

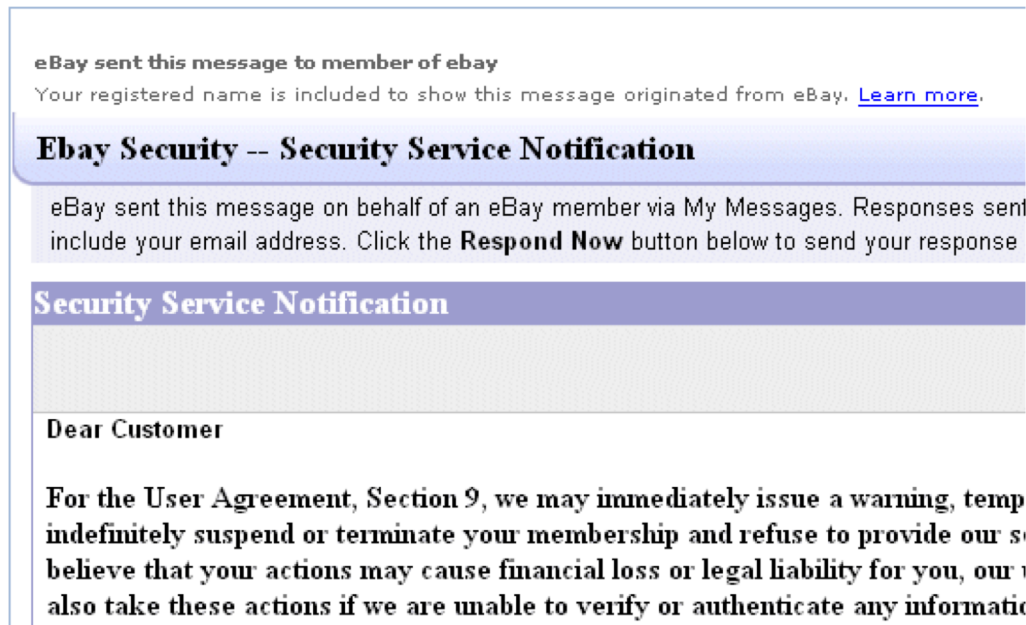
Introduction to Machine Learning

Introduction

Prof. Andreas Krause
Institute for Machine Learning
(las.ethz.ch)

What is Machine Learning I: An example

Classify email messages as “Spam” or “Non Spam”



Classical Approach: manual rules

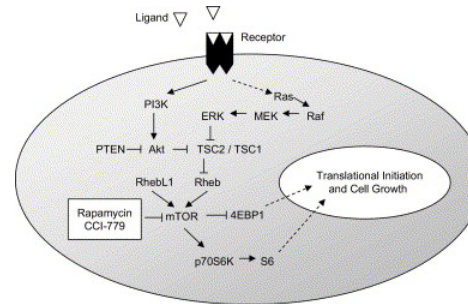
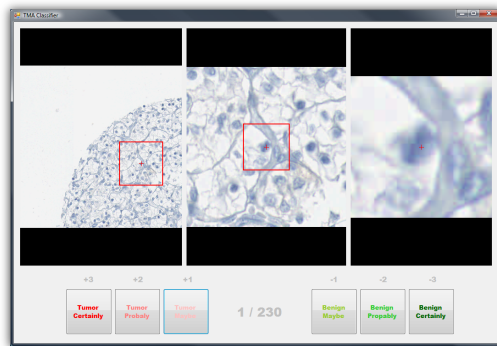
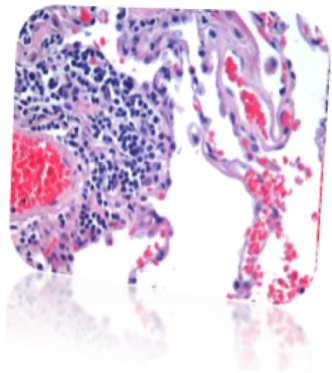
IF text body contains “Please login here”
THEN classify as “spam” ELSE “non-spam”

Machine Learning: Automatic discovery of rules
from training data (examples)

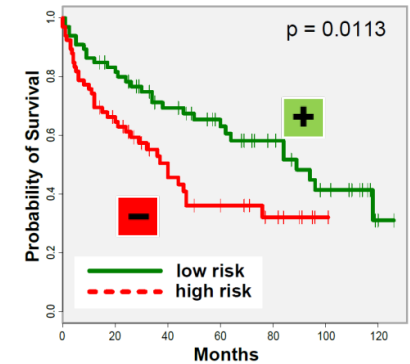
What is ML II: One Definition [Tom Mitchell]

„A computer program is said to learn from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in T, as measured by P, improves with experience E“

Our Digital Society and the Information Technology value chain



Activation of the mTOR Signaling Pathway in Renal Clear Cell Carcinoma. Robb et al., J Urology 177:346 (2007)



Data → **Information** → **Knowledge** → **Value**

Machine Learning plays a core role in this value chain

Related disciplines

information
theory

philosophy
epistemology
causality

statistics

**machine
learning**

algorithms
& optimization

neuro-
informatics

Overview

- Introductory course
- Preparation for M.Sc. Level ML courses
- Two main topics
 - Supervised learning
 - Unsupervised learning
- Algorithms, models & applications
- Handouts etc. on course webpage
 - <https://las.ethz.ch/teaching/introml-s20>
 - Old slides available at <.../introml-s19>
 - Password can be retrieved from within ETH network
- Textbooks listed on course webpage (some available online)

Prerequisites

- Basic knowledge in linear algebra, calculus and probability
- If you need a refresher:
 - Part I of "Mathematics for Machine Learning" by Deisenroth, Faisal, Ong
 - Available online at <https://mml-book.com/>
- Basic programming (in Python)
 - Links to tutorials on website
- If you plan not to complete the course, please deregister!

Syllabus

- Linear regression
- Linear classification
- Kernels and the kernel trick
- Neural networks & Deep Learning
- Unsupervised learning
- The statistical perspective
- Statistical decision theory
- Discriminative vs. generative modeling
- Bayes' classifiers
- Bayesian approaches to unsupervised learning
- Generative modeling with neural networks

After participating in this course you will

- Understand **basic machine learning ideas & concepts**
- Be able to **apply** basic machine learning **algorithms**
- Know how to **validate** the output of a learning method
- Have some experience using machine learning on **real data**
- Learn what role machine learning plays in **decision making** under uncertainty

Relation to other ML Courses @ ETHZ

- Advanced Machine Learning (Fall)
 - Continuation and advanced topics
- Deep Learning (Fall)
 - Deep neural networks and their applications
- Probabilistic Artificial Intelligence (Fall)
 - Reasoning and decision making under uncertainty
- Computational Intelligence Lab (Spring)
 - Matrix Factorization, Recommender Systems, projects
- Statistical Learning Theory (Spring)
 - Theoretical foundations; model validation
- Guarantees for Machine Learning (Spring)
- Computational Statistics (D-MATH, Spring)

People

- *Instructor:*

Andreas Krause (krausea@ethz.ch)

- *Teaching assistants:*

Head TA: Philippe Wenk (wenkph@ethz.ch)

Andisheh Amrollahi, Nemanja Bartolovic, Ilija Bogunovic, Zalán Borsos, Charlotte Bunne, Sebastian Curi, Radek Danecek, Gideon Dresdner, Joanna Ficek, Vincent Fortuin, Carl Johann Simon Gabriel, Shubhangi Gosh, Nezihe Merve Gürel, Matthias Hüser, Jakob Jakob, Mikhail Karasikov, Kjong Lehmann, Julian Mäder, Mojmír Mutný, Harun Mustafa, Anastasia Makarova, Gabriela Malenova, Mohammad Reza Karimi, Max Paulus , Laurie Prelot, Jonas Rothfuss, Stefan Stark, Jingwei Tang, Xianyao Zhang

Video-recording

- Lectures are **video-recorded**, and will be available at <https://video.ethz.ch/lectures/d-infk.html>
- Videos, slides etc. from last year are still available <https://video.ethz.ch/lectures/d-infk/2019/spring/252-0220-00L.html>

