

CS 6140: Machine Learning

Ehsan Elhamifar

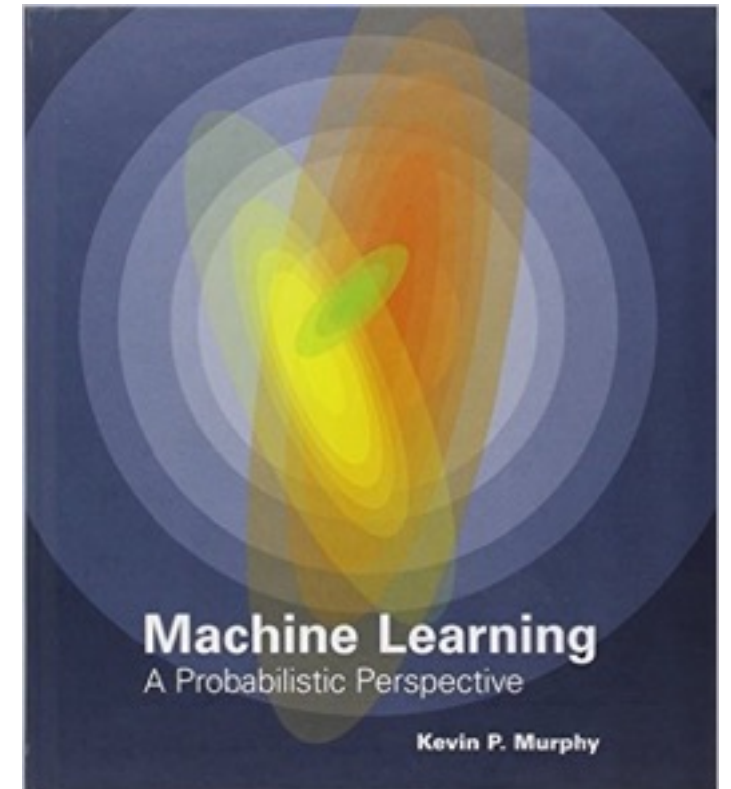
eelhami@ccs.neu.edu

Logistics

- Instructor: Ehsan Elhamifar
 - Email: eelhami@ccs.neu.edu
 - Office hours: Thursdays, 4pm—5pm, 310E WVH
- Teaching Assistants:
 - Rui Dong (dongrui@ccs.nedu.edu)
 - Haiyi Mao (harrymao@ccs.neu.edu)
- Course website: http://www.ccs.neu.edu/home/eelhami/courses_nu_cs6140.htm
- Discussions on Piazza

Textbook

- **Machine Learning: A Probabilistic Perspective**
Kevin Murphy, MIT Press 2013
- Optional:
 - Pattern recognition and machine learning, Christopher Bishop, Springer 2007
 - Machine learning, Tom Mitchell, McGraw Hill, 1997
 - Probabilistic graphical models, Daphne Koller and Nir Freedman, MIT Press 2009



Grading

- **Homeworks** (50% of the grade)
 - 4 HWs: analytical + programming assignments (Python)
 - **No late** HWs, submit at the **beginning** of class
- **Exam** (25% of the grade)
 - Will cover most of materials learned in the class, 1 cheat sheet
- **Final Project** (25% of the grade)
 - Topics: application (related to research), algorithm, theory
 - Project proposals: Thursday, October 27
 - Project report: Monday, December 5
- **Class participation** (-7% to +7%)

What is machine learning?

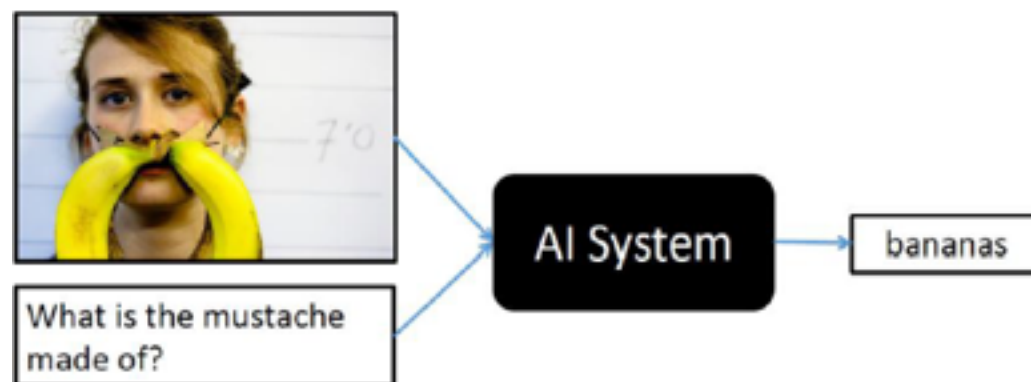
- Traditional Programming



- Machine Learning



What can ML do?



Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 <p>A person riding a motorcycle on a dirt road.</p>	 <p>Two dogs play in the grass.</p>	 <p>A skateboarder does a trick on a ramp.</p>	 <p>A dog is jumping to catch a frisbee.</p>
 <p>A group of young people playing a game of frisbee.</p>	 <p>Two hockey players are fighting over the puck.</p>	 <p>A little girl in a pink hat is blowing bubbles.</p>	 <p>A refrigerator filled with lots of food and drinks.</p>
 <p>A herd of elephants walking across a dry grass field.</p>	 <p>A close up of a cat laying on a couch.</p>	 <p>A red motorcycle parked on the side of the road.</p>	 <p>A yellow school bus parked in a parking lot.</p>

What is Machine Learning ?

