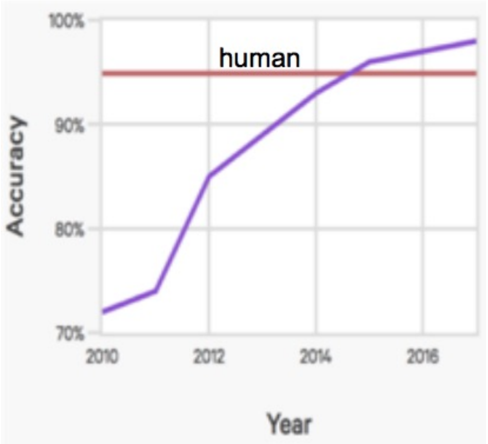
  10-605, 10-805 – Machine Learning with Large Datasets

  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

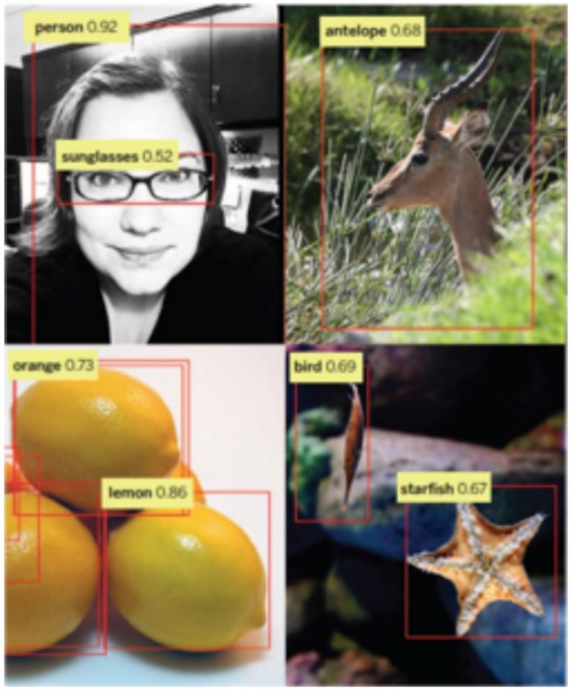
Machine Learning in Action



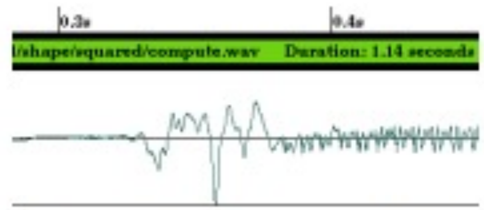


Computer vision

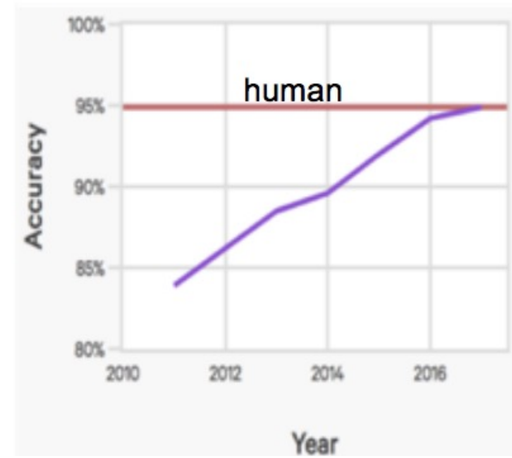
Machine Learning in Action



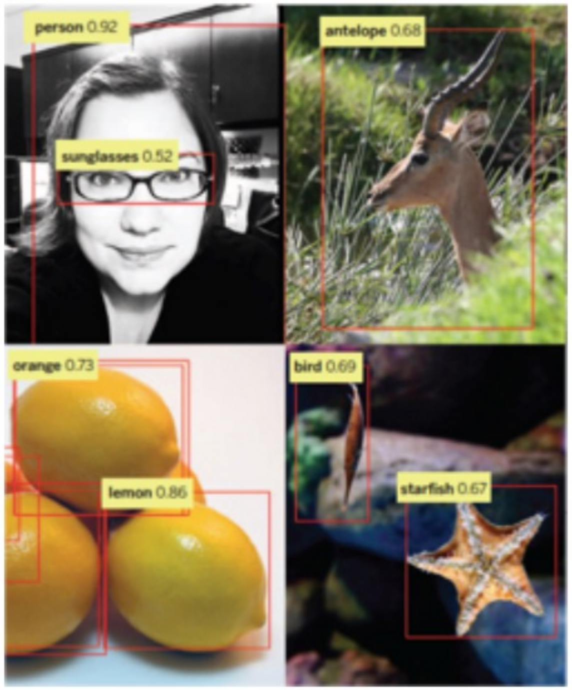
Computer vision



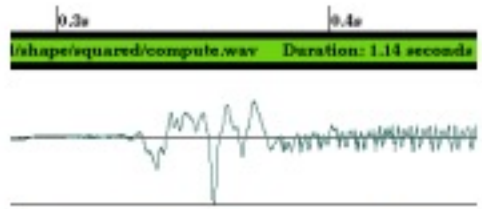
Speech Recognition



Machine Learning in Action



Computer vision



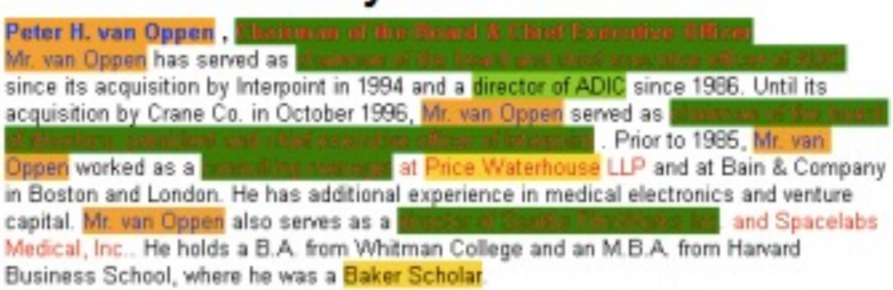
Speech Recognition

Robotic control

Games & Reasoning

Protein folding

Text analysis


Peter H. van Oppen , [illegible]
Mr. van Oppen has served as [illegible]
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 [illegible]
[illegible] . Prior to 1985, Mr. van
Oppen worked as a [illegible] 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 [illegible] 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.

Machine Learning in Action

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

Recommender systems

⋮

Enablers :