Waitlist situation

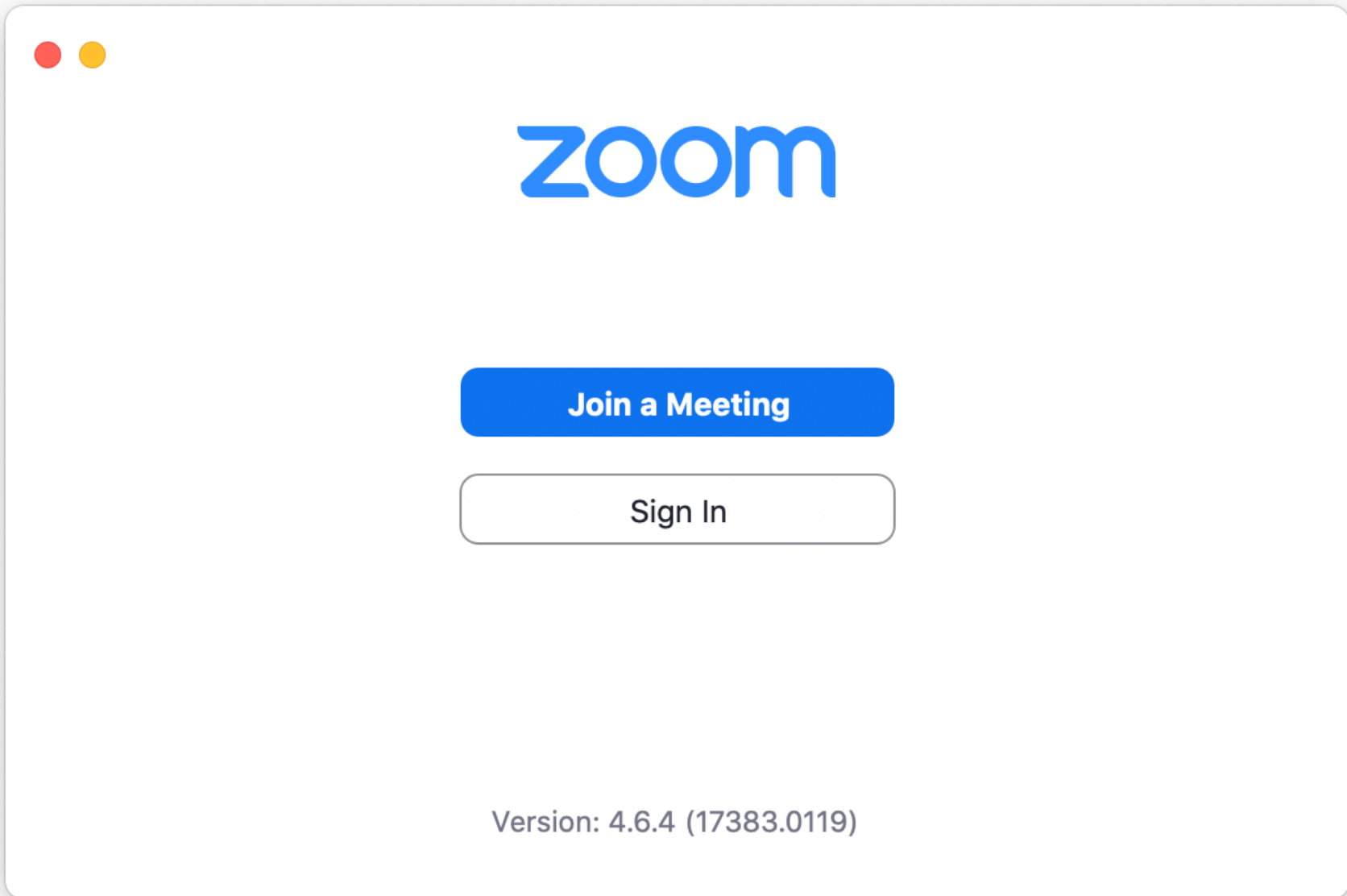
- We are currently trying to create **extra capacity** and allow more students to register for the course
- If you are on the waitlist, **please keep following** the course – there will be more information next week

Exercises

- Take them seriously if you want to pass the exam...
- Published and partially corrected in moodle
- More involved solutions on website
- This week: Optional refresher on basic linear algebra, calculus and probability

Online tutorials

- Every Wednesday, 15:00-18:00
- 1-2 hours of presentation, 1-2 hours open Q&A
Participate actively via Q&A feature
- Presentation will be recorded
- Public viewing at CAB G61
No TAs present. LIMITED CAPACITY



Zoom client:

<https://ethz.zoom.us/j/869018193>

Join Meeting

Meeting ID or Personal Link Name ▼

IML Tutorial TEST

135-421-738

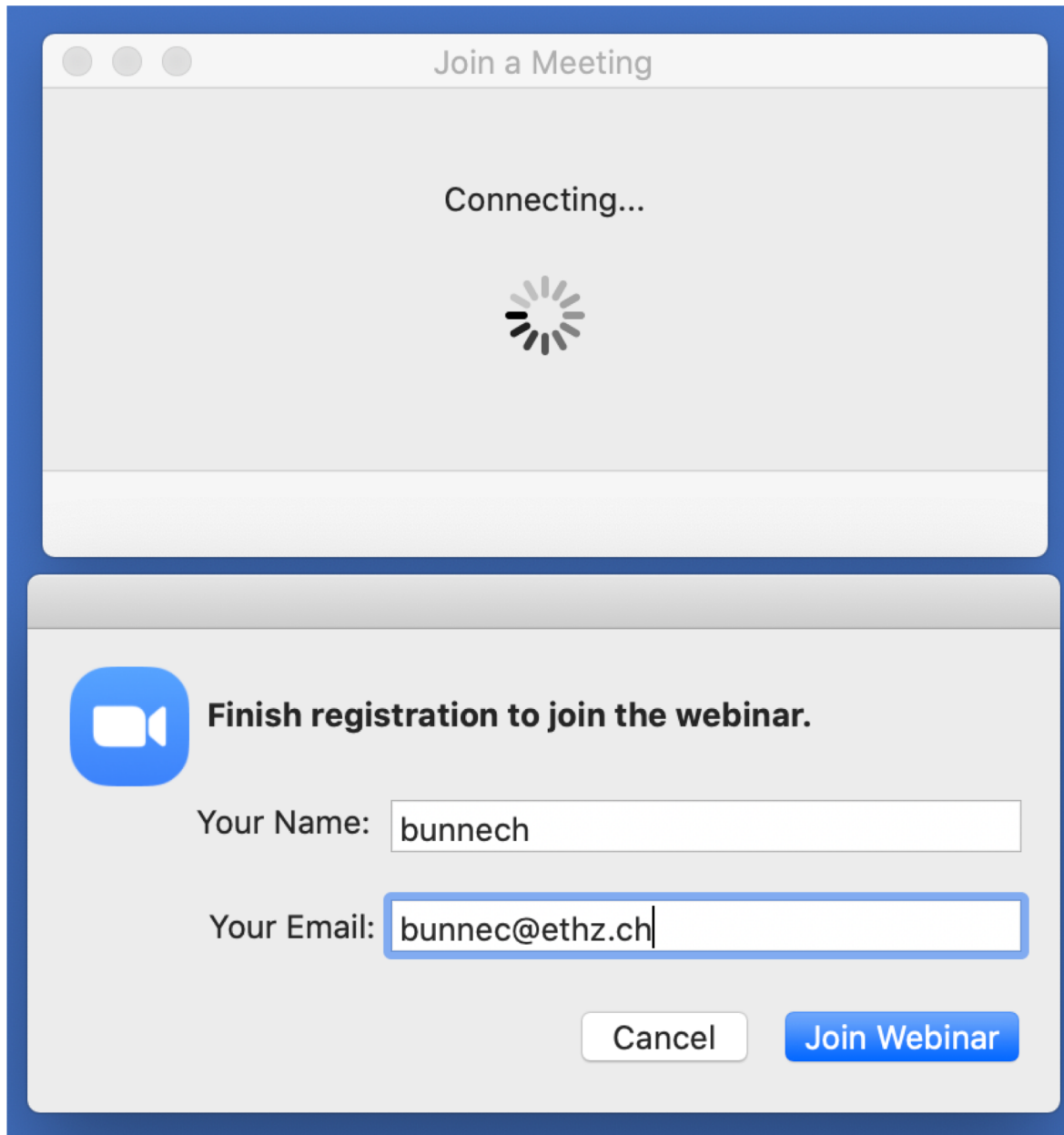
[Clear History](#)

- Remember my name for future meetings
- Don't connect to audio
- Turn off my video

Cancel

Join

Meeting ID: **869-018-193**



Use true ethz email when registering



 Please wait. The webinar will begin soon.

Start at 4:00 PM

IML Tutorial TEST

If you are the host, [sign in](#) to start this meeting

Test My Speaker



PRESENTATION SLIDE or DOCU CAM



All quest

My Question 02:37

Wie geht das?

This question has

Type your question



Questions

- Main resource: [Piazza](#)
- <https://www.piazza.com/ethz.ch/spring2020/252022000l/home>
- During tutorials via Q&A feature (live)
Limited Capacity
- Office hours, Fridays, ML D28, 13:00-15:00
Very limited Capacity

Course Project

- In a course project, you will apply basic learning methods to **make predictions on real data**
- Submit predictions on test data
- To do now:
 - Team up in groups of (up to) three students
 - Will send instructions on how to register by end of week
- More details to follow in the tutorials

- Contributes to 30% of final grade
- Project **must be passed on its own** and has a bonus/penalty function

Project server: <https://project.las.ethz.ch>

IML20 Project

Overview

Task 0

Task 1a

Task 1b

Task 2

Task 3

Task 4

Logout

HI NETHZ!

