

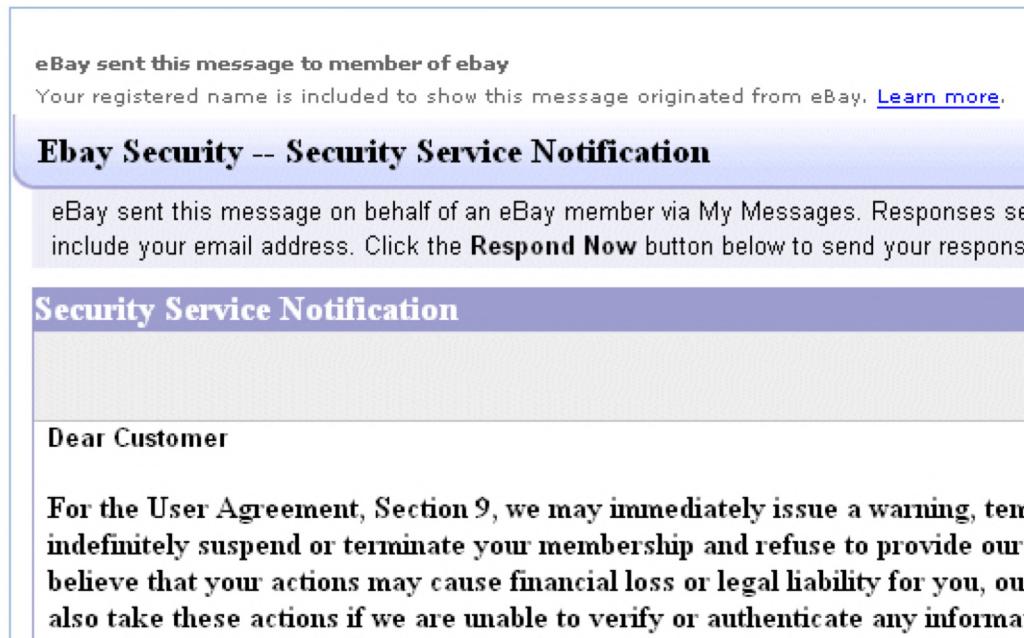
# Introduction to Machine Learning

## Introduction

Prof. Andreas Krause  
Institute for Machine Learning  
([las.ethz.ch](http://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”

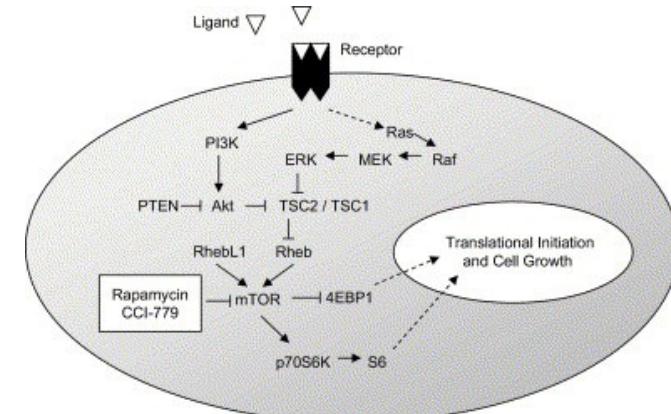
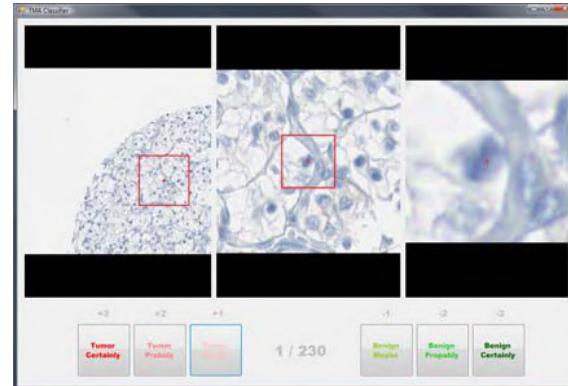
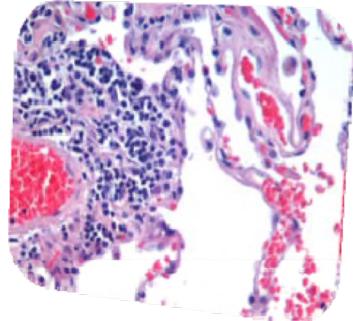
THEN classify as “spam” ELSE “non-spam”

**Machine Learning:** Automatic identification 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

statistics

algorithms  
& optimization

information  
theory

**machine  
learning**

philosophy  
epistemology  
causality

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.inf.ethz.ch/teaching/introml-s19>
  - Old slides available at [.../introml-s18](#)
  - 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.github.io>
- 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

# 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 Foundations of AI (Fall)
  - Reasoning and decision making under uncertainty
- Computational Intelligence Lab (Spring)
  - Matrix Factorization, Recommender Systems, projects
- Statistical Learning Theory (Spring)
  - Theoretical foundations; model validation
- Computational Statistics (D-MATH, Spring)

# Overview

- *Instructor:*

Andreas Krause ([krausea@ethz.ch](mailto:krausea@ethz.ch))

- *Teaching assistants:*

Head TA: Mojmir Mutny ([mojmir.mutny@inf.ethz.ch](mailto:mojmir.mutny@inf.ethz.ch))

Andisheh Amrollahi, Mohammad Karimi, Prashanth Chandran, Joanna Ficek, Vincent Fortuin, Gürel Nezihe Merve, Harun Mustafa, Jingwei Tang, Kjong Lehmann, Natalie Davidson, Olga Mineeva, Laurie Prelot, Stefan Stark, Johannes Kirschner, Sebastian Curi, Max Paulus, Phillippe Wenk, Aytunc Sahin, Kfir Levy, Anastasiia Makarova, Maysam Haghda