(by examples)

Classification

(from data to discrete classes)

Spam filtering

data

prediction

☆ **Osman Khan** to Carlos [show details](#) Jan 7 (6 days ago) [Reply](#)

sounds good
+ok

Carlos Guestrin wrote:
Let's try to chat on Friday a little to coordinate and more on Sunday in person?

Carlos



Welcome to New Media Installation: Art that Learns

☆ **Carlos Guestrin** to 10615-announce, Osman, Michel [show details](#) 3:15 PM (8 hours ago) [Reply](#)

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.
You can contact the instructors by emailing: 10615-instructors@cs.cmu.edu



Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rik Spam | X

☆ **Jaquelyn Halley** to nherlein, bcc: thehorney, bcc: an [show details](#) 9:52 PM (1 hour ago) [Reply](#)

=== 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
- * More Self Confidence
- * Cleanse and Detoxify Your Body
- * Much More Energy
- * BetterSexLife
- * A Natural Colon Cleanse



Spam
vs
Not Spam

Object classification

			
lens cap	abacus	slug	hen
<div>reflex camera</div> <div>Polaroid camera</div> <div>pencil sharpener</div> <div>switch</div> <div>combination lock</div>	<div>abacus</div> <div>typewriter keyboard</div> <div>space bar</div> <div>computer keyboard</div> <div>accordion</div>	<div>slug</div> <div>zucchini</div> <div>ground beetle</div> <div>common newt</div> <div>water snake</div>	<div>hen</div> <div>cock</div> <div>cocker spaniel</div> <div>partridge</div> <div>English setter</div>
			
tiger	chambered nautilus	tape player	planetarium
<div>tiger</div> <div>tiger cat</div> <div>tabby</div> <div>boxer</div> <div>Saint Bernard</div>	<div>lampshade</div> <div>throne</div> <div>goblet</div> <div>table lamp</div> <div>hamper</div>	<div>cellular telephone</div> <div>slot</div> <div>reflex camera</div> <div>dial telephone</div> <div>iPod</div>	<div>planetarium</div> <div>dome</div> <div>mosque</div> <div>radio telescope</div> <div>steel arch bridge</div>

Weather prediction



Regression

(predicting numeric values)

Stock market prediction



Weather prediction (revisited)