The project consists of an ungraded dummy task and several graded tasks.



TASK 0 SUBMISSIONS END ON SUNDAY 08 MAR 2020 20:00



TASK 1A STARTS ON TUESDAY 03 MAR 2020 05:00



TASK 1B STARTS ON TUESDAY 03 MAR 2020 05:00



TASK 2 STARTS ON TUESDAY 24 MAR 2020 05:00



TASK 3 STARTS ON TUESDAY 21 APR 2020 05:00



TASK 4 STARTS ON TUESDAY 05 MAY 2020 05:00

Some FAQs

- Distance exams
 - are possible (as exception), but need to officially request with study administration
- Doctoral students for whom a “Testat” or 2 ECTS credits suffice:
 - Can take unit “[Introduction to Machine Learning \(only project\)](#)”
- Repeating the exam
 - requires repeating the project
- Will maintain an FAQ list on webpage

Introduction to Machine Learning

A brief tour of supervised and
unsupervised learning

Prof. Andreas Krause
Institute for Machine Learning (las.ethz.ch)

Machine Learning Tasks

Supervised Learning

- Classification
- Regression
- Structured Prediction, ...

Unsupervised Learning

- Clustering
- Dimension reduction
- Anomaly detection, ...

Many other specialized tasks

Supervised Learning

$$f : X \rightarrow Y$$

Example: E-Mail Classification

eBay sent this message to member of ebay
Your registered name is included to show this message originated from eBay. [Learn more.](#)

Ebay Security -- Security Service Notification

eBay sent this message on behalf of an eBay member via My Messages. Responses sent include your email address. Click the **Respond Now** button below to send your response

Security Service Notification

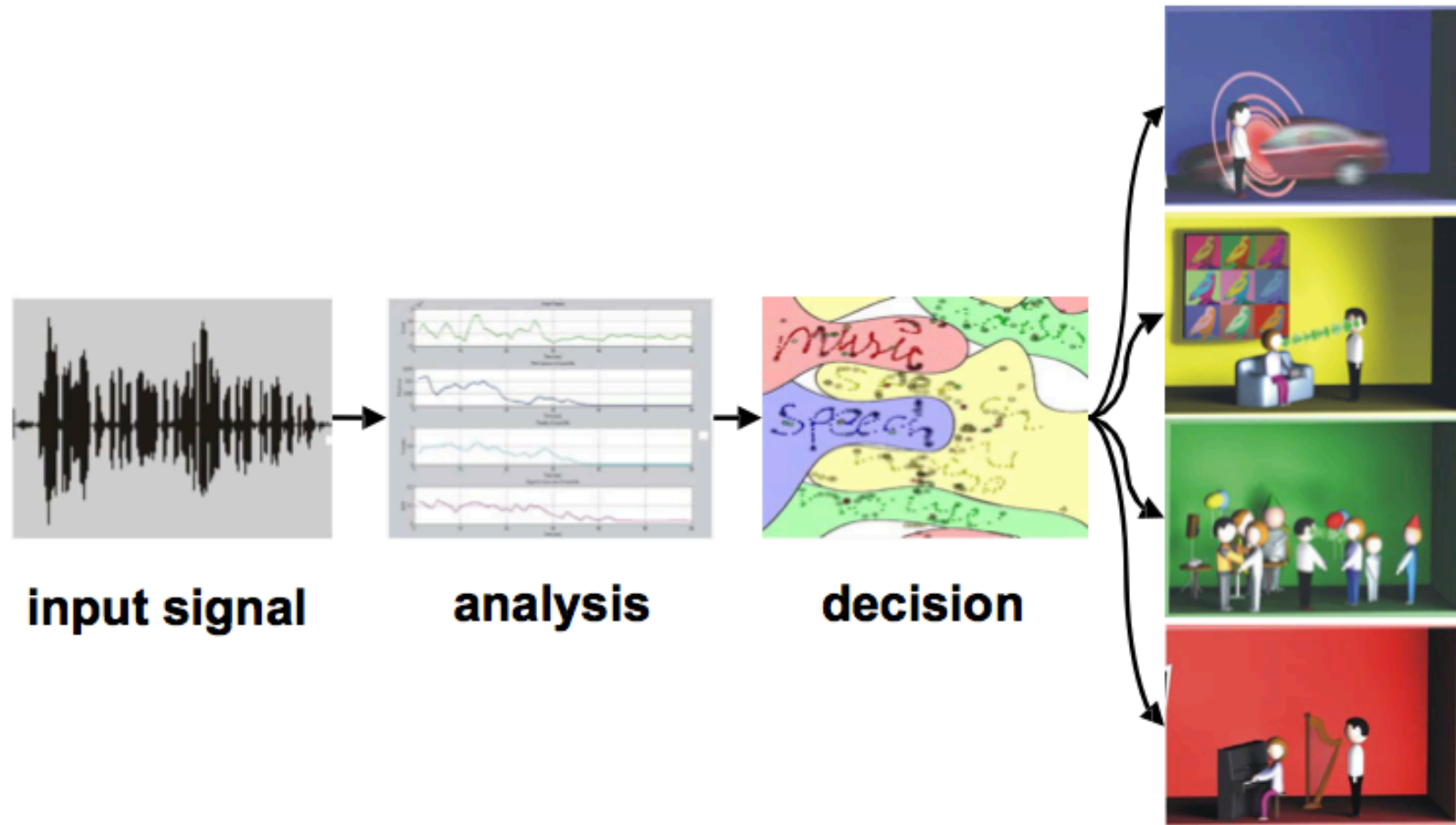
Dear Customer

For the User Agreement, Section 9, we may immediately issue a warning, temporarily suspend or terminate your membership and refuse to provide our services if we believe that your actions may cause financial loss or legal liability for you, our members, or eBay. We also take these actions if we are unable to verify or authenticate any information.

- **X:** E-Mail Messages
- **Y:** label: “spam” or “non-spam”

Example: Improving Hearing Aids

[Buhmann et al]



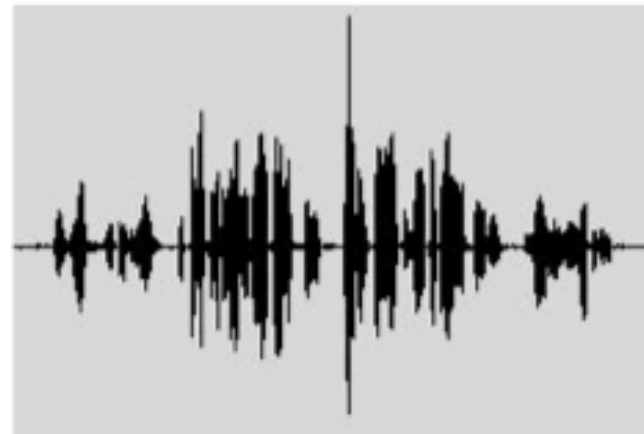
- **X**: Acoustic waveforms
- **Y**: label *speech, speech in noise, music, noise*

Example: Improving Hearing Aids

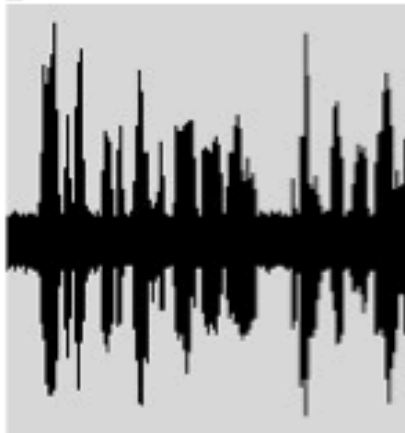
Music



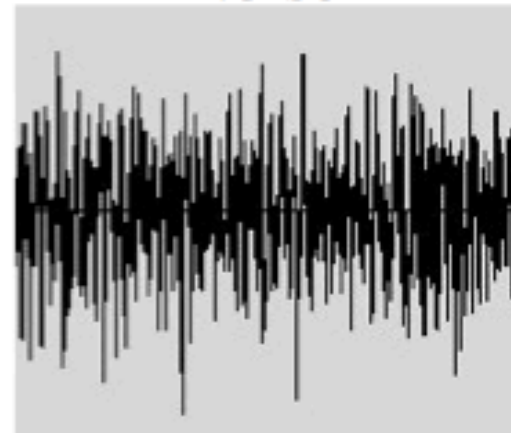
Speech



Speech in noise



Noise



Example: Image Classification

X:



Y:

	mite	container ship	motor scooter	leopard
	mite	container ship	motor scooter	leopard
	black widow	lifeboat	go-kart	jaguar
	cockroach	amphibian	moped	cheetah
	tick	fireboat	bumper car	snow leopard
	starfish	drilling platform	golfcart	Egyptian cat

X:



Y:

	grille	mushroom	cherry	Madagascar cat
	convertible	agaric	dalmatian	squirrel monkey
	grille	mushroom	grape	spider monkey
	pickup	jelly fungus	elderberry	titi
	beach wagon	gill fungus	ffordshire bullterrier	indri
	fire engine	dead-man's-fingers	currant	howler monkey

Regression

- **Goal:** Predict **real valued** labels (possibly vectors)
- **Examples:**

X

Flight route

Real estate objects

Patient & drug

....