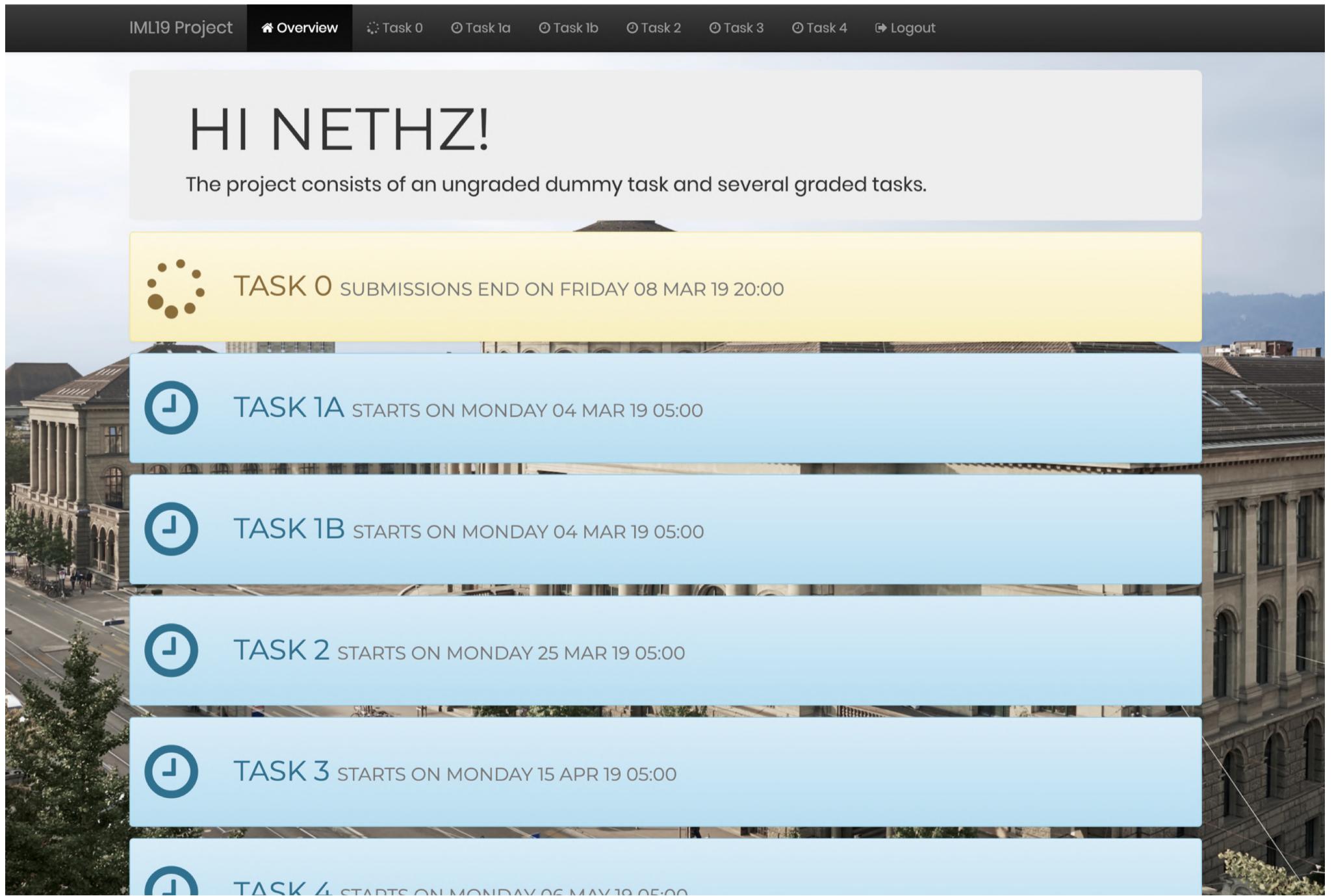
# Exercises

- Take them seriously if you want to pass the exam...
- Four sessions (attend according to initial of last name)
  - Mon      15-17   HG D 1.2      Last Names **A-C**
  - Tue      15-17   HG D 1.2      Last Names **D-H**
  - Wed      15-17   CAB G 11      Last Names **I-M**
  - Fri      13-15   ML D 28      Last Names **N-Z**
- This week: Optional refresher on basic linear algebra, calculus and probability

# 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>



IML19 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 FRIDAY 08 MAR 19 20:00

**TASK 1A** STARTS ON MONDAY 04 MAR 19 05:00

**TASK 1B** STARTS ON MONDAY 04 MAR 19 05:00

**TASK 2** STARTS ON MONDAY 25 MAR 19 05:00

**TASK 3** STARTS ON MONDAY 15 APR 19 05:00

**TASK 4** STARTS ON MONDAY 06 MAY 19 05:00

# Some FAQs

- Distance exams
  - are possible (as exception), but need to officially request with study administration
- Doctoral students who need a “Testat”
  - need to receive a passing grade (4.0 or higher)
- 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  
Learning and Adaptive Systems ([las.ethz.ch](http://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**

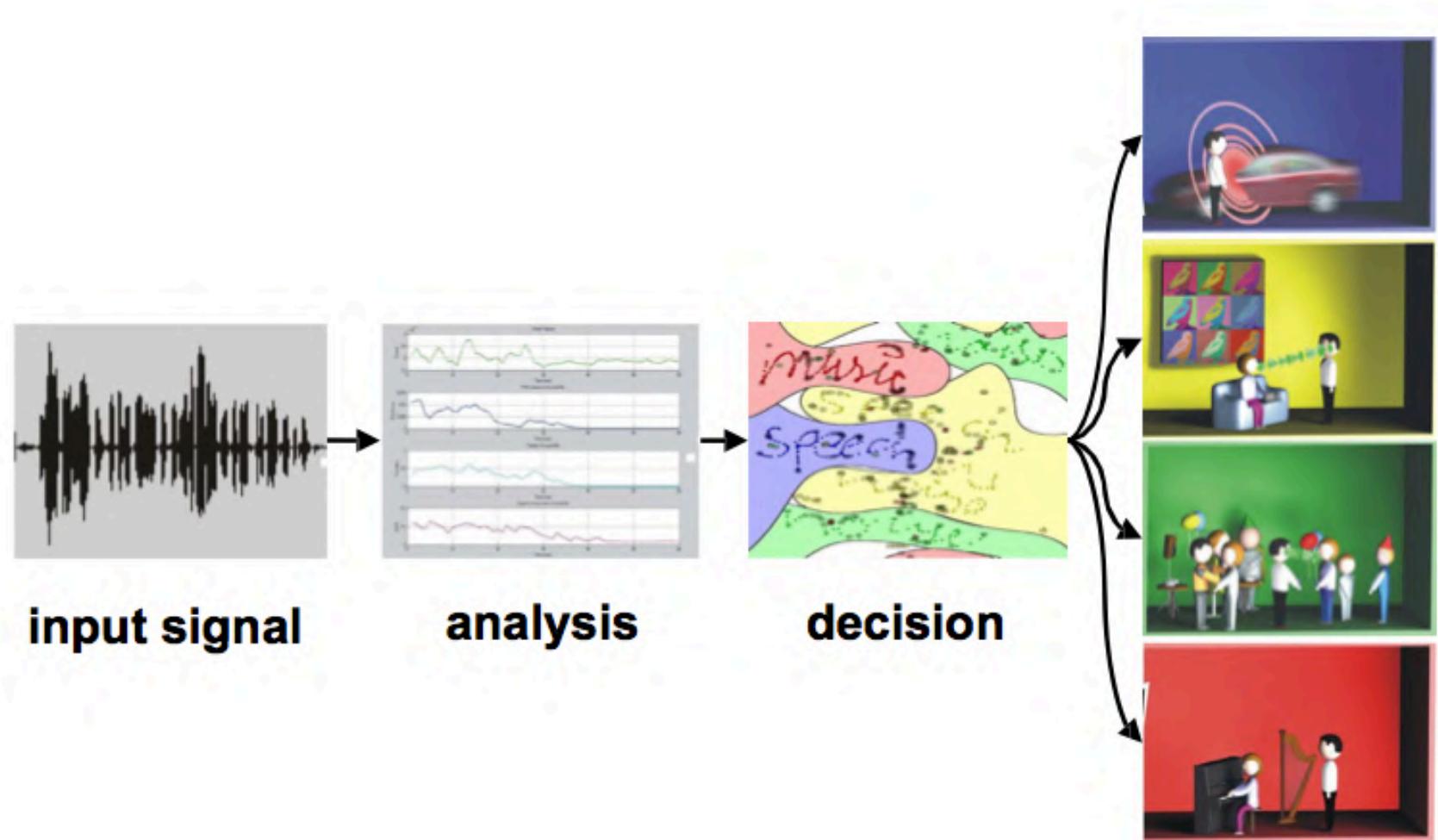
Dear Customer

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

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

# Example: Improving Hearing Aids

[Buhmann et al]