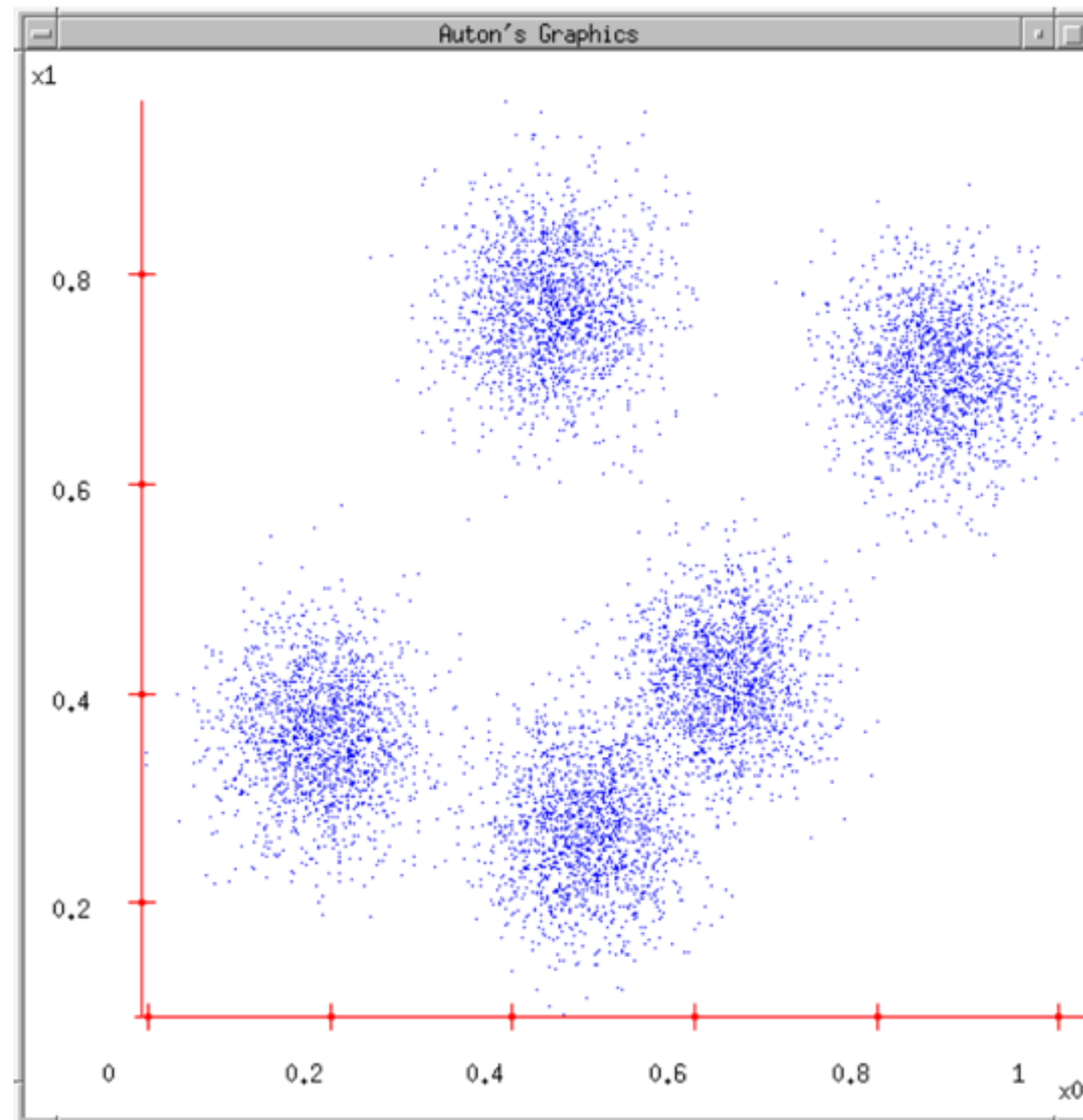
Temperature

72° F

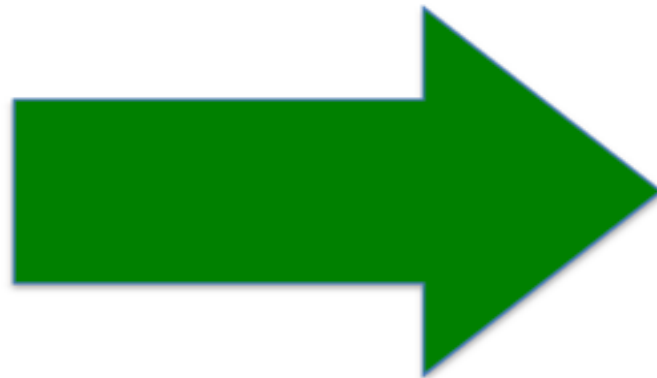
Clustering

(discovering structure in data)

Clustering: group similar items




Clustering images



Clustering news

U.S. edition Modern

Top Stories



CNN International

Saudi execution of Shia cleric threatens to deepen regional sectarian crisis

CNN International - 3 hours ago

(CNN) Sheikh Nimr al-Nimr was not among the "A-list" of Shia clerics in Saudi Arabia. But his execution has provoked a regional crisis, sparking condemnation from Iraq, Iran and even senior U.N.

[Oil Rises in Asia Due to Iran-Saudi Arabia Tensions](#) Wall Street Journal

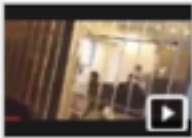
[A reckless regime](#) Washington Post

Highly Cited: [Iranian Protesters Ransack Saudi Embassy After Execution of Shiite Cleric](#) New York Times

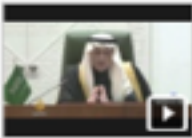
From Saudi Arabia: [Saudi Arabia severs Iran ties](#) Arab News

Wikipedia: [Nimr al-Nimr](#)


[See realtime coverage](#)




CNN



Aljazeera.com



YouTube




Washington Post

Armed activists in Oregon touch off unpredictable chapter in land-use feud

Washington Post - 2 hours ago

BURNS, Ore. - An unpredictable new chapter in the wars over federal land use in the West unfolded Sunday after a group of armed activists split off from an earlier protest march and occupied a national wildlife refuge in remote southeastern Oregon.




Firstpost

One dead as 6.8 magnitude quake strikes eastern India - police

Reuters - 1 hour ago

GUWAHATI, India At least one person was killed and a dozen injured when an earthquake measuring 6.8 struck near Imphal in eastern India on Monday, sending people running from their homes and knocking out power to the city near the Myanmar border.




CBS News

ISIS threatens UK in new execution video

CBS News - 5 hours ago

BEIRUT -- A video circulated online Sunday purported to show the Islamic State of Iraq and Syria (ISIS) killing five men accused of spying for Britain in Syria.




Press Herald

NTSB releases haunting video of El Faro wreckage on ocean floor

Press Herald - 23 minutes ago

The merchant ship carrying 33 crew members, including four from Maine, sank off the Bahamas last fall. By Dennis Hoey Staff Writer.



The Boston Globe

In NH, Clinton hits on opioid abuse as a top concern

The Boston Globe - 2 hours ago

DERRY, N.H. - Hillary Clinton, who arrived to loud applause here at one of three New Hampshire campaign stops Sunday, said prohibitively expensive education, lack of support for families coping with Alzheimer's disease, and the rising tide of opioid ...

Related

[Saudi Arabia »](#)

[Sheikh Nimr »](#)