Y

Delay (minutes)

Price

Treatment effectiveness

...

Example: Recommender systems

Your Recent History (What's this?)

Recently Viewed Items



[Probabilistic Graphical Models: Principles and...](#)
by Daphne Koller

Continue shopping: Customers Who Bought Items in Your Recent History Also Bought



Knatsch unter den Atomfreunden

Basler Zeitung - vor 2 Stunden

Die AKW-Betreiber stecken kein Geld in den Abstimmungskampf gegen die Atomausstiegsinitiative. Besonders einer hat «null Verständnis» dafür. Müsste bei einem Ja zur Atomausstiegsinitiative spätestens 2024 stillgelegt werden: Das Atomkraftwerk in ...



Der «Dschungel» wird geräumt

Tages-Anzeiger Online - vor 2 Stunden

Die Zeltstadt in Calais ist für Tausende Flüchtlinge Endstation auf ihrem Weg nach Grossbritannien. Jetzt reissen die Behörden einen grossen Teil ab. Ein Flüchtling im «Dschungel» bei Calais: Am Mittwoch soll der südliche Teil des Flüchtlingscamps ...



Graubünden: Vier Alpinisten von Lawinen mitgerissen

20 Minuten - vor 8 Stunden

In den Bündner Bergen sind am Sonntag vier Sportler von Lawinen mitgerissen worden. Zwei Skitourengänger und zwei Eiskletterer wurden verletzt in Spitäler geflogen. storybild. Lawine am Sentisch Horn: Zwei Männer wurden mitgerissen und stürzten 200 ...



«Es ist falsch, Mütter aus dem Arbeitsprozess auszugrenzen»

Tages-Anzeiger Online - vor 10 Minuten

Frischgebackene Mütter werden nach ihrer Rückkehr an den Arbeitsplatz immer häufiger entlassen. Eine Betroffene erzählt. Gesetzlich benachteiligt? Eine werdende Mutter sitzt an ihrem Arbeitsplatz. (Symbolbild) Bild: Gaetan Bally/Keystone. Yannick Wiget



Referendum in Bolivien: Der Glanz von Morales ist verblasst

Neue Zürcher Zeitung - vor 27 Minuten

Evo Morales hat in den zehn Jahren an der Macht nicht nur Bolivien verändert, sondern auch sich selbst. Eine weitere Amtszeit ab 2020 dürfte es für Morales aller Voraussicht nach nicht geben. von Tjerk Brühwiller, São Paulo; 22.2.2016, 20:26 Uhr ...

- **X:** User & article / product features
- **Y:** Ranking of articles / products to display

Example: Image captioning

Y

A person riding a motorcycle on a dirt road.



Two dogs play in the grass.



A skateboarder does a trick on a ramp.



X

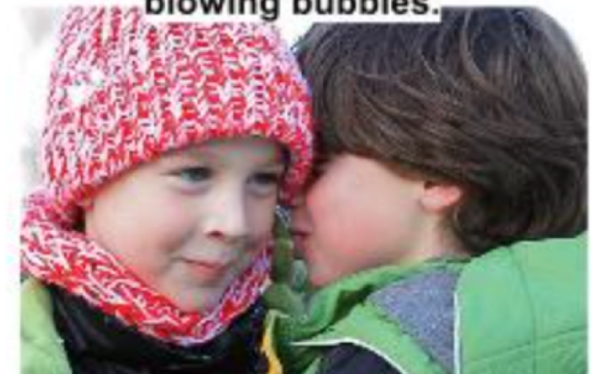
A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



A little girl in a pink hat is blowing bubbles.



Vinyals et al. *Show and Tell: A Neural Image Caption Generator* '14

Example: Translation

The screenshot shows the Google Translate web interface. At the top left is the Google logo. To the right are icons for a grid, a notification bell, and a user profile picture. Below the logo, the word "Translate" is written in red. On the right side of this bar, there is a link "Turn off instant translation" and a star icon. The main interface has two language selection boxes. The left box shows "English", "Spanish", "French", and "Detect language" with a dropdown arrow. The right box shows "English", "Spanish", and "German" with a dropdown arrow. A blue "Translate" button is positioned between the two boxes. Below the language boxes, there are two text input areas. The left area contains the text "Machine learning is getting more accurate" and is highlighted with a blue border. Below this text are icons for a speaker, a keyboard, and a character count "41/5000". The right area contains the German translation "Maschinelles Lernen wird immer genauer". Below this text are icons for a star, a copy icon, a speaker, a share icon, and a pencil icon.

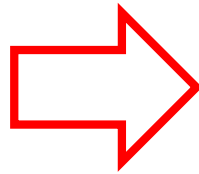
X

Y

Example: Predicting program properties

[Raychev, Vechev, Krause POPL '15]

```
function chunkData(e, t) {  
  var n = [];  
  var r = e.length;  
  var i = 0;  
  for (; i < r; i += t) {  
    if (i + t < r) {  
      n.push(e.substring(i, i + t));  
    } else {  
      n.push(e.substring(i, r));  
    }  
  }  
  return n;  
}
```



```
/* str: string, step: number, return: Array */  
function chunkData(str, step) {  
  var colNames = []; /* colNames: Array */  
  var len = str.length;  
  var i = 0; /* i: number */  
  for (; i < len; i += step) {  
    if (i + step < len) {  
      colNames.push(str.substring(i, i + step));  
    } else {  
      colNames.push(str.substring(i, len));  
    }  
  }  
  return colNames;  
}
```

X

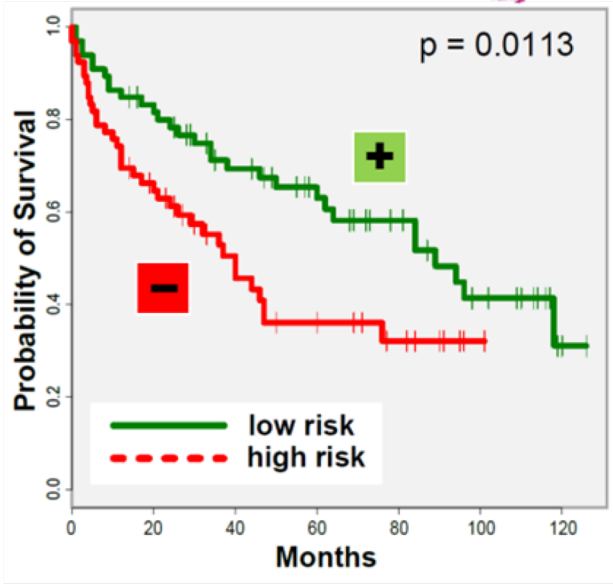
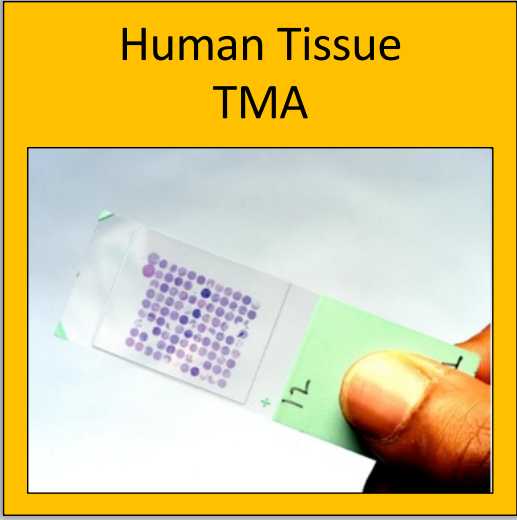
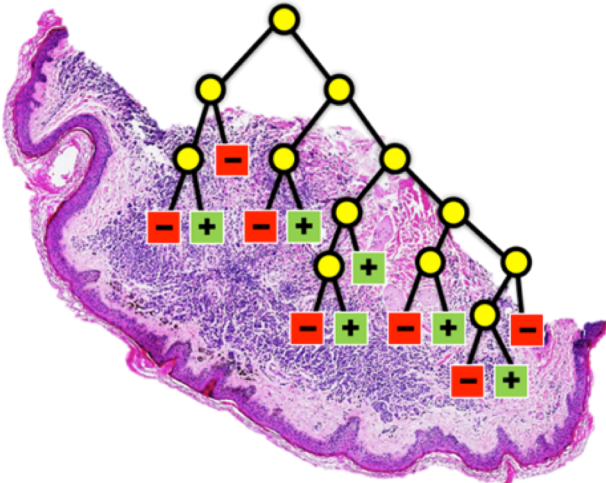
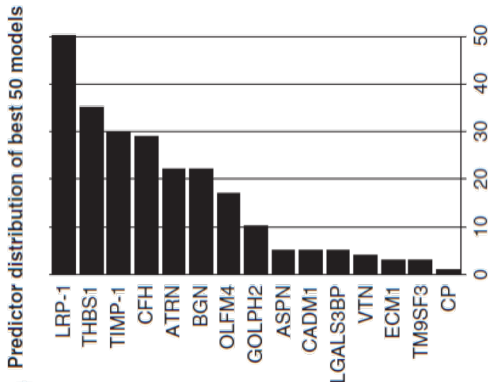
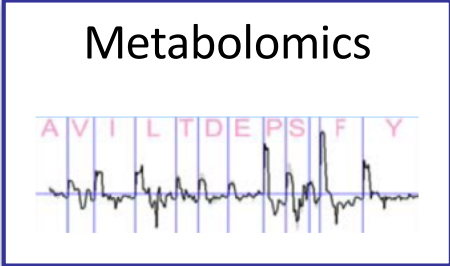
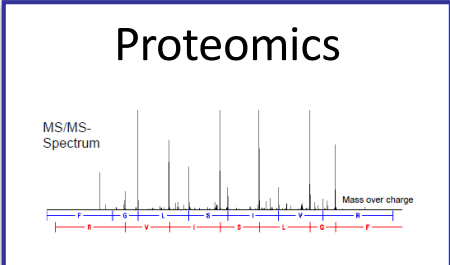
Y