→ Data

→ GPU + hardware

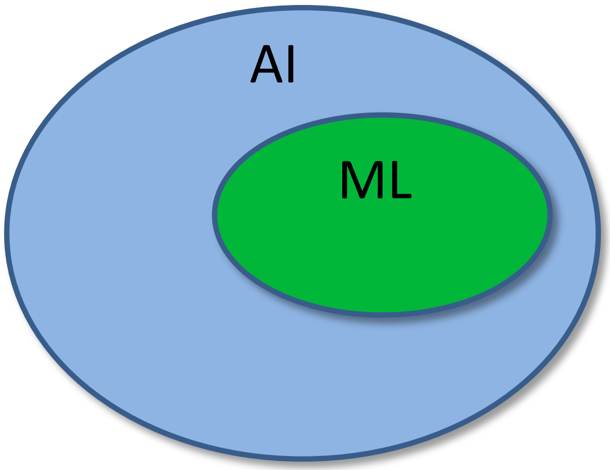
→ Better algos

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

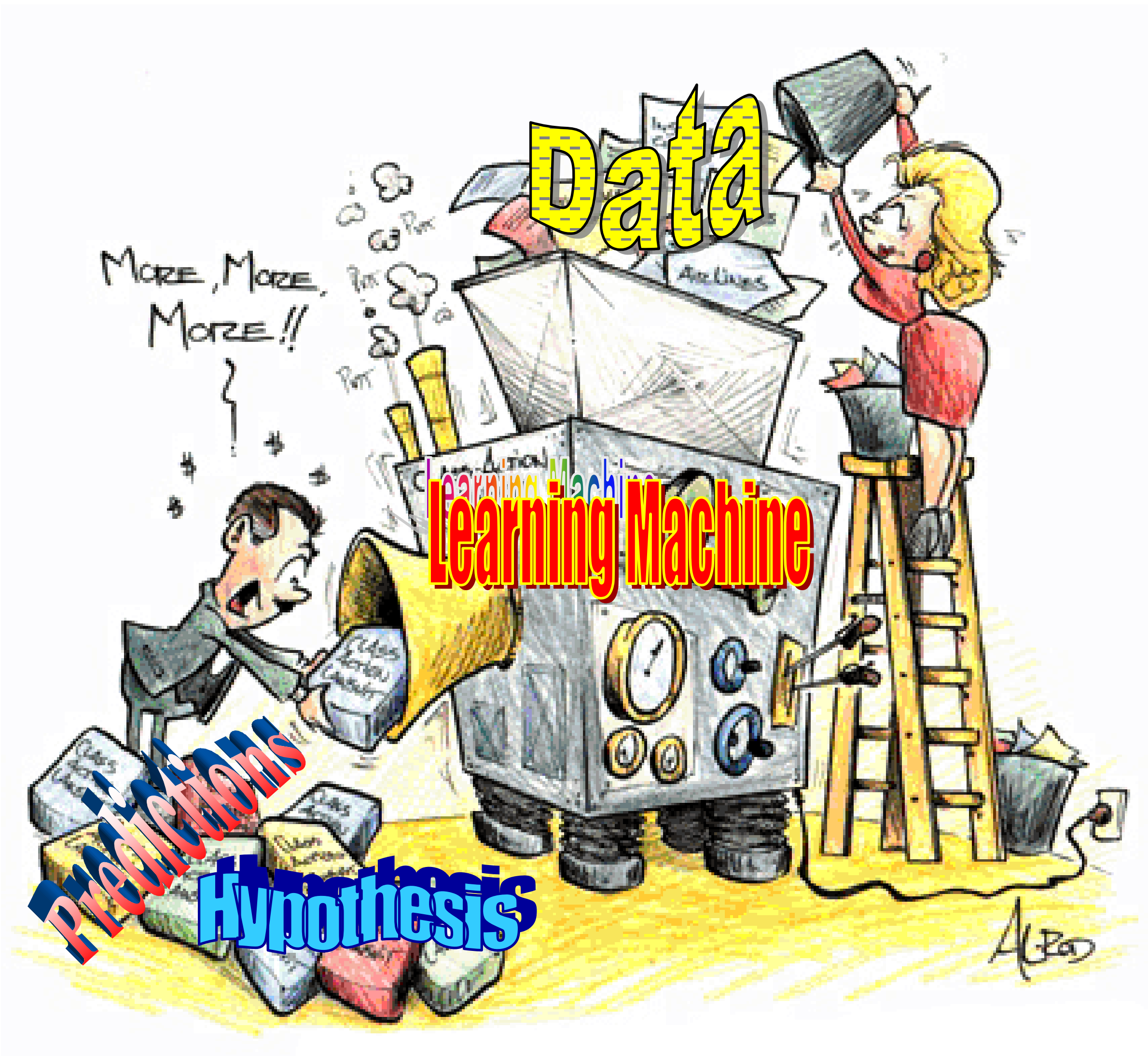
AI: develop intelligent agents

ML: learn to generalize using data



Fun begins …

What is Machine Learning?



Data

⇩

Learning algorithm

⇩

Knowledge

What is Machine Learning?

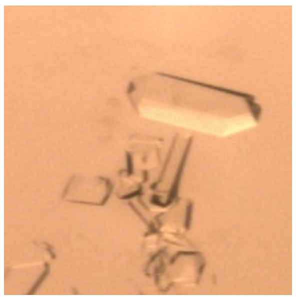
Design and Analysis of algorithms that

- improve their performance
- at some task
- with experience

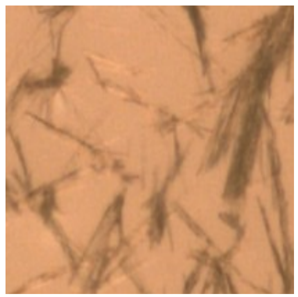
(experience) **(performance on task)**

Human learning

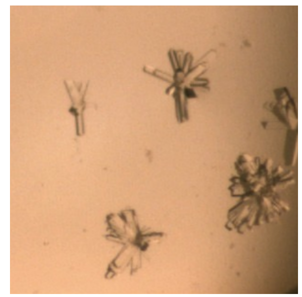
Task: Learning stage of protein crystallization