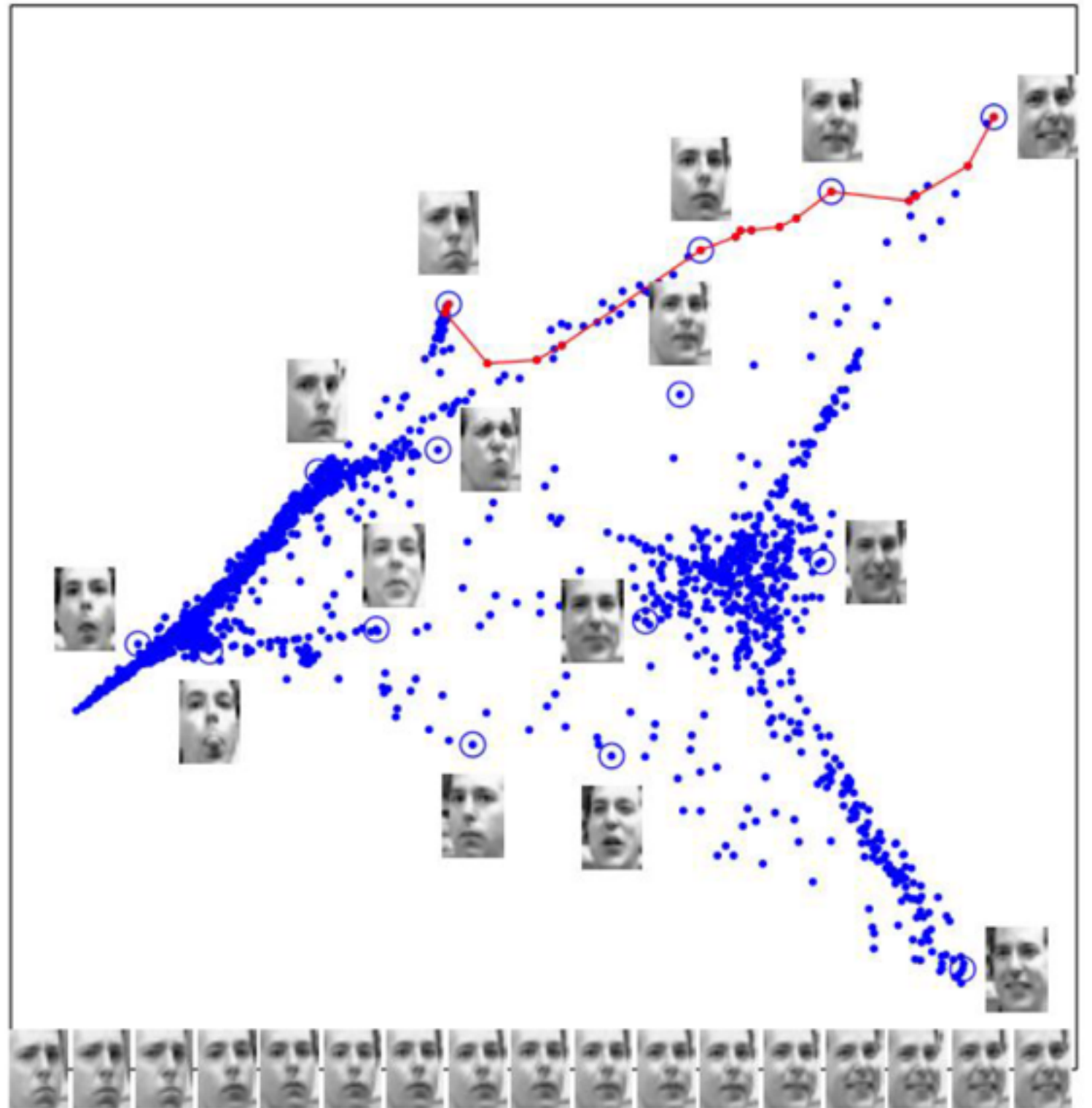
[Iran »](#)

Dimensionality reduction

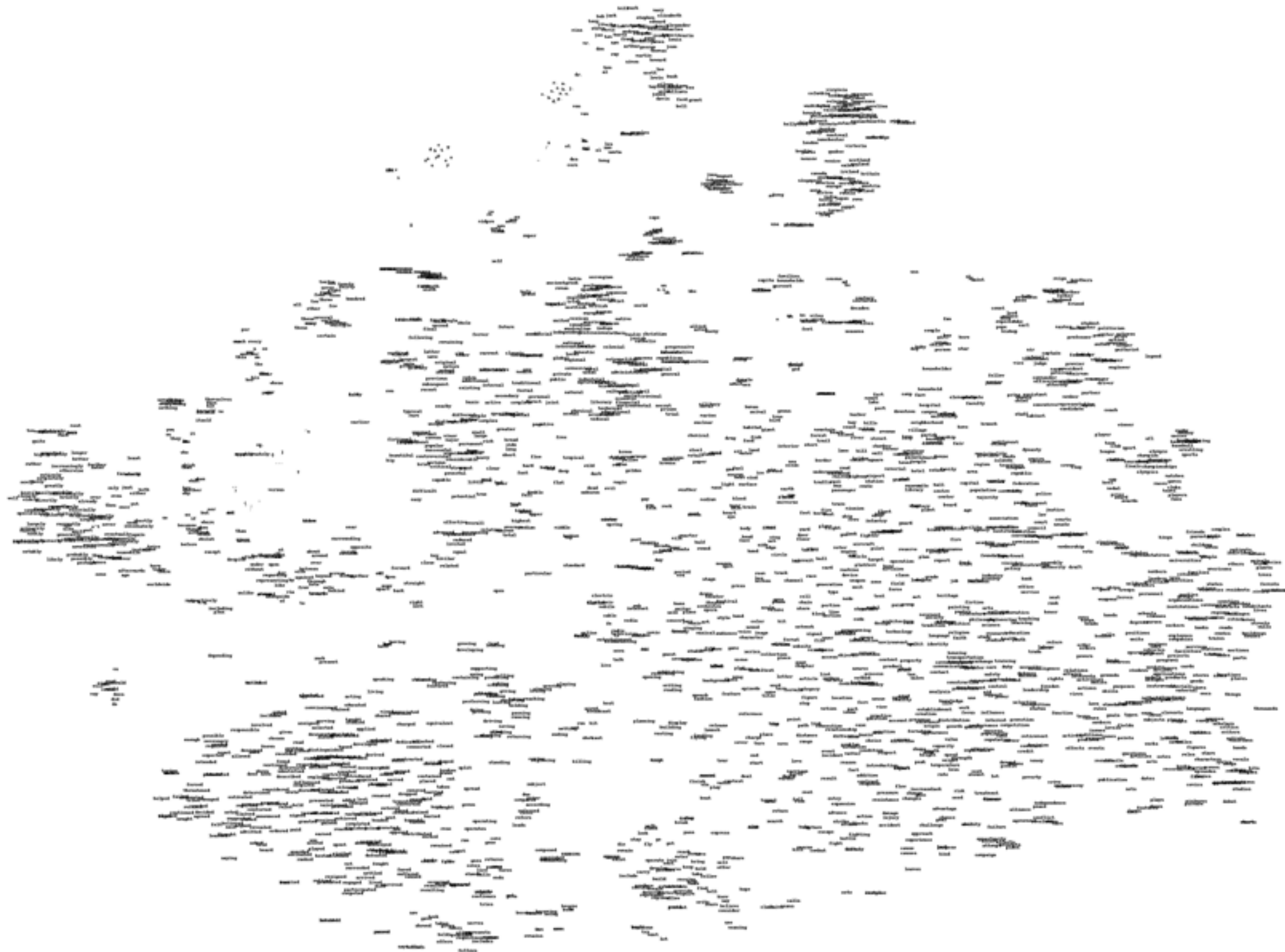
(visualizing data)