jsnice.org

Example: Computational Pathology

[Buhmann, Fuchs et al.]

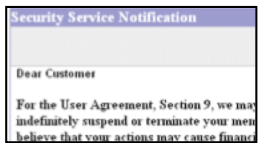
X



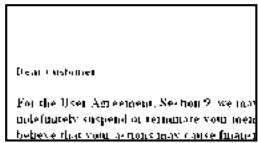
Y

Basic Supervised Learning Pipeline

Training Data



“spam”



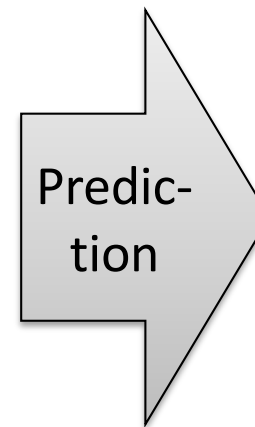
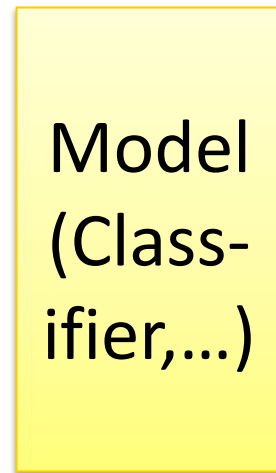
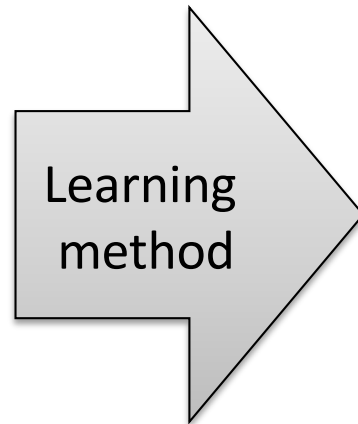
“ham”



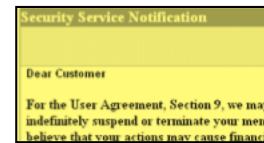
“spam”

\mathcal{X}

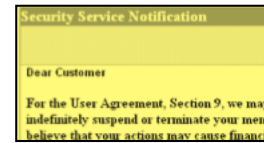
\mathcal{Y}



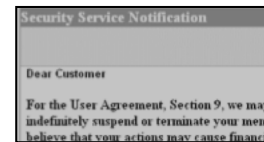
Test Data



?



?



?

\mathcal{X}

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Representation

Model fitting

Prediction and Generalization

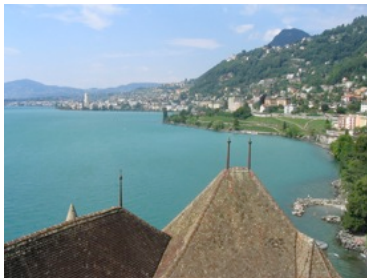
Representing Data

- Learning methods expect **standardized representation** of data (e.g., Points in vector spaces, nodes in a graph, similarity matrices ...)

The quick brown
fox jumps over
the lazy dog ...



[0 1 0 0 0 3 2 0 1 0 0 0]



[.3 .01 .1 2.3 0 0 1.1 ...]

- Concrete choice of representation („features“) is **crucial** for successful learning
- This class (typically): **feature vectors** in \mathbb{R}^d

Example: Bag-of-words

- Suppose language contains at most $d=100000$ words
- Represent each document as a vector \mathbf{x} in \mathbb{R}^d
 - i -th component x_i counts occurrence of i -th word

Word	Index
a	1
abandon	2
ability	3
...	
is	578
...	
test	2512
...	
this	2809
....	

$x =$ "this is a test"

$$\mapsto [1 \ 0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0] \in \mathbb{R}^d$$

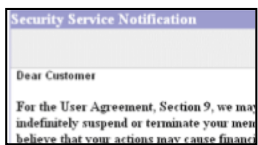
↑ ↑ ↑ ↑ ↑
a abandon is test this

Bag-of-words: Improvements

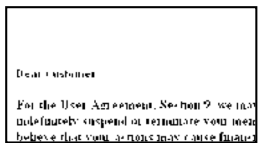
- Length of the document should not matter
 - Replace counts by binary indicator (yes/no)
 - Normalize to unit length
- Some words more „important“ than others
 - Remove „stopwords“ (the, a, is, ...)
 - Stemming (learning, learner, learns -> learn)
 - Discount frequent words (tf-idf)
- Bag-of-words ignores order
 - Consider pairs (n-grams) of consecutive words
- Does not differentiate between similar and dissimilar words (ignores semantics)
 - Word embeddings (e.g., word2vec, GloVe)

Basic Supervised Learning Pipeline

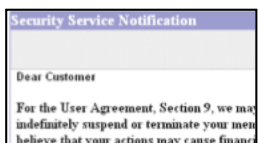
Training Data



“spam”



“ham”

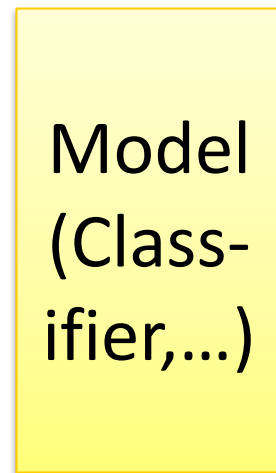
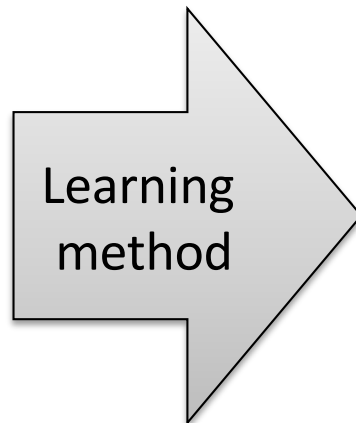


“spam”

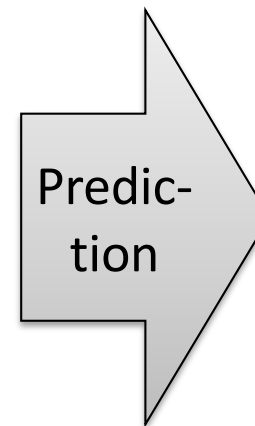
\mathcal{X}

\mathcal{Y}

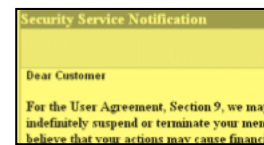
Representation



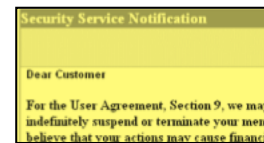
Model fitting



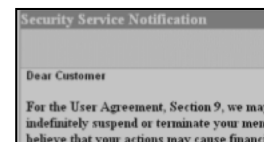
Test Data



?



?