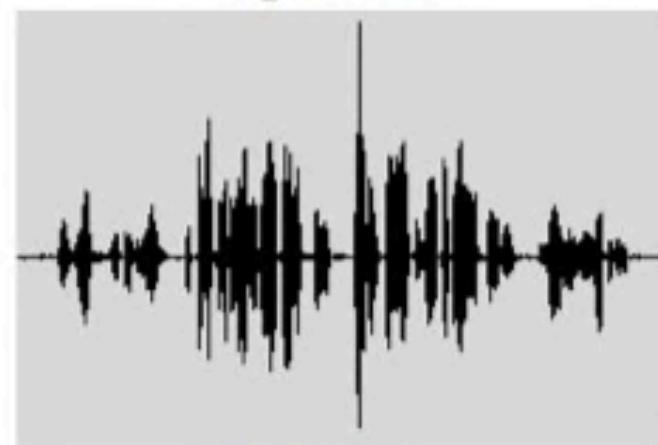
- **X:** Acoustic waveforms
- **Y:** label *speech, speech in noise, music, noise*

# Example: Improving Hearing Aids

Music



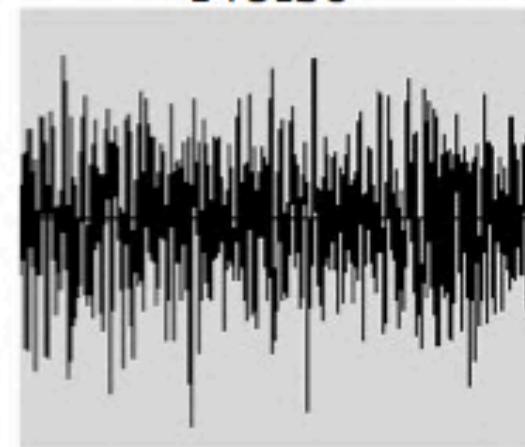
Speech



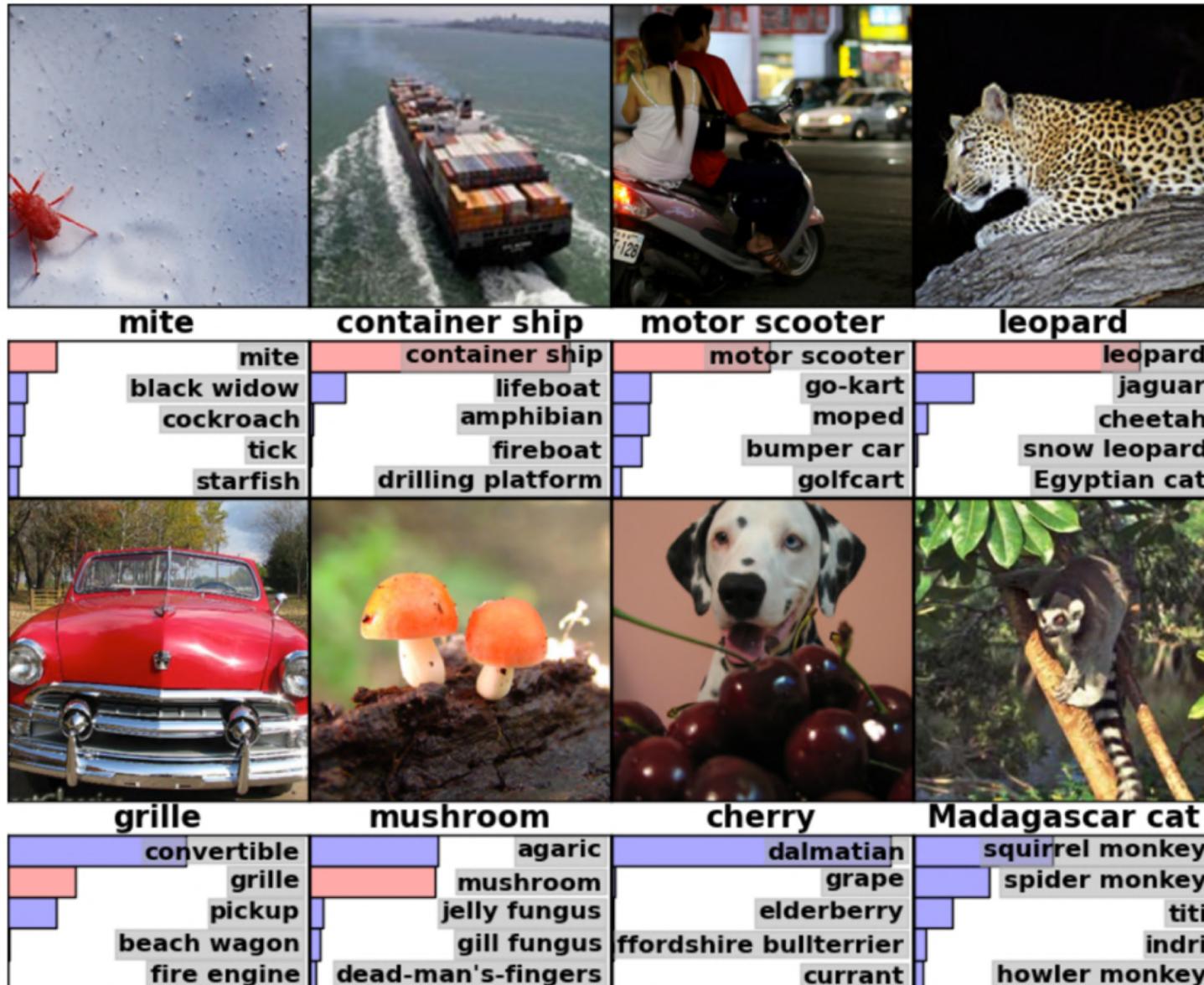
Speech in noise



Noise



# Example: Image Classification



# Regression

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

X	Y
Flight route	Delay (minutes)
Real estate objects	Price
Patient & drug	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 Skitouengänger und zwei Eiskletterer wurden verletzt in Spitäler geflogen. storybild. Lawine am Sentsch Horn: Zwei Männer wurden mitgerissen und stürzten 200 ...

**«Es ist falsch, Mütter aus dem Arbeitsprozess auszgrenzen»**  
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 Brühwiler, 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.



X

A group of young people playing a game of frisbee.



Two dogs play in the grass.