"Crystal"

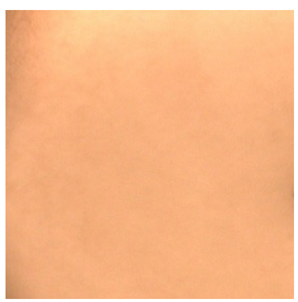


"Needle"

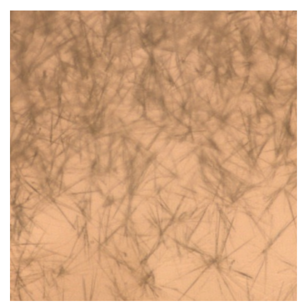


"Tree"

"Tree"



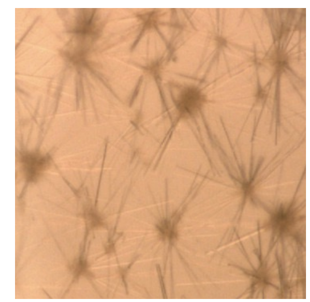
"Empty"



"Needle"

Experience

- Predict the label of the test image?



?

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

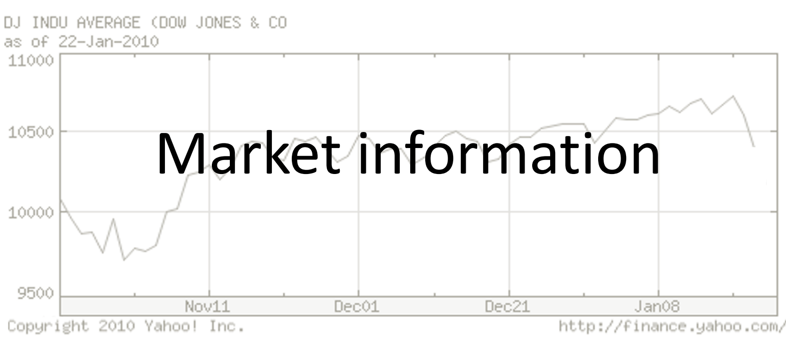
Input $X \in \mathcal{X}$

Label $Y \in \mathcal{Y}$

Document/Article

"Sports"
"News"
"Science"
...

Discrete Labels
Classification



Market information

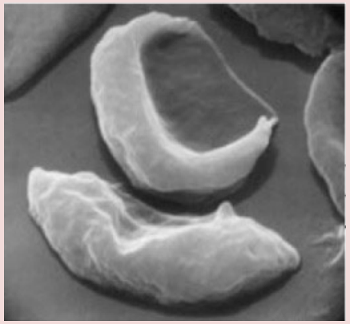
Share Price
"$ 24.50"

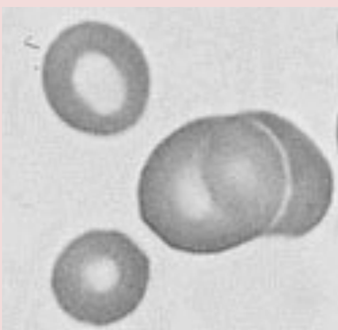
Continuous Labels
Regression

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

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

Classification or Regression?

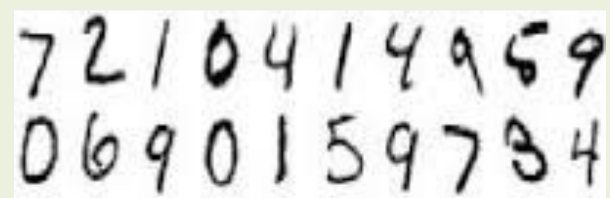




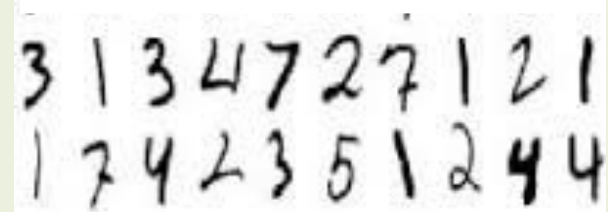
Medical Diagnosis

“Anemic”
“Healthy”