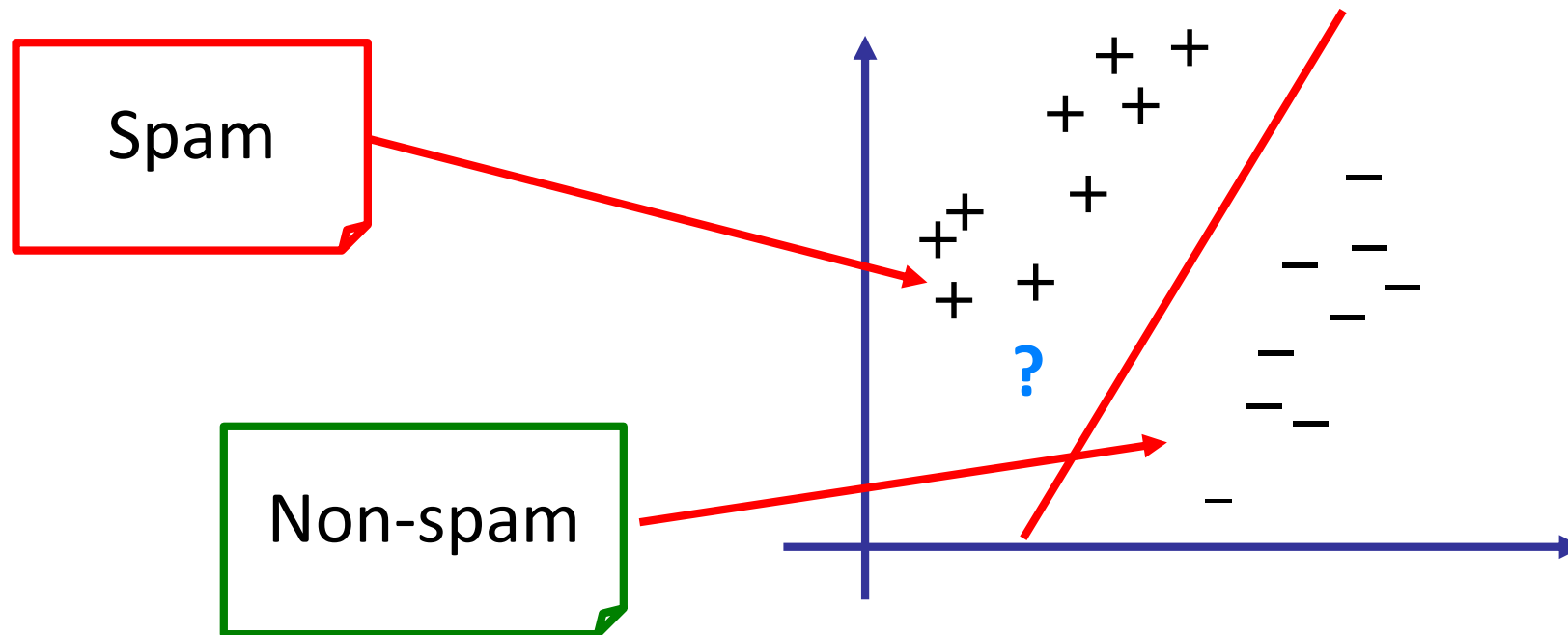
?

\mathcal{X}

Prediction and Generalization

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

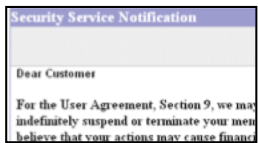
Example: Classifying Documents



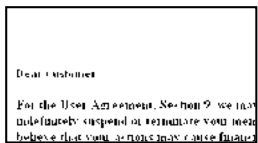
- **Input:** Training examples (e.g., “bag-of-words”) with positive (+) and negative (-) labels
- **Goal:** Decision rule (aka hypothesis, e.g., linear, decision tree, random forest, deep neural network ...)

Basic Supervised Learning Pipeline

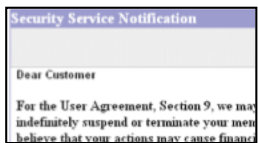
Training Data



“spam”



“ham”

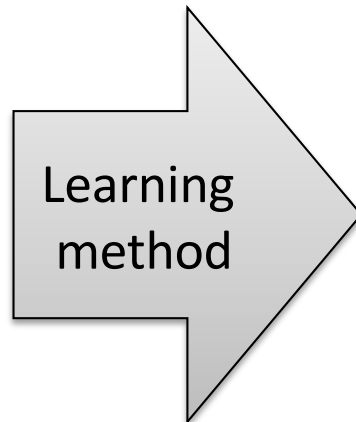


“spam”

\mathcal{X}

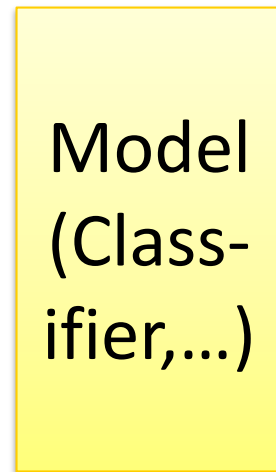
\mathcal{Y}

Representation



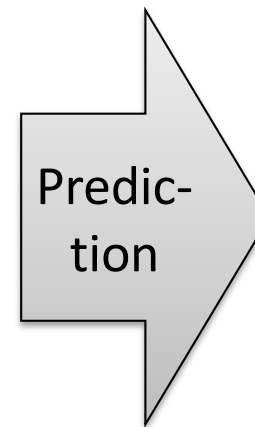
Learning
method

Model fitting



Model
(Class-
ifier,...)

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



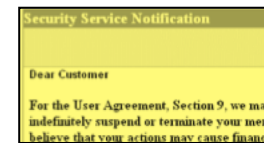
Predic-
tion

Prediction and
Generalization

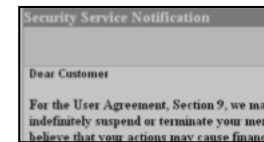
Test Data



?



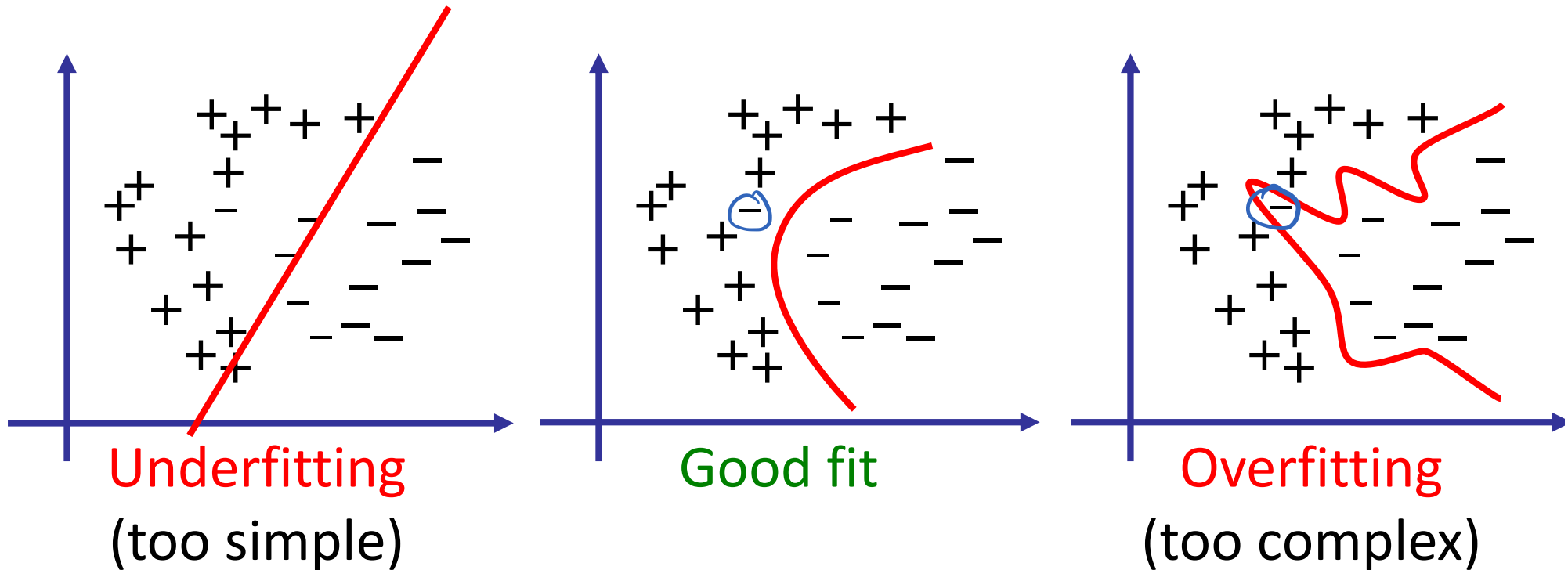
?



?

\mathcal{X}

Model selection and validation



- Automatic model-selection and validation of crucial importance (→ statistical learning theory)
- **Goal:** Balance of “Goodness of Fit” and complexity
- Ideal models are simultaneously statistically and computationally efficient

Machine Learning Tasks

Supervised Learning

- Classification
- Regression
- Structured Prediction, ...