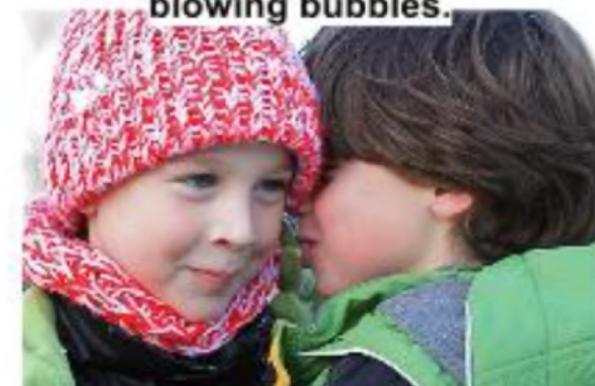
A skateboarder does a trick on a ramp.



Two hockey players are fighting over the puck.



A little girl in a pink hat is blowing bubbles.



# Example: Translation

The screenshot shows the Google Translate web interface. At the top left is the Google logo. On the right are icons for a grid, a bell, and a user profile. Below the logo, the word "Translate" is written in red. To its right are links for "Turn off instant translation" and a star icon. The main area has two language selection bars. The left bar shows "English" as the source language and "Spanish", "French", "Detect language" as options. The right bar shows "English" as the source language and "Spanish", "German" as options, with a "Translate" button to its right. Below these bars are two text boxes. The left text box contains the English sentence "Machine learning is getting more accurate" and has a character limit of 41/5000. The right text box displays the German translation "Maschinelles Lernen wird immer genauer". At the bottom of each text box are small icons for audio, keyboard, and sharing.