Image embedding

- Images have millions/ billions of pixels
- Can we give each image a few coordinates so that similar images are near each other?



Word embedding




Word embedding (zoom in)



Machine learning types

- **Supervised** learning
 - regression
 - classification
- **Unsupervised** learning
 - clustering
 - dimensionality reduction
 - matrix completion
- **Reinforcement** learning
 - learning by weak supervision: reward function

Supervised learning

- **Given**: training data $\{(x_i, y_i), i = 1, \dots, N\}$
- **Find**: a good approximation to $f : \mathcal{X} \rightarrow \mathcal{Y}$
 **hypothesis**
- Examples:
 - Spam detection: \mathcal{X} = email, \mathcal{Y} = spam/ham
 - Digit recognition: \mathcal{X} = digit images, $\mathcal{Y} = \{0, 1, 2, \dots, 9\}$
 - Stock prediction: \mathcal{X} = new/historic prices, etc, \mathcal{Y} = real numbers

Example: spam filter

- **Input:** email
- **Output:** spam/ham
- **Setup:**
 - Get a large collection of example emails, each labeled “spam” or “ham”
 - Note: someone has to hand label all this data!
 - Want to learn to predict labels of new, future emails
- **Features:** The attributes used to make the ham / spam decision
 - Words: FREE!
 - Text Patterns: CAPS
 - Non-text: SenderInContacts
 - ...



Dear Sir.

First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret. ...



TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES
FOR ONLY \$99



Ok, I know this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Example: digit classification

- **Input:** images / pixel grids
- **Output:** a digit 0-9
- **Setup:**
 - Get a large collection of example images, each labeled with a digit
 - Note: someone has to hand label all this data!
 - Want to learn to predict labels of new digit images
- **Features:** The attributes used to make the digit decision
 - Pixels: (6,8)=ON
 - Shape Patterns: NumComponents, AspectRatio, NumLoops
 - ...