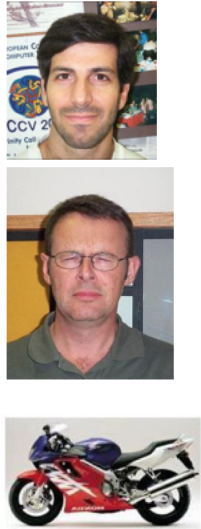
Unsupervised Learning

- Clustering
- Dimension reduction
- Anomaly detection, ...

Many other specialized tasks

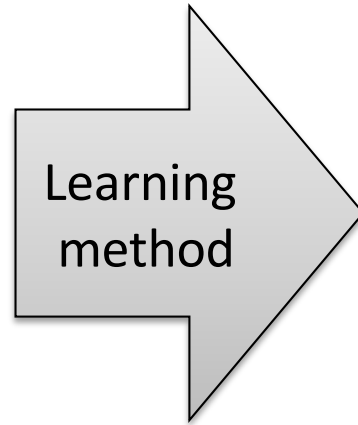
Basic Unsupervised Learning Pipeline

Training Data

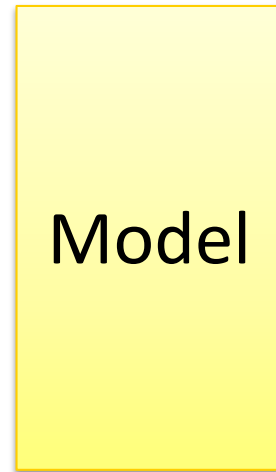


\mathcal{X} ~~\mathcal{Y}~~

Representation



Model fitting



$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

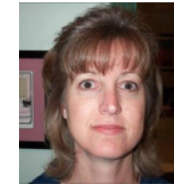
Test Data



?



?



?

Prediction

Unsupervised learning

- „Learning without labels“
- Examples:
 - Clustering (e.g., unsupervised classification)
 - Dimension reduction (e.g., unsupervised regression)
 - Generative modeling (topic models, autoencoders, GANs etc.)
- Common goals:
 - Compact representation / compression of data sets
 - Identification of latent variables
- Use-cases:
 - Exploratory data analysis
 - Feature learning / embedding
 - Anomaly detection of „unusual“ data points

Example: Clustering



- **Input:** Data set without labels
- **Goal:** Assignment to clusters (infer labels)

Example: Dimension Reduction

[Roweis & Saul, Nonlinear dimensionality reduction by locally linear embedding, Science '00]

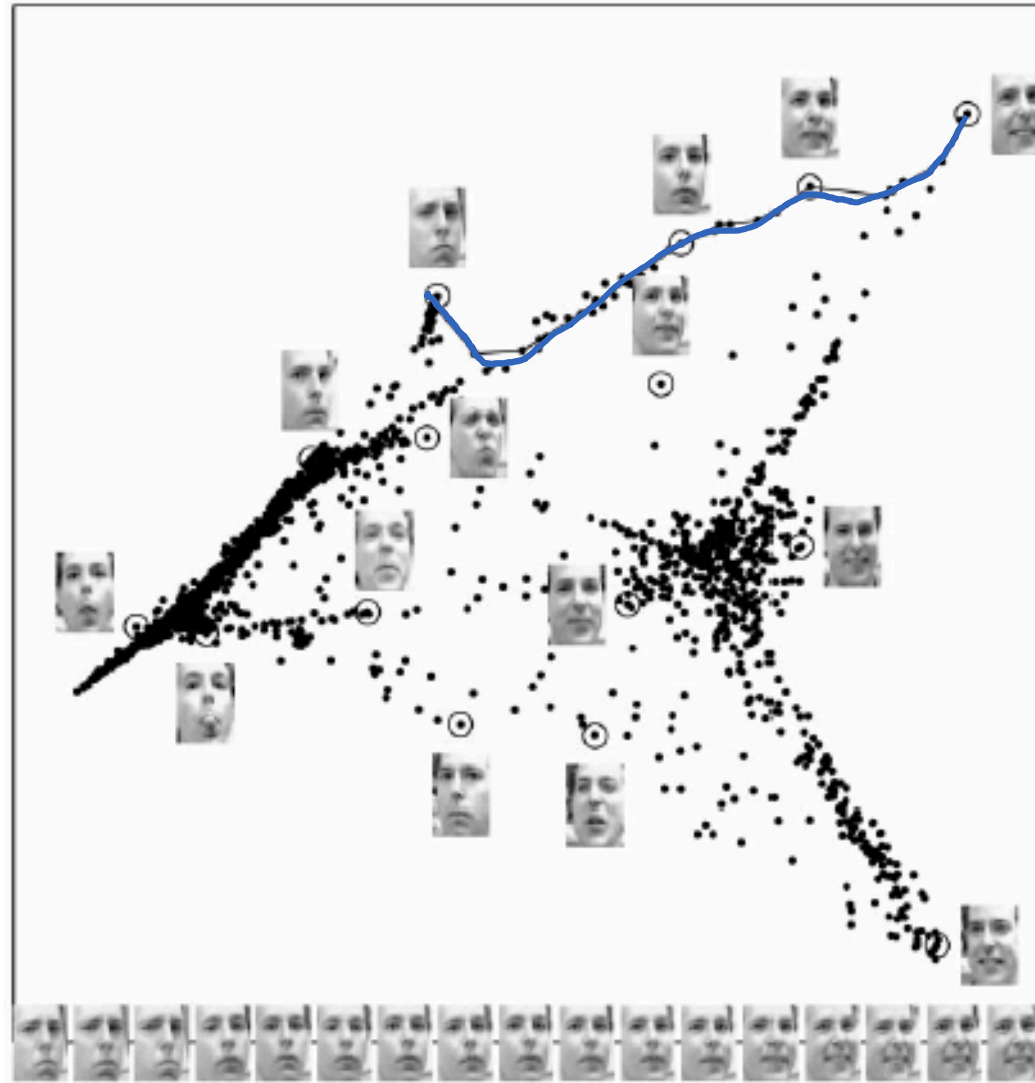


Fig. 3. Images of faces (11) mapped into the embedding space described by the first two coordinates of LLE. Representative faces are shown next to circled points in different parts of the space. The bottom images correspond to points along the top-right path (linked by solid line), illustrating one particular mode of variability in pose and expression.

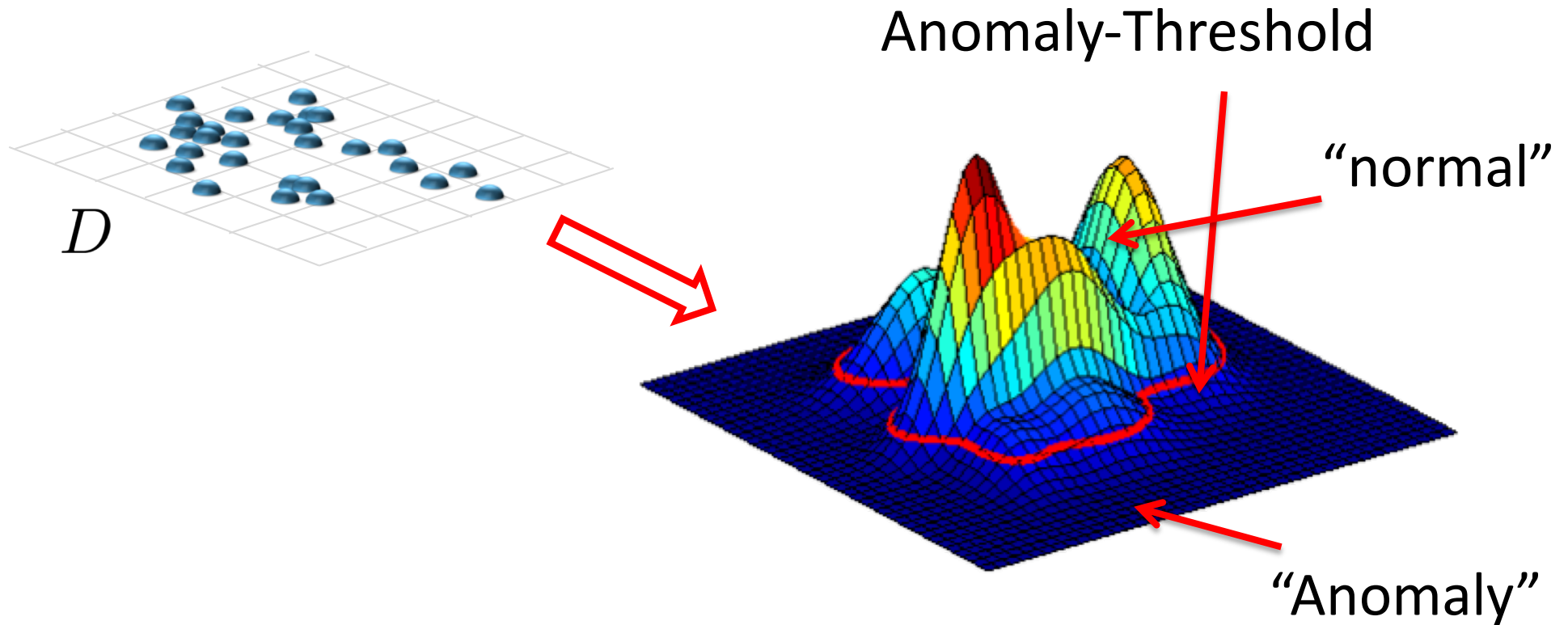
Example: Dimension reduction

- Often, high-dimensional data can be well approximated in low dimensions
- Very useful for visualization!
- Many methods available, e.g.,
 - *Linear* (Principal Component Analysis, Linear Discriminant Analysis, ...)
 - *Non-linear* (ISOMAP, Kernel-PCA, Max. variance unfolding, t-SNE, autoencoders based on neural networks, ...)
 - *Sparse modeling / inference*