Machine learning is getting more accurate

41/5000

Maschinelles Lernen wird immer genauer

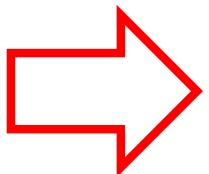
X

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;  
}
```

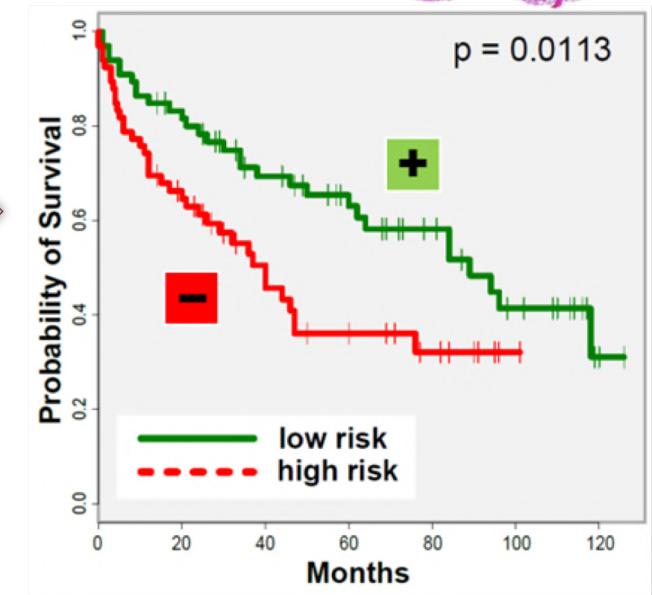
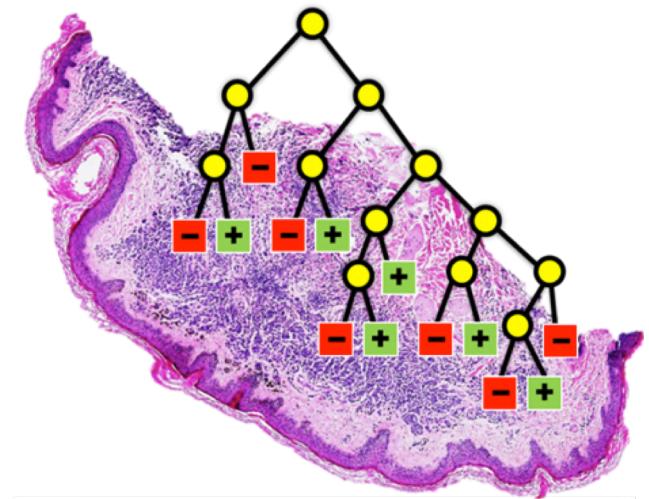
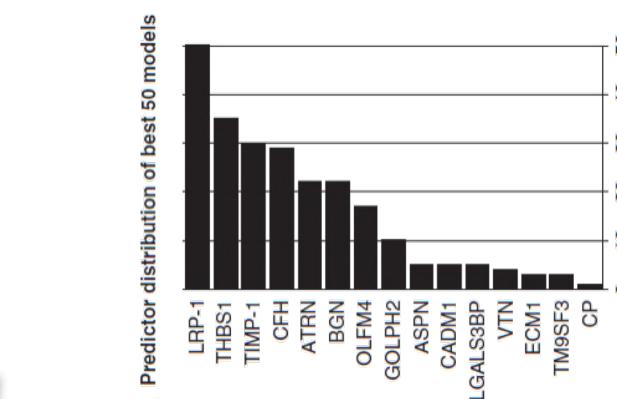
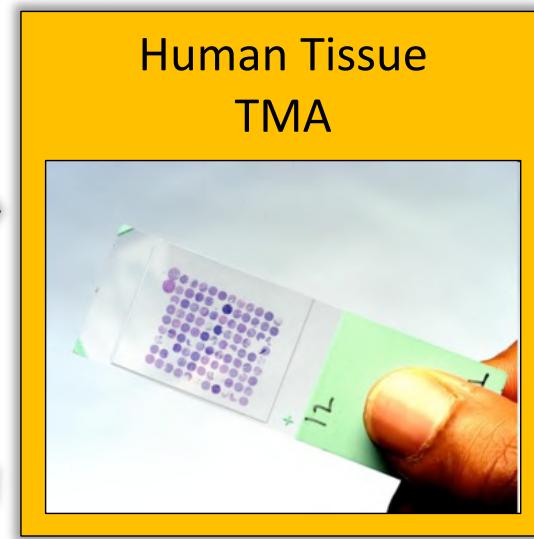
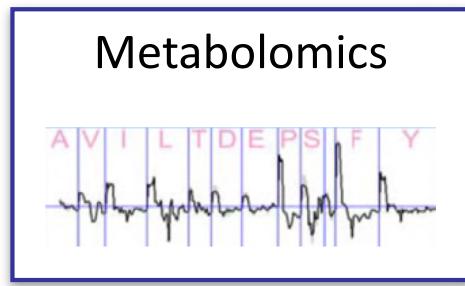
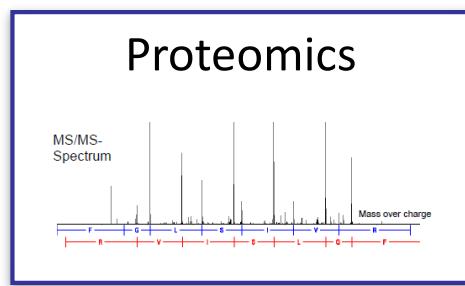
X

Y

jsnice.org

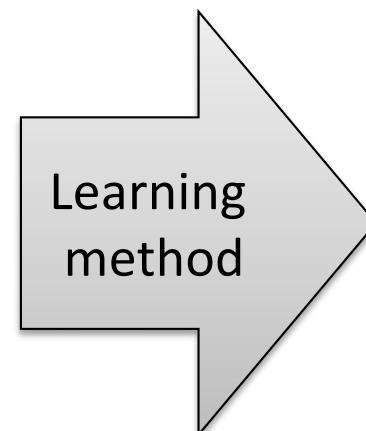
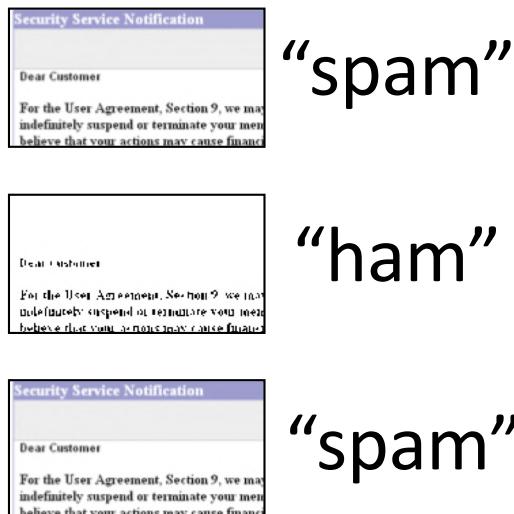
# Example: Computational Pathology

[Buhmann, Fuchs et al.]



# Basic Supervised Learning Pipeline

## Training Data



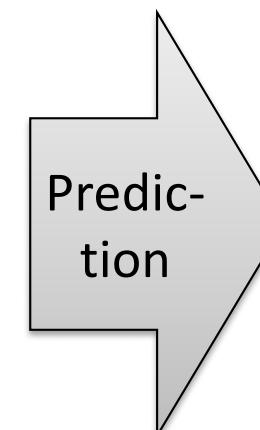
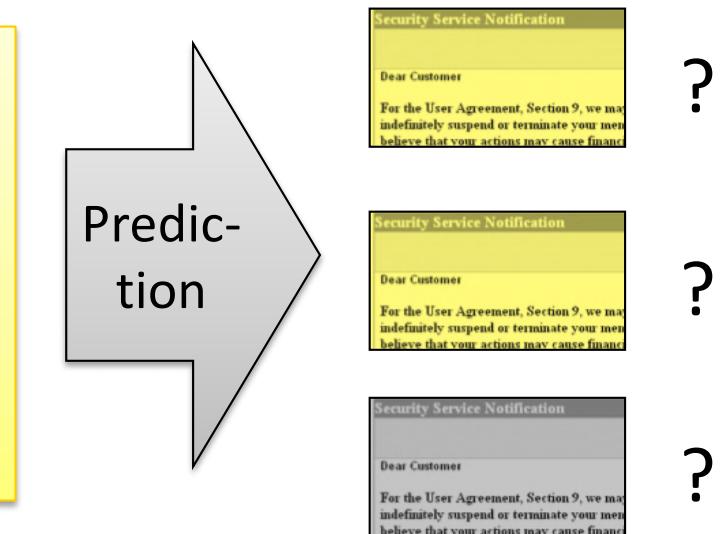
$\mathcal{X}$      $\mathcal{Y}$

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



Model fitting

## Test Data



$\mathcal{X}$

Prediction and  
Generalization

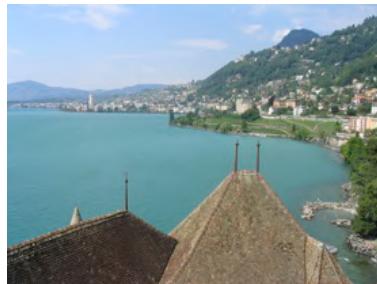