0



1



2



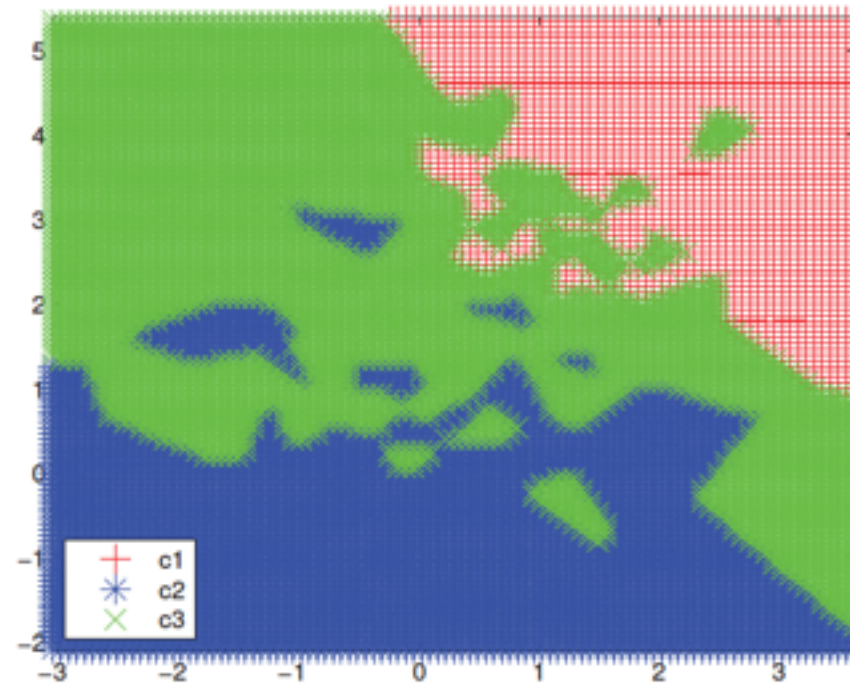
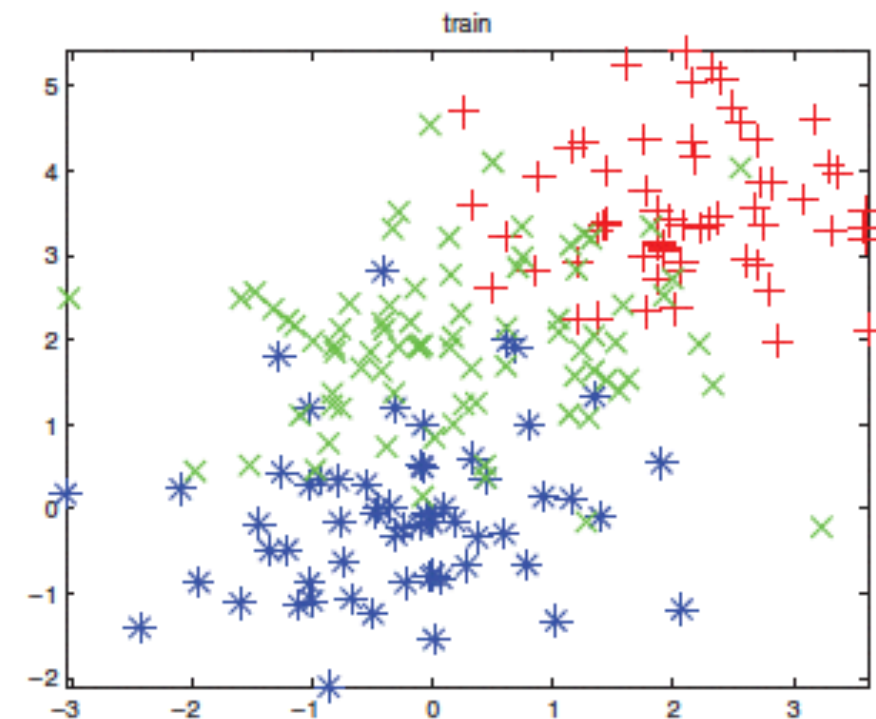
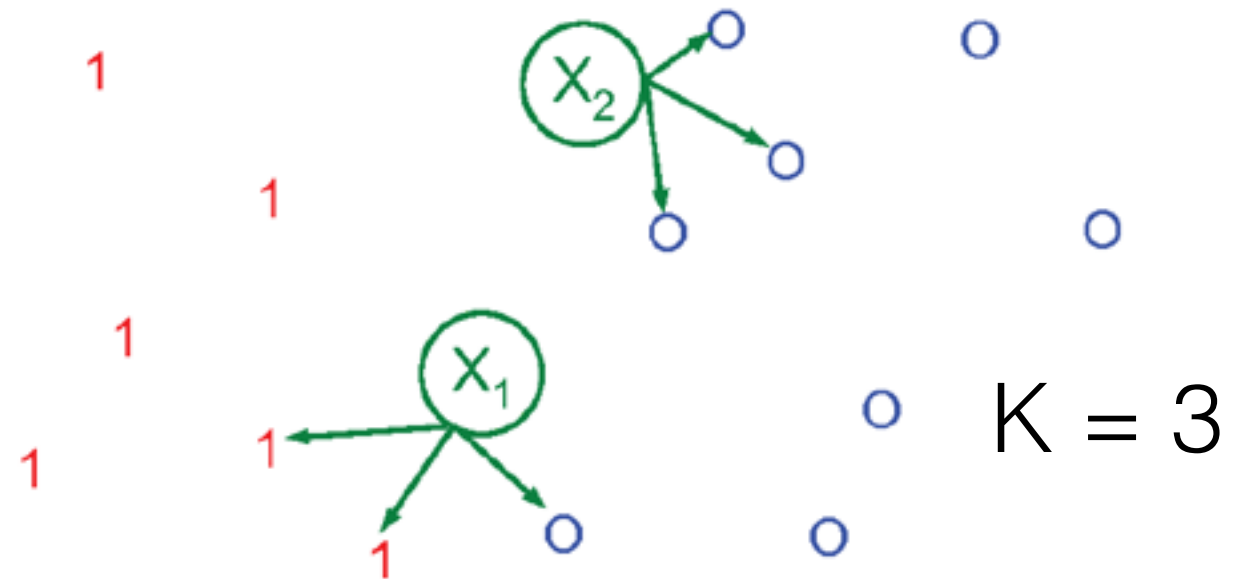
1



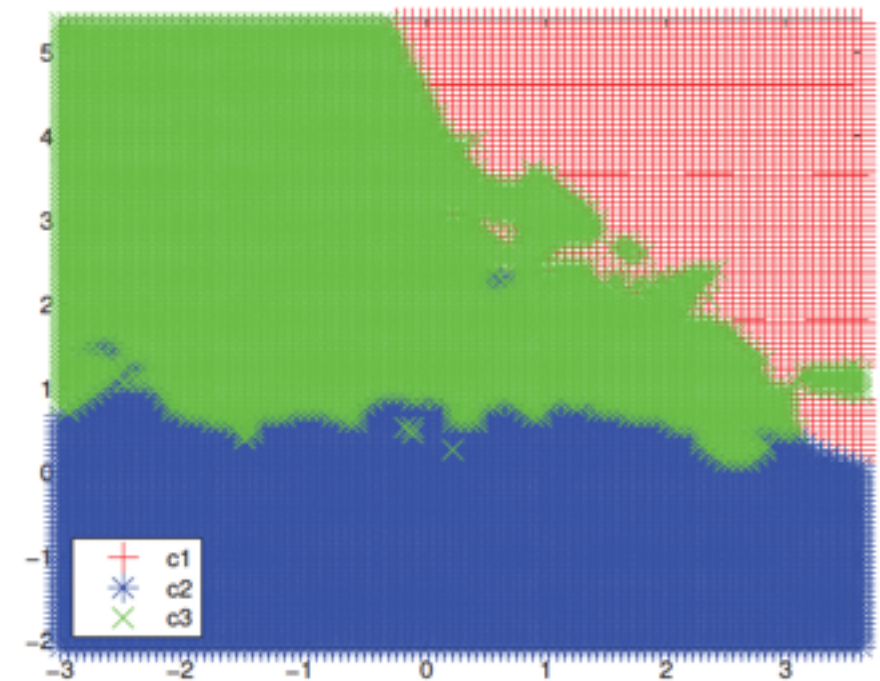
??