Weather prediction



Handwriting recognition



Unsupervised Learning

Aka "learning without a teacher"

Input $X \in \mathcal{X}$

Word distribution
(Probability of a word)

$p(X = \text{"android"})$

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

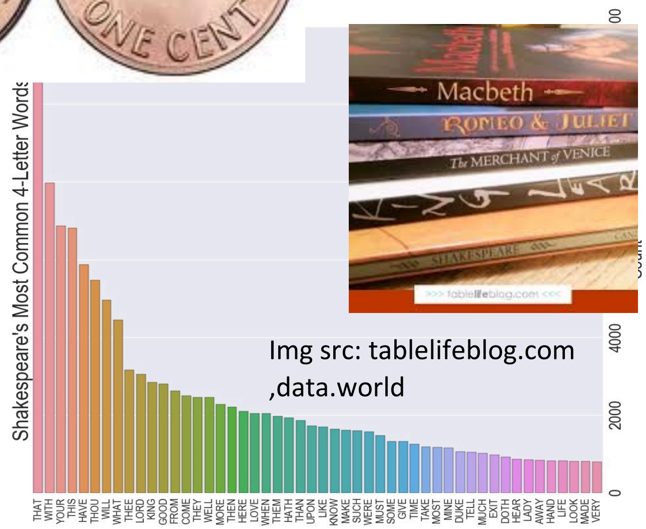
Unsupervised Learning

Learning a Distribution

Bias of a coin

p(H)
p(T)

Distribution of words in text



Img src: tablelifeblog.com ,data.world

Clustering

➢ What other distribution would be interesting to learn?