# Representing Data

- 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
....	

This is a test  
↳ [ 0 ... 0 0 1 0 ... 0 0 ]  
↑ ↑ ↑ ↑  
a 578 2512 2809  
a test is this

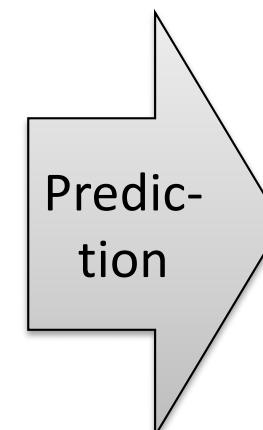
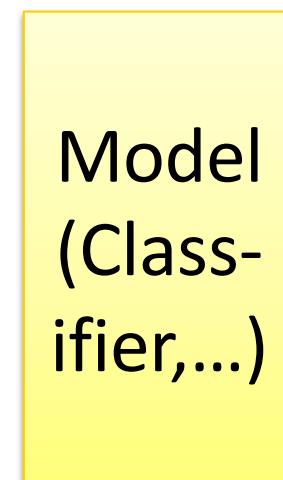
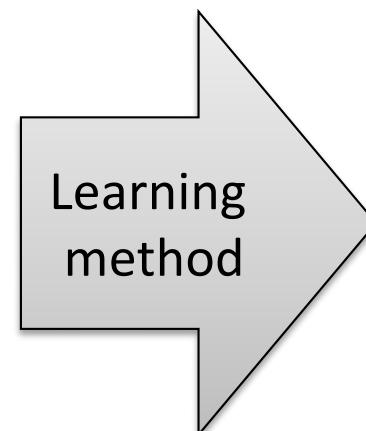
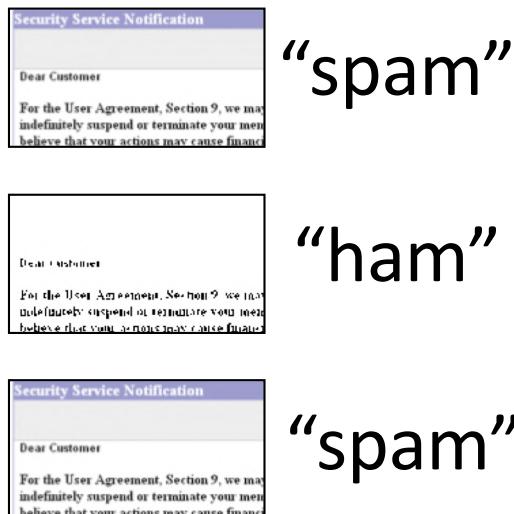
this → [ 0 ... 0 1 0 ... 0 ]  
↑  
2809

# Bag-of-words: Improvements

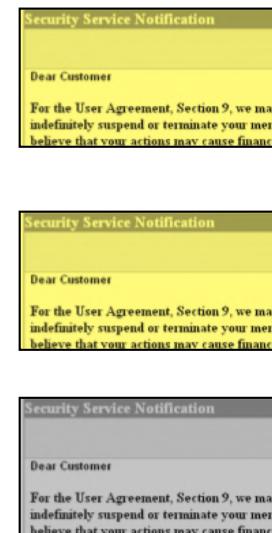
- Length of the document should not matter
  - Replace counts by binary indicator (yes/no)
  - Normalize to unit length
- Some words more „important“ than others
  - Remove „stopwords“ (the, a, is, ...)
  - Stemming (learning, learner, learns -> learn)
  - Discount frequent words (tf-idf)
- Bag-of-words 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



## Test Data



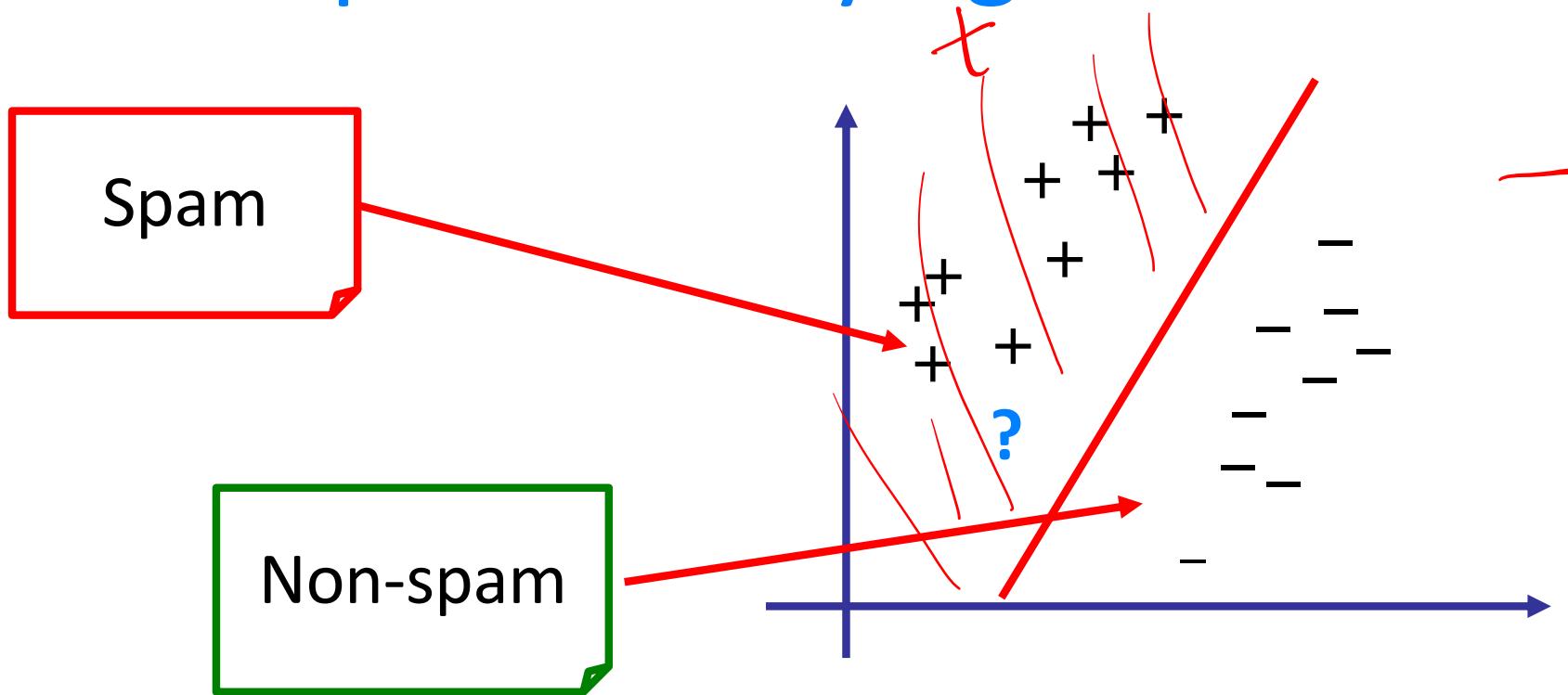
$$\mathcal{X} \quad \mathcal{Y} \quad f : \mathcal{X} \rightarrow \mathcal{Y}$$

Representation

Model fitting