KNN classifier

- Assign the test sample to the class with largest #votes
- How to choose K?



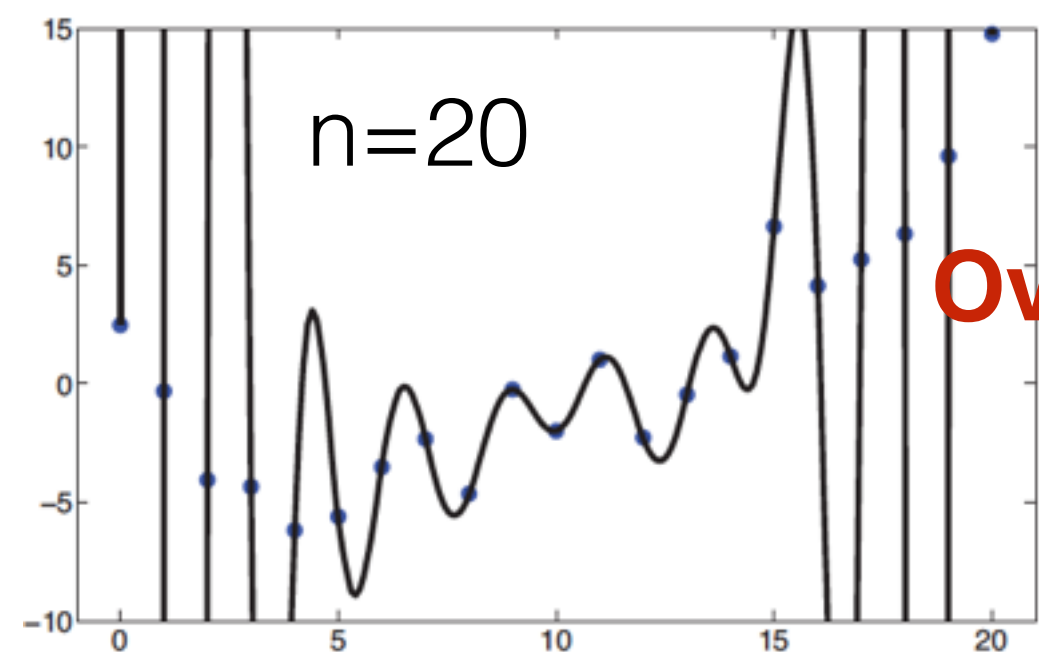
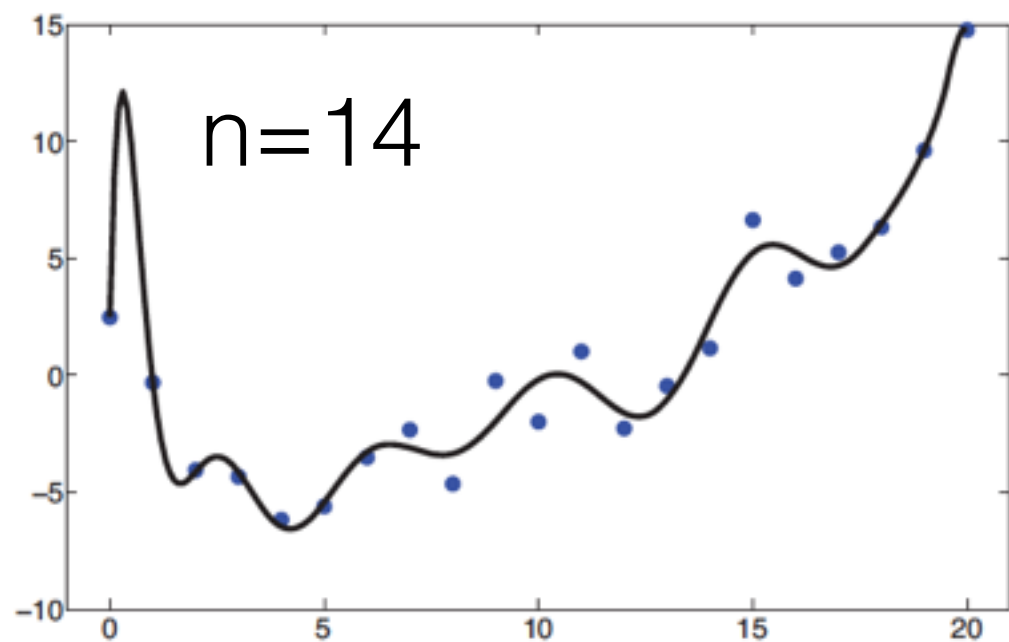
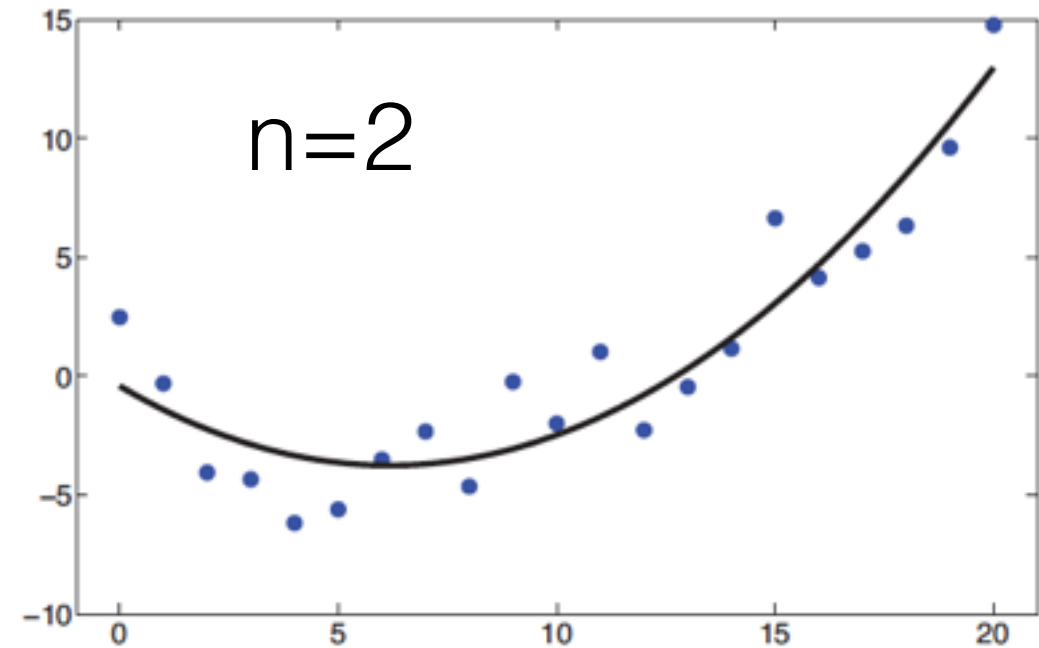
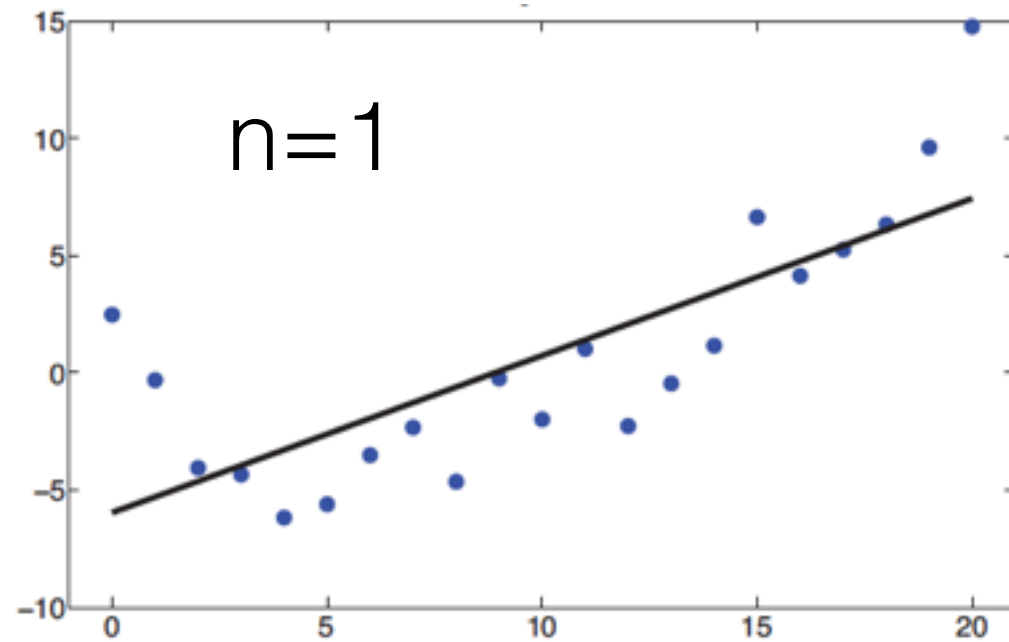
$K = 1$