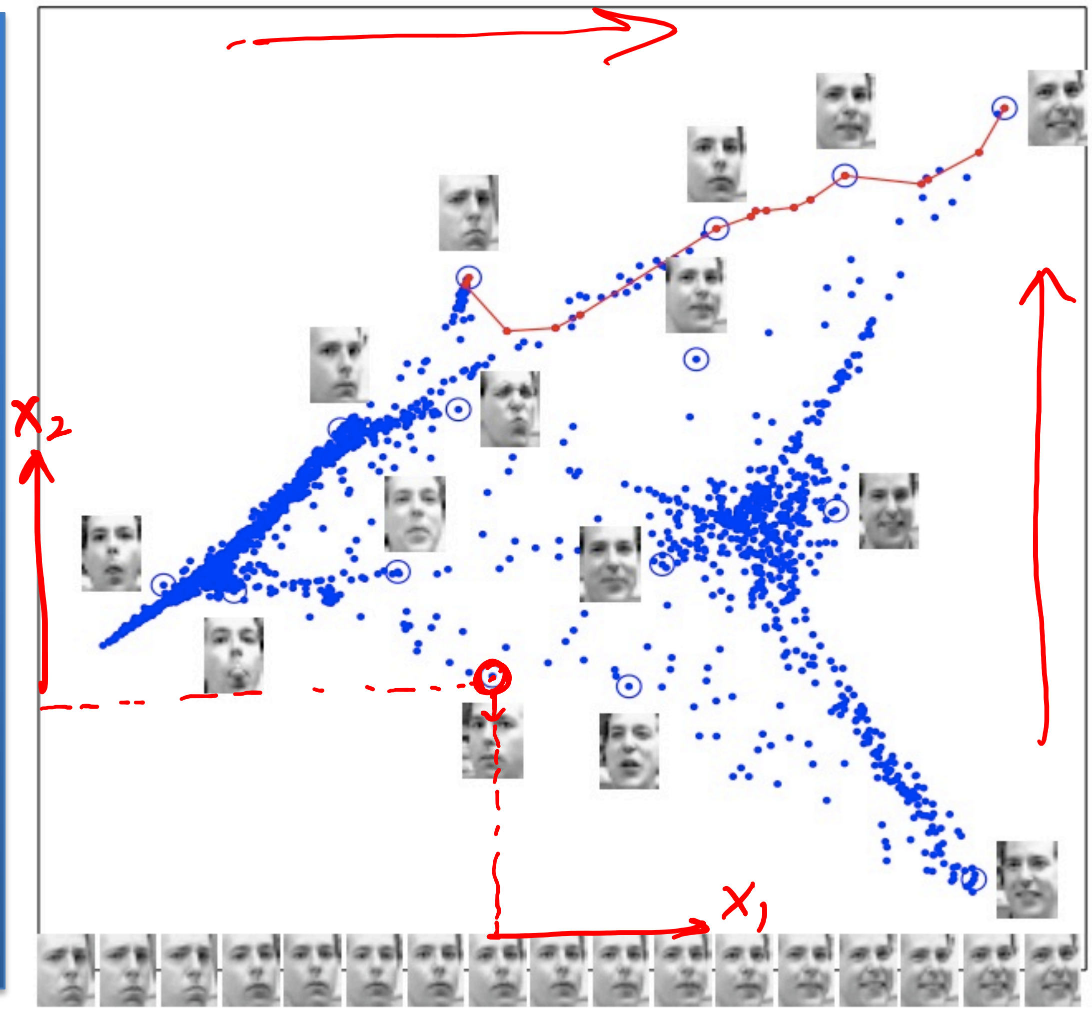
Unsupervised Learning

Dimensionality Reduction/Embedding

[Saul & Roweis '03]

Images have thousands or millions of pixels.

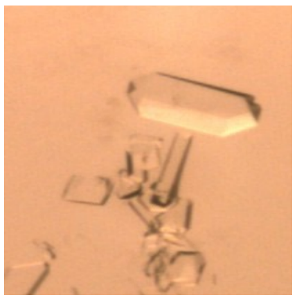
Can we give each image a small set of coordinates, such that similar images are near each other?



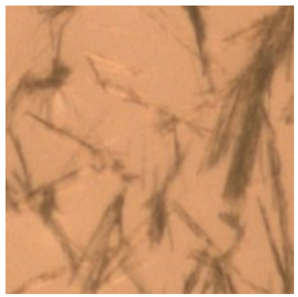
Tasks, **Experience**, Performance

Experience = Training Data

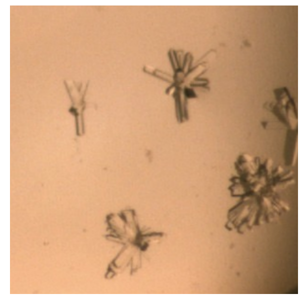
Task: Learning stage of protein crystallization



Crystal

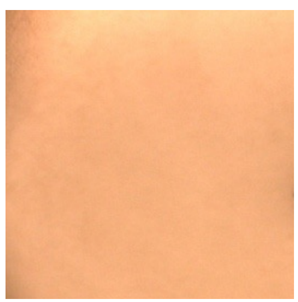


Needle

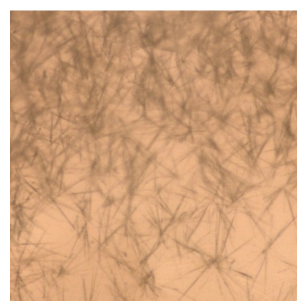


Tree

Tree

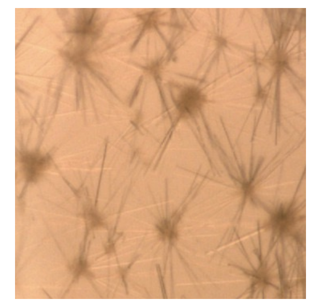


Empty



Needle

Experience



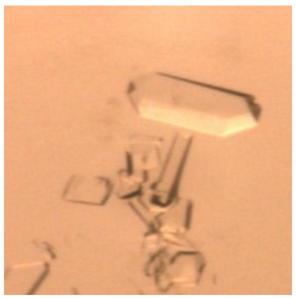
?