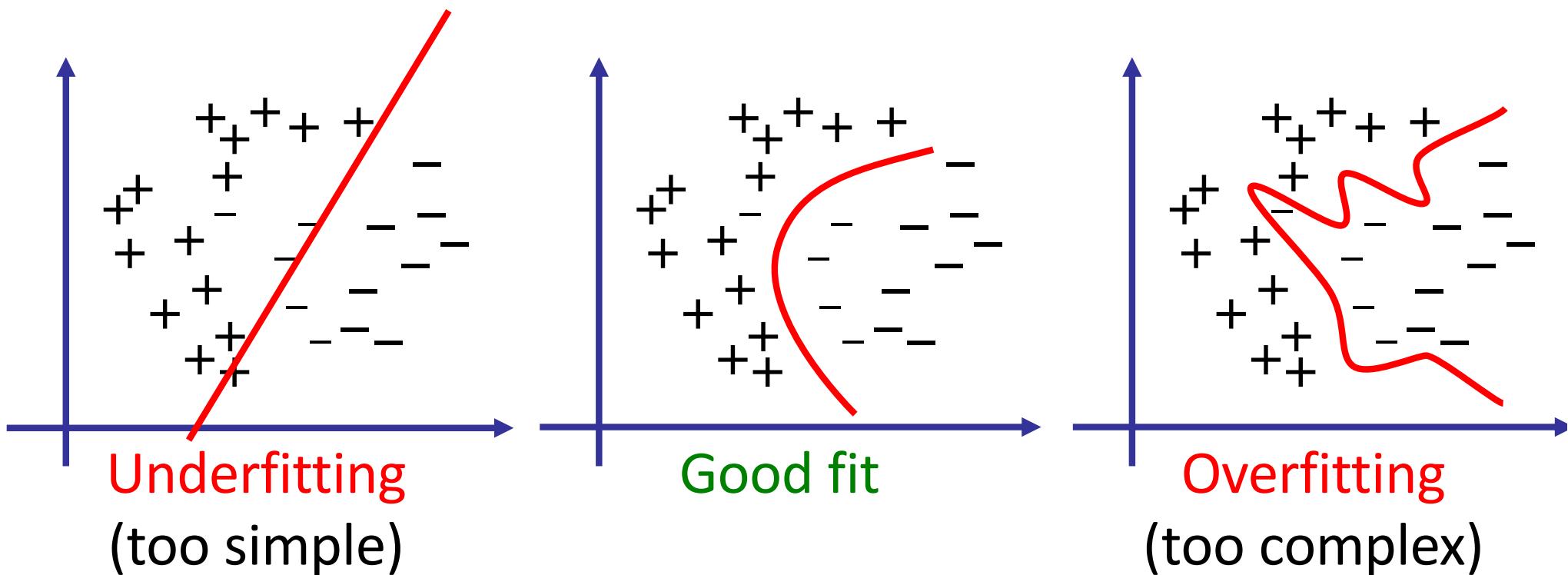
Prediction and  
Generalization

# 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 forests, graphical models, deep neural networks...)

# 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, ...

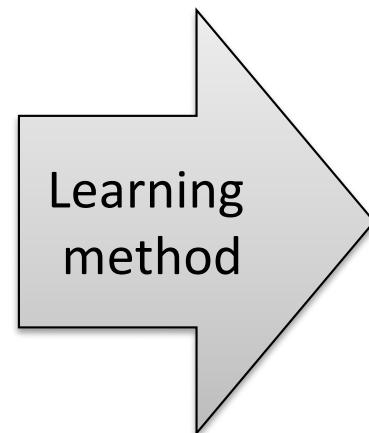
## *Unsupervised Learning*

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

*Many other specialized tasks*

# Basic Unsupervised Learning Pipeline

## Training Data



$\mathcal{X}$  

Representation

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

Model fitting

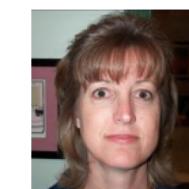
## Test Data



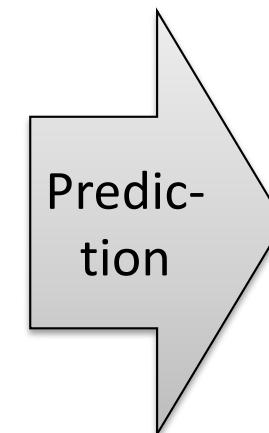
?



?



?



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 [Saul & Roweis]

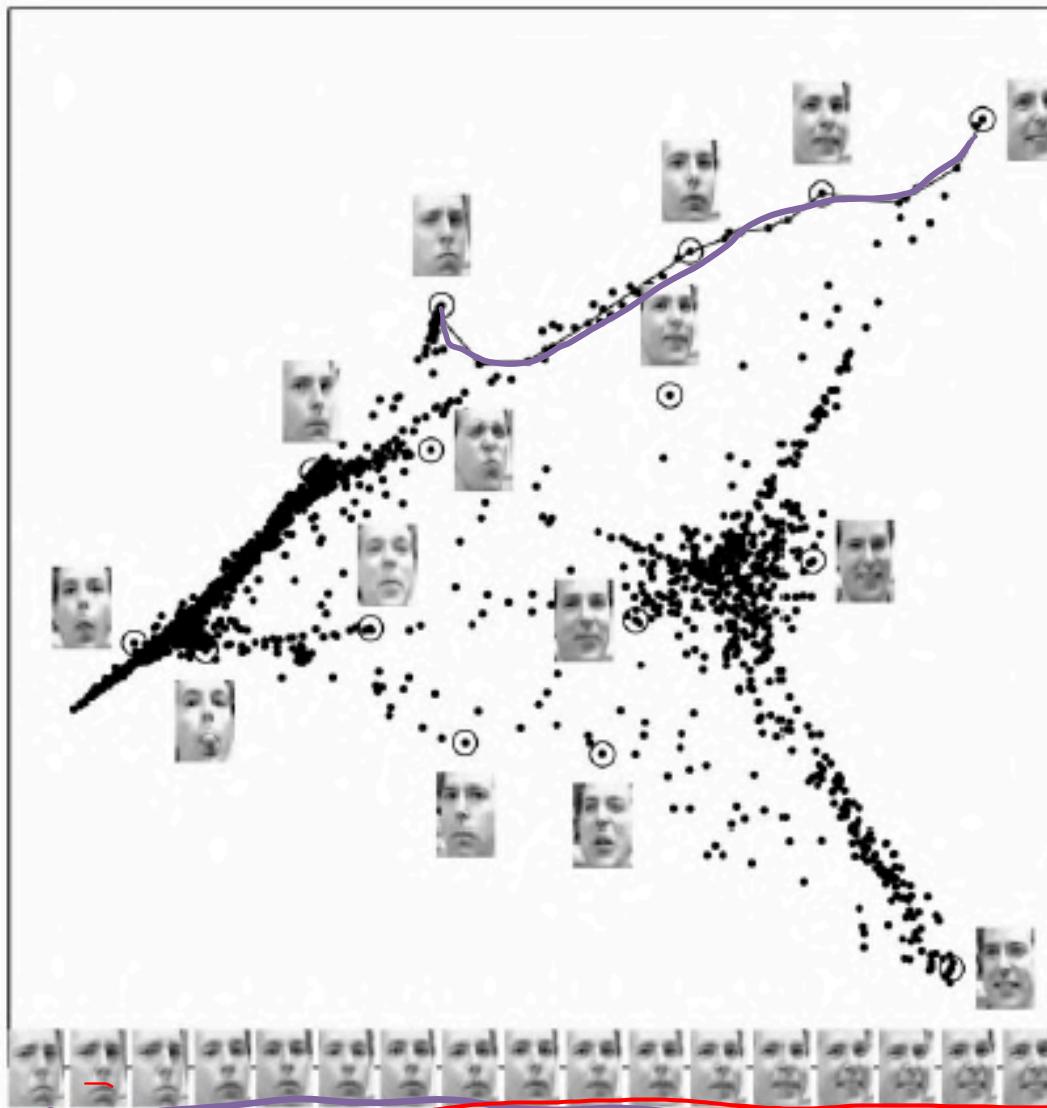
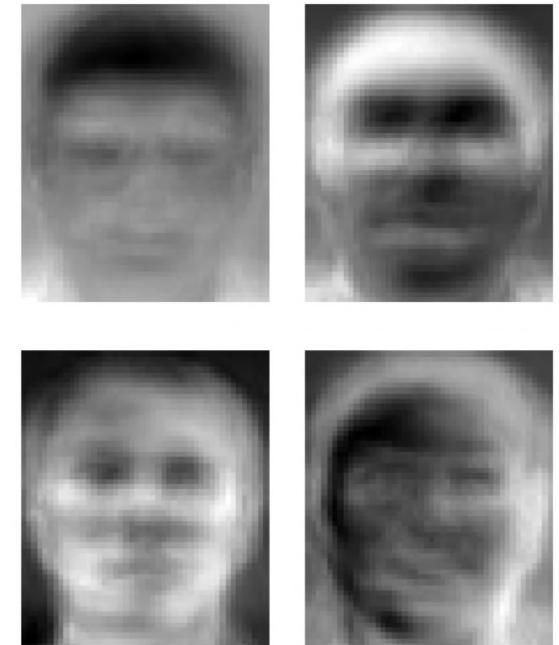