$K = 5$

Regression

- What degree polynomial (n) to fit to data?



Overfitting

Key issues in ML

- What are good **hypotheses**? finding best hypothesis?
- How to optimize for accuracy of **unseen data** (avoid overfitting)?
- How much data is needed to be **confident** in results?
- How to deal with **data nuisances**: missing data, errors, outliers
- How to deal with **high-dimensionality** of data / large datasets?
- How to **model applications** as machine learning problems?

Summary

- Components of a machine learning system
 - **Data** (input: x , output: y)
 - **Hypothesis space** (e.g. polynomial degree)
 - **Objective/cost functions**: what makes one 'incorrect answer' better/worse than another 'incorrect answer'
- Important concepts
 - Want high accuracy on unseen test data, but have only see training data
 - Central challenge in ML: **generalization**

Course syllabus

- Introduction, regression
- Logistic regression, Naive Bayes
- Decision Trees, ensemble methods (boosting, bagging)
- Support vector machines, kernels
- Neural networks: Deep NNs, CNN, RNN
- Graphical models: Hidden Markov Models, Linear Dynamical Sys
- Learning theory: VC dimension
- Unsupervised learning: PCA, CCA, Kmeans, Spectral clustering

What you should already know

- Concepts in **probability**:
 - Conditioning, marginals, expectations
 - Gaussian/Laplace/Bernoulli/Multinomial/... distributions
- Concepts from **linear algebra** and matrix analysis
 - Eigenvalues, eigenvectors, SVD
 - Cholesky/QR decomposition, projections, subspace

What you should already know

- Concepts in **probability**:
 - Conditioning, marginals, expectations
 - Gaussian/Laplace/Bernoulli/Multinomial/... distributions
 - Concepts from **linear algebra** and matrix analysis
 - Eigenvalues, eigenvectors, SVD
 - Cholesky/QR decomposition, projections, subspace
- 1) **Read chapter 2 of the book**
 - 2) **Read tutorials on the course webpage**