Eigenfaces
[AT&T Labs Cambridge]

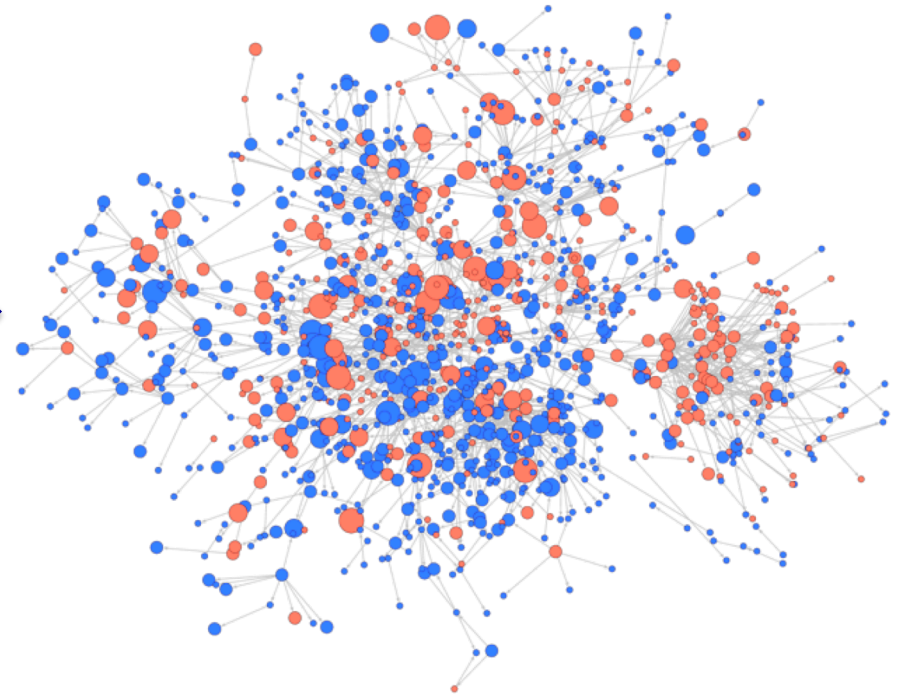
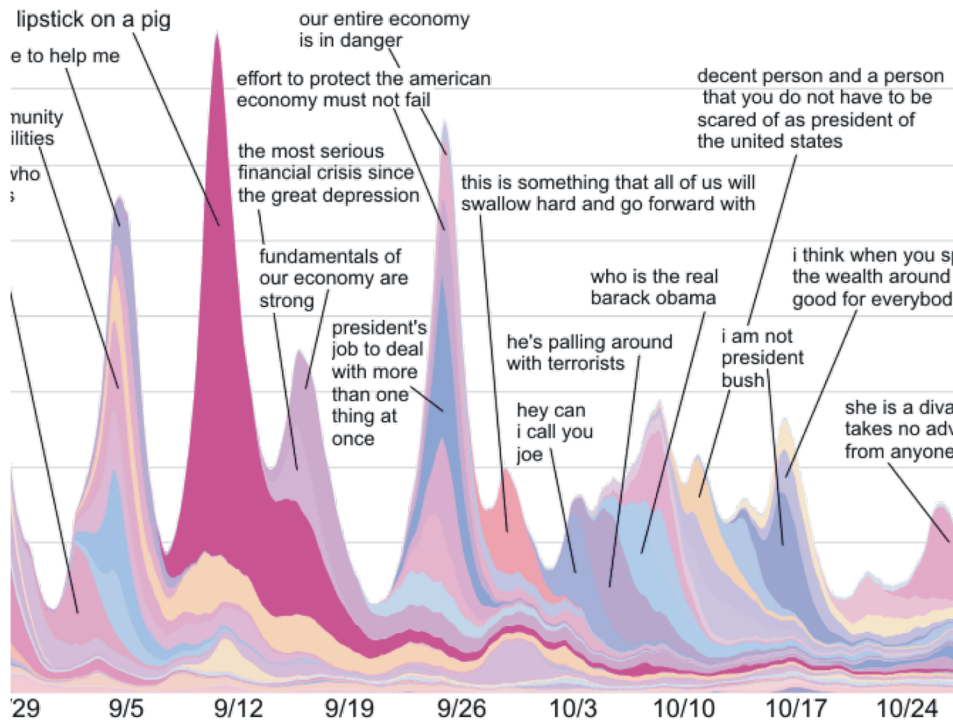
Example: Anomaly detection



- **Application:** Quality control, fraud detection, ...
- Fit statistical model of “normal” data
- Declare “unusual” (low prob.) data as anomaly

Example: Network inference

[Gomez Rodriguez, Leskovec, Krause ACM TKDE 2012]



- Estimate flow of information and influence in the „blogosphere“ (ecosystem of blogs and social media)

Example: Never Ending Language Learning

[Mitchell et al.]

(Mostly) unsupervised acquisition of facts by „reading“
the internet

Recently-Learned Facts



Refresh

instance	iteration	date learned	confidence	
gold_coast_casino is a visualizable attribute	896	24-jan-2015	94.4	
stranger_software is a tool	896	24-jan-2015	99.1	
regent_beach_resort is a trail	896	24-jan-2015	100.0	
squitieri_studio_theatre is a stadium or event venue	896	24-jan-2015	100.0	
fish_river_seaplane_base is an airport	896	24-jan-2015	100.0	
john_lucas plays the athletic team position player	901	14-feb-2015	93.8	
european_architects is a generalization of walter_gropius	901	14-feb-2015	100.0	
young is a person who moved to the state california	901	14-feb-2015	100.0	
justine_henin is an athlete who beat svetlana_kuznetsova	901	14-feb-2015	100.0	
public_administration is an academic program at the university louisiana_state_university	899	05-feb-2015	96.9	

[rtw.ml.cmu.edu]

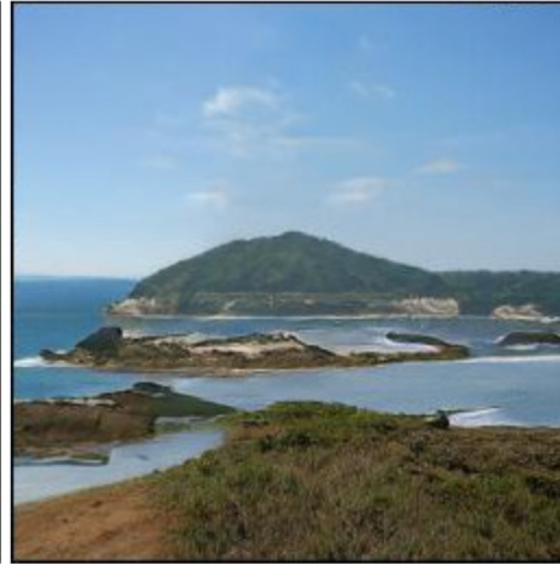
Example: GANs

[Goodfellow et al'14, Salimans et al'16]



BigGAN

[Brock, Donahue, Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis ICLR '19]



Machine Learning Tasks

Supervised Learning

- Classification
- Regression
- Structured Prediction, ...

Unsupervised Learning

- Clustering
- Dimension reduction
- Anomaly detection, ...

Many other specialized tasks

Other models of learning

- Semi-supervised learning
 - Learning from both labeled and unlabeled data
- Transfer & meta learning
 - Learn on one domain and test on another
- Active learning
 - Acquiring most informative data for learning
- Online / lifelong / continual learning
 - Learning from examples as they arrive over time
- Reinforcement learning
 - Learning by interacting with an unknown environment
- ...

Summary so far

- Two basic forms of learning:
 - Supervised vs. Unsupervised learning
- Key challenge in ML
 - Trading goodness of fit and model complexity
- Representation of data is of key importance