Fig. 3. Images of faces (111) mapped into the embedding space described by the first two coordinates of LLE. Representative faces are shown next to circled points in different parts of the space. The bottom images correspond to points along the top-right path (linked by solid line), illustrating one particular mode of variability in pose and expression.

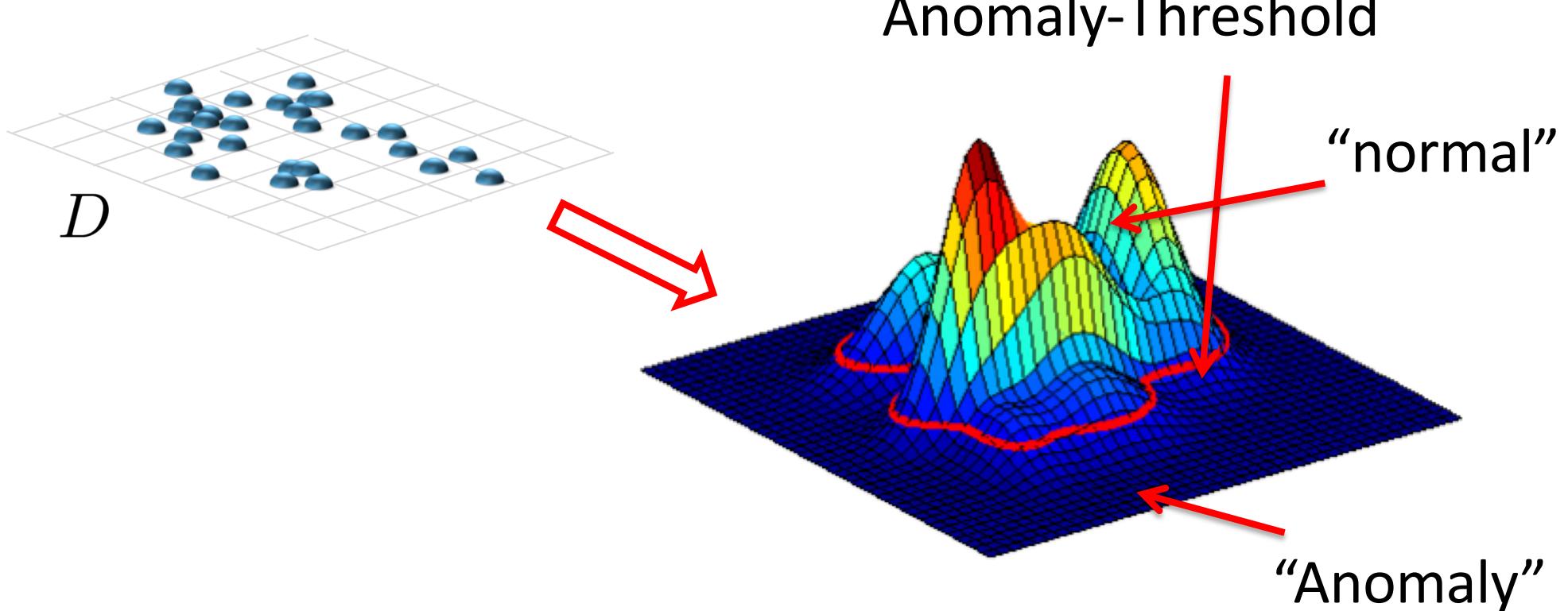
# Example: Dimension reduction

- 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]

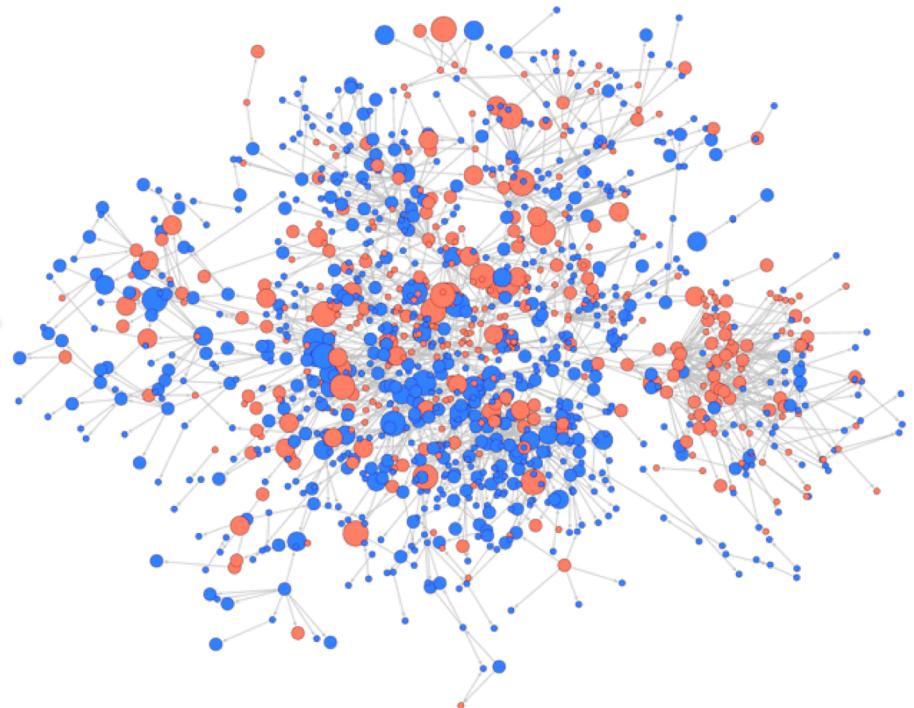
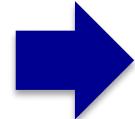
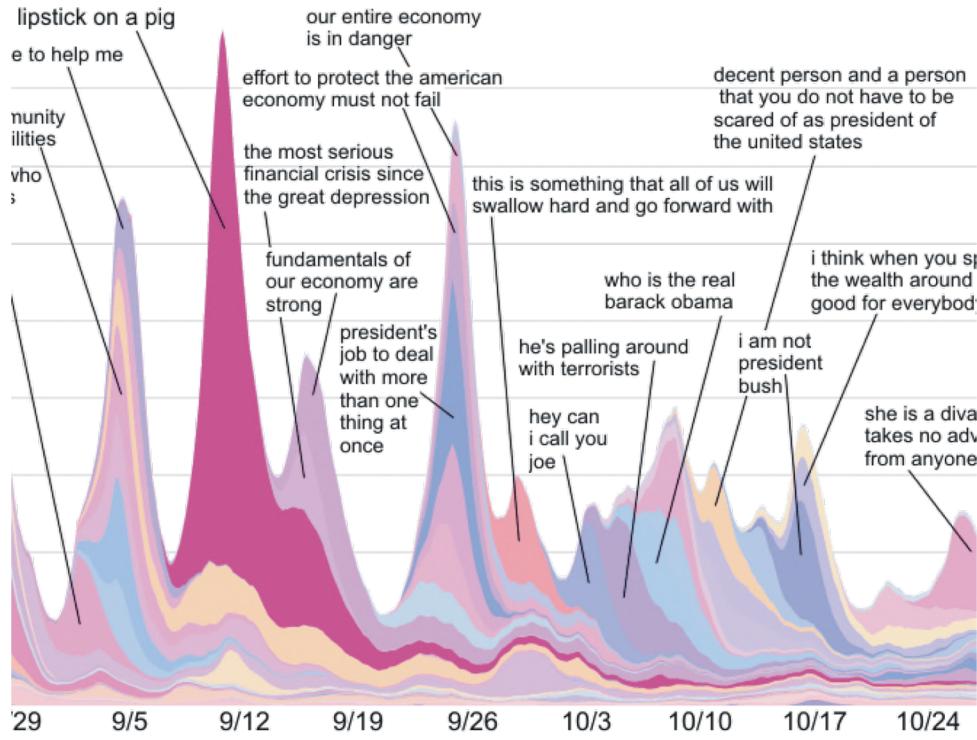
# Example: Anomaly detection



- **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

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

# Example: GANs

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



# BigGAN

[Brock, Donahue, Simonyan 09/2018]



# 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 learning
  - Learn on one domain and test on another
- Active learning
  - Acquiring most informative data for learning
- Online 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