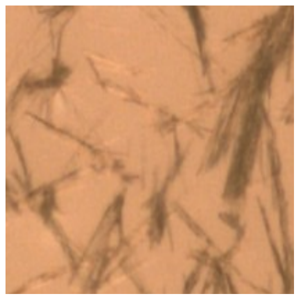
Performance

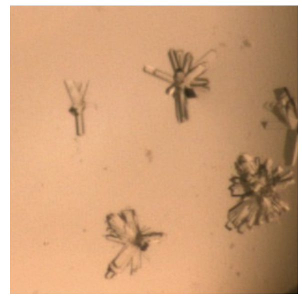
Training Data vs. Test Data

Task: Learning stage of protein crystallization

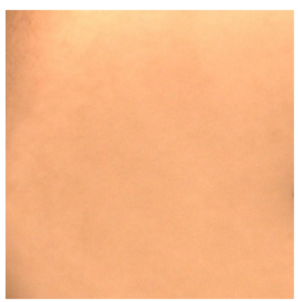
Training Data

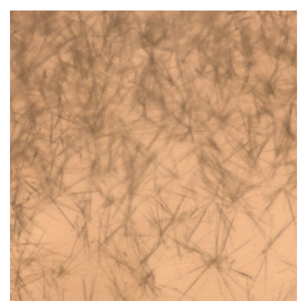

Crystal


Needle


Tree

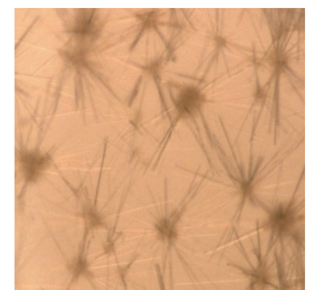

Tree


Empty


Needle

Experience

Test Data


?

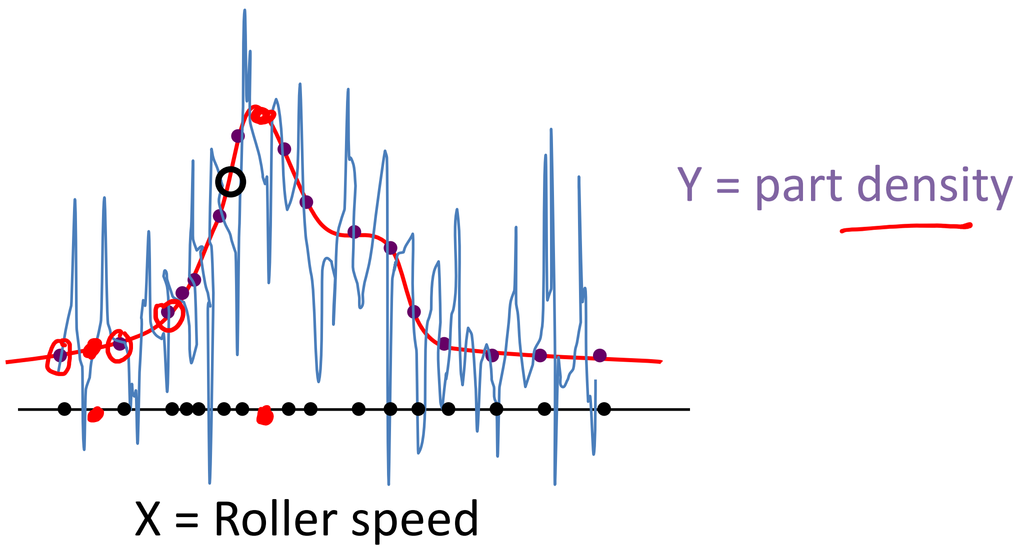
Performance

Generalization & Overfitting

A good ML algorithm

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

should not: **overfit** the training data



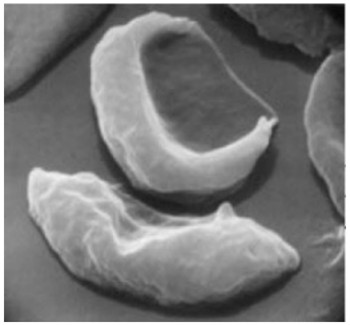
- ➢ Critical to report test and NOT training data accuracy

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

$\mathbb{1}_A = \begin{cases} 1 & A \text{ true} \\ 0 & \text{o.w.} \end{cases}$

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

Performance Measure

$\min_f \text{loss}(Y, f(x))$

Performance:

$\underbrace{|y - f(x)|}$ $\underbrace{(Y - f(x))^2}$

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

X	Share price, *Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
		23.00	
Past performance, trade volume etc. as of Sept 8, 2010	“$24.50”	“$24.50”	0
		“$26.00”	1?
		“$26.10”	2?